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] and Mark Stoneking[1,*]

[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

* To whom correspondence should be addressed. Tel: +49 (0) 341 3550-504; Email: stonekg@eva.mpg.de; alexander_huebner@eva.mpg.de

[§] These authors contributed equally to the manuscript.

Keywords: mtDNA, heteroplasmy, liver, sharing

ABSTRACT

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 coding 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 coding region, but significantly more sharing of synonymous heteroplasmies. These contrasting patterns for the control vs. the coding region, and for non-synonymous vs. synonymous heteroplasmies, suggest that selection plays a role in heteroplasmy sharing.

INTRODUCTION

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 (Larsson 2010). 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 (Larsson 2010; Wallace and Chalkia 2013; Stewart and Chinnery 2015). 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 (Stewart and Chinnery 2015). The

total number of heteroplasmic sites strongly correlates with age and several studies have shown that heteroplasmic sites are tissue specific (Michikawa et al. 1999; Calloway et al. 2000; Wang et al. 2001; Samuels et al. 2013; Li et al. 2015; Naue et al. 2015), 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. Therefore, the question arises, whether cells from the same tissue share a similar profile of heteroplasmic sites despite being separated 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 a pattern expected under random genetic drift (Jenuth et al. 1996). This result is expected when the underlying heteroplasmies are evolving neutrally. However, it has been shown that mtDNA variants do not behave fully neutrally (Nachman et al. 1996; Jenuth et al. 1997) 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 (Raap et al. 2012; Jayaprakash et al. 2015). This suggests that there might be further population genetic forces, e.g. within-cell mtDNA population structure (Kowald and Kirkwood 2011) or selection, that shape the distribution of heteroplasmies within as well as between tissues.

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 (Li et al. 2015). This result is remarkable because these coding region mutations are likely to have a functional effect (Li et al. 2015), and coding region mutations are strongly selected against during transmission from one generation to another (Stewart et al. 2008). 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 (Li et al. 2015). This makes liver a good candidate tissue to analyze 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 (Raap et al. 2012; Jayaprakash et al. 2015), 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. MtDNA heteroplasmy was evaluated by capture-enrichment sequencing (Li et al. 2010; Maricic et al. 2010; Li and Stoneking 2012; Li et al. 2015), and we analyzed sharing of mtDNA heteroplasmy between the liver lobes for different regions of the mitochondrial genome. We find a high correlation in the minor allele frequency (MAF) at heteroplasmic sites in the control region between the two liver samples, but a much weaker correlation in the coding region.

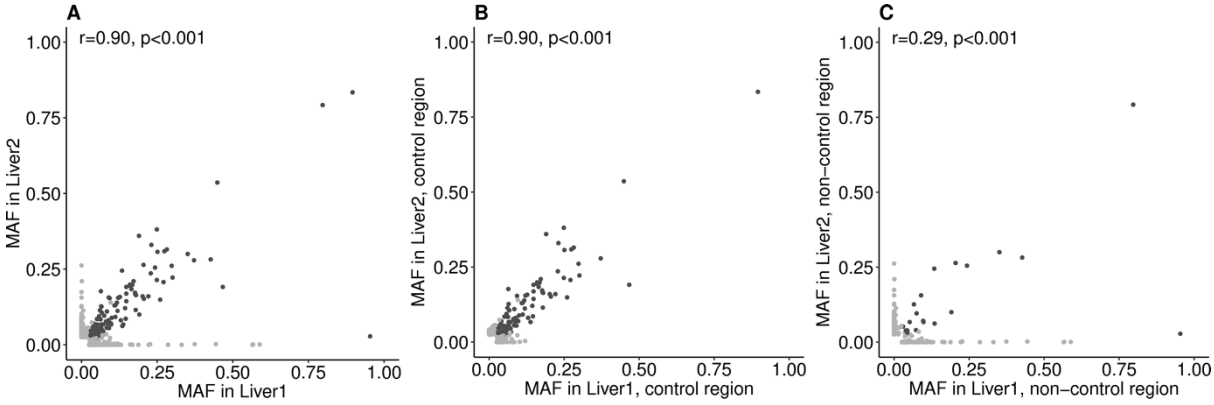
RESULTS

Heteroplasmy sharing within the liver

We investigated mitochondrial DNA heteroplasmy in liver and blood tissue samples of 83 individuals. For liver, two samples taken from different lobes were analyzed 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 coding region 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 in blood (site 72: 1 individual, site 60: 1 individual, site 94: no individuals). These

98 data are in accordance with results from a previous study ((Li et al. 2015), Supplementary
99 Figure S1), indicating that heteroplasmy is tissue specific, with different individuals exhibiting
100 similar heteroplasmic patterns.

101 Virological tests revealed that three individuals had active hepatitis B virus infection, one had
102 active hepatitis C virus infection and one individual was HIV positive, with low viral load
103 (Supplementary Table S2). Those individuals were kept in all downstream analyses, as the
104 number of positive cases was too low to analyze separately. There was no effect of liver fat
105 content on either the total number or the MAFs of heteroplasmic sites (Supplementary Figure
106 S2, p>0.05) and the mitochondrial DNA copy numbers, estimated for each liver sample as
107 described before (Wachsmuth et al. 2016), were highly correlated between corresponding liver
108 samples of an individual, suggesting no functional differences between the liver lobes
109 (Supplementary Figure S3, r=0.81, p<0.001).

110 

111 **Figure 1: Correlation of MAFs at heteroplasmic sites in liver lobes.** Each dot is one
112 heteroplasmic site in one individual. Pearson's correlation coefficient r is given.
113 Heteroplasmic sites are compared in **A** liver sample 1 and 2, **B** the control region of liver
114 sample 1 and 2, **C** the non-control region of liver sample 1 and 2. Heteroplasmic sites
115 shared between liver sample 1 and 2 are in dark grey, non-shared sites in light grey.

116

117 We next asked whether heteroplasmy was correlated between the three different samples
118 from an individual. For blood and each liver sample, there is a low but significant correlation
119 (r=0.35 and 0.38, p<0.001, Supplementary Figure S4) with only 7.5% of the heteroplasmies
120 shared between the tissues. However, the correlation between the two liver samples was
121 higher (Figure 1A, r=0.90, p<0.001). While 355 sites were heteroplasmic in only one of the
122 two liver samples of an individual, 136 sites (28%) were heteroplasmic in both liver samples
123 and these exhibited similar MAFs (Figure 1A). Moreover, there were more shared
124 heteroplasmies from the control region than from the rest of the genome (Table 1, p<0.001):
125 55% of control region heteroplasmies were shared, vs. 7.5% of heteroplasmies outside the
126 control region. The high correlation between the MAFs in the two liver samples was mainly
127 driven by the control region (Figure 1B, r=0.90, p<0.001), while there was a lower, but still
128 significant correlation outside the control region (Figure 1C, r=0.29, p<0.001). The difference
129 in correlation coefficients is significant, based on random partitions of all of the heteroplasmic
130 sites into two sets with the same number of sites as observed for the control region and the
131 non-control region (p<0.001, Supplementary Figure S5). Thus, not only are more
132 heteroplasmies shared in the control region than outside the control region, but control region
133 heteroplasmies also have more similar MAFs.

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

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

136

137 **Potential mechanisms underlying the more frequent sharing in the control region**

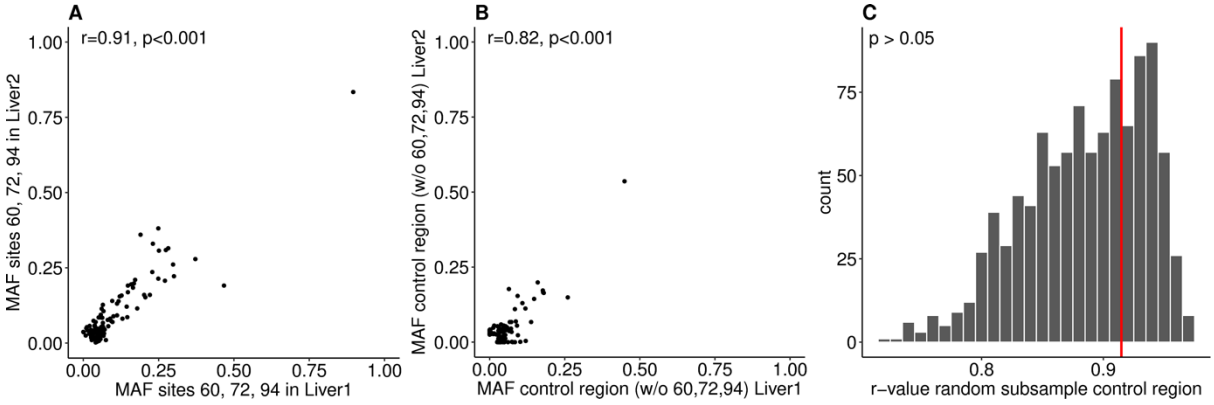


Figure 2: Heteroplasmy sharing in liver samples and correlation with single sites. MAFs per site in liver sample 1 and 2 for: **A** sites 60, 72 and 94; and **B** for all other heteroplasmic sites in the control region. Each dot is one heteroplasmic site in one individual. Pearson's correlation coefficient 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.

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 analyzed them separately. While these three sites did indeed show a high correlation in MAF between the two liver samples (Figure 2A, r = 0.91, p<0.001), the remaining sites in the control region still showed a significant correlation (Figure 2B, r=0.82, p<0.001) that is higher than the correlation for the non-control 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 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

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

161 *7S DNA*. The control region includes the D-loop region, in which a third strand, the 7S DNA,
162 displaces the heavy strand and binds to the light strand. Inferred heteroplasmies in the D-loop
163 region might therefore reflect mutations in the 7S DNA rather than mutations in the mtDNA
164 itself. To see if sequences from 7S DNA were likely to be present in the sequencing libraries,
165 we estimated the relative mtDNA copy number from the capture-enrichment sequencing
166 coverage of the D-loop region and the rest of the mtDNA molecule separately, as described
167 before (Wachsmuth et al. 2016). As the 7S DNA has several starting and end points, we used
168 the outer limits reported in the literature, namely from site 16,097 to site 191 (Roberti et al.
169 1998; Nicholls and Minczuk 2014). The D-loop did not exhibit a higher copy number than the
170 other parts of the mtDNA genome, indicating that 7S DNA is unlikely to be present in the DNA
171 libraries ($p=0.30$ and 0.41 for blood and liver, respectively, Supplementary Figure S6A).
172 Furthermore, the correlation between the MAF for the two liver samples is almost as high for
173 the D-loop region as it for the rest of the control region (Supplementary Figure S6 B,C: $r=0.90$
174 vs. $r = 0.89$). Hence, the significant correlation in MAFs in the control region is likely a
175 phenomenon of the entire control region.

176 *Higher mutation rate*. We also tested if heteroplasmic sites showed a higher mutation rate than
177 non- heteroplasmic sites by comparing the number of heteroplasmies to the inferred mutation
178 rate at each site, based on observed polymorphism data (Soares et al. 2009). Heteroplasmic
179 sites had significantly higher mutation rates than sites that were not heteroplasmic ($p<0.001$,
180 Supplementary Figure S7) with the mutation rate being higher in the control region than
181 outside the control region. However, mutation rates did not differ between shared vs. non-
182 shared heteroplasmies ($p>0.05$, Supplementary Figure S7), suggesting that the mutation rate
183 does not increase the probability of heteroplasmies to be shared.

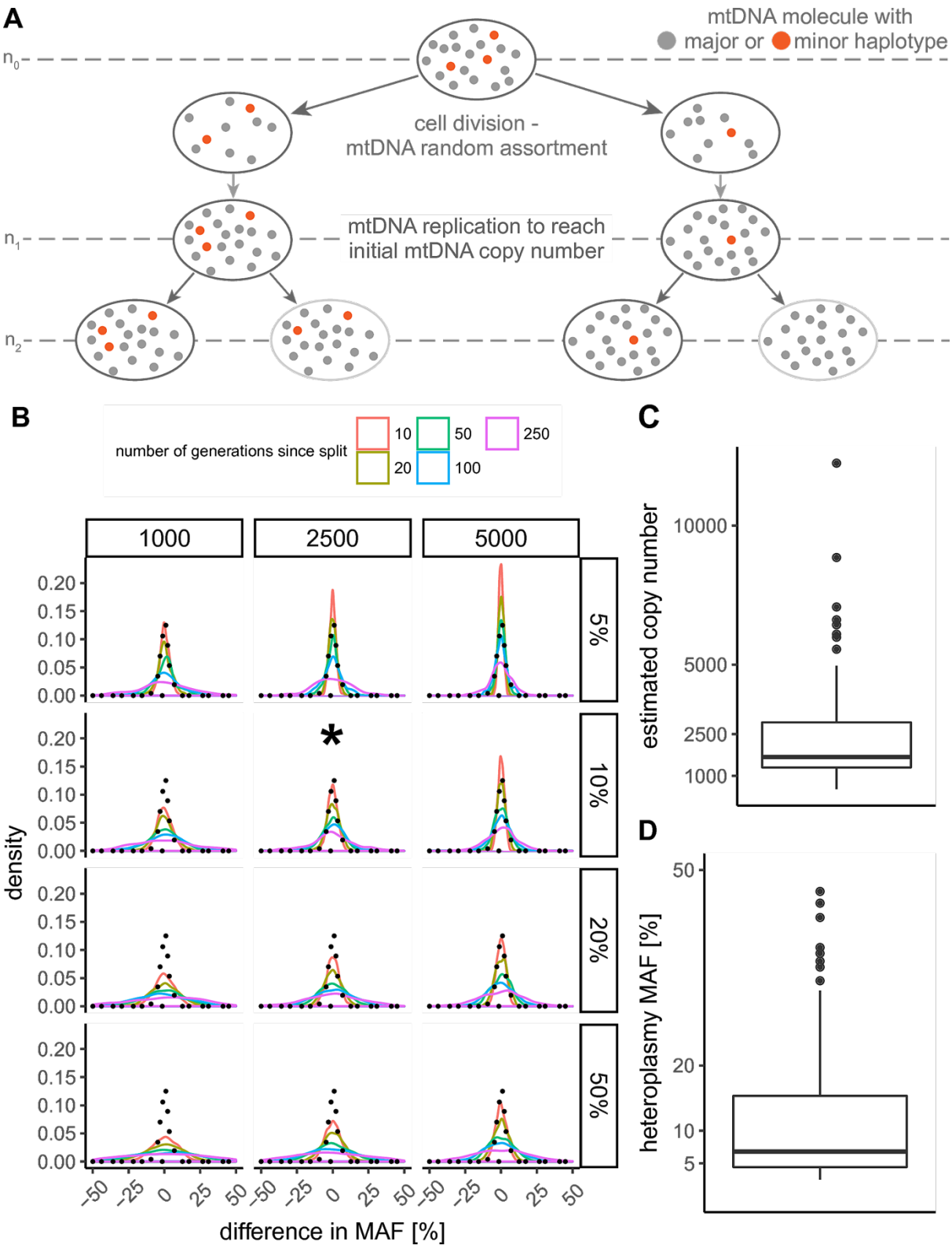


Figure 3: Expected difference in minor allele frequency (MAF) assuming random genetic drift. A schematic scheme of the two-step mtDNA replication model used for the simulations: first, all mtDNA molecules are segregated equally into two daughter cells followed by replication step to re-gain the same number of mtDNA copies as in the

189 progenitor cell. After each generation n_i, the MAF frequency of one random cell on each
190 site of the pedigree is determined and the MAF difference between the two cells
191 calculated. **B** expected difference in MAF after 10, 20, 50, 100, and 250 generations
192 since the split of the two cells. Different mtDNA copy numbers (1,000, 2,500, and 5,000;
193 kept constant throughout simulation) and initial MAFs (5%, 10%, 20%, and 50%) were
194 used to simulate random genetic drift between two cells. For each combination of mtDNA
195 copy number and initial MAF, 1000 replicates were simulated. The black, dotted line
196 indicates the observed MAF difference between the shared liver heteroplasmies in the
197 data set. The asterisk highlights the simulation with the parameters closest to the ones
198 observed in the data set. **C** the mtDNA copy number distribution estimated from the
199 capture-enriched sequencing data of the liver samples. The values were corrected using
200 a correction "ratio" of 1/150 (Wachsmuth et al. 2016) to convert the relative to absolute
201 mtDNA copy numbers. **D** the distribution of MAF in the liver heteroplasmies.

202

203 *Random genetic drift*. If control region heteroplasmies are more likely to have arisen in a
204 common progenitor cell of the cells sampled in the two liver lobes, then the higher correlation
205 in MAF might reflect this common ancestry. We therefore tested if the correlation in MAF is
206 compatible with expected amount of random genetic drift from a common progenitor cell by
207 simulating mtDNA as a two-step model (Figure 3A) following (Jayaprakash et al. 2015). We
208 started with different MAFs at the initial generation n_0 (5%, 10%, 20%, 50%) and different
209 constant mitochondrial DNA copy numbers (1,000, 2,500, and 5,000) and sampled the
210 difference in MAF 10, 20, 50, 100, and 250 generations after the initial split of the cells (Figure
211 3B). For all simulations, we observed a higher MAF difference with an increasing number of
212 cell replications. This effect was more pronounced for higher MAF at generation n_0 and lower
213 mtDNA copy number. Based on the observed mtDNA copy number (Figure 3C; mean=2,528)
214 and MAF of shared heteroplasmies with the same consensus allele (Figure 3D; mean=10.8%),
215 the most similar corresponding simulation is with a copy number of 2,500 and a MAF at

216 generation n_0 of 10%; for this simulation, the variance in observed MAF differences was
217 smaller than would be expected after 50 generations of random genetic drift (one-sided *F*-test
218 for the equality of two variances, $p<0.001$). With an increasing number of cell replications, the
219 simulated MAF difference further diverged from the observed values, which suggests that
220 random genetic drift cannot explain the observed sharing.

221 **Differences in MAF between corresponding liver samples are not correlated with age**

222 We next investigated the influence of age on heteroplasmy sharing between different liver
223 regions of an individual. Overall, the total number of different heteroplasmic sites in an
224 individual increased with age for both the control region and the coding region (Supplementary
225 Figure S8A, $r=0.42$ and adjusted $p<0.001$ both within and outside of the control region).
226 However, the MAF at heteroplasmic sites did not increase with age (Supplementary Figure
227 S8B, $r=0.09$ for the control region, -0.01 for outside of the control region, adjusted $p>0.05$).
228 Although some heteroplasmies exhibit high MAFs only at ages above 50, many sites remain
229 at low frequencies even at older ages (Supplementary Figure S8B). The hypothesis that
230 random genetic drift has an effect on the difference in MAF of two corresponding liver samples
231 would suggest that with increasing age the difference in MAF would increase, too. Yet, when
232 testing for a correlation between the difference in MAF between two corresponding liver
233 samples and age of the individual, we did not observe a significant correlation either within or
234 outside the control region (Figure 4, $r=0.06$ within control region, $r=-0.01$ outside the control
235 region, both $p>0.05$). In order to test whether, in contrast, specific sites have a significant
236 correlation between the MAF difference and age, we separately calculated the correlation
237 between MAF difference and age for the three most frequent heteroplasmy sites (sites 60, 72,
238 and 94), for which at least 20 individuals were heteroplasmic (Supplementary Figure S9). The
239 correlation between MAF and age was not significant for any of these three sites (adjusted
240 $p>0.05$), which further supports the hypothesis that the difference in MAF does not increase
241 with age, contrary to what would be expected from random genetic drift.

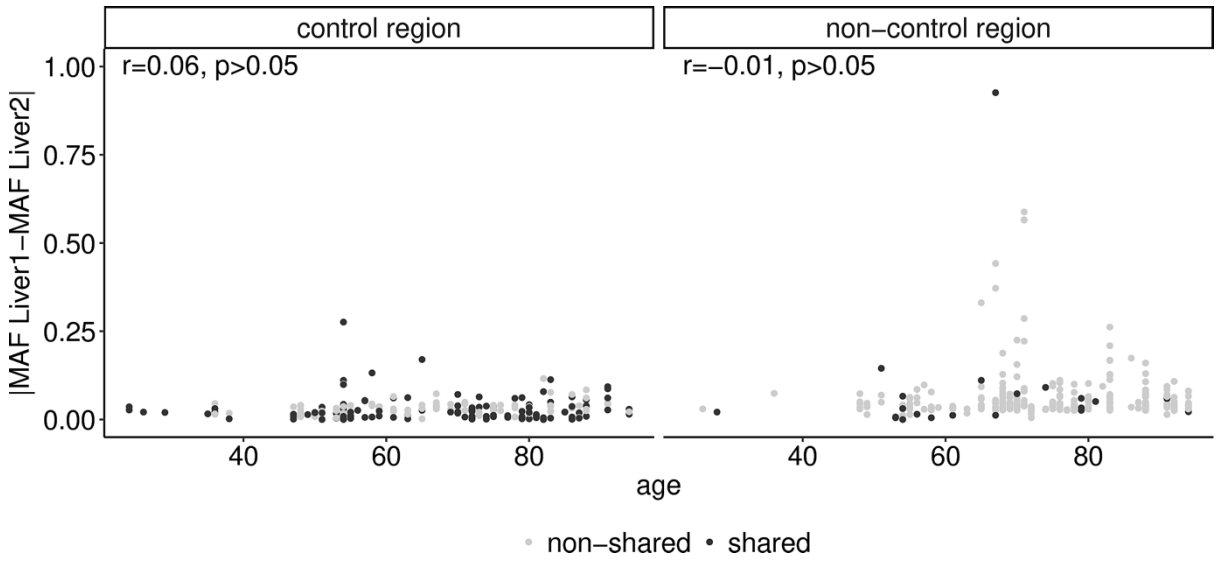


Figure 4: Minor allele frequency difference and age. The difference in minor allele frequency between liver sample 1 and 2 versus age plotted separately for sites within and outside the control region. Dark grey dots indicate sites that were shared (i.e. heteroplasmic in both liver samples from an individual), light grey dots indicate sites that were not shared.

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 (Li et al. 2015). We investigated this by calculating the ratio of non-synonymous heteroplasmies per non-synonymous site vs. synonymous heteroplasmies per synonymous site (hN/hS). 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 (Li et al. 2015), the hN/hS ratio is significantly greater than 1 (adjusted $p<0.05$, Supplementary Figure S10A), indicating positive selection for non-synonymous heteroplasmies rather than relaxation of functional constraints (Li et al. 2015). Additionally, we observed a strong increase in the number of heteroplasmies occurring the coding region compared to the control region in individuals older than 60 years (Supplementary Figure S8A). We, therefore, investigated whether the excess in non-synonymous heteroplasmies is age-dependent (Supplementary

261 Figure S10B). While both synonymous and non-synonymous heteroplasmies increased in
262 number with increasing age, only non-synonymous heteroplasmies showed a significant
263 difference between the age groups below 60 years and above 60 years (two-sided Mann-
264 Whitney U test, adjusted $p<0.05$). We then asked if either synonymous or non-synonymous
265 heteroplasmic sites were more likely to be shared between different liver samples. Although
266 there were more than twice as many non-synonymous than synonymous heteroplasmies in
267 the data set (163 vs. 62), there were significantly more synonymous sites shared than non-
268 synonymous sites (11 vs. 6, Table 2). Accordingly, the median MAF difference between
269 corresponding liver samples was on average higher for non-synonymous heteroplasmies than
270 synonymous ones (median MAF difference of 4.4% to 3.8%), albeit not significantly so ($p>0.05$,
271 Supplementary Figure S11A). These results suggest that non-synonymous heteroplasmies
272 more often arise independently in different cells. However, neither non-synonymous nor
273 synonymous heteroplasmies showed a significant correlation for the difference in MAF
274 between corresponding liver samples with age ($p>0.5$, Supplementary Figure S11B).

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

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

277

278 **DISCUSSION**

279 **Heteroplasmies are shared across liver lobes**

280 In this study, we compared mtDNA heteroplasmy patterns in samples collected from the two
281 primary liver lobes of 83 individuals sampled at autopsy and found a significant correlation in
282 heteroplasmy MAF between the two liver samples from an individual. Moreover, if sites are
283 heteroplasmic in both liver samples of an individual, the MAFs at those sites tend to be similar

284 independent of age. Previous studies found similar correlations in MAF between cell colonies
285 derived from single cells and grown in culture (Raap et al. 2012; Jayaprakash et al. 2015), so
286 our results extend these findings to different parts of the same organ within living individuals.
287 In addition, we analyzed patterns of heteroplasmy sharing and MAF correlation across
288 different parts of the mtDNA genome, which provide further information concerning the
289 potential mechanism(s) behind these observations. In particular, we found significantly more
290 sharing of heteroplasmic sites and higher MAF correlations for the control region than for the
291 rest of the genome. We showed that these differences are not due to the presence of 7S DNA,
292 nor are they driven by just a few sites. Moreover, within the coding region, we found
293 significantly more sharing of heteroplasmies and higher MAF correlations at synonymous sites
294 than at nonsynonymous sites.

295 These observations allow us to evaluate several potential explanations for this heteroplasmy
296 sharing within liver. First, a trivial potential explanation is that a heteroplasmy could appear to
297 be shared between liver lobes when it is actually present in blood. About 1 liter of blood per
298 minute flows through the blood vessels of the liver (Wynne et al. 1989), so the DNA from the
299 liver samples also contains DNA arising from blood cells. Heteroplasmies present in the blood
300 cells could therefore be detected as heteroplasmies in the liver samples and hence seem to
301 be shared across lobes. There are ~10^{10} liver cells in 10 g of liver tissue and 5 x 10^{10} blood
302 cells in 10 ml of blood (Sender et al. 2016), and in liver autopsy samples there is approximately
303 10 ml blood per 100 g of tissue (Greenway and Stark 1971). The ratio of liver to blood cells in
304 a liver autopsy sample is therefore about 1.8. As we additionally observed for our samples
305 that the average mitochondrial copy number is five times higher in liver than in blood samples
306 (Supplementary Figure S6a), the overall ratio of mitochondrial genomes from liver cells to
307 those from blood cells is approximately 9 to 1. Therefore, a heteroplasmy in blood would need
308 a MAF of at least 22.5% to be called a heteroplasmy in liver (minimum MAF = 2.5%) when it
309 was actually totally absent in the liver samples. Only 8 out of the 541 heteroplasmies in blood
310 have a MAF of ≥ 22.5%, of which four are shared across blood and both liver lobes

311 (Supplementary Table S1). Moreover, the heteroplasmies that are commonly observed in liver
312 (at positions 60, 72, and 94) are practically absent from blood; these results rule out
313 experimental contamination with blood DNA as the primary driver for heteroplasmy sharing
314 across liver lobes.

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

324 Third, it has been shown that colonic stem cells with non-synonymous mtDNA mutations can
325 expand clonally from a few cells and spread throughout a tissue by crypt fission (Greaves et
326 al. 2006). During this process they retain the same level of mutant DNA on a cellular level.
327 While a similar mechanism in liver could explain the results presented here, this process would
328 require a cellular turnover on a huge scale throughout the entire liver, starting from just a few
329 stem cells. This does not seem very likely, as clonal expansion was shown for single intestinal
330 crypts only and is supposed to be rather slow (Greaves et al. 2006). Moreover, it is not clear
331 to what extent liver regeneration is driven by stem cells vs. mature hepatocytes (Grompe 2014).

332 Fourth, heteroplasmy sharing could also derive from a pre-existing, inherited heteroplasmy
333 (Guo et al. 2013; Payne et al. 2013; Rebolledo-Jaramillo et al. 2014; Li et al. 2015) that
334 remains at similar frequencies across the tissue because drift (random changes in MAF) is
335 limited. We tested this possibility, using a simulation scheme from a previous study
336 (Jayaprakash et al. 2015), with different initial heteroplasmy frequencies, and mtDNA copy
337 numbers that were kept constant (Berk and Clayton 1974). As expected, we observed a larger

338 variance in MAF between daughter cells with increasing starting MAF and decreasing mtDNA
339 copy number (Figure 3B). For the simulation results closest to our observed average mtDNA
340 copy number and average heteroplasmy MAF (2,500 and 10%, Figure 3C-D), the observed
341 difference in MAF between shared heteroplasmies was smaller than would be expected after
342 random genetic drift acting on 50 mtDNA replication steps. However, more than these 50
343 mtDNA replication steps are expected to have taken place during the life of an average
344 individual in our study. Assuming 3.61×10^{11} liver cells in an adult liver (Bianconi et al. 2013),
345 a total of 39 hepatocyte cell replications (including mtDNA replication) are needed to obtain a
346 full-size, adult liver from a single hepatocyte cell. After the development stage, the post-mitotic
347 cells continue to replicate their mtDNA independently of cell replication ("relaxed" replication
348 (Poovathingal et al. 2009)). The estimated half-life of mtDNA in post-mitotic cells ranges
349 between 2-10 days (Miwa et al. 2008) and 30-300 days (Poovathingal et al. 2012). Assuming
350 an mtDNA half-life of 30 days, there would be complete replacement of all mtDNA molecules
351 of a cell within a year, given a mtDNA copy number of 2,500 per cell, so approximately one
352 "relaxed" replication cycle occurs within the liver of an adult for every year of age. Thus, for
353 the age range in our data set of 24 to 94 years, the liver samples would have gone through
354 about 62 - 132 cell replications prior to sampling, and so the observed difference in MAF in
355 the liver samples is significantly smaller than expected. Moreover, liver samples accumulated
356 significantly more heteroplasmies with age (Supplementary Figure S8A), further arguing that
357 shared heteroplasmies do not reflect pre-existing, inherited heteroplasmies. Also, there was
358 no significant correlation of MAF difference with age, as would be expected with random
359 genetic drift. In sum, our results extend to liver tissues the previous observations (Raap et al.
360 2012; Jayaprakash et al. 2015) that random genetic drift alone cannot explain heteroplasmy
361 sharing between cells in culture.

362 Finally, an equilibrium of heteroplasmy across an entire tissue could be explained by an
363 exchange of genetic material from mitochondria between cells. Cells can donate whole
364 mitochondria to adjacent cells through nanotubes, but this has been suggested for distances

365 up to 100 μM only and the exchange is often triggered by functional impairments in the
366 acceptor cell (Rogers and Bhattacharya 2013). An additional way for cells to exchange DNA
367 material could be the uptake of extracellular DNA material that is either secreted by healthy
368 cells or is present as the remains of apoptotic cells (van der Vaart and Pretorius 2008). While
369 the uptake and integration of cell-free nuclear DNA material has been shown (Basak et al.
370 2016), it is unclear whether cells would also accept mitochondrial DNA. However, studies of
371 heteroplasmy at the single cell level (reviewed in (Yao et al. 2015)) do suggest the possibility
372 of transfer between cells. Experiments with cell culture mixes of fluorescently labelled cell lines
373 suggested the exchange of mtDNA between co-cultured partner cell lines, although the
374 specific mechanism, either transfer of mitochondrial organelles or transfer of free mtDNA,
375 could not be identified (Jayaprakash et al. 2015). Overall, such intercellular DNA exchange,
376 followed by incorporation of mtDNA fragments into the mtDNA of the recipient cells, could
377 account for the significant correlation we observe in MAF between liver lobes.

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

387 Instead, our data suggest that even if intercellular exchange is occurring, selection must be
388 involved in the sharing of heteroplasmies and correlation in MAF between liver lobes. Several
389 aspects of the data suggest that selection influences heteroplasmies. In the coding region of
390 the mtDNA, we observed a significant excess of non-synonymous vs. synonymous
391 heteroplasmies (Table 2), more so than can be explained by relaxation of functional

392 constraints on non-synonymous mutations. Moreover, the number of non-synonymous
393 heteroplasmies increased significantly in individuals above 60 years. Overall, these results
394 strongly suggest positive selection for nonsynonymous heteroplasmies in liver, as found
395 previously (Li et al. 2015), and possibly reflecting the hypothesis of the "survival of the slowest"
396 (deGrey 1997), which postulates that mitochondria with reduced respiratory function due to
397 increasing mutations suffer less degradation from the production of reactive oxygen species
398 (ROS). Hence, mitochondria that lack these mutations suffer ROS-related damage and are
399 removed from cells, thereby resulting in an increase in frequency in mtDNAs with non-
400 synonymous mutations that decrease mitochondrial function.

401 However, our results indicate a more complex role for selection in the different patterns of
402 heteroplasmy sharing and MAF correlation across different regions of the mtDNA genome, in
403 keeping with evidence from other studies. In mice and humans, significantly more synonymous
404 than non-synonymous heteroplasmies are transmitted to the next generation (Stewart et al.
405 2008; Rebolledo-Jaramillo et al. 2014; Floros et al. 2018), suggesting selection against non-
406 synonymous heteroplasmies during transmission. The notable tissue-specificity and allele-
407 specificity of particular heteroplasmic sites in the control region also suggests a role for
408 positive selection on heteroplasmies during aging (Samuels et al. 2013; Li et al. 2015). The
409 increasing evidence for both purifying and positive selection acting on heteroplasmic variants
410 warrants further investigation, particularly into the potential health-related consequences.

411 **MATERIAL AND METHODS**

412 **Tissue collection and DNA extraction**

413 Blood and liver were sampled at autopsy from 94 individuals (57 males, 37 females, age range:
414 24-94, mean: 63, median: 63). Two samples were taken from each liver, one from the right
415 lobe and one from the left lobe. DNA was extracted as previously described (Li et al. 2015).
416 The collection of samples and the experimental procedures were approved by the Ethics

417 Commissions of the Rheinische Friedrich Wilhelms University Medical Faculty (Lfd. Nr. 097/15)
418 and the University of Leipzig Medical Faculty (Az. 305-15-24082015).

419 **Virological assays and histological investigation**

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

438 Fat content of the liver was determined by histological investigations and Sudan staining
439 (Mulisch M 2015). Tissues with <10% hepatocytes including fat droplets were considered low,
440 10-30% were medium, 31-50% were high fat and >50% were considered adipohepatic.

441 **Illumina library preparation and sequencing**

442 Double-barcoded DNA libraries for sequencing were prepared and capture-enriched for
443 mtDNA as previously described (Li et al. 2015). DNA was sequenced on the Illumina HiSeq
444 platform in rapid mode with 95 bp paired-end reads. Bases were called with FreeIbis (Renaud
445 et al. 2013) and reads were subsequently trimmed and merged using leeHom (Renaud et al.
446 2014).

447 **Heteroplasmy detection**

448 Heteroplasmy was detected according to the DREEP pipeline (Li et al. 2010; Li and Stoneking
449 2012). First, heteroplasmies were called if: the minor allele frequency (MAF) for the most
450 frequent minor allele was at least 2.5% on both the forward and reverse strand; the sequencing
451 depth was at least 500-fold at a candidate site; and there were at least 10 reads supporting
452 the minor allele on each strand. Additionally, we required a minimum heteroplasmic quality
453 score of 10 on each strand. In order to discriminate a true heteroplasmy from sequencing error,
454 the DREEP pipeline compares the minor allele pattern of any inferred heteroplasmic site to a
455 database that comprises the minor allele patterns at this site from all other individuals in the
456 study. When the majority of the samples are from a single tissue like liver in this study (two
457 liver samples and one blood sample per individual), DREEP is prone to considered commonly
458 heteroplasmic sites as elevated sequencing error and therefore under-estimates the
459 heteroplasmic quality score for these. Thus, we used the information about commonly
460 heteroplasmic sites from (Li et al. 2015) to flag these for both blood and liver tissue in this
461 study, respectively. A site was considered commonly heteroplasmic in a tissue when at least
462 five individuals were heteroplasmic for it. Based on this criterion, we ignored the heteroplasmic
463 quality for sites 60, 72, 94, 185, 189, 203, 11,126, 16,093, and 16,126 for liver and 12,705 for
464 blood and considered a site heteroplasmic if all other criteria were fulfilled. The following
465 regions were excluded for heteroplasmy analysis: 302-316, 513-526, 566-573, and 16,181-
466 16,194. We confirmed that all heteroplasmies were within a coverage between 20 % and 200 %
467 of the average coverage of the sample. In addition, all samples that could have been
468 contaminated with other samples during library preparation/extraction were removed. To

469 detect such contamination, pairwise comparisons of all liver samples with each other as well
470 as all blood samples with each other were performed. All three samples of an individual were
471 removed if all of the following criteria were fulfilled for any pairwise comparison between
472 individuals across all sites of the mitochondrial genome: 1) for at least 80% of the sites, for
473 which two samples had different consensus alleles, the minor allele in the recipient sample
474 was identical to the major allele in the donor sample; 2) the average MAF across these sites
475 was at least 1%; and 3) at least 60% of all sites in the recipient sample, for which a minor
476 allele was observed, were identified as heteroplasmies by the DREEP pipeline (Li et al. 2010;
477 Li and Stoneking 2012). Furthermore, an additional filter for potential contamination was
478 applied, in which the heteroplasmic sites for each sample were checked to see if five or more
479 sites could be explained by contamination from another haplogroup. In total, ten individuals
480 were removed in these contamination filter steps. Finally, we removed a single individual
481 because its samples had more than 2% of the MT genome below a sequencing depth of 500-
482 fold, the cut-off for being considered a heteroplasmy, and thus would have had a higher false
483 negative rate than the other samples. Overall, we retained 83 individuals for the subsequent
484 analyses. Minor allele frequencies were calculated with respect to the major allele in blood.

485 **Correlation analysis**

486 Statistical analysis was performed using R (https://www.R-project.org), with analyses
487 performed for the entire mtDNA genome and separately for the coding region (577-16,023),
488 the control region (16,024-576) and the D-loop region (16,097-191). For correlation of MAFs,
489 we selected only sites that were identified as heteroplasmies and passed our quality filters in
490 at least one of the tissues. We then compared the MAFs of these heteroplasmies to the MAF
491 in the other tissues of an individual, even if the site was not detected as a heteroplasmy in the
492 other tissues. Pearson correlation coefficients were calculated for correlations between MAFs
493 among samples as well as for correlations with age assuming a two-sided alternative
494 hypothesis; the significance of the correlation was tested by randomly permuting the data.
495 Permutation tests were also used to assess the association between specific sites or regions

496 and minor allele sharing between liver lobes; all permutations were carried out 1000 times.
497 Whenever categorical data were compared (e.g. synonymous vs. non-synonymous sites),
498 two-sided Mann-Whitney U tests were used to test for significant differences. Fisher's exact
499 test was used to test for an association of sharing of heteroplasmic sites between liver lobes
500 for the control region vs. non-control region and for synonymous vs. non-synonymous
501 heteroplasmies using a two-sided alternative hypothesis. P-values were adjusted for multiple
502 testing using Benjamini-Hochberg correction (Benjamini and Hochberg 1995) and highlighted
503 in the text.

504 **Coverage across the mitochondrial genome**

505 We used per-site coverage determined by the *filter_and_summary.pl* script of the DREEP
506 pipeline (Li and Stoneking 2012) for each sample and calculated the average coverage across
507 the D-loop region and across the rest of the mtDNA genome.

508 **Non-synonymous heteroplasmies**

509 The hN/hS ratio (Li et al. 2015) was calculated by calculating the Ka/Ks ratio using
510 *KaKs_Calculator 2.0* (Wang et al. 2010) between the revised Cambridge Reference Sequence
511 (rCRS; Andrews, 1999 doi:10.1038/13779) and a mock sequence created by introducing all
512 minor alleles of heteroplasmies into the coding region of the rCRS. A significance test was
513 performed by randomly introducing the same substitutions as observed for heteroplasmies at
514 any site of the coding region of rCRS and calculating the hN/hS ratio in comparison to the non-
515 altered rCRS.

516 The potential functional impact of non-synonymous heteroplasmies was analyzed by
517 overlapping the position of the heteroplasmy and its minor allele with the MitImpact database
518 (Castellana et al. 2015) and comparing the *MutationAssessor* (Reva et al. 2011) score.

519

520 **DATA ACCESS**

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

525 **ACKNOWLEDGEMENTS**

526 We thank Anne Haubner, Antje Weihmann, Barbara Höber and Melanie Grabmüller for
527 technical support. We thank Enrico Macholdt, Michael Dannemann, and Martin Petr for
528 valuable discussion. The research was funded by the Max Planck Society

529 **AUTHOR CONTRIBUTIONS**

530 MS, BM, ML, AH and MW designed the study. BM collected the samples. RS performed DNA
531 isolation and sequencing library preparation, AMEH performed virological assays and BM
532 performed histological analyses. ML and AH wrote pipeline and executed heteroplasmy
533 detection, AH and MW performed statistical analyses. AH, MW, AMEH and MS wrote the
534 manuscript. All authors read and approved the final manuscript.

535 **DISCLOSURE DECLARATION**

536 The authors declare that they have no competing interests.

537 **REFERENCES**

538 Basak R, Nair NK, Mittra I. 2016. Evidence for cell-free nucleic acids as continuously arising
539 endogenous DNA mutagens. *Mutation Research-Fundamental and Molecular*
540 *Mechanisms of Mutagenesis* **793**: 15-21.
541 Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate - a practical and powerful
542 approach to multiple testing. *Journal of the Royal Statistical Society Series B-*
543 *Methodological* **57**: 289-300.
544 Berk AJ, Clayton DA. 1974. Mechanism of mitochondrial DNA replication in mouse L-cells:
545 asynchronous replication of strands, segregation of circular daughter molecules,

546 aspects of topology and turnover of an initiation sequence. *Journal of molecular biology*
547 **86**: 801-824.
548 Bianconi E, Piovesan A, Facchin F, Beraudi A, Casadei R, Frabetti F, Vitale L, Pelleri MC,
549 Tassani S, Piva F et al. 2013. An estimation of the number of cells in the human body.
550 *Annals of human biology* **40**: 463-471.
551 Calloway CD, Reynolds RL, Herrin GL, Anderson WW. 2000. The frequency of heteroplasmy
552 in the HVII region of mtDNA differs across tissue types and increases with age.
553 *American Journal of Human Genetics* **66**: 1384-1397.
554 Castellana S, Ronai J, Mazza T. 2015. MitImpact: an exhaustive collection of pre-computed
555 pathogenicity predictions of human mitochondrial non-synonymous variants. *Hum*
556 *Mutat* **36**: E2413-2422.
557 deGrey A. 1997. A proposed refinement of the mitochondrial free radical theory of aging.
558 *Bioessays* **19**: 161-166.
559 Floros VI, Pyle A, Dietmann S, Wei W, Tang WWC, Irie N, Payne B, Capalbo A, Noli L,
560 Coxhead J et al. 2018. Segregation of mitochondrial DNA heteroplasmy through a
561 developmental genetic bottleneck in human embryos. *Nature cell biology* **20**: 144-144.
562 Greaves LC, Preston SL, Tadrous PJ, Taylor RW, Barron MJ, Oukrif D, Leedham SJ,
563 Deheragoda M, Sasieni P, Novelli MR et al. 2006. Mitochondrial DNA mutations are
564 established in human colonic stem cells, and mutated clones expand by crypt fission.
565 *Proceedings of the National Academy of Sciences of the United States of America* **103**:
566 714-719.
567 Greenway C, Stark R. 1971. Hepatic vascular bed. *Physiological reviews* **51**: 23-65.
568 Grompe M. 2014. Liver stem cells, where art thou? *Cell stem cell* **15**: 257-258.
569 Guo Y, Li CI, Sheng QH, Winther JF, Cai QY, Boice JD, Shyr Y. 2013. Very Low-Level
570 Heteroplasmy mtDNA Variations Are Inherited in Humans. *Journal of Genetics and*
571 *Genomics* **40**: 607-615.
572 Jayaprakash AD, Benson EK, Gone S, Liang R, Shim J, Lambertini L, Toloue MM, Wigler M,
573 Aaronson SA, Sachidanandam R. 2015. Stable heteroplasmy at the single-cell level is
574 facilitated by intercellular exchange of mtDNA. *Nucleic Acids Research* **43**: 2177-2187.
575 Jenuth JP, Peterson AC, Fu K, Shoubridge EA. 1996. Random genetic drift in the female
576 germline explains the rapid segregation of mammalian mitochondrial DNA. *Nature*
577 *genetics* **14**: 146-146.
578 Jenuth JP, Peterson AC, Shoubridge EA. 1997. Tissue-specific selection for different mtDNA
579 genotypes in heteroplasmic mice. *Nature genetics* **16**: 93-93.
580 Kowald A, Kirkwood TBL. 2011. Evolution of the mitochondrial fusion--fission cycle and its role
581 in aging. *Proceedings of the National Academy of Sciences* **108**: 10237-10242.

582 Larsson NG. 2010. Somatic mitochondrial DNA mutations in mammalian aging. *Annual review*
583 *of biochemistry* **79**: 683-706.
584 Li M, Schoenberg A, Schaeferd M, Schroeder R, Nasidze I, Stoneking M. 2010. Detecting
585 Heteroplasmy from High-Throughput Sequencing of Complete Human Mitochondrial
586 DNA Genomes. *American Journal of Human Genetics* **87**: 237-249.
587 Li M, Schroeder R, Ni S, Madea B, Stoneking M. 2015. Extensive tissue-related and allele-
588 related mtDNA heteroplasmy suggests positive selection for somatic mutations.
589 *Proceedings of the National Academy of Sciences of the United States of America* **112**:
590 2491-2496.
591 Li M, Stoneking M. 2012. A new approach for detecting low-level mutations in next-generation
592 sequence data. *Genome Biology* **13**: R34.
593 Maricic T, Whitten M, Paeaebo S. 2010. Multiplexed DNA Sequence Capture of Mitochondrial
594 Genomes Using PCR Products. *Plos One* **5**: e14004.
595 Michikawa Y, Mazzucchelli F, Bresolin N, Scarlato G, Attardi G. 1999. Aging-dependent large
596 accumulation of point mutations in the human mtDNA control region for replication.
597 *Science* **286**: 774-779.
598 Miwa S, Lawless C, Von Zglinicki T. 2008. Mitochondrial turnover in liver is fast in vivo and is
599 accelerated by dietary restriction: application of a simple dynamic model. *Aging cell* **7**:
600 920-923.
601 Mulisch M WU. 2015. *Romeis - Mikroskopische Technik*. Springer Spektrum, Berlin
602 Heidelberg.
603 Nachman MW, Brown WM, Stoneking M, Aquadro CF. 1996. Nonneutral mitochondrial DNA
604 variation in humans and chimpanzees. *Genetics* **142**: 953-963.
605 Naue J, Horer S, Sanger T, Strobl C, Hatzer-Grubwieser P, Parson W, Lutz-Bonengel S. 2015.
606 Evidence for frequent and tissue-specific sequence heteroplasmy in human
607 mitochondrial DNA. *Mitochondrion* **20**: 82-94.
608 Nicholls TJ, Minczuk M. 2014. In D-loop: 40 years of mitochondrial 7S DNA. *Experimental*
609 *Gerontology* **56**: 175-181.
610 Payne BAI, Wilson IJ, Yu-Wai-Man P, Coxhead J, Deehan D, Horvath R, Taylor RW, Samuels
611 DC, Santibanez-Koref M, Chinnery PF. 2013. Universal heteroplasmy of human
612 mitochondrial DNA. *Human Molecular Genetics* **22**: 384-390.
613 Poovathingal SK, Gruber J, Halliwell B, Gunawan R. 2009. Stochastic drift in mitochondrial
614 DNA point mutations: a novel perspective ex silico. *PLoS computational biology* **5**:
615 e1000572.
616 Poovathingal SK, Gruber J, Lakshmanan L, Halliwell B, Gunawan R. 2012. Is mitochondrial
617 DNA turnover slower than commonly assumed? *Biogerontology* **13**: 557-564.

618 Raap AK, Tafrechi RSJ, van de Rijke FM, Pyle A, Wählby C, Szuhai K, Ravelli RBG, de Coo
619 RFM, Rajasimha HK, Nilsson M et al. 2012. Non-random mtDNA segregation patterns
620 indicate a metastable heteroplasmic segregation unit in m. 3243A> G cybrid cells.
621 *PLoS One* **7**: e52080-e52080.
622 Rebolledo-Jaramillo B, Su MS-W, Stoler N, McElhoe JA, Dickins B, Blankenberg D,
623 Korneliussen TS, Chiaromonte F, Nielsen R, Holland MM et al. 2014. Maternal age
624 effect and severe germ-line bottleneck in the inheritance of human mitochondrial DNA.
625 *Proceedings of the National Academy of Sciences of the United States of America* **111**:
626 15474-15479.
627 Renaud G, Kircher M, Stenzel U, Kelso J. 2013. freeIbis: an efficient basecaller with calibrated
628 quality scores for Illumina sequencers. *Bioinformatics* **29**: 1208-1209.
629 Renaud G, Stenzel U, Kelso J. 2014. leeHom: adaptor trimming and merging for Illumina
630 sequencing reads. *Nucleic Acids Research* **42**: e141.
631 Reva B, Antipin Y, Sander C. 2011. Predicting the functional impact of protein mutations:
632 application to cancer genomics. *Nucleic Acids Res* **39**: e118.
633 Roberti M, Musicco C, Polosa PL, Milella F, Gadaleta MN, Cantatore P. 1998. Multiple protein-
634 binding sites in the TAS-region of human and rat mitochondrial DNA. *Biochemical and*
635 *Biophysical Research Communications* **243**: 36-40.
636 Rogers RS, Bhattacharya J. 2013. When Cells Become Organelle Donors. *Physiology* **28**:
637 414-422.
638 Samuels DC, Li C, Li B, Song Z, Torstenson E, Clay HB, Rokas A, Thornton-Wells TA, Moore
639 JH, Hughes TM et al. 2013. Recurrent Tissue-Specific mtDNA Mutations Are Common
640 in Humans. *Plos Genetics* **9**: e1003929.
641 Sender R, Fuchs S, Milo R. 2016. Revised estimates for the number of human and bacteria
642 cells in the body. *PLoS biology* **14**: e1002533.
643 Soares P, Ermini L, Thomson N, Mormina M, Rito T, Rohl A, Salas A, Oppenheimer S,
644 Macaulay V, Richards MB. 2009. Correcting for Purifying Selection: An Improved
645 Human Mitochondrial Molecular Clock. *American Journal of Human Genetics* **84**: 740-
646 759.
647 Stewart JB, Chinnery PF. 2015. The dynamics of mitochondrial DNA heteroplasmy:
648 implications for human health and disease. *Nat Rev Genet* **16**: 530-542.
649 Stewart JB, Freyer C, Elson JL, Wredenberg A, Cansu Z, Trifunovic A, Larsson N-G. 2008.
650 Strong purifying selection in transmission of mammalian mitochondrial DNA. *PLoS*
651 *biology* **6**: e10-e10.
652 van der Vaart M, Pretorius PJ. 2008. Circulating DNA - Its origin and fluctuation. In *Circulating*
653 *Nucleic Acids in Plasma and Serum V*, Vol 1137 (ed. PB Gahan, R Swaminathan), pp.
654 18-26.

655 Wachsmuth M, Huebner A, Li M, Madea B, Stoneking M. 2016. Age-Related and
656 Heteroplasmy-Related Variation in Human mtDNA Copy Number. *Plos Genetics* **12**:
657 e1005939.
658 Wallace DC, Chalkia D. 2013. Mitochondrial DNA Genetics and the Heteroplasmy Conundrum
659 in Evolution and Disease. *Cold Spring Harbor Perspectives in Biology* **5**: a021220.
660 Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. KaKs_Calculator 2.0: a toolkit incorporating
661 gamma-series methods and sliding window strategies. *Genomics, proteomics &*
662 *bioinformatics* **8**: 77-80.
663 Wang Y, Michikawa Y, Mallidis C, Bai Y, Woodhouse L, Yarasheski KE, Miller CA, Askanas
664 V, Engel WK, Bhasin S et al. 2001. Muscle-specific mutations accumulate with aging
665 in critical human mtDNA control sites for replication. *Proceedings of the National*
666 *Academy of Sciences of the United States of America* **98**: 4022-4027.
667 Wynne HA, Cope LH, Mutch E, Rawlins MD, Woodhouse KW, James OF. 1989. The effect of
668 age upon liver volume and apparent liver blood flow in healthy man. *Hepatology* **9**:
669 297-301.
670 Yao YG, Kajigaya S, Young NS. 2015. Mitochondrial DNA mutations in single human blood
671 cells. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis*
672 **779**: 68-77.

673