Bi-phasic dynamics of the mitochondrial DNA mutation m.3243A>G in blood: An unbiased, mutation level-dependent model implies positive selection in the germline.

Zoe Fleischmann[1], Sarah J. Pickett[2], Melissa Franco[1], Dylan Aidlen[1], Mark Khrapko[1], David Stein[1], Natasha Markuzon[3], Konstantin Popadin[4,5], Maxim Braverman[6], Dori C. Woods[1], Jonathan L. Tilly[1], Doug M. Turnbull[2], and Konstantin Khrapko[1]*

[1]Department of Biology, Northeastern University, Boston, Massachusetts, USA.
[2]Wellcome Centre for Mitochondrial Research and Institute for Translational and Clinical Research, Newcastle University and Newcastle Medical School, Newcastle-upon-Tyne, United Kingdom.
[3]AstraZeneca Pharmaceuticals LP, Waltham, Massachusetts, USA.
[4]School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.
[5]Center for Mitochondrial Functional Genomics, Immanuel Kant Baltic Federal University, Kaliningrad, Russian Federation.
[6]Department of Mathematics, Northeastern University, Boston, Massachusetts, USA.
* Correspondence: k.khrapko@northeastern.edu

Key Words: mitochondrial DNA, mtDNA, mtDNA mutation, m.3243A>G, mitochondrial dynamics

This study was supported by a grant from the U.S. National Institutes of Health (R01-HD091439 to J.L.T., D.C.W. and K.K.), the Russian Academic Excellence Project at the Immanuel Kant Baltic Federal University (K.P.), a grant from the Russian Foundation of Basic Research (18-29-13055 to K.P.). This research was funded in part, by the Wellcome Trust [204709/Z/16/Z to S.J.P. and 203105/Z/16/Z to the Wellcome Centre for Mitochondrial Research]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. We would like to acknowledge those involved in the collection of data for the Grady et al. 2018 (Grady et al., 2018a) and Pickett et al. 2019 (Pickett et al., 2019) manuscripts
.

Abstract

The A-to-G point mutation at position 3243 in the human mitochondrial genome (m.3243A>G) is the most common pathogenic mtDNA variant responsible for disease in humans. It is widely accepted that m.3243A>G levels decrease in blood with age, and correction representing ~2% annual decline is often applied to account for this change in mutation level. Here we report that recent data indicate the dynamics of m.3243A>G are far more complex and depend on the blood mutation level in a bi-phasic way. As a consequence, the traditional 2% correction, which is adequate 'on average', creates opposite predictive biases at high and low mutation levels. Thus, overall accuracy of traditional correction depends on the proportion of individuals with high and low mutant levels in the dataset. Unbiased age correction is needed to circumvent these drawbacks of the standard model. We propose to abolish both biases by using an approach where correction depends on mutation level in biphasic way, to account for the biphasic dynamics of m.3243A>G in blood. The significance of removing bias was further tested using germline selection as a model, in which we detected mutation patterns consistent with the possibility of positive selection for m.3243A>G. We conclude that use of bi-phasic approach will greatly improve the predictive accuracy of modelling data for changes in mtDNA mutations in the germline and in somatic cells during aging.

Introduction

Pathogenic variants in the mitochondrial genome are responsible for a wide range of diseases that affect mitochondrial function (Gorman et al., 2016). The multi-copy nature of mitochondrial DNA (mtDNA) means that it is possible for more than one species of mtDNA to co-exist within the same cell, termed heteroplasmy. By far the most common heteroplasmic mtDNA mutation is an A to G transition at position 3243 (m.3243A>G) within *MT-TL1*, which encodes mitochondrial $tRNA^{Leu(UUR)}$ (Goto et al., 1990). Estimates of m.3243A>G carrier frequency range from 140 to 250 people per 100,000 (Manwaring et al., 2007), (Elliott et al., 2008), although the point prevalence for adult disease is much lower than this, at 3.5 per 100,000 (Gorman et al., 2015), suggesting that many carriers are either asymptomatic or have mild, undiagnosed symptoms.

Originally identified within a cohort of patients presenting with a severe syndrome characterized by mitochondrial encephalopathy, lactic acidosis and stroke-like episodes (MELAS), m.3243A>G is associated with extremely varied clinical presentations. Patients can experience a variety phenotypes including ataxia, diabetes, deafness, ptosis, chronic progressive ophthalmoplegia, cardiomyopathy, cognitive dysfunction and severe psychiatric manifestations (Koga et al., 2000), (de Laat et al., 2012), (Nesbitt et al., 2013), (Mancuso et al., 2014), (Fayssoil et al., 2017), (Pickett et al., 2018). Disease burden can be partly explained by an individual's m.3243A>G mutation level, but this relationship is not simple; other factors are likely to also play a role (Pickett et al., 2018), (Grady et al., 2018b), (Boggan et al., 2019). These complexities make offering prognostic advice to patients very difficult and, coupled with its high frequency, means that m.3243A>G-related disease is one of the biggest challenges in the mitochondrial disease clinic.

Although levels of m.3243A>G are relatively stable in post-mitotic tissues such as muscle, there is strong evidence of negative selection against the G allele in mitotic tissues, even in relatively asymptomatic patients; this loss of m.3243A>G in mitotic tissues has been studied in detail in blood (de Laat et al., 2012), (Grady et al., 2018b), (Sue et al., 1998), (Pyle et al., 2007), (Rajasimha et al., 2008) , (Mehrazin et al., 2009), (Langdahl et al., 2018). Initial simulation studies suggested that this decline could be exponential (Rajasimha et al., 2008), although more recent studies using larger amounts of data point to a more complex process (Grady et al., 2018b), (Veitia, 2018). To study the dynamics of this mutation, we took advantage of the large quantity of longitudinal data that are available for m.3243A>G levels in blood and have developed a new, empirical model that better describes the dichotomous pattern that we observe.

Furthermore, the m.3243A>G variant is maternally inherited and, similar to other heteroplasmic mtDNA mutations, undergoes a genetic bottleneck in development leading to offspring often with very different levels of m.3243A>G than their mothers (Chinnery et al., 2000) (Pickett et al., 2019). Different mtDNA mutations segregate at different rates, demonstrating that the dynamics of this bottleneck are dependent on the mtDNA variant being transmitted (Wilson et al., 2016). We postulated that studying this bottleneck may help us understand why the m.3243A>G variant is so common in the population; using our new model of the dynamics of blood heteroplasmy/mutation level, we explored whether there is a transmission bias for m.3243A>G between mother and child.

Results

Longitudinal changes of m.3243A>G levels in blood with age reveal a dichotomous pattern (Fig. 1).

A large amount of data on the changes of the level of m.3243A>G mutation in blood with the age of the individual have been compiled in a recent report (Grady et al., 2018b). We have reproduced these data in Fig. 1A. Traditionally the dynamics of m.3243A>G mutation in blood is considered to follow an exponential decay with age of approximately 2% per year. The data in Fig. 1A, however, appears visually dichotomous, i.e., at higher m.3243A>GA>G levels, we see more cases where the mutation level decreases (red), whereas at lower levels is little evidence of decline and more cases with an increase or stable levels (blue). The dichotomy can be further revealed by plotting the ratio of the numbers of the ascending and the descending mutation cases to the total number of cases above and below a variable threshold of mutation level, correspondingly. principle of this "ratio curve" approach is to estimate the probability for a case to progress to a lower or higher mutation level depending on the initial mutation level.

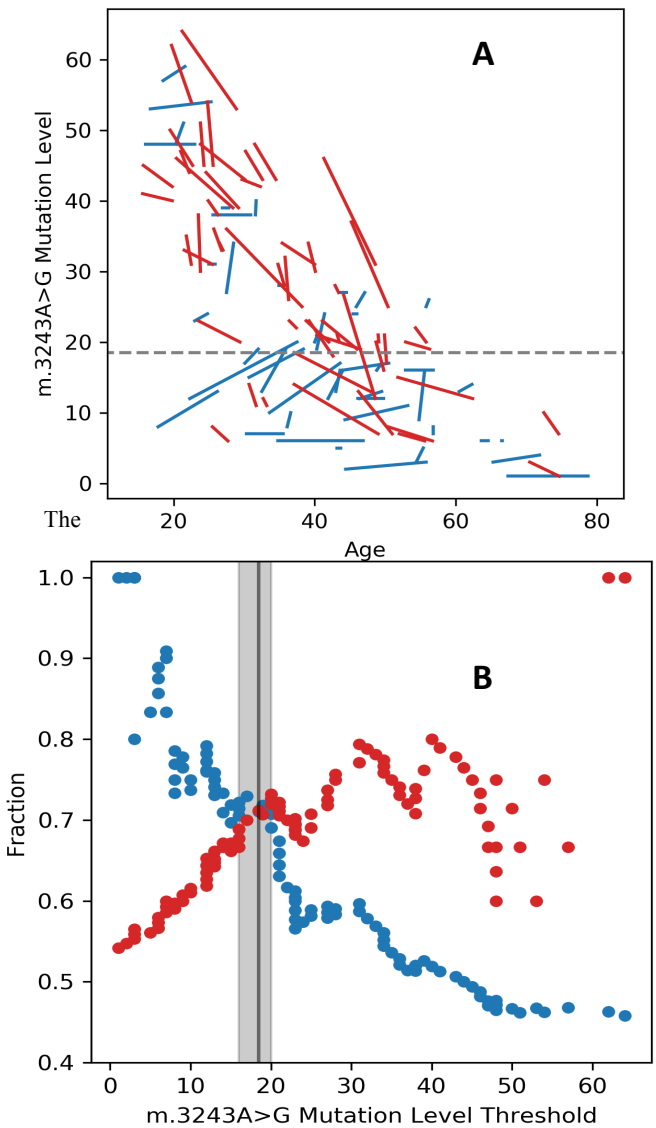

Fig. 1. Analysis of longitudinal blood m.3243A>G levels. (**A**) Each segment represents the change in m.3243A>G level in blood for a single individual over a longitudinal period of follow-up. Red and blue lines are cases of age-related increased or decreased/stable m.3243A>G levels, respectively. (**B**) Red: Proportion of descending cases with initial mutation levels above the indicated threshold. Blue: Proportion of ascending cases with initial mutation levels above the threshold. Shaded region: vicinity of the crossover point.

For more details on the ratio curves see Materials and Methods and Supplementary Fig. S1. Fig. 1B shows that while individuals with higher m.3243A>G levels tend to, in accord with the conventional model, decrease their levels with age, individuals with lo levels, in contradiction to the convention, tend to increase or stabilize mutation level. The crossover point (a mutation level of ~15-20%) on the graph represents the "point of the best discrimination", and also the m.3243A>G level where cases on average do not change their mutation level with age. Monte-Carlo simulations demonstrate that this distribution is significantly non-random (p<0.0001).

Mother-child m.3243A>G levels also show a dichotomous pattern

The dichotomous pattern of m.3243A>G dynamics in blood is further supported by the analysis of independent, though related, datasets. We note that in the mother-child pairs dataset that inheritance of m.3243A>G between mothers and children show a similar dichotomous pattern (Fig. 3A) as the dynamics in blood with age (Fig. 2A). Indeed, in the high child mutation level range (above ~10%) the mother/child relationships are mostly strongly descending. In contrast, in the low child mutation level region (below ~10%), the ascending child>mother pattern is prevailing. In accordance with this visual appearance, the "ratio curve" approach shows, in the mother-child data, the same cross-over pattern as the longitudinal data (Fig. 2B). This cross-over is also highly significant (p<0.0001, Monte-Carlo simulations). We conclude that both the longitudinal and the mother/child datasets each consist of two "domains" (high and low mutation level) with opposite dynamics of age-related change in the m.3243A>G mutation levels. Increase of mutational level (positive rate) prevails in the low mutation level "domain" and decrease (negative rate) prevails in the high mutation level domain. In conclusion, the rate and the direction of change of m.3243A>G level in blood appears to depend on the mutation level.

2% annual decline model biased at high and low mutational levels

The observed dichotomy of m.3243A>G (Figs 1 and 2) implies that, because the conventional 2% annual decline model cannot account for two opposite processes, it must be making biased predictions among individuals with low or high levels of m.3243A>G mutation, or both. To test this hypothesis, we split the longitudinal dataset into the low and the high-level subsets and evaluated the 2% annual decline model for bias in each of the two subsets. We noted, however, that while analysis presented in Figs 2 and 3 does demonstrate the existence of dichotomy in mutation dynamics in blood, it does not permit to precisely determine the threshold separating the two subsets with different mutation dynamics. To address this, we use 4 different thresholds (10, 15, 20, and 30%) arbitrarily chosen in the vicinity of what visually appears as a region separating the two domains (~15-20% mutation, shaded area in Fig. 3B) to generate 8 subsets of the longitudinal dataset were generated, 4 above the thresholds ('low mutant level subsets') and 4 above the thresholds ('high mutant level subsets'). Using multiple thresholds excludes the possibility of a spurious result that stems from the selection of a particular threshold value. We then used a 2% annual decline model (Rajasimha et al., 2008) to predict, for each individual in the longitudinal dataset, mutant levels at their last measurement given mutant levels at the first measurement and the age difference between measurements. More specifically, we used the conventional annual 2% decline model (Rajasimha et al., 2008):

$$MF2_{pred}=MF1 \times (1+R)^{\Delta A} \qquad (1)$$

where R is the rate of change on mutant fraction per year, *MF1 is* the *actual* mutation level (MF stands for Mutant Fraction) at the first measurement and $MF2_{pred}$ is the *predicted* level the last measurement, and ΔA is the difference in age between the two measurements. We then calculated the 'error ratios' $MF2_{pred}/MF2$ of these individual predictions and calculated geometric average of ratios for each given subset (see also Fig. 3A caption). We then subtracted 1 from the averages to obtain

"average relative error," which is a fair measure of the bias of the model in the given subset. The result is shown in Fig. 3A. As shown in forward predictions, the standard (2% annual decline) model underestimates (negative average error) in the low mutant domain (blue) and overestimates (positive average error) in the high fraction domain (red). This implies that the m.3243A>G decline rate of 2% per year is too slow for the high fraction individuals, in whom mutations apparently decline on average faster than 2%, and too fast for the low fraction domain, where m.3243A>G level is more stable or actually increases with age on average.

Bi-phasic models alleviate biases of the single rate 2% annual decline model

To eliminate the biases of the 2% annual decline model, we propose to use a bi-phasic model, which reflects the dichotomous pattern of mutation decline with age (Fig. 3). This approach is dependent on the specification of one parameter: a threshold level of mutation to separate the dataset into high and low mutation level subsets; the 2 rates of annual decline of mutation level are then determined unambiguously by the requirement of the model being unbiased. We used the longitudinal dataset as a 'training set' to build a model as this is the most inclusive dataset available that reports the actual decline on the mutation levels in blood with age. As far as separating threshold is concerned, we used the same thresholds and generated the same eight subsets as those used for testing the conventional 2% annual decline model described in the previous section (Fig. 3A and 3B). For each of the eight subsets, we determined the 'unbiased' rate of decline. To do so, we predicted mutation level at the second measurement time point for each individual in the dataset as previously, using equation (1), except the rate R was not fixed at -2%, but scanned in steps of 0.1 % from -5% to 5% to pinpoint the rate at which the average bias of the prediction across the given subset was zero (Materials and Methods; Fig. 1). Reassuringly, the 'unbiased' model for each subset was also close to the best fit model, meaning that the sum of squared errors was close to a minimum (see Supplementary Fig. S2 for a full set of graphs).

As expected, the unbiased rates (Fig. 3B) look like a "mirror image" of the average relative error bars (Fig. 3A). Note that the 'line of reflection' in Fig. 3B is not the X-axis, but the dotted line at -0.02, which is because the 2% annual decline model (a.k.a. -0.02) is the 'reference' in this case. Interestingly, unbiased rates tend to be positive (i.e., mutant fraction increases in the blood with age) in individuals with a low mutant fraction, and negative in individuals with a high mutant fraction. This distinction remains relatively stable as the threshold separating the high and the low mutant fraction subsets is varied from 10 to 30%. The fair consistency of the neutral decline rates in subsets that lie predominately within the high (>15%, '>20%','>30%') or within the low mutation level domain ('<10%', '<15%', '<20%') imply that the unbiased decline rates are innate characteristics of each domain. We therefore conclude that the high and the low mutation level domains should be analysed differently/separately. This is particularly true as the unbiased decline rates in the two domains have opposite signs, which are likely to compensate each other and obviate the details of the dynamics of m.3243A>G if they are (incorrectly) treated jointly.

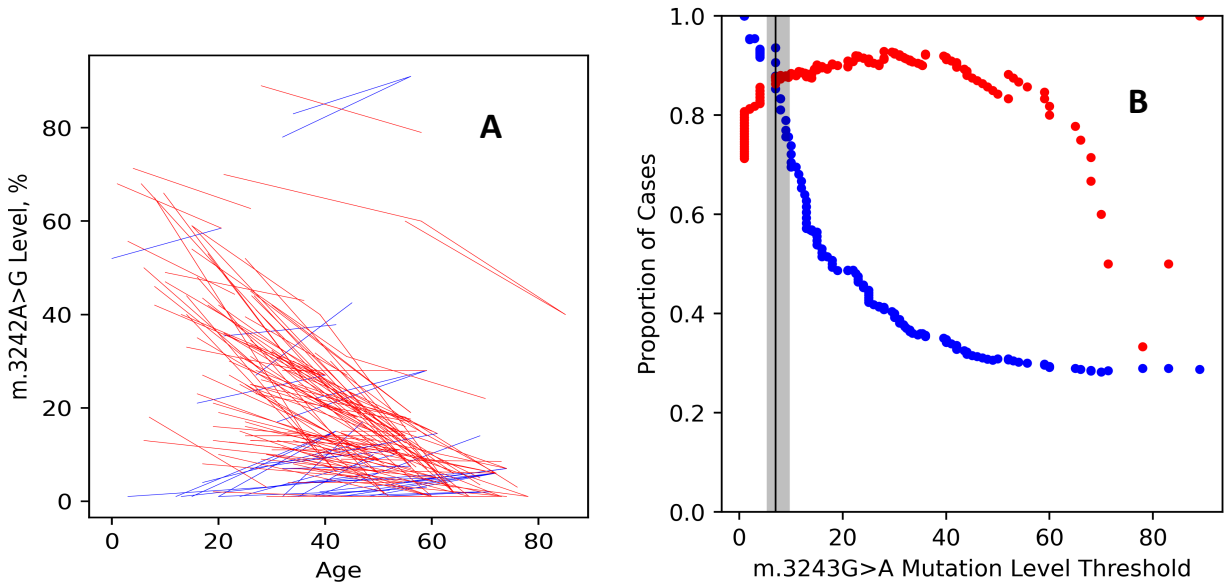

Fig. 2: Analysis of mother-child m.3243A>G levels in blood. (**A**) Every segment connects the mutant m.3243A>G level of a child (left end of the segment) to the mutant level of their mothers (right end). Descending segments are coded red, ascending - blue. (**B**) Ascent/descent ratio curve for the mother-child m.3243A>G levels. Plotted are proportions, (**A**), of descending (red) and ascending (blue) segments with child's mutation level above and below, respectively, the indicated mutation level threshold (these are merely arbitrary thresholds used for partial analysis, not to be confused with the fundamental 'phenotypic threshold' of mtDNA mutation). Note that the exact position of crossover point (i.e., ~10% here vs. ~20% in Fig. 2b) depends on several factors including the balance between low and high mutant level, so the position of the point should not be interpreted as a fundamental mutational level separating the domains of ascending and descending dynamics; it merely proves that such domains do exist.

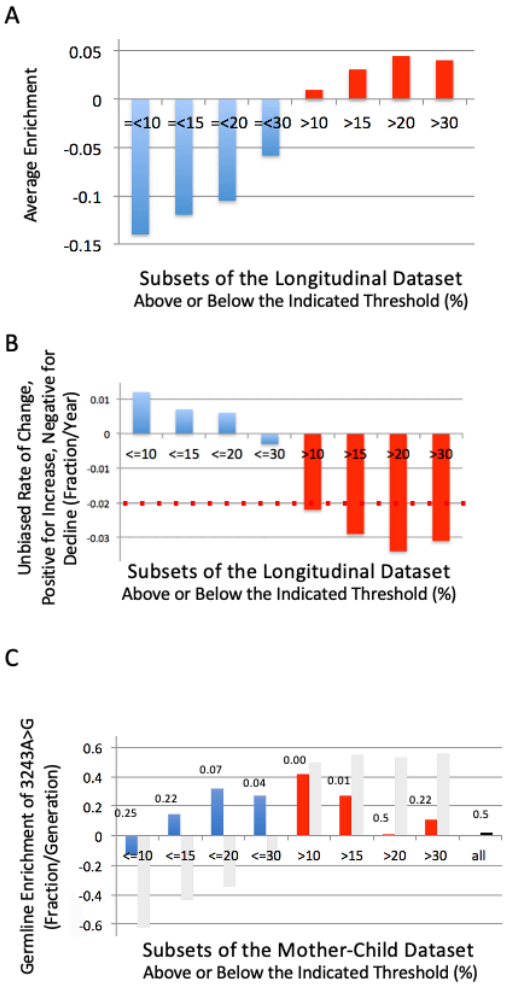

Fig 3: (**A**) The standard 2% annual decline model, is negatively biased at low mutation levels and positively biased at high mutation levels. To show the existence and consistent dependence of bias on mutation levels, four high-mutation and four low-mutation subsets of the longitudinal dataset were created by splitting the dataset into two subsets at four arbitrary mutation level thresholds (10, 15, 20, and 30%; at first measurement; the subset sizes are as follows: <10%,18; <15%, 31; <20%, 39; < 30%, 61; >10%, 78; >15%, 65; >20%, 57; > 30%, 35; all, 96).). The 2% annual decline model was used to predict the level of m.3243A>G in each individual at last measurement based on the first measurement and age difference. The average *relative errors* within each subset were calculated (for details see Results and Discussion) and were plotted as blue bars (low mutation subsets) or red bars (high mutation level subsets). (**B**) Subset-specific 'unbiased' mutation decline rates that neutralize biases. 'Unbiased' rates were determined for each of 8 subsets described in (a), as the rate of decline R (equation (1)) such that error of prediction of the resulting 'unbiased' model averaged within the subset was zero. Blue bars – low mutation level subsets, red bars – high mutation level subsets. Red dotted line represents the conventional -0.02 (2% annual decline) rate. (**C**) 'Unbiased' models reveal positive germline selection. To estimate 'unbiased' enrichment of m.3243A>G per generation due to germline selection, mother–child dataset was split into 8 overlapping subsets in the same way as longitudinal dataset in panel a. The unbiased rates derived from the 8 subsets of the longitudinal dataset shown in panel b were used to predict child's mutational level at mother's age and the ratio of adjusted child's to mother's mutation level was considered estimate of germline selection. Bars represent enrichments (i.e., geometric averages of child/mother ratios within each of the 8 subsets minus 1). Blue bars represent low mutation level subsets, red – high mutation level subsets. Black bar ('All') represents enrichment (i.e., essentially lack thereof) estimated based in the standard -0.02 decline rate within the entire mother-child dataset, which is consistent with the previous studies. Grey bars represent germline selection predicted if the conventional 2% decline was used in each of the 8 subsets instead of the specific unbiased decline rates. Numbers above the bars denote level of significance of being positive (one tailed p-value). See Materials and Methods for details of calculations.

Estimating germline selection using the bi-phasic model.

To determine whether the bias we observe affected estimates of germline selection, and if so, to obtain adjusted estimates, we split the mother-child dataset 4-ways into eight subsets, mirroring our methodology for the longitudinal dataset. We then used, for each of the subsets, the specific unbiased decline rates devised from the longitudinal data (Fig. 3B), to factor out the mother-child age differences within the corresponding subsets of the mother-child dataset. Children's mutational levels were projected to mothers' age using the 'unbiased' rate of m.3243A>G decline and divided by the mothers' mutation levels. Of note, such enrichments in individual mother-child pairs are subject to high variance resulting in part from the action of the mtDNA bottleneck, so an average of enrichments across the subset best characterizes the selection in the germline. The details of the calculations for these estimates are presented in Materials and Methods. Germline enrichments tend to be positive in both the low and high mutant fraction subsets (Fig. 3C; color bars); four of the eight subsets show significant enrichment ($p<0.05$). The apparent variability of estimates is, in part, related to the small size of the dataset which exacerbates the influence of outlier datapoints. This especially affects the leftmost (<=10) and rightmost (>30) subsets, which contain the least datapoints (numbers of datapoints are in Figure 3A legend).

For direct comparison to the standard model, we also calculated the expected germline selection for each of the subsets, and for the entire mother-child dataset, under assumption of the uniform decline rate of 2% per year. For the entire dataset, this produces a germline selection estimate which is close to zero (Fig. 3C; black bar "All"). The subset-based analysis (Fig. 3C; grey bars), unlike our bi-phasic approach which tends to predict weak positive selection estimates in most subsets, clearly predicts strong negative selection in the low mutation level domain, and strong positive selection in the high mutation level domain. The explanation for this paradoxical behaviour of the standard approach is as follows: when a flat -2% rate is being used for age adjustment of the low mutation level subsets the mother-child pairs are being adjusted with a very incorrect mutation decline rate (2%/year, i.e., *decrease* instead of the unbiased 2%/year *increase* as in our calculations). This excessively negative rate results in excessive negative (instead of the needed positive) adjustment of the child's mutation level projection to the mother's age, which results, in a very negative germline selection estimate. In contrast, in the high mutant fraction subsets, the use of decline rate of 2%/year is not sufficient to compensate for the even more rapid, and real, decline in these subsets. By our estimates, decline rates in high mutant subsets are about 3%/year, (average of the red bars in Fig. 3C). Thus, insufficient negative adjustment leaves child's levels too high, which results in an overestimate of positive germline selection.

Discussion

Dynamics of m.3243A>G in blood is dichotomous.

Negative selection of the m.3243A>G pathogenic variant in human blood is well established; previous attempts to model this decline with age have pointed towards an exponential or sigmoidal process but the dynamics of m.3243A>G are more complex and these models do not fully-explain the data (Rajasimha et al., 2008) (Grady et al., 2018b), (Veitia, 2018). Using a large quantity of recently compiled longitudinal data (Grady et al., 2018b), we report that the dynamics of m.3243A>G in blood are dichotomous; levels predominantly decline in individuals with high mutation levels and are predominantly stable or increasing at low mutation levels. As a consequence, the standard 2% annual decline correction, which is adequate on average, creates bias both at high and low heteroplasmy and its predictions depend on the proportion of individuals with high and low mutant levels in the dataset. Interestingly, we detected a similar dichotomy in a large dataset of blood m.3243A>G levels in mother-child pairs, providing further support for a model in which the dynamics of m.3243A>G decline are dependent upon mutation level. Unbiased age correction is needed to circumvent the drawbacks of the 2% annual decline model; we used our observations to develop a new, empirical model of the decline of m.3243A>G in blood which better accounts for this dichotomy. We then used this unbiased age correction to explore the transmission of m.3243A>G from mother to child, detecting patterns consistent with positive germline selection of this variant, which may be a contributing factor in the comparatively high frequency of m.3243A>G compared to other pathogenic mtDNA point variations (Gorman et al., 2015).

Dichotomous dynamics cannot be described by a single decline rate, therefore, any model that does not account for this dichotomy will be biased. To reveal this, we tested a 2% annual decline model as a predictor of subsequent measurements of mutation level based on preceding measurements and the age difference in different subsets of the longitudinal dataset. Average error is positive in the subsets of individuals with high mutation levels and negative in low mutation level individuals, confirming that the 2% annual decline model is biased. We note however, consistent with previous reports (Wilson et al., 2016), (Otten et al., 2018), that it appears unbiased overall; the average error of its predictions across the entire longitudinal dataset is very small and not significantly different from zero. This can be explained by the neutralizing effect of the two partial biases, which occurs when averaging across both the high and low mutation subsets of the data. Interestingly, this means that overall bias of the 2% decline model is not universal – it depends on the composition of the dataset, i.e., how many individuals are in the high and in the low mutation groups.

Bi-phasic model – an unbiased approach to dichotomous dynamics of m.3243A>G.

Adjustment of the blood levels of m.3243A>G in mother and child for the age difference using the 2% annual decline model has been shown to have a dramatic effect on the estimates of germline selection of the m.3243A>G mutation as compared to unadjusted estimates (Rajasimha et al., 2008). We therefore reasoned that this bias, would significantly affect current estimates of selection as well. We therefore sought to modify the 2% model to eliminate this bias while keeping the elegant framework of the 2% model maximally unchanged. We chose to use a bi-phasic model to reflect the dichotomy of the m.3243A>G dynamics. In the interest of keeping the number of parameters in the model to a minimum, we made the simplest assumption that there are two domains, i.e., the high and the low mutation level domains each with their specific exponential dynamics. It should, however, be noted that individual variability decline rate still exists within these domains. We postulate that this could arise from differences in the threshold for biochemical expression of the variant allele between individuals, thus affecting negative selection against cells with high levels of the variant allele. This theory is compatible with the high level of variability in disease burden and phenotypic spectrum that is seen in individuals carrying m.3243A>G (Nesbitt et al., 2013), (Mancuso et al., 2014), (Kaufmann et al., 2011), (de Laat et al., 2012), (Chin et al., 2014), (Pickett et al., 2018) (Grady et al., 2018b).

The development of this model revealed important features of the dynamics of the mutation in blood. Contrary to the conventional view that blood m.3243A>G levels universally decrease with age, in the blood of individuals with low mutation levels, m.3243A>G levels tend to stabilise and even increase. Conversely, unbiased rates in the high mutation level individuals are actually more aggressively negative than conventionally accepted decline rate of 2%. Thus, the perceived overall 2% decline rate of the standard model probably stems from *de facto* "averaging" of the rates at high and low mutational levels, which has been possible to detect due to the recent availability of larger, longitudinal datasets.

The similarity that we see between the longitudinal and mother-child datasets is reassuring but not unexpected; m.3243A>G levels in child and mother can be viewed as two sequential samplings of the same germline. By 'sampling', we mean that the continuous germline gave rise to somatic tissue (blood cells in our case) in the mother and then in the child, each of which were used to infer mutation level of the germline. These two samplings, however, systematically differ in that in the child the mutation spends fewer years in the blood cells than in the mother. Similarly, in the longitudinal dataset, every pair of sequential measurements represent two samplings of the germline. The difference between the two being that in the mother-child dataset the two samplings are in different generations and in the longitudinal dataset they are from one. The only consistent difference expected between these two datasets would be related to germline selection between generations in the mother-child dataset, if we admit that germline selection is nonzero. Interestingly, the crossover point in the mother-child dataset appears to be shifted to the right, an effect expected if germline selection were positive. This is because positive selection makes the slopes of the child-to-mother segments more negative than expected if there were no selection. It is difficult to estimate the significance of this observation, so it should not be considered an argument in favour of positive selection. The presence of the dichotomous pattern in the mother-child dataset is an indication that this pattern is a general feature of m.3243A>G dynamics in blood.

Unbiased bi-phasic approach reveals positive selection in the germline.

Germline selection can be estimated by comparing mutation levels in children and mothers. However, because of the systematic changes in the blood levels of m.3243A>G with age (Rajasimha et al., 2008), the relative levels in blood in mothers and children must be adjusted for the age difference. Previously, this has been achieved by applying the annual decline model of ~2% per year (Wilson et al., 2016). We sought to explore the transmission dynamics of m.3243A>G in the context of the bi-phasic model. To address this question, we divided mother-child pairs into high and low mutation level subsets and used the subset-specific unbiased rates as the best empirical means to correct for the mother-child age difference needed for estimation on intergenerational germline selection. We show that, although the unbiased m.3243A>G decline rates are opposite in the high and low mutant fraction subsets, germline enrichment appears to be positive in both. We do note a variability in the estimates; obtaining more precise estimates of the magnitude of this apparent germline selection will require even larger and more balanced datasets.

Bi-Phasic vs. 2% annual decline approach: independence of the mutation level.

To put these results in the context of previous studies, we performed additional analysis to explore the sources of differences between our estimates and those reported previously, which found no evidence of positive selection (Otten et al., 2018), (Wilson et al., 2016). First, we directly compared our approach to the 2% annual decline model by applying this correction to exactly the same mother-child dataset that we used for the bi-phasic estimates, to exclude the possibility that differences are caused by differences in the dataset. Reassuringly, in agreement with previous studies, the predicted overall germline selection when calculated over the entire mother-child dataset using the standard 2% decline rate is very low and not significantly different from zero. Most notably, unlike our bi-phasic approach, which tends to predict weak positive selection estimates in most subsets, the 2% annual decline approach clearly predicts strong negative selection in the low mutation level domain, and strong positive selection in the high mutation level domain, an effect that we believe is due to insufficient adjustment in the high m.3243A>G level domain and excessive adjustment in the low m.3243A>G level domain.

This illustrates an important drawback of the standard model compared to the bi-phasic approach. Estimates based on the 2% annual decline model are highly sensitive to the composition of the mother-child dataset. Indeed, from Fig 3C (grey bars) one can deduce that the higher the proportion of low mutation level mother-child pairs included in the dataset, the lower the expected overall estimate of the germline selection using 2% adjustment. In fact, the near zero estimate of selection in the current mother-child dataset is merely the result of the particular proportion of low and high mutation level pairs that happened to be in this dataset. The bi-phasic model is poised to deliver much higher stability of the estimate with respect to the changing proportion of high and low mutation level patients.

Comparison to other mutations.

Positive germline selection of the m.3243A>G has not been reported previously, but has been observed for other detrimental mtDNA variants. An apparently milder variant, m.8993T>G, may be under positive germline selection throughout the mutant fraction range (Otten et al., 2018)). Interestingly, our re-analysis of the data from Freyer et al (2012) (Freyer et al., 2012) shows that a specific highly detrimental mutation, the m.3875delC in the tRNAMet in the mouse, at an intermediate mutant fraction (45-60%) systematically increases its presence in offspring compared to the mother, implying positive selection of a detrimental mutation in the mouse germline, which, however, is is reversed at higher (>60%) levels). Of note, no data is available for mutational levels below 45%, so it is possible that selection decreases at lower mutant fraction as we observe for 3243 (Fig 3C). Caution should be exercised when comparing our results to those of Freyer et al. Not only there may be mouse/human differences, but also Freyer measured mutation loads directly in the germline, while we have to infer them from the levels in blood. As shown by some of us, average blood 3243 levels at any age are lower than those of the germline, at least by 20% (Grady et al., 2018a). Thus percentages in Fig 3C should be corrected upwards to represent levels of mutation in the germline, though more precise estimates will require more data and more research. In conclusion, it appears that detrimental mutations, at least in some cases, are under positive selection in the germline. It would be interesting to determine how general such a phenomenon is.

Potential applications of the bi-phasic approach.

From a practical angle of view, this study will hopefully lead to models that better account for the dynamics of pathogenic mtDNA variation by including the effect of mutation level. The simple bi-phasic model proposed hereby may serve as a practical alternative to the 2% annual decline model in cases where, like in studies of germline selection, unbiased prediction and versatility of the model (applicability to datasets that are unbalanced with respect to cases with low and/or high mutational levels) is essential. This can be realized simply by applying one of the two rates: ~3% annual *decline* for cases with mutation levels above 20% and ~1% annual *increase* for cases with levels below 20%. Of note, in this study we used the mutation level at first measurement, or mutation level of the child (which is analogous to the first measurement). We have chosen this convention because the dynamics of the mutation converges with time (phase lines become denser with time), so, in general, prediction of the succeeding mutation level is more stable than prediction of the preceding mutation level. As of now, this bi-phasic model is not a finished working instrument for clinicians to predict mutational levels in individual patients. Rather, this research reveals that dynamics of m.3243A>G in blood is more complex than previously considered and highlights the need for more detailed correction methods, especially in the lower range of mutation levels. With more data available for models to build upon and test, our approach may be further optimized or replaced by a more detailed and precise model. Our current model, however, is appropriate to make conclusions about average behaviour of mutations, such as positive selection, with the data currently at hand.

Materials and Methods

Data sources:

Longitudinal data

Longitudinal measurements of m.3243A>G levels in blood were obtained from Grady et al. (Grady et al., 2018b). This dataset is comprised of 96 individuals who were recruited into the Mitochondrial Disease Patient Cohort UK and had two or more measurements of m.3243A>G blood level taken between 2000 and 2017. The median time between first and last measurement was 2.5 years (IQR = 4.35, range = 0.01 – 15.20), the median age at first measurement was 36.25 years (IQR = 20.58, range = 15.60 – 72.50) and at last measurement was 40.25 years (IQR = 21.00, range = 19.80 – 78.80).

Mother-child data

Measurements of m.3243A>G levels in mother-child pairs were obtained from Pickett et al. (Pickett et al., 2019). This dataset contains 183 mother-child pairs (from 113 different mothers), comprised of 67 pairs from the Mitochondrial Disease Patient Cohort UK (Nesbitt et al., 2013) and 116 pairs obtained from a literature search and previously published by Wilson and colleagues (Wilson et al., 2016). To minimize ascertainment bias, pairs in which the child was the proband had previously been removed, as had one pair from the literature where the m.3243A>G variant was thought to have arisen *de novo* in the child (Ko et al., 2001). We identified an additional 30 pairs where the mother was the proband; to reduce bias these were also excluded, leaving a total of 153 pairs (from 97 mothers) that were taken forward into the analysis. The mean age at m.3243A>G level assessment was 53.9 years (SD = 12.4, range = 23.0, 85.0) for the mothers and 27.8 years (SD = 12.2, range = 0.4, 58.0) for the children.

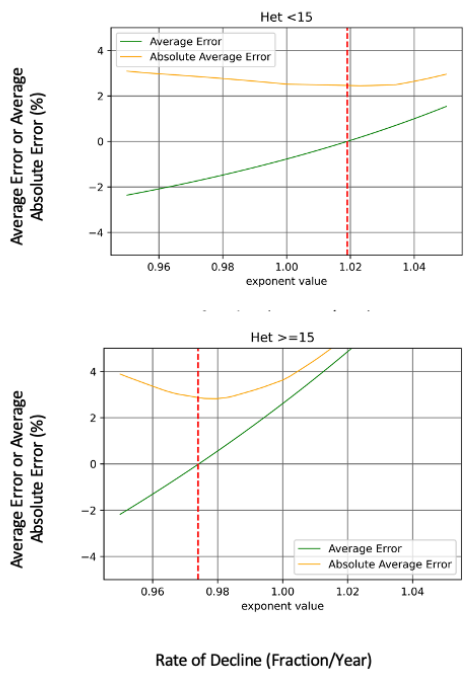

Fig. 4: Example of the calculation of an unbiased rate of decline for the two subsets: <15% (top) >=15% (bottom). See **Supplementary Fig. S2** for a complete set of curves. Red vertical lines indicate the unbiased rate.

Construction of the "ratio curves" and demonstration of the cross-over point significance

To estimate the probability for a case to progress to a lower or higher mutation level at the next mutation measurement depending on the initial mutation level we plotted, for each mutation level at the first measurement X (X indicates any value one can encounter in the dataset), the ratio of all descending cases with initial mutational level above X to the total number of cases with initial mutational level above X. Plotting these ratios for every mutational level from 0 to 70 resulted in the red curve in for 1B. Similarly, the blue curve has been plotted using ratios of ascending cases with initial heteroplasmy below X for all cases with initial mutational level below X. This has been done for the longitudinal dataset. For the mother-child dataset the same processing was done, except instead of initial mutational level we used child's mutational level and instead of the final mutational level we used mother's level. The calculations used to construct the "ratio curves" are further illustrated in **Supplementary Fig. 1S**.

To determine whether biphasic behavior that is visually apparent from ratio curves is indeed statistically significant, we performed a Monte-Carlo style simulation. Our null hypothesis was that there was no dichotomy and instead the pattern emerged by pure chance. To eliminate any potential systematic dichotomy, we randomized the longitudinal and the mother-child datasets, i.e., randomly assigned the segments as being "decreasing" or "increasing". The proportion of increasing and decreasing segments was the same as in the actual dataset. We then calculated the simulated/randomized data in the same way as the actual data. The simulated dataset was considered "equally or more extreme" for the p-value purposes, if the curves were separable for more points (moving from the left) than the original dataset (38 and 36 datapoints, correspondingly). "Separable" here and above means that the difference between the curves' points continuously remained positive or negative. The process was repeated 10,000 times and no equally or more extreme cases were discovered. Note that "separability" of the curves from the right is a natural consequence of descending segments being more prevalent in this dataset than the ascending ones, therefore "separability from the right" was not taken into account.

Determination of the unbiased mutation decline rates

The approach we used to calculate the unbiased mutation decline rates in the various subsets of the longitudinal dataset (Grady et al., 2018b) is illustrated in Fig. 4. First, we limited analysis to the first and the last measurement for each person, so that there were only two measurements for each person in the dataset, mf1 and mf2 ('mf ', for mutant fraction). For every individual (of 96 individuals in the dataset) and thus a pair of data points (mf1, mf2), the mf2 was predicted based on mf1 using an exponential model with variable parameter R, the fractional decline of mutational load per year (R is positive for an increase, and negative for decline). R is called here the "rate of decline" –

$$mf2_{pred}(R) = mf1 \times (1+R)^{\Delta A},$$

where ΔA is the age difference between the times of mutation level measurements. R was varied between 0.05 decrease to 0.05

increase per year (i.e., from -0.05 to 0.05) as shown in Fig. 4 in the steps of 0.001, 100 steps overall. Thus, we obtained 100 values of $mf2_{pred}$ (R) for each R, for each individual (i.e., 100x96 total). For each $mf2_{pred}$ (R), an error ratio ($mf2_{pred}$ (R)/mf2) and absolute error (sqrt[(mf2pred/mf2)]^2) was calculated and then both error and absolute error were averaged among individuals within each of 8 subsets. We obtained 100x2x8 of the averaged data points and plotted them for each value of R in 8 graphs shown in Supplementary Fig. 2 (one graph per subset). The minimum of the absolute error (corresponding to the best fit rate) and the zero of the average non-absolute error ("unbiased rate") were determined graphically. Unbiased rates were then used in for plotting. The best fit rates were used to make sure that the unbiased model was close to the best fit model.

Calculation of the enrichments/selection of m.3243A>G per generation in the germline

The enrichments/selection per generation for a specific subset of the data S is determined as geometric average of ratios of age-corrected child mutation levels to those of their mothers. In practice, we calculated the average of logarithms of the ratios and then converted the average back from logarithm to real ratio/geometric mean. That is, we first determine the germline selection for each child-mother pair within each subset using Equation 2 –

$$MF_{child} \times (1+R_S)^{\Delta A} / MF_{mother} \quad (2)$$

where R_s is the estimated unbiased rate for the subset S to which the mother-child pair has been assigned (as determined by the value of MF_{child}) and both MF_{child} and MF_{mother} are expressed as proportions between 0 and 1. ΔA is the age difference between mother and child.

For example, for the 20% threshold, a child with a m.3243A>G mutation level of 40% would be assigned to the '>20%' subset with an R_s of -0.029 (see Figure 3A). If their mother was 30 years older and had a m.3243A>G level of 10%, the germline selection estimate in that individual mother/child pair would be: $0.4 \times 0.971^{30} / 0.1 = 1.65$

We then averaged log ratios across all mother-child pairs in the given data subset and used one-sample one tail t-test to determine the p-value associated with this average of logarithms being different from 0 (the null hypothesis of no selection). Finally, the mean log ratios for the various data subsets were converted into mean mutation level ratios and 1 was subtracted from average ratios to produce "enrichment" (i.e., fractional increase, positive for enrichment, negative for depletion, analogous to the rate of decline used describe the dynamics of m.3243A>G in blood).

Conflict of Interest Statement

The authors declare no conflict of interest.

Abbreviations:

MF: Mutant fraction

References:

Boggan, R.M., Lim, A., Taylor, R.W., McFarland, R., and Pickett, S.J. (2019). Resolving complexity in mitochondrial disease: Towards precision medicine. Mol Genet Metab *128*, 19–29.

Chin, J., Marotta, R., Chiotis, M., Allan, E.H., and Collins, S.J. (2014). Detection rates and phenotypic spectrum of m.3243A>G in the MT-TL1 gene: a molecular diagnostic laboratory perspective. Mitochondrion *17*, 34–41.

Chinnery, P.F., Thorburn, D.R., Samuels, D.C., White, S.L., Dahl, H.M., Turnbull, D.M., Lightowlers, R.N., and Howell, N. (2000). The inheritance of mitochondrial DNA heteroplasmy: random drift, selection or both? Trends Genet. *16*, 500–505.

de Laat, P., Koene, S., van den Heuvel, L.P.W.J., Rodenburg, R.J.T., Janssen, M.C.H., and Smeitink, J.A.M. (2012). Clinical features and heteroplasmy in blood, urine and saliva in 34 Dutch families carrying the m.3243A > G mutation. J Inherit Metab Dis *35*, 1059–1069.

Elliott, H.R., Samuels, D.C., Eden, J.A., Relton, C.L., and Chinnery, P.F. (2008). Pathogenic mitochondrial DNA mutations are common in the general population. Am. J. Hum. Genet. *83*, 254–260.

Fayssoil, A., Laforet, P., Bougouin, W., Jardel, C., Lombes, A., Bécane, H.M., Berber, N., Stojkovic, T., Béhin, A., Eymard, B., et al. (2017). Prediction of long-term prognosis by heteroplasmy levels of the m.3243A>G mutation in patients with the mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes syndrome. Eur J Neurol *24*, 255–261.

Freyer, C., Cree, L.M., Mourier, A., Stewart, J.B., Koolmeister, C., Milenkovic, D., Wai, T., Floros, V.I., Hagstrom, E., Chatzidaki, E.E., et al. (2012). Variation in germline mtDNA heteroplasmy is determined prenatally but modified during subsequent transmission. Nature Genet *44*, 1282–1285.

Gorman, G.S., Chinnery, P.F., DiMauro, S., Hirano, M., Koga, Y., McFarland, R., Suomalainen, A., Thorburn, D.R., Zeviani, M., and Turnbull, D.M. (2016). Mitochondrial diseases. Nature Reviews Disease Primers *2*, 16080.

Gorman, G.S., Schaefer, A.M., Ng, Y., Gomez, N., Blakely, E.L., Alston, C.L., Feeney, C., Horvath, R., Yu-Wai-Man, P., Chinnery, P.F., et al. (2015). Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease. Ann Neurol *77*, 753–759.

Goto, Y., Nonaka, I., and Horai, S. (1990). A mutation in the tRNA(Leu)(UUR) gene associated with the MELAS subgroup of mitochondrial encephalomyopathies. Nature *348*, 651–653.

Grady, J.P., Pickett, S.J., Ng, Y.S., Alston, C.L., Blakely, E.L., Hardy, S.A., Feeney, C.L., Bright, A.A., Schaefer, A.M., Gorman, G.S., et al. (2018a). mtDNA heteroplasmy level and copy number indicate disease burden in m.3243A>G mitochondrial disease. EMBO Mol Med *10*.

Grady, J.P., Pickett, S.J., Ng, Y.S., Alston, C.L., Blakely, E.L., Hardy, S.A., Feeney, C.L., Bright, A.A., Schaefer, A.M., Gorman, G.S., et al. (2018b). mtDNA heteroplasmy level and copy number indicate disease burden in m.3243A>G mitochondrial disease. EMBO Mol Med *10*, e8262.

Kaufmann, P., Engelstad, K., Wei, Y., Kulikova, R., Oskoui, M., Sproule, D.M., Battista, V., Koenigsberger, D.Y., Pascual, J.M., Shanske, S., et al. (2011). Natural history of MELAS associated with mitochondrial DNA m.3243A>G genotype. Neurology *77*, 1965–1971.

Ko, C.H., Lam, C.W., Tse, P.W., Kong, C.K., Chan, A.K., and Wong, L.J. (2001). De novo mutation in the mitochondrial tRNALeu(UUR) gene (A3243G) with rapid segregation resulting in MELAS in the offspring. J Paediatr Child Health *37*, 87–90.

Koga, Y., Akita, Y., Takane, N., Sato, Y., and Kato, H. (2000). Heterogeneous presentation in A3243G mutation in the mitochondrial tRNA(Leu(UUR)) gene. Arch Dis Child *82*, 407–411.

Langdahl, J.H., Larsen, M., Frost, M., Andersen, P.H., Yderstræde, K.B., Vissing, J., Dunø, M., Thomassen, M., and Frederiksen, A.L. (2018). Lecocytes mutation load declines with age in carriers of the m.3243A>G mutation: A 10-year Prospective Cohort. Clinical Genetics *93*, 925–928.

Mancuso, M., Orsucci, D., Angelini, C., Bertini, E., Carelli, V., Comi, G.P., Donati, A., Minetti, C., Moggio, M., Mongini, T., et al. (2014). The m.3243A>G mitochondrial DNA mutation and related phenotypes. A matter of gender? J Neurol *261*, 504–510.

Manwaring, N., Jones, M.M., Wang, J.J., Rochtchina, E., Howard, C., Mitchell, P., and Sue, C.M. (2007). Population prevalence of the MELAS A3243G mutation. Mitochondrion *7*, 230–233.

Mehrazin, M., Shanske, S., Kaufmann, P., Wei, Y., Coku, J., Engelstad, K., Naini, A., De Vivo, D.C., and DiMauro, S. (2009). Longitudinal changes of mtDNA A3243G mutation load and level of functioning in MELAS. American Journal of Medical Genetics

Part A *149A*, 584–587.

Nesbitt, V., Pitceathly, R.D.S., Turnbull, D.M., Taylor, R.W., Sweeney, M.G., Mudanohwo, E.E., Rahman, S., Hanna, M.G., and McFarland, R. (2013). The UK MRC Mitochondrial Disease Patient Cohort Study: clinical phenotypes associated with the m.3243A>G mutation--implications for diagnosis and management. J Neurol Neurosurg Psychiatry *84*, 936–938.

Otten, A.B.C., Sallevelt, S.C.E.H., Carling, P.J., Dreesen, J.C.F.M., Drüsedau, M., Spierts, S., Paulussen, A.D.C., de Die-Smulders, C.E.M., Herbert, M., Chinnery, P.F., et al. (2018). Mutation-specific effects in germline transmission of pathogenic mtDNA variants. Hum. Reprod. *33*, 1331–1341.

Payne, B.A., Wilson, I.J., Hateley, C.A., Horvath, R., Santibanez-Koref, M., Samuels, D.C., Price, D.A., and Chinnery, P.F. (2011). Mitochondrial aging is accelerated by anti-retroviral therapy through the clonal expansion of mtDNA mutations. Nature Genet *43*, 806–810.

Pickett, S.J., Blain, A., Ng, Y.S., Wilson, I.J., Taylor, R.W., McFarland, R., Turnbull, D.M., and Gorman, G.S. (2019). Mitochondrial Donation — Which Women Could Benefit? New England Journal of Medicine *380*, 1971–1972.

Pickett, S.J., Grady, J.P., Ng, Y.S., Gorman, G.S., Schaefer, A.M., Wilson, I.J., Cordell, H.J., Turnbull, D.M., Taylor, R.W., and McFarland, R. (2018). Phenotypic heterogeneity in m.3243A>G mitochondrial disease: The role of nuclear factors. Ann Clin Transl Neurol *5*, 333–345.

Pyle, A., Taylor, R.W., Durham, S.E., Deschauer, M., Schaefer, A.M., Samuels, D.C., and Chinnery, P.F. (2007). Depletion of mitochondrial DNA in leucocytes harbouring the 3243Aâ†'G mtDNA mutation. J Med Genet *44*, 69–74.

Rajasimha, H.K., Chinnery, P.F., and Samuels, D.C. (2008). Selection against Pathogenic mtDNA Mutations in a Stem Cell Population Leads to the Loss of the 3243A→G Mutation in Blood. Am J Hum Genet *82*, 333–343.

Sue, C.M., Quigley, A., Katsabanis, S., Kapsa, R., Crimmins, D.S., Byrne, E., and Morris, J.G. (1998). Detection of MELAS A3243G point mutation in muscle, blood and hair follicles. J Neurol Sci *161*, 36–39.

Veitia, R.A. (2018). How the most common mitochondrial DNA mutation (m.3243A>G) vanishes from leukocytes: a mathematical model. Hum. Mol. Genet. *27*, 1565–1571.

Wilson, I.J., Carling, P.J., Alston, C.L., Floros, V.I., Pyle, A., Hudson, G., Sallevelt, S.C.E.H., Lamperti, C., Carelli, V., Bindoff, L.A., et al. (2016). Mitochondrial DNA sequence characteristics modulate the size of the genetic bottleneck. Hum. Mol. Genet. *25*, 1031–1041.

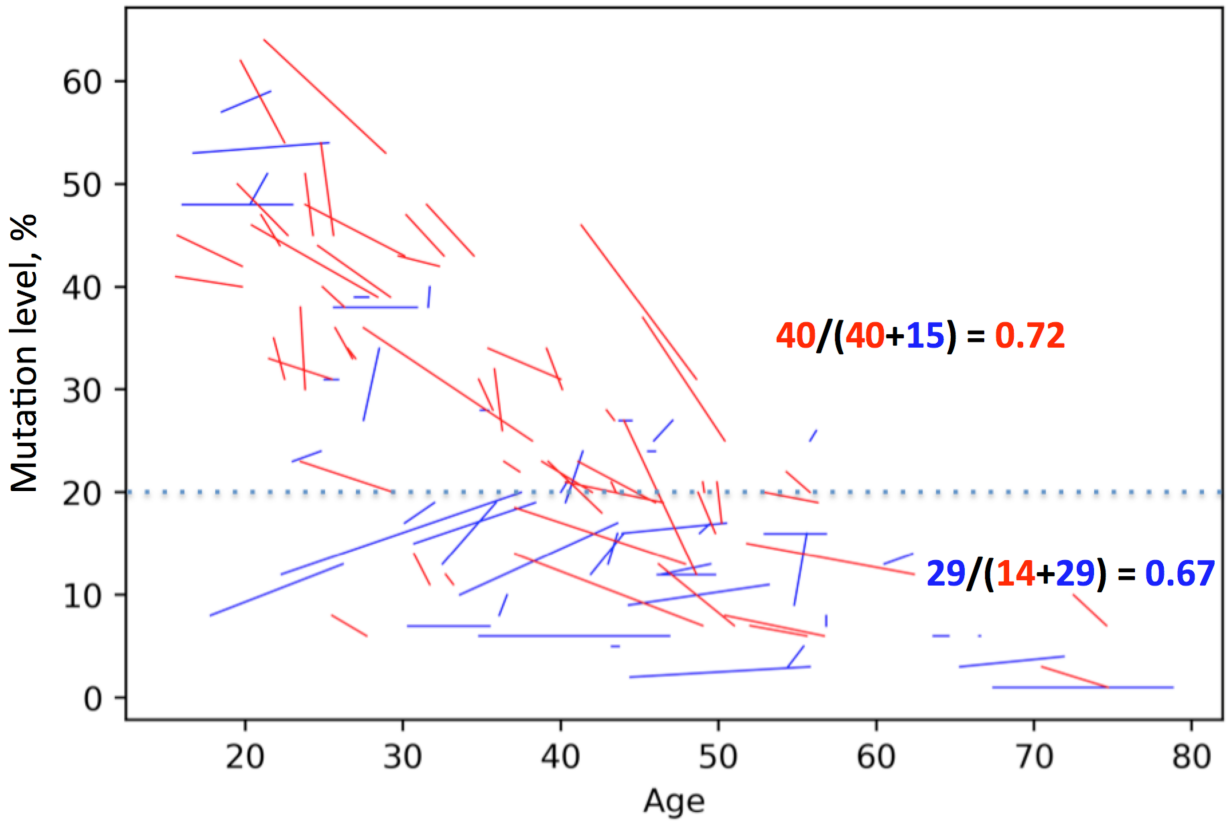

Figure S1: Example of calculation for ascending/descending ratio curves (shown in Figures 1B and 2B). Calculation is shown for a single point (the 20% threshold value) of the red curve (0.72) and a single point of the blue curve (0.67) in Fig1B. There are 40 descending measurements and 15 ascending measurements above a threshold of 20% (dotted line), giving a "descending ratio" of 0.72 for a 20% threshold. Likewise, 29 measurements were ascending and 14 descending below this threshold, giving an "ascending ratio" of 0.67. This type of calculation was used to produce Figure 1B and 2B.

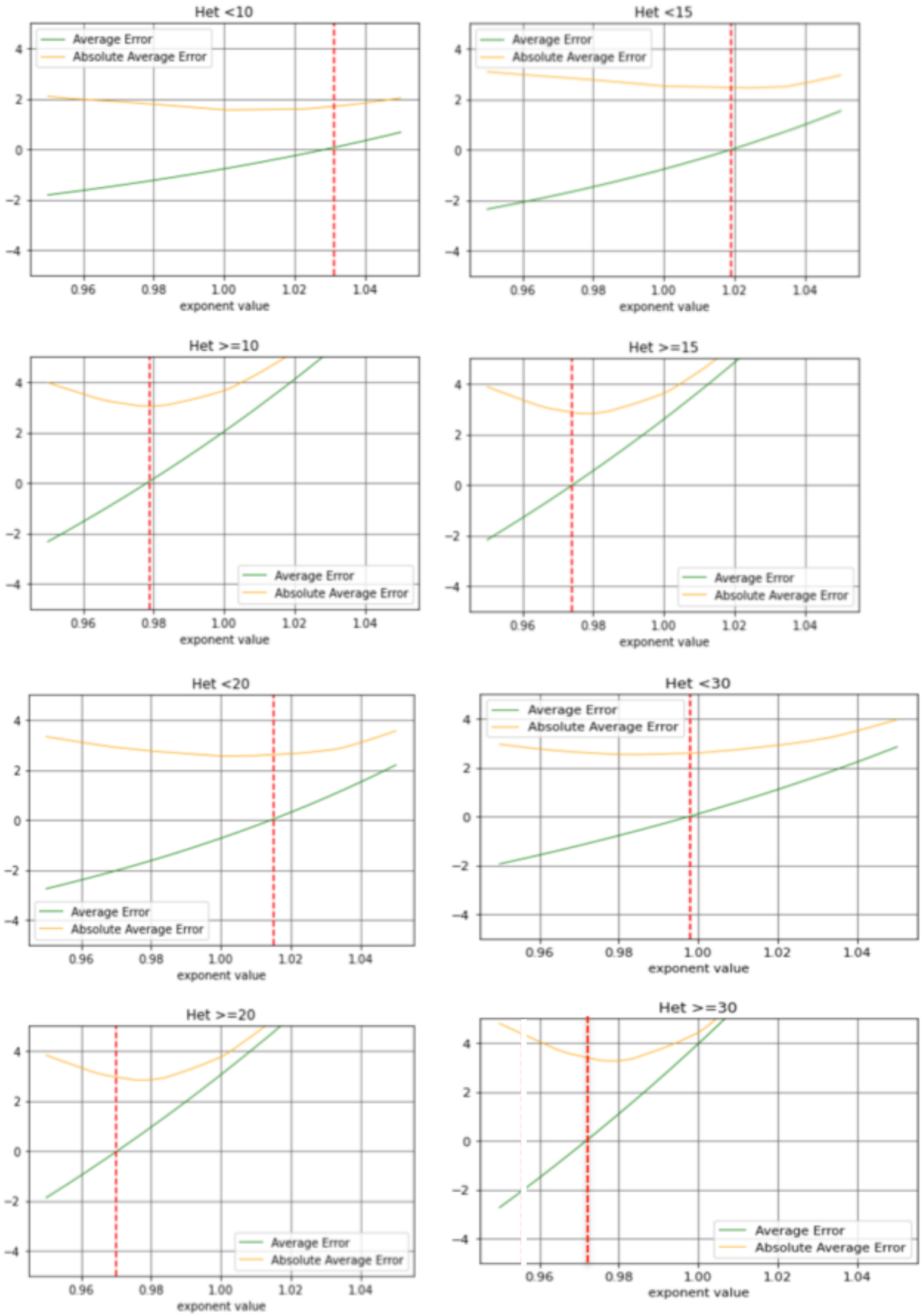

Figure S2. **Full set of graphs used to determine the unbiased decline rates**.