



Mitochondrial DNA somatic mutation burden and heteroplasmy are associated with chronological age, smoking, and HIV infection

Adam S. Ziada^{1,2} | Meng Ying Lu^{1,2} | Jarek Ignas-Menzies³ | Elijah Paintsil^{4,5,6} | Min Li⁴ | Onyema Ogbuagu⁷ | Sara Saberi^{1,2} | Anthony Y. Y. Hsieh^{1,2} | Beheroze Sattha^{1,2} | P. Richard Harrigan⁸ | Steve Kalloger¹ | H el ene C. F. C ot e^{1,2,9} | the CIHR team grant on cellular aging, HIV comorbidities in women, children (CARMA)*

¹Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada

²Centre for Blood Research, University of British Columbia, Vancouver, BC, Canada

³Department of Mechanical Engineering, University of British Columbia, Vancouver, BC, Canada

⁴Department of Pediatrics, Yale School of Medicine, New Haven, CT, USA

⁵School of Public Health, Yale University, New Haven, CT, USA

⁶Department of Pharmacology, Yale School of Medicine, New Haven, CT, USA

⁷Department of Medicine, Yale School of Medicine, New Haven, CT, USA

⁸Department of Medicine, University of British Columbia, Vancouver, BC, Canada

⁹Women's Health Research Institute, Vancouver, BC, Canada

Correspondence

H el ene C. F. C ot e, Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada.
Email: helene.cote@ubc.ca

Funding information

Canadian Institutes of Health Research, Grant/Award Number: HOP-123801 and TCO-125269; National Institutes of Health, Grant/Award Number: K08AI074404

Abstract

The gradual accumulation of mitochondrial DNA (mtDNA) mutations is implicated in aging and may contribute to the accelerated aging phenotype seen with tobacco smoking and HIV infection. mtDNA mutations are thought to arise from oxidative damage; however, recent reports implicate polymerase γ errors during mtDNA replication. Investigations of somatic mtDNA mutations have been hampered by technical challenges in measuring low-frequency mutations. We use primer ID-based next-generation sequencing to quantify both somatic and heteroplasmic blood mtDNA point mutations within the D-loop, in 164 women and girls aged 2–72 years, of whom 35% were smokers and 56% were HIV-positive. Somatic mutations and the occurrence of heteroplasmic mutations increased with age. While transitions are theorized to result from polymerase γ errors, transversions are believed to arise from DNA oxidative damage. In our study, both transition and transversion mutations were associated with age. However, transition somatic mutations were more prevalent than transversions, and no heteroplasmic transversions were observed. We also measured elevated somatic mutations, but not heteroplasmy, in association with high peak HIV viremia. Conversely, heteroplasmy was higher among smokers, but somatic mutations were not, suggesting that smoking promotes the expansion of preexisting mutations rather than de novo mutations. Taken together, our results are consistent with blood mtDNA mutations increasing with age, inferring a greater contribution of polymerase γ errors in mtDNA mutagenesis. We further suggest that smoking and HIV infection both contribute to the accumulation of mtDNA mutations, though in different ways.

KEYWORDS

aging, heteroplasmy, HIV, mitochondria, mtDNA, smoking, somatic mutations

*CIHR team grant on cellular aging and HIV comorbidities in women and children (CARMA) see in Appendix A

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

  2019 The Authors. *Aging Cell* published by the Anatomical Society and John Wiley & Sons Ltd.

1 | INTRODUCTION

Mitochondria contain their own DNA, which code for 22 tRNAs, 2 rRNAs, and 13 proteins that are essential components of the electron transport chain. Mitochondrial DNA (mtDNA) mutations, whether inherited or acquired, have been linked to several degenerative diseases, cancers, and aging (Polyak et al., 1998; Wallace, 1999). The free radical theory of aging proposes that mtDNA, given its close physical proximity to the electron transport chain, is susceptible to oxidative damage by reactive oxygen species (ROS), leading to mtDNA damage as well as other cellular injury (Gerschman, Gilbert, Nye, Dwyer, & Fenn, 1954; Harman, 1956; Miquel, Economos, Fleming, & Johnson, 1980). Therefore, it was often assumed that ROS would increase newly acquired (somatic) mtDNA mutations, as opposed to inherited ones. However, more recent studies suggest that somatic mtDNA mutations may be predominantly introduced through replication errors by mitochondrial polymerase γ during mtDNA replication (Kennedy, Salk, Schmitt, & Loeb, 2013; Trifunovic et al., 2004). Over time, these low-frequency mutations may undergo clonal expansion and reach the pathologic levels that contribute to aging (Payne et al., 2011). When more than one mtDNA species is observed, it is referred to as heteroplasmy. Although the association between mtDNA mutations and aging is well accepted, few studies have been able to examine the dynamics of somatic and heteroplasmic mtDNA mutations in a large cohort, in part because detecting and quantifying de novo somatic mtDNA mutations is technically challenging.

A previous study has proposed that people acquire mtDNA mutations at a rate of 6 per 10^8 bp per year (Marcelino & Thilly, 1999). If this is true, an assay with an extremely low background error rate would be required to measure the random accumulation of such rare mutations. Currently, only primer ID (PID)-based sequencing methods (Hiatt, Patwardhan, Turner, Lee, & Shendure, 2010; Jabara, Jones, Roach, Anderson, & Swanstrom, 2011) have achieved sufficiently low background rates (e.g., $\leq 10^{-5}$ mutations per bp) for detection of somatic mtDNA mutations. Above this threshold, detected mutations are more likely to be inherited or heteroplasmic, having arisen from clonal expansion of previously existing mutations. Given this, the definition of somatic versus heteroplasmic mtDNA mutations in the current literature may be inconsistent. Within the current study, mtDNA mutations refer to mutations present on the single strand analyzed.

People living with HIV (HIV-positive) appear to experience premature aging (Bhatia, Ryscavage, & Taiwo, 2012). Most studies report a decrease in lifespan of ≥ 10 years among HIV-positive individuals compared to HIV-negative controls (Antiretroviral Therapy Cohort Collaboration, 2008), as well as earlier onset and higher prevalence of age-related comorbidities (Guaraldi et al., 2011), such as cardiovascular disease (Hsue et al., 2004), some non-AIDS-defining cancers (Shiels, Pfeiffer, & Engels, 2010), neurocognitive decline (Ances et al., 2010), and osteoporosis (Serrano et al., 1995). Previous studies on mtDNA mutation burden in HIV have had limited sample size and focused on HIV and/or HIV therapy (Martin et al., 2003;

Payne et al., 2011) without considering other potential confounders. Larger studies have either concentrated on heteroplasmic mtDNA mutations (Li et al., 2017) or used assays not sensitive enough to detect low-frequency somatic mutations (Jitratkosol et al., 2012).

Tobacco smoking is associated with accelerated aging (Bernhard, Moser, Backovic, & Wick, 2007). Smoking increases oxidative damage, which is in turn hypothesized to cause acceleration of aging (Kiyosawa et al., 1990; Loft et al., 1992). However, no next-generation sequencing (NGS) studies have examined the association between mtDNA mutations and smoking to date. Using other methods, increased mtDNA heteroplasmy has been linked to smoking (Lewis, Fradley, Griffiths, Baxter, & Parry, 2002; Tan et al., 2008), but evidence for mtDNA somatic point mutations is generally lacking (Coller et al., 1998). Again, these studies focused on smoking and were not powered to consider other factors that may be associated with mtDNA mutations.

The goal of the present study was to use a low background NGS assay to measure both somatic and heteroplasmic mtDNA mutation burden in relation to aging, taking into consideration important confounders of the aging process. We hypothesized that somatic mtDNA mutations increase with age, smoking, and HIV infection.

2 | RESULTS

2.1 | Blood somatic mtDNA substitution frequencies

A total of 92 HIV-positive and 72 HIV-negative Children and women: AntiRetrovirals and Markers of Aging (CARMA) participants were included in the study. Participants were all females aged 1–62 years, either current or never (but not past) smokers, and had no history of hepatitis C or B virus infection. Among all participants, chronological age was similar between the HIV-positive (median [range] 35 [2–61]) and the HIV-negative (32 [1–62]) groups (Table 1). The same was true for their smoking status, with 36% of HIV-positive and 35% of HIV-negative participants being current smokers (Table 1). However, the HIV-positive group had lower alcohol use and a higher body mass index (BMI) (26 [23–29] vs. 22 [21–27], $p = .01$), with a different ethnic makeup ($p < .001$) from that of the HIV-negative group. For example, 41% of the HIV-positive individuals were African/Caribbean/Black, compared to none in the HIV-negative group (Table 1). Among the HIV-positive participants, 67 (73%) were currently receiving combination antiretroviral therapy (cART), 8 (9%) were cART-naïve, and 58 (63%) had an undetectable HIV plasma viral load (pVL) at study visit.

As the somatic mtDNA substitution frequencies did not follow a normal distribution, data transformation was used and $\ln(x + 1)$ best normalized their distribution for linear regression modeling. All results presented hereafter refer to transformed data unless specified. Among all participants, a median [IQR] of 0.42 [0.28–0.61] transformed blood somatic mtDNA substitutions per 10,000 bp was observed. This 0.42 would translate into 0.86 mutations per mtDNA genome. Similar analyses of untransformed data are shown in Figure S1 in our Supporting Information. The assay's background error rate was calculated based on the repeated inclusion ($n = 12$) of a single cloned plasmid DNA

TABLE 1 Demographic characteristics of study participants showing mtDNA substitution mutations

Parameters	All participants			Adult participants		
	HIV + (n = 92)	HIV - (n = 72)	p-value	HIV + (n = 80)	HIV - (n = 59)	p-value
Age (years)	35 [27–42] (2–61)	32 [24–41] (1–62)	0.134	36 [31–43] (20–61)	33 [27–43] (21–62)	0.240
BMI (kg/m ²)				26 [23–29] (17–46)	22 [21–27] (16–53)	0.010
Smoking status						
Current smoker	33 (36)	25 (35)	1	33 (41)	25 (42)	1.000
Never smoker	59 (64)	47 (65)		47 (59)	34 (58)	
Alcohol use (drink-year)	0 [0–7] (0–352)	2 [0–12] (0–424)	0.004	0 [0–9] (0–352)	3 [1–24] (0–424)	<0.001
Drug use (weekly/daily)						
Current or Past	26 (28)	24 (33)	0.609	26 (33)	24 (41)	0.474
Never	64 (70)	48 (67)		52 (65)	35 (59)	
Ethnicity						
Caucasian	33 (36)	32 (44)	<0.001	31 (39)	32 (54)	<0.001
African/Caribbean/ Black	38 (41)	0 (0)		29 (36)	0 (0)	
Indigenous	12 (13)	14 (19)		11 (14)	14 (24)	
Asian	8 (9)	10 (14)		8 (10)	10 (17)	
Unknown	1 (1)	16 (22)		1 (1)	3 (5)	
CD4 ⁺ cell count (cells/μl)						
Current	510 [330–705] (30–2350)			460 [310–655] (30–1570)		
Nadir	225 [138–303] (0–2350)			210 [143–285] (0–1110)		
HIV plasma viral load (copies/ml)						
Current pVL (<50)	59 (64)			47 (59)		
Peak pVL > 100,000	40 (43)			31 (39)		
HIV treatment status						
PI-based	46 (50)			38 (48)		
NNRTI-based	19 (21)			16 (20)		
On other cART	2 (2)			1 (1)		
cART-naive	8 (9)			8 (10)		
Off cART	17 (18)			17 (21)		

Note: Data presented as median [IQR] (range) or n (%); ethnicity is unknown for all pediatric HIV-negative participants; cART, combination antiretroviral therapy; PI, protease inhibitor; NNRTI, non-nucleoside reverse transcriptase inhibitor; BMI, body mass index is not presented for groups that include pediatric participants. BMI was unknown for 5 adults, current CD4⁺ cell count was unknown for 1 adult. pVL, HIV plasma viral load < 50 copies/ml is “undetectable.” Current pVL was unknown for 2 adults. Mann-Whitney test was used for age, alcohol use, and BMI. Chi-square test was used for smoking status, drug use, and ethnicity.

control that functioned as a “no mutation” negative control for the assay. The assay's background somatic mtDNA substitution frequency was estimated at 0.063 [0.00–0.24] or 0.01 mutations per mtDNA genome, which was significantly lower than clinical samples ($p < .0001$; Figure S2). The assay's coefficient of variation was estimated at 6.3%, based on the inclusion of the same clinical sample in both runs (total $n = 3$).

The primary finding of our study is that these somatic mtDNA substitution mutations were significantly associated with age ($p < .0001$), whereby older individuals showed higher mutation frequencies (Figure 1). This association with age was seen among all participants ($n = 138$, $r = .37$, $p < .001$; Figure 1) and adult participants ($n = 113$, $r = .27$, $p = .025$; Figure 1).

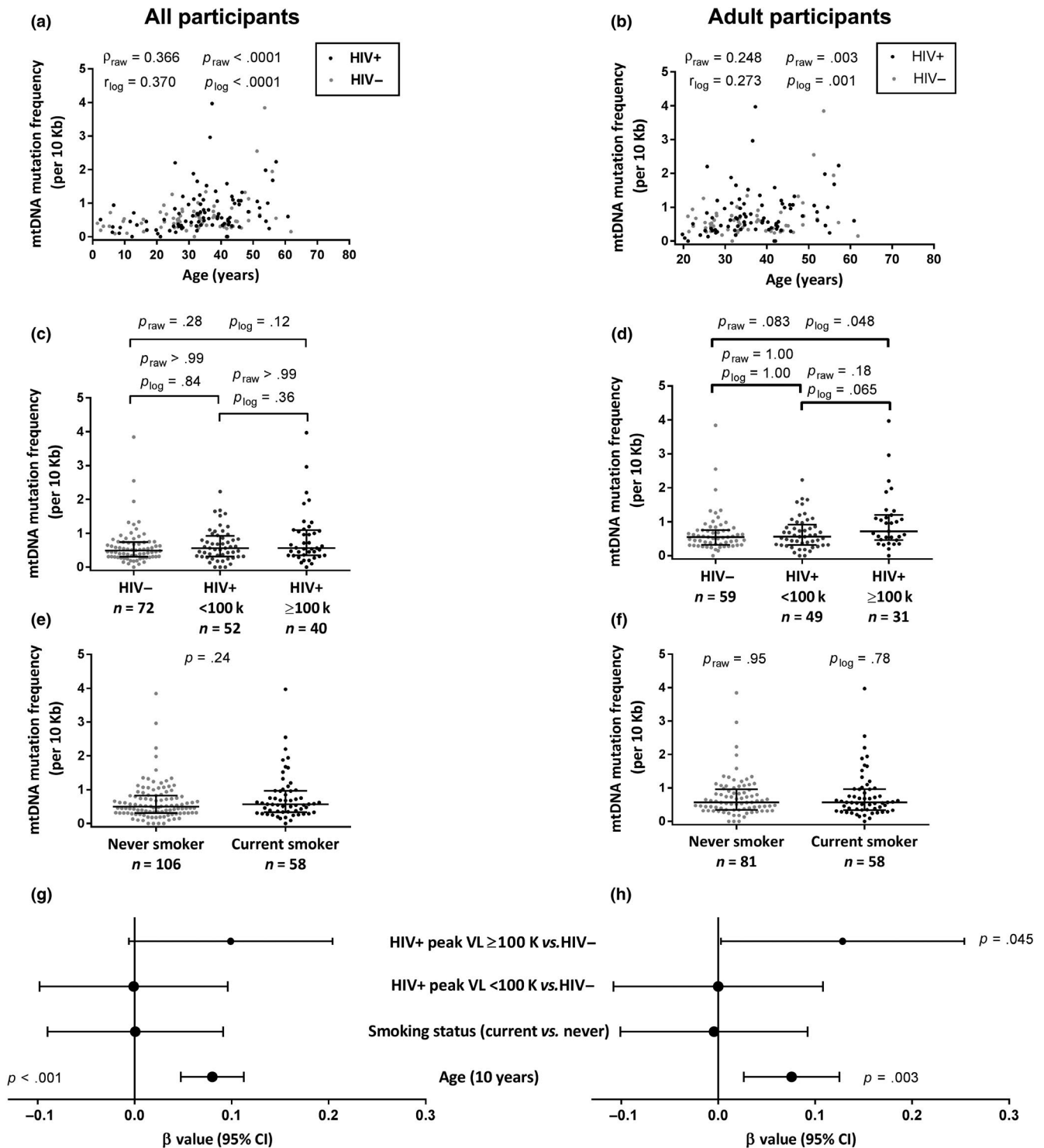
Furthermore, transition mutations, namely $A \leftrightarrow G$ and $C \leftrightarrow T$ mutations, were significantly more frequent than transversion mutations ($A \leftrightarrow C$, $G \leftrightarrow T$, $T \leftrightarrow A$, and $C \leftrightarrow G$) ($p < .0001$; Figure S3b). Similar frequencies and types of mutations were seen in both HIV groups (Figure S3d). This remained true upon trichotomizing the groups according to their HIV status and peak HIV pVL (Figure S3e). The background substitutions, among the 12 cloned plasmids, also consisted primarily of transition mutations (Figure S3a).

Among participants who had nonzero somatic mtDNA substitution frequencies ($n = 160$), plotting each type of mutation against age individually reveals that both transition ($n = 159$, $\rho = .38$, $p < .0001$) and transversion ($n = 31$, $\rho = .42$, $p = .020$)

somatic mutation burden were positively associated with age (Figure 2b,d), after removing participants with a corresponding mutation (transition or transversion) frequency equal to zero, respectively.

Having established the association between somatic mtDNA substitution frequency and age, we then investigated whether HIV, a chronic infection, was associated with increased mtDNA

mutation frequencies. HIV-positive status did not show any association with somatic mtDNA substitution frequencies univariately (Table S1). However, HIV-positive participants with a peak pVL $\geq 100,000$ copies/ml had a higher somatic mtDNA substitution frequency than HIV-negative participants ($p = .048$; Figure 1d and Table S1). Other variables investigated for possible inclusion in these models were ethnicity and BMI. Neither showed an



log_e: statistical test performed on log_e(x + 1) somatic mtDNA mutation data

raw: statistical test performed on untransformed somatic mtDNA mutation data

FIGURE 1 Older chronological age and high peak HIV viral load are associated with increased blood somatic mtDNA substitutions. The measured mtDNA mutation frequencies (expressed as mutations per 10,000 bp) are presented (a-f) and p-values for both raw and transformed ($\ln [x + 1]$) data are shown. (a, b) Blood somatic mtDNA substitutions are positively correlated with chronological age among all participants (a) and adult participants (b). (c, d) With respect to HIV, only adult participants with a peak HIV plasma viral load $\geq 100,000$ copies/ml showed a marginally higher somatic mtDNA substitution frequency compared to HIV-negative controls. (e, f) Tobacco smoking was not associated with somatic mtDNA mutation substitutions. (g, h) Forest plot showing the estimated size of the effect (β value) and the 95% confidence interval on that estimate, based on an analysis of covariance for all ($R^2 = .159$) and adult ($R^2 = .107$) participants. The models suggest that older age remains independently associated with increased somatic mtDNA mutations, after adjusting for HIV and smoking. Furthermore, among adults, having a peak pVL $\geq 100,000$ copies/ml is also associated with a higher mtDNA mutation frequency although the 95% confidence interval is wide. All univariate tests on raw data were nonparametric and used (c, d) Dunn's correction for multiple comparisons where appropriate. (c, d) ANOVA tests with Tukey's corrections. A single p-value is reported if the same nonparametric test was used for raw and transformed data

association ($p \geq .1$) with the somatic mtDNA substitution frequency (Table S1). Of note, there was also no evidence of an association between somatic mtDNA substitution frequency and smoking. Similarly, among HIV + adults, only HIV peak pVL and CD4 nadir showed some association (Table S1).

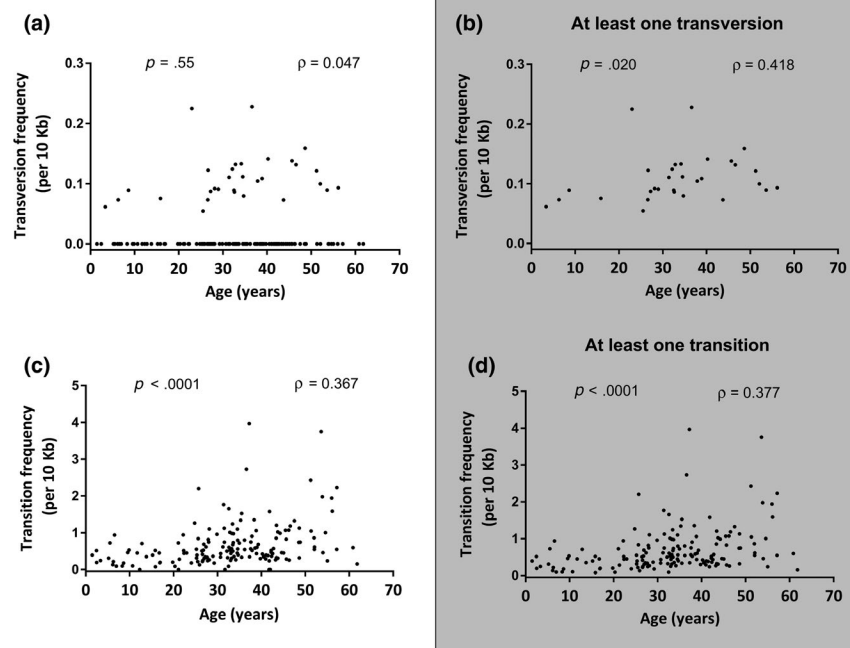
2.2 | Multivariable correlates of somatic mtDNA substitution frequency

In a linear regression model of adult participants ($n = 139$, $R^2 = .107$) that included age, smoking status, and trichotomized HIV status, both older age and having HIV with pVL $\geq 100,000$ copies/ml (vs. HIV-negative) remained independently associated with somatic mtDNA mutation burden, although the latter exhibited a large confidence interval (Figure 1h). The same model among all participants ($n = 164$, $R^2 = .159$; Figure 1g) only showed age as being associated with somatic mtDNA substitution frequencies. Our models describe 11% and 16% of the variance in somatic mtDNA substitutions among adult and all participants, respectively.

2.3 | Blood mtDNA heteroplasmic substitution frequencies

Heteroplasmy was detected in 39 of the 164 participants, and information regarding heteroplasmic positions is provided in Table S2. All heteroplasmic variants were transition mutations, and the pattern of the heteroplasmic positions was different for most participants. HIV-specific variables were also explored and again showed no association (Table S3). Only smoking status was associated with the occurrence of heteroplasmy among all ($p = .007$; Figure 3a), but not adult-only ($p = .055$; Figure 3d) participants in univariate analysis (Table S2). An interaction was noted between age and smoking status with respect to the presence of mtDNA heteroplasmy, whereby the frequency of heteroplasmy increased with age among nonsmokers, but decreased with age among smokers (Figure 4a). Binary logistic regression models were built that included age, trichotomized HIV and peak pVL status, smoking status, and an age* smoking interaction term, for all participants ($n = 164$, $R^2 = .339$; Figure 3g) and adult-only participants ($n = 139$, $R^2 = .31$; Figure 3h). In both models, smoking

FIGURE 2 Blood mtDNA transition and transversion mutations increase with older chronological age. The measured mtDNA mutation frequencies (expressed as mutations per 10,000 bp) are presented. (a) Among all participants, no correlation between age and transversion mutation is observed; however (b) in a subanalysis removing participants with no transversion mutations, a mild correlation with increased age is seen. (c, d) Among all participants, transition mutations are associated with older age, before (c) and after (d) removing participants with no transition mutations. (a-d) Spearman's correlations were performed on both raw and ($\ln [x + 1]$) transformed data; a single P- and ρ values are presented



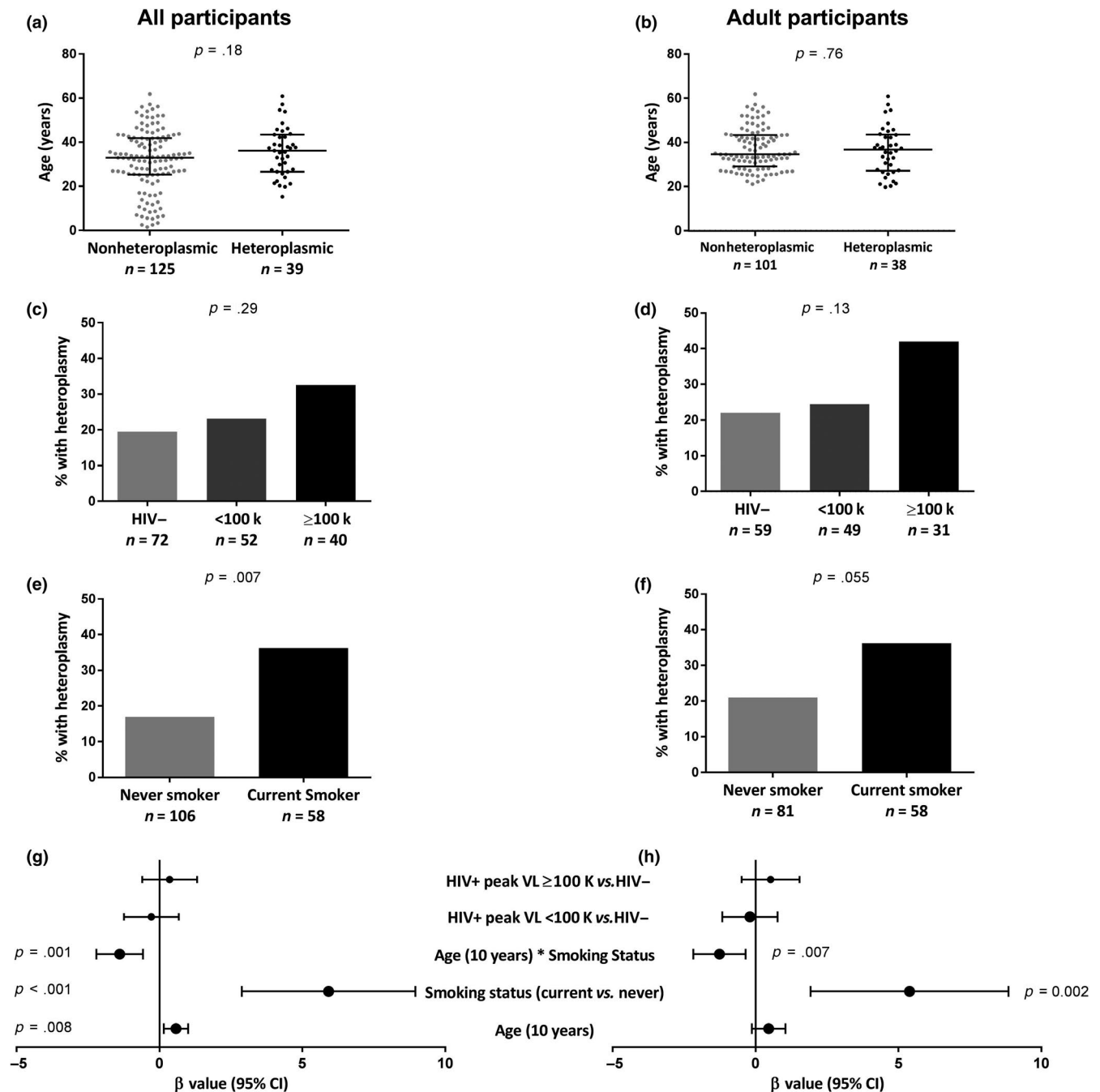


FIGURE 3 Older chronological age and tobacco smoking are associated with increased blood mtDNA heteroplasmy. Heteroplasmy was categorized as a yes/no variable. (a, b) The occurrence of heteroplasmy is not univariately associated with chronological age. (c, d) No association is observed between HIV and the occurrence of heteroplasmy. (e, f) Among all participants, current smokers show an increased occurrence of heteroplasmy (e); however, this difference is not seen among adult participants ($p = .055$) (f). (g, h) Forest plot showing the estimated size of the effect (β value) and the 95% confidence interval on that estimate, based on binary logistic regressions of all ($R^2 = .339$) and adult ($R^2 = .308$) participants. The model among all participants suggests that older age and smoking are associated with the occurrence of heteroplasmy (g); however among adults, smoking showed an association with heteroplasmy (h). Furthermore, a significant interaction between smoking and age is observed in both models

status ($p \leq .002$) and the age*smoking interaction ($p \leq .007$) were independently associated with heteroplasmy. Older age was also independently associated with heteroplasmy ($p = .008$) among all participants (Figure 3g). These models describe 31% and 34% of the variance in mtDNA heteroplasmy among adult and all

participants, respectively. In subanalyses within smoker groups, nonsmokers showed a significant increase in heteroplasmy with increasing age ($p = .004$; Figure 4a,c), while current smokers showed the reverse, namely a decrease in heteroplasmy with age ($p = .025$, Figure 4a,b).

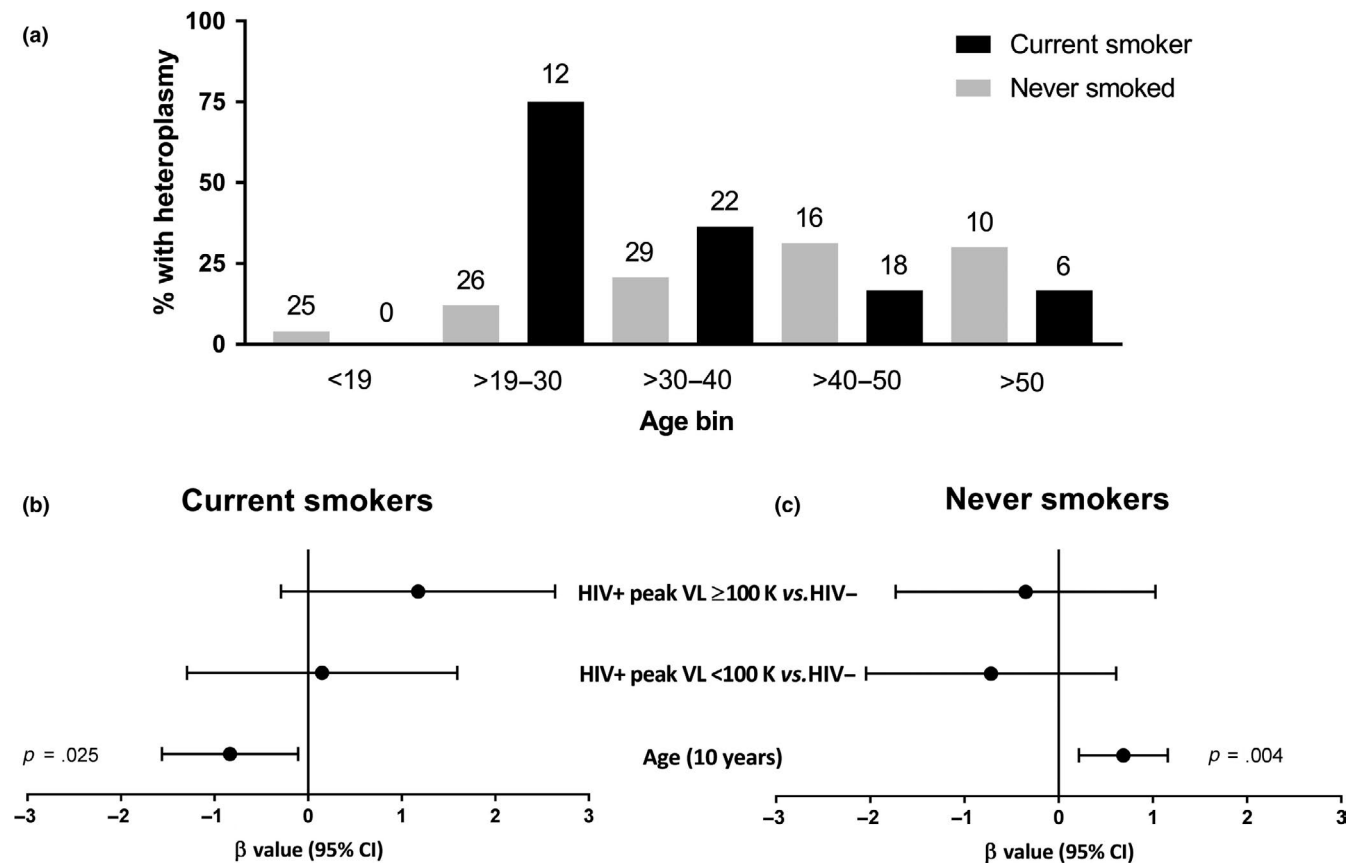


FIGURE 4 Current tobacco smokers experience a decrease in heteroplasmy with older age, while never smokers show the opposite. Heteroplasmy was categorized as a yes/no variable. (a) Percentage of study participants with heteroplasmy binned by age and smoking status, with total number of participants for each bin (above bar), and percentage of those participants (bar height). (b, c) Binary logistic regression models among current smokers ($R^2 = .225$) and never smokers ($R^2 = .155$) to further understand the effect of age on heteroplasmy. Older age is associated with a decrease in the occurrence of heteroplasmy among current smokers (b), while the reverse is seen for never smokers (c). This indicates an interaction whereby the effect of age on heteroplasmy is modulated by smoking status

In an effort to independently validate this observation, data were obtained from an independent cohort at Yale New Haven Hospital (Li et al., 2017). After removing past smokers and participants with unknown smoking status, the narrow age range 53 [50–56] (30–66), and smaller sample size ($n = 40$) of the remaining participants made repeating our models within this separate cohort difficult (Figure S4). When the two cohorts were combined (results presented in Supporting Information), heteroplasmy was independently associated with older age ($p < .001$), smoking ($p < .001$), and the same age*smoking interaction term described above (Figure S5g,h) ($p < .001$), among all participants ($R^2 = .24$, $n = 204$; Figure S5g) and all adults ($R^2 = .18$, $n = 159$; Figure S5h). Further exploring the interaction term (Figure S6a), participants who currently smoked showed a nonsignificant trend toward lower heteroplasmy while they aged ($p = .098$; Figure S6b), and participants who never smoked had more heteroplasmy as they aged ($p < .001$, Figure S6c). While HIV status was univariably associated with heteroplasmy, we were unable to include it in the final multivariable models, due to the lack of peak pVL data for half of the HIV + participants in the Yale New Haven Hospital cohort.

3 | DISCUSSION

3.1 | Somatic mutations

We quantified somatic mtDNA substitution burden in human blood for the first time and show that somatic mtDNA substitutions increase with age. Our results therefore support the general theory that mtDNA mutations accumulate with age (Harman, 1956; Michikawa, Mazzucchelli, Bresolin, Scarlato, & Attardi, 1999). Age alone explained 14% of the variance in somatic mtDNA mutation frequency and remained the strongest independent predictor after adjusting for covariates. Our study builds on the work of a dual primer ID-based study showing higher mtDNA somatic mutation frequencies in the brain of five elderly individuals compared to five infants (Kennedy et al., 2013). Of note, that study suggested that mutations accumulate asymmetrically on the two strands of mtDNA and that they increase with age on both strands (Kennedy et al., 2013). As our single PID method measured putative somatic mtDNA mutations on one mtDNA strand, we cannot ascertain whether these variants are present in the other strand. It is possible that some of our signal

is due to chemical lesions that could be repaired or become a double-stranded mutation. Nevertheless, our observation that these increase with age is consistent with the dual PID study (Kennedy et al., 2013).

We did not observe this age effect in our previous study restricted to mtDNA transversion (A ↔ C & T ↔ G) mutations in the blood of HIV-positive and HIV-negative mothers and their infants (Jitratkosol et al., 2012). This is likely related to the low sample size, the young age of the mothers, and the relative scarcity of transversion mutations seen to increase with age in our own study.

Indeed, with respect to the nature of the substitutions, transition mutations (A ↔ G, C ↔ T) were the most frequently observed, while transversion mutations (A ↔ C, A ↔ T, C ↔ G, G ↔ T) were rarely seen in our study, echoing the results seen in the above brain mtDNA study (Kennedy et al., 2013). Transition mutations are generally believed to arise from polymerase γ errors (Longley, Nguyen, Kunkel, & Copeland, 2001; Spelbrink et al., 2000), while transversion mutations are considered the signature of DNA oxidative damage. The 7,8-dihydro-8-oxo-deoxyguanosine (8-oxodG) lesion (Yasui et al., 2014), one of the most commonly studied DNA oxidative lesions, has been shown to increase with age (Fraga, Shigenaga, Park, Degan, & Ames, 1990). However, several other lesions may also lead to transition mutations (Basu, Loechler, Leadon, & Essigmann, 1989; Kreutzer & Essigmann, 1998), and 8-oxodG is only one of the 37 reported major products of oxidative damage to DNA (Evans, Dizdaroglu, & Cooke, 2004). Therefore, although our results do support the growing body of literature suggesting that polymerase γ errors are the major source of age-related mtDNA mutations and variants (Kennedy et al., 2013; Trifunovic et al., 2004), we cannot rule out the potential contribution of oxidative damage as a source of mtDNA transition mutations in our study (Fraga et al., 1990; Harman, 1956; Schriener et al., 2005). In fact, among the subset of participants (19%) with detectable somatic transversion mutations, these mutations showed a weak but significant increase with age.

In addition to the effect of chronological age on somatic mtDNA substitutions, we sought to explore whether these were influenced by other factors including tobacco smoking and chronic immune activation/inflammation as seen among people living with HIV. In our sample, we failed to detect any association between somatic mtDNA substitutions and smoking. To date, no study of smoking and somatic mtDNA mutations has been conducted in blood; however, a small ($n = 4$ vs. 4) study of bronchial epithelial tissue also failed to detect any association with smoking, although the study was not sequencing-based (Coller et al., 1998).

With respect to HIV, we only detected a significantly increased somatic mtDNA substitution burden among participants who had a high peak HIV pVL. Thus far, only two studies have investigated somatic mtDNA point mutation frequencies and HIV status. The first found significantly higher ($p = .012$) blood A ↔ C and T ↔ G transversion mutations among new mothers living with HIV ($n = 42$) than in HIV-negative controls ($n = 39$) (Jitratkosol et al., 2012). The second investigated skeletal mtDNA and found no difference in low-frequency mtDNA point variants (exceeding 0.2% frequency) between

persons living with HIV (ART-treated, $n = 8$, and untreated, $n = 4$) and HIV-negative controls ($n = 4$) (Payne et al., 2011). Our study builds on these by suggesting that HIV status alone is not necessarily associated with somatic mtDNA mutation frequency, as those with a peak pVL < 100,000 copies/ μ l had comparable mutation frequencies to HIV-negative individuals. Higher peak pVL may reflect poorer host control of the virus or delayed initiation of cART, both of which would increase exposure to HIV viremia and immune activation, and could potentially lead to mtDNA mutations.

Given that mtDNA mutations are estimated to accumulate at a frequency of one per 6×10^8 bp per year (Marcelino & Thilly, 1999), without the application of PIDs, most sequencing-based assays measure inherited or clonally expanded mutations rather than true somatic mtDNA mutations. Our approach exploits PIDs to distinguish sequencing errors and postprimer extension PCR errors from mutations in the original template, and is based on the prior work by other groups. These applications use either a single PID (Hiatt et al., 2010; Jabara et al., 2011) (as seen in this study) or a dual PID approach, which further eliminates possible mutations introduced during primer extension step (Schmitt et al., 2012). Since their pioneering work, PID-based methods have been used in several fields of research including HIV (Jabara et al., 2011), cancer (Narayan et al., 2012), microbial populations (Schmitt et al., 2012), and aging (Kennedy et al., 2013). Among studies using single PIDs, reported background rates range from 1.4×10^{-2} to 7.6×10^{-3} mutations per 10,000 bp (Fox, Reid-Bayliss, Emond, & Loeb, 2014). We measured a background error rate of 6.3×10^{-3} mutations per 10,000 bp, which compares favorably to the lowest single PID error rates achieved to date.

3.2 | Heteroplasmy

The majority of studies to date have measured mtDNA heteroplasmy, defined herein as a given mutation present at >2% frequency for a given position. We report a positive relationship between older age and the presence of heteroplasmic point mutations, consistent with the literature (Michikawa et al., 1999; Munscher, Muller-Hocker, & Kadenbach, 1993).

Furthermore, unlike somatic mutations, mtDNA heteroplasmy was significantly more prevalent among participants who were current smokers as opposed to never smokers. We also failed to detect any heteroplasmy among pediatric participants. There is an extensive body of literature linking smoking and oxidative DNA damage (Kiyosawa et al., 1990; Loft et al., 1992). Two non-sequencing-based studies suggest that mtDNA heteroplasmy is increased in the parotid gland (Lewis et al., 2002) and buccal cells (Tan et al., 2008) of tobacco smokers. This association of smoking with mtDNA heteroplasmy, along with the lack of association with somatic mtDNA substitutions, suggests that smoking promotes the clonal expansion of mtDNA mutations rather than the generation of de novo somatic mutations. In support of this, both increased leukocyte count (Petitti & Kipp, 1986) and mtDNA content fluctuations (Lee, Lu, Fahn, & Wei, 1998) have been associated with tobacco smoking.

The observed interaction between age and smoking with respect to mtDNA heteroplasmy, whereby the occurrence of heteroplasmy increases with age among nonsmokers but decreases among smokers has not been reported. Although our study cannot ascertain the mechanism behind this, we speculate that the higher heteroplasmic mtDNA mutation burden, or related factors in smokers, may more readily lead to the elimination of damaged cells carrying heteroplasmic mtDNA mutations. This highlights the importance of considering lifestyle confounders that may influence the frequency of detected mtDNA mutations. Unlike smoking, we did not find evidence that stressors associated with living with HIV lead to increased heteroplasmy. However, larger studies are needed to confirm this, especially since the vast majority of our participants living with HIV were treated with antiretroviral therapy and had well-controlled HIV viremia.

Although we were unable to replicate the observed interaction between smoking and age in the Yale New Haven Hospital cohort alone, this was attributed to the narrow age range and lower sample size. When the CARMA cohort was combined with the Yale New Haven Hospital cohort, we were able to fully replicate the observations seen in the CARMA cohort alone. Large cohorts with a wide age range, information on smoking status, and next-generation mtDNA sequencing data are very uncommon. In context, our study builds upon current knowledge on smoking and mtDNA mutation by utilizing primer ID NGS and a large cohort of individuals with a wide age range; however, further studies will be needed to confirm this interaction in an independent cohort.

To date, two small studies (Martin et al., 2003; Payne et al., 2011) and one larger study (Li et al., 2017) have investigated mtDNA heteroplasmic point mutations in persons living with HIV. The first one suggested that HIV therapy-experienced participants showed a higher frequency of blood mtDNA heteroplasmy ($n = 5/16$) than HIV therapy-naïve participants ($n = 0/10$). The second study demonstrated increased mtDNA deletions in muscle tissue from participants treated with HIV therapy ($n = 21$) compared to those who were therapy-naïve ($n = 11$), but showed no evidence of increased mtDNA heteroplasmy (Payne et al., 2011). Finally Li et al. (2017) reported a significantly higher frequency of mtDNA variants among persons living with HIV and treated with cART ($n = 47$) compared to HIV-negative controls ($n = 24$).

In our study, we found no association between HIV status and mtDNA heteroplasmy. Although the presence of heteroplasmy was somewhat more frequent among persons with a high HIV peak pVL, this did not reach significance in any model. A longitudinal study is required to confirm whether somatic mtDNA mutations, which we found more frequently among those with high peak pVL, eventually lead to heteroplasmy over time. Our study was not designed or powered to investigate the effect of cART.

3.3 | Strengths and limitations

Our study has several strengths: (a) The use of a novel mtDNA substitution burden assay allows the quantification of very low-frequency

mutations in blood cells; (b) a large sample size facilitated multivariable modeling; (c) the broad age range of our participants, the similar behavioral characteristics between HIV groups, and the exclusion of other chronic infections (HCV and HBV) enhanced our ability to detect independent associations with age, smoking, and/or HIV.

However, the study has some limitations. As this is an observational cross-sectional study, we cannot infer causality. Furthermore, we cannot tease apart the possible effect of HIV versus that of cART. Moreover, our analysis was restricted to mtDNA substitutions; hence, we cannot comment on mtDNA insertions and deletions. Since our analysis was restricted to the D-loop region, we cannot generalize the observed mutation rates to the rest of the mitochondrial genome. Because we studied whole blood, we cannot address the distribution of mutations among blood cell subsets or platelets, and these could change with age, smoking, or HIV. Finally, because our study only included female participants, it may not be generalizable to males.

4 | CONCLUSIONS

Low-frequency mtDNA somatic substitution mutations can be quantified in blood cells, a tissue with fast turnover and lower mutation rates (Li, Schröder, Ni, Madea, & Stoneking, 2015). Our data suggest that the blood somatic mtDNA mutation burden increases with age, consistent with current thinking on biological aging, and may be increased among persons living with uncontrolled HIV. Our findings with respect to mtDNA heteroplasmy are consistent with the hypothesis that smoking promotes clonal expansion of mtDNA mutations, which may play a role in accelerating biological aging among smokers.

5 | MATERIALS AND METHODS

5.1 | Study sample

Study participants were women and girls living with HIV (HIV-positive) and HIV-uninfected controls (HIV-negative) enrolled in the prospective CARMA cohort. The study protocol was approved by the University of British Columbia (UBC) Clinical Research Ethics Board (H09-02867 and H08-02028). Each participant provided written informed consent prior to enrollment in the study. Details of the CARMA cohort study have been previously published (Cote et al., 2012; Zanet et al., 2014). Blood samples were collected between December 2008 and July 2011 in addition to relevant clinical, behavioral, and demographic information. For HIV-negative participants under 19 years of age, leftover blood from routine testing was obtained anonymously from British Columbia (BC) Children's Hospital, and only age and sex were known for these participants (Cote et al., 2012). The inclusion criteria are presented in Figure 5.

Among all adult CARMA cohort participants, all cis-gender females who were either current tobacco smokers or never smoked

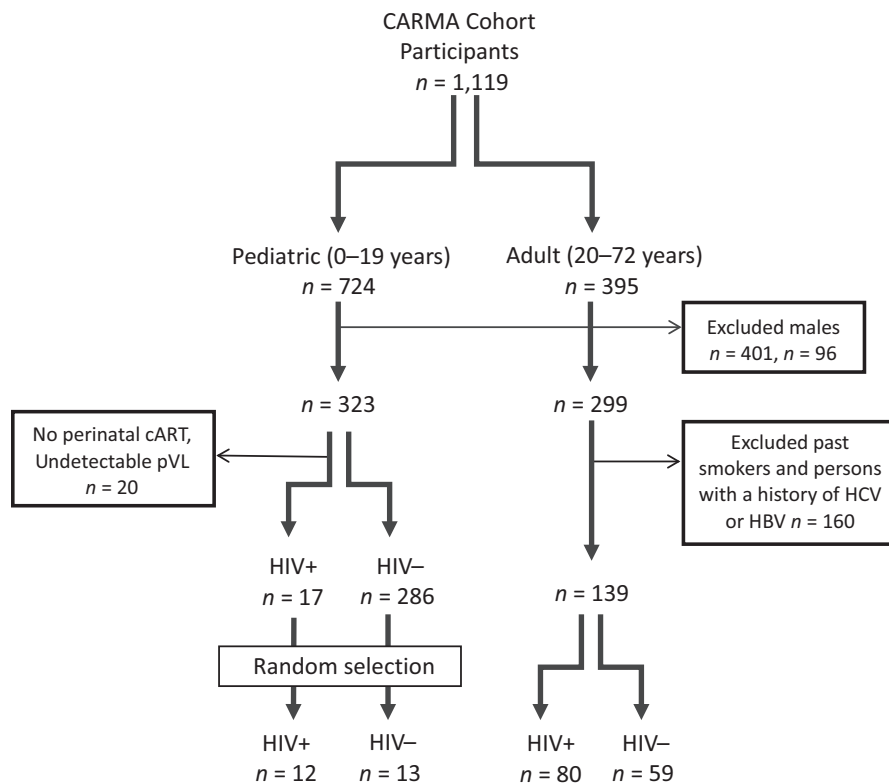


FIGURE 5 Study participants and the inclusion/exclusion criteria applied to samples analyzed. Of 1,119 CARMA cohort participants, 724 children and 395 adults were available for inclusion. As the cohort was predominantly female (75%), only females were included. Among the adults, past smokers were excluded, as were participants with a history of hepatitis C virus (HCV) or hepatitis B virus (HBV) infection, to focus on the effect of HIV. Among the 323 pediatric participants, due to restricted space in the sequencing runs, 12 HIV-positive and 13 HIV-negative participants were randomly selected for inclusion, all of whom were assumed HCV and HBV uninfected

(no history of ever smoking), and with no history of infection with hepatitis B or C viruses (as per self-report and/or serology), were included. The remaining 25 participants were selected among CARMA female pediatric participants: 12 of 37 HIV-positive individuals who were not exposed to antiretroviral agents perinatally and 13 of 286 anonymous HIV-negative controls. Pediatric participants were assumed to be hepatitis B- and C-uninfected, and to not drink, use drugs, or smoke. In addition to HIV infection status, age, BMI, self-reported ethnicity, drug use, alcohol use, and smoking status were also included in analyses presented here.

For HIV-positive participants, CD4 count at time of sampling, CD4 nadir (lowest CD4 count on record), HIV plasma viral load (pVL) at time of sampling, and peak viral load (highest pVL on record) were also considered. Combination antiretroviral therapy regimen was categorized as protease inhibitor (PI)-based, non-nucleoside reverse transcriptase inhibitor (non-NRTI)-based, other class-based, cART-naïve, or off cART.

In this study, HIV status was trichotomized: HIV-negative, HIV-positive with peak pVL < 100,000 copies/ml, and HIV-positive with peak pVL ≥ 100,000 copies/ml. This categorization was chosen based on our previous observation that having a peak HIV pVL > 100,000 copies/ml was associated with shorter leukocyte telomere length (Zanet et al., 2014), another marker of aging.

5.2 | Somatic mtDNA substitution mutation assay

5.2.1 | DNA extraction and mtDNA quantification

All blood specimens were stored at -80°C until DNA extraction. Total DNA was extracted from 0.1 ml whole blood using the

QIAamp DNA Mini Kit (QIAGEN, Mississauga, Ontario, Canada) on the QIAcube (QIAGEN) and eluted in 100 μl of kit AE buffer, according to the manufacturer's protocol. Each extract was then diluted 1:100 in AE (elution buffer), and mtDNA copy number was quantified by qPCR on a LightCycler 480 (Roche) with the LightCycler 480 SYBR Green I Master Kit (Roche). In brief, 2 μl of DNA was assayed in a 10 μl reaction that contained primers MT325F and MT474R (see Table S4) at a final concentration of 1 μM . The qPCR conditions were as follows: one cycle of $95^{\circ}\text{C}/10$ min, followed by 45 cycles of $95^{\circ}\text{C}/5$ s; $60^{\circ}\text{C}/10$ s; and $72^{\circ}\text{C}/5$ s, with acquisition at the end of the 72°C step. Temperature ramping was $4.4^{\circ}\text{C}/\text{s}$ for all steps except the 95 to 60°C step which ramped at $2.2^{\circ}\text{C}/\text{s}$. Each extract was quantified in duplicate, and the mtDNA copy numbers were determined based on a standard curve consisting of a plasmid containing the cloned amplicon, serially diluted 1:10 with a linear range extending from 3.34×10^7 to 3.34×10^1 copies/ μl .

5.2.2 | Labeling with Primer IDs via single-cycle primer extension

The 25 μl primer extension reaction contained 5×10^5 copies of mtDNA, 0.5 μl of PfuUltra II fusion HS DNA polymerase and 2.5 μl of its 10 \times Buffer (Agilent Technologies), 0.25 mM deoxyribonucleotide triphosphate (dNTP) (Invitrogen), and 0.125 μM of a given HPLC-purified Extension Primer (Integrated DNA Technologies) which binds the H strand of the D-loop and extends from bp16560 of the mtDNA genome (see Table S5) according to the revised Cambridge reference sequence (<https://www.mitomap.org/foswiki/bin/view/MITOMAP/HumanMitoSeq>). The D-loop region was chosen as it was less likely

to be deleted compared to the rest of mitochondrial genome or have pseudogenes. A negative control designed to estimate assay background consisted of 5×10^5 copies of a clonal plasmid DNA containing the mtDNA region of interest. Both sequencing experiments also included one or two internal controls consisting of whole blood DNA extracted from the same individual(s). Amplification conditions were 1 cycle of 95°C/140s, 60°C/20s, and 72°C/195s. The reactions were then stored at -80°C.

For every 12 participant DNA extracts assayed, three negative controls were generated by replacing the mtDNA, the extension primer, or the DNA polymerase with water. These controls were used to estimate the background incorporation of the extension primers and were taken through the subsequent steps of the assay. To prevent cross-contamination, DNA extract labeled with the same multiplex identifier (MID) was never manipulated at the same time, and surfaces/instruments were thoroughly cleaned between handlings.

5.2.3 | AMPure Purification

To remove unincorporated extension primer, each reaction was combined with 50 μ l of UltraPure distilled water (Invitrogen) and 50 μ l of AMPure beads (Beckman Coulter) for a 10-min incubation at room temperature, then placed into the DynaMag-2 Magnet (Life Technologies) for 5 min, as per the manufacturer's protocol. In brief, the supernatant was removed and 200 μ l of 70% ethanol was added before lightly vortexing for 5 s and placing the sample on the magnetic particle collector for 1 min. This step was repeated once more, after which the remaining supernatant was removed and the tube was placed in a 37°C heating block for 5 min. The purified product was eluted by addition of 20 μ l of 10 mM Tris and 2 mM EDTA pH 7.5 for a 2-min incubation on the magnetic particle collector. The supernatant contains the purified mtDNA, which was stored at -80°C.

5.2.4 | qPCR quantification of PID-labeled mtDNA templates

The undiluted AMPure-purified and PID-labeled mtDNA templates were quantified via qPCR as described above using the short KSF and MT48F primers (Table S4), both at a final concentration of 1 μ M. The standard curve consisted of serial 1:10 dilutions of plasmid containing the cloned amplicon, with a linear dynamic range of 1.45×10^7 to 1.45×10^1 copies/ μ l. The controls lacking DNA template, enzyme, and primer were assayed together with the respective study participant DNA extracts. The highest copy number observed among these controls (typically $\leq 10\%$ of the participant DNA extracts) was subtracted from the copy number of the same amplified DNA sample to estimate the PID-labeled mtDNA concentration.

5.2.5 | Preparation of amplicons for next-generation sequencing

Each 50 μ l PCR contained 1 μ l of PfuUltra II fusion HS DNA polymerase and 5 μ l of its 10 \times Buffer (Agilent Technologies), 0.25 mM

dNTP, 0.5 μ M of primers LAKSF and LBDLR (Table S4), and 2,000 copies of PID-labeled mtDNA. Amplification conditions were 1 cycle at 95°C/2 min, followed by 35 cycles of 95°C/20s, 60°C/20s, and 72°C/15s, and 1 cycle of 72°C/2 min. The PCR products were stored at -20°C until gel extraction.

A 2% agarose gel with a 0.5 \times final concentration of GelRed Nucleic Acid Gel Stain 10,000 \times (Biotium) was run until clear separation of the 606 bp amplicon from the lower molecular weight primer-dimer-like material was observed. The desired band was rapidly excised under low-intensity UV illuminator, and DNA was extracted from the gel using the QIAquick Gel Extraction Kit (QIAGEN), according to the manufacturer's protocol. Columns were wiped dry prior to elution, and incubation with elution buffer was extended to 10 min prior to the final spin. To avoid cross-contamination within and between gel extractions, no two participants' DNA extracts with the same MID were run in the same gel chamber, and new scalpels were used to excise each band. Between gels, running buffer was changed and all combs, gel trays, and chambers were thoroughly washed.

Each gel-extracted product was diluted 1:40,000 in QIAGEN AE buffer and quantified by qPCR as described above, using primers MT325F and MT474R (Table S4) at 1 μ M. Purified PCR products were diluted to the lowest common concentration, and 10 μ l of amplicons with different PIDs was pooled. Typically, a sequencing run has approximately 10,000 reads per DNA extract. Assuming 120,000 reads per lane and 8 lanes per run, 12 amplicons were pooled in each of the 8 libraries for a total of 96 DNA samples per run.

5.2.6 | Single read GS FLX Sequencing

The pooled libraries were AMPure-purified (Beckman Coulter) twice according to the manufacturer's protocol. DNA quality was then assessed via the Agilent High Sensitivity DNA Kit according to the manufacturer's instructions (Agilent Technologies). To assess the quality of the purification, each library was quantified using the KAPA Library Quantification Kit for the Roche 454 GS titanium platform as per the manufacturer's instructions (KAPA Biosystems). The libraries were then amplified following the Roche emPCR amplification method manual—Lib L SV for GS FLX + series -XL+ (May 2011)—and unidirectional sequencing was performed according to the Roche sequencing method manual for the GS FLX + instrument (May 2011). Data were processed via the default shotgun pipeline of the 454 Sequencing System Software v2.9 and output in FNA file format. All data for the present study were obtained from two full GS FLX runs, each run containing specimens from 82 study participants and 16 assay-specific controls. The number of adults and children, HIV-positive and HIV-negative, and smokers/never smokers were not significantly different between the two runs. For each run, samples were selected such that the median age, and the number of smokers and participants living with HIV were similar in each run although the second run tended to have fewer children.

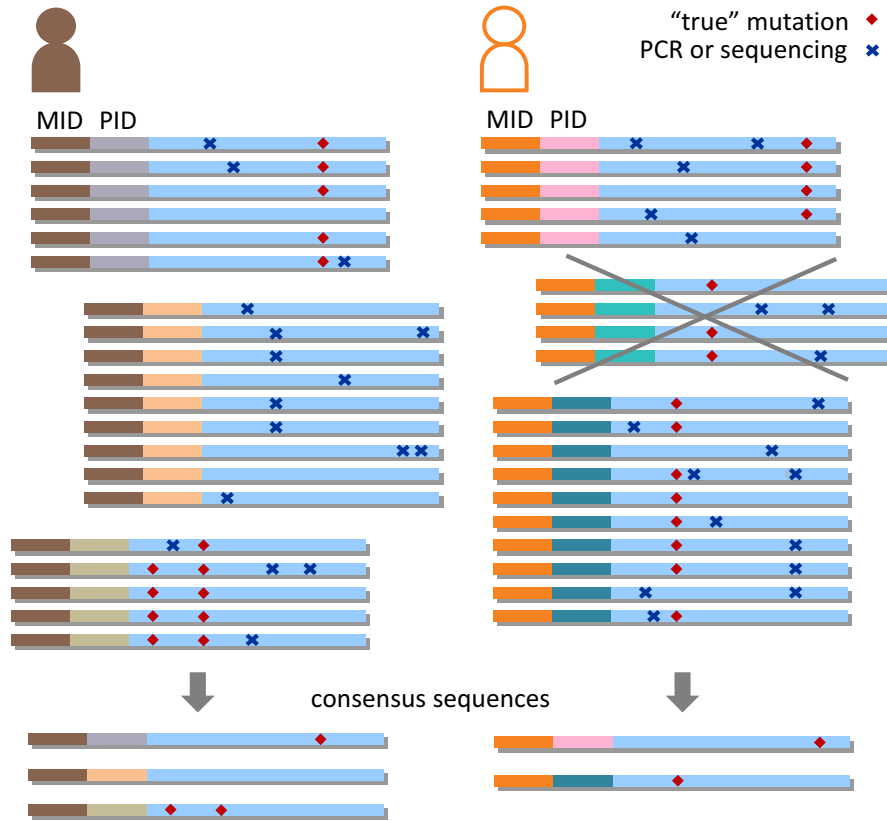


FIGURE 6 Schematic of the mtDNA mutation assay. Sequences belonging to a given individual are grouped according to their multiplex identifier (MID) and according to their mtDNA primer identifier (PID). A consensus sequence is generated for each individual, based on all sequence groups containing at least five reads. Each group's consensus sequence is then compared to the individuals' consensus sequence. If a mutation (i.e., not present in the individual's consensus sequence) is present in $\geq 75\%$ of all reads belonging to a given group, it is categorized as a "true" mutation which is presumed to have been present in the original mtDNA molecule, at the time of primer labeling. All other mutations are deemed to represent PCR errors that occurred after the initial single-cycle primer extension step or sequencing errors

5.2.7 | Data analysis with primer IDs

FNA files were divided into 8 lanes, then split into multiple participant files based on their MID. These were subsequently sorted into groups based on their PID sequence. Consensus sequences were built for each group of sequences with identical PIDs, defined as the 19 bp immediately following the MID. These were then used to build a consensus sequence for each participant, based on all sequences sharing the same MID in a given lane. Sequences that did not reach position 279 in the mtDNA genome or did not contain one of the 12 MIDs used (Table S5) within the first 100 bp were removed from the analysis.

Mafft sequence aligner within Unipro UGENE v1.24.1 (<http://ugene.net/>) was used, with the opening penalty set at 0.05, the extension penalty at 0.05, and the max iterations at 10. PID consensus sequences were built from ≥ 5 distinct sequences or reads. Each participant's consensus sequence was built from ≥ 100 PID consensus sequences. A mutation was defined as any difference between the participant DNA extract's consensus sequence and the PID's consensus sequence, if the mutation was present in $\geq 75\%$ of the ≥ 5 sequences used to build the PID consensus sequence. These thresholds were established a priori, based on pilot work. Only substitution mutations between bp 16,560 and 279 (see Figure S7) of the mitochondrial D-loop, based on the revised Cambridge reference sequence, were analyzed. Somatic mtDNA substitution mutations were defined as mutations occurring in $\leq 2\%$ of PID consensus sequences for a given participant. Mutations detected in the same position in $>2\%$ of PID consensus sequences were defined as heteroplasmic. Within the current study, mtDNA mutations refer to mutations present on the single strand

analyzed. Insertions and deletions were not considered in this analysis. Both code and data used in this manuscript are available upon request.

Our main outcome of interest was somatic mtDNA substitution mutation burden, reported as substitutions per 10,000 bp, extrapolated from the 289-bp interrogated region in the mitochondrial D-loop. Only PID groups with ≥ 5 sequences were considered for analysis. As such, the somatic mtDNA substitution frequency = (number of somatic mtDNA substitutions \times 10,000) divided by (number of distinct PID groups with 5 or more sequences \times 289 bp).

mtDNA heteroplasmy was reported as a binary variable (yes/no). Figure 6 summarizes the principle of the assay, which exploits the fact that true mutations present in the original mtDNA template will be present in a majority of sequences sharing the same PID (established as $\geq 75\%$ here). Mutations caused by PCR errors after the single-cycle primer extension step or sequencing errors will result in a scattered pattern within PID groups.

5.3 | Validation in an independent cohort

Furthermore, we validated our results using data from an independent cohort enrolled at Yale New Haven Hospital, New Haven, Connecticut, USA (Li et al., 2017) (Li et al., 2017). We included in the current analysis data from study participants living with HIV who were either current smokers or never smokers. Although the original study included HIV-negative controls, smoking status was not available for those individuals. Only substitution mutations within the same 289-bp region of the mitochondrial D-loop were considered.

5.4 | Statistics for somatic mtDNA substitutions

Given the non-normal distribution of the raw data, somatic mtDNA substitution frequencies were $\ln(x + 1)$ -transformed, where x is the somatic mtDNA substitution frequency per 10,000 bp. Comparisons between the HIV-positive and HIV-negative groups were done using the Mann–Whitney U and Fisher's exact tests as appropriate. Univariate associations between (transformed) somatic mtDNA substitution mutation burdens and possible explanatory variables were investigated with linear regressions (age, log-transformed BMI, current CD4 count), Spearman's correlation (CD4 nadir), two-sample t or Mann–Whitney U tests (smoking status, current HIV pVL, peak pVL), one-way ANOVA, or Kruskal–Wallis (trichotomized HIV status, HIV treatment status, and ethnicity), as appropriate. An a priori decision was made to include age, smoking status, and HIV status in all models. Other variables were included in the ANCOVA multivariable model if they were important univariately ($p < .1$), and improved the fit of the multivariable model.

5.5 | Statistics for heteroplasmic mtDNA substitutions

Mann–Whitney U or Fisher's exact tests were used to compare groups and assess univariate associations in the presence or absence of heteroplasmy. For the multivariable models, variables were selected as above for inclusion in the binary logistic regression model.

5.6 | Statistics used to analyze specific mtDNA substitution types

Mann–Whitney U tests or Kruskal–Wallis tests with a Dunn's correction for multiple comparisons were used to compare the somatic mutation frequency of specific types of mutations. Spearman's correlation was used to assess their relationship with age.

5.7 | Statistical software

Statistical analyses were performed with XLSTAT Version 2013.1.01 (Addinsoft) and JMP software, v. 12.2.0 (SAS Institute), with the exception of Dunn's and Tukey's corrections, which were done with GraphPad Prism version 7.01 (GraphPad Software).

ACKNOWLEDGEMENTS

Funding for the study was from the Canadian Institutes of Health Research (CIHR) grants HOP-123801 (to HCFC) and TCO-125269 (to HCFC and co-PIs). ASZ was supported by scholarships from the University of British Columbia (UBC) Department of Pathology and Laboratory Medicine, the UBC Centre for Blood Research, the Genome Sciences and Technology (GSAT), National Science and Engineering Research Council (NSERC), and the UBC Faculty of Medicine. EP was supported by grant from the National Institutes of Health (K08AI074404). We thank all the CARMA participants,

the CARMA investigators, and research staff. Special thanks to Edmund Gu and Abhinav Ajaykumar. All sequencing was done at the Vancouver Prostate Centre, and our thanks go to Dr. Stephane Le Bian and Brian McConeghy for their support.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

A.S.Z., MY.L., S.S., P.R.H., and H.C.F.C. designed research; A.S.Z., M.L., O.O., and B.S. performed research; A.S.Z. and J.I.M. contributed new reagents/analytic tools; A.S.Z., S.K., and H.C.F.C. analyzed data; A.S.Z., E.P., A.Y.Y.H., and H.C.F.C. wrote the paper; A.S.Z., MY.L., S.S., P.R.H., and H.C.F.C. designed the NGS assay; E.P. designed the validation study; M.L. and O.O. revised the manuscript.

ORCID

Elijah Paintsil  <https://orcid.org/0000-0002-8283-4564>

Onyema Ogbuagu  <https://orcid.org/0000-0001-7342-2608>

P. Richard Harrigan  <https://orcid.org/0000-0003-2631-3390>

Steve Kalloger  <https://orcid.org/0000-0003-1332-7881>

Hélène C. F. Côté  <https://orcid.org/0000-0002-0399-5141>

REFERENCES

- Ances, B. M., Vaida, F., Yeh, M. J., Liang, C. L., Buxton, R. B., Letendre, S., ... Ellis, R. J. (2010). HIV infection and aging independently affect brain function as measured by functional magnetic resonance imaging. *The Journal of Infectious Diseases*, 201(3), 336–340. <https://doi.org/10.1086/649899>
- Antiretroviral Therapy Cohort Collaboration (2008). Life expectancy of individuals on combination antiretroviral therapy in high-income countries: a collaborative analysis of 14 cohort studies. *Lancet*, 372(9635), 293–299. [https://doi.org/10.1016/S0140-6736\(08\)61113-7](https://doi.org/10.1016/S0140-6736(08)61113-7)
- Basu, A. K., Loechler, E. L., Leadon, S. A., & Essigmann, J. M. (1989). Genetic effects of thymine glycol: site-specific mutagenesis and molecular modeling studies. *Proceedings of the National Academy of Sciences*, 86(20), 7677–7681. <https://doi.org/10.1073/pnas.86.20.7677>
- Bernhard, D., Moser, C., Backovic, A., & Wick, G. (2007). Cigarette smoke—an aging accelerator? *Experimental Gerontology*, 42(3), 160–165. <https://doi.org/10.1016/j.exger.2006.09.016>
- Bhatia, R., Ryscavage, P., & Taiwo, B. (2012). Accelerated aging and human immunodeficiency virus infection: Emerging challenges of growing older in the era of successful antiretroviral therapy. *Journal of Neurovirology*, 18(4), 247–255. <https://doi.org/10.1007/s13365-011-0073-y>
- Coller, H. A., Khrapko, K., Torres, A., Frampton, M. W., Utell, M. J., & Thilly, W. G. (1998). Mutational spectra of a 100-base pair mitochondrial DNA target sequence in bronchial epithelial cells: A comparison of smoking and nonsmoking twins. *Cancer Research*, 58(6), 1268–1277.
- Côté, H. C. F., Soudeyns, H., Thorne, A., Alimenti, A., Lamarre, V., Maan, E. J., ... Forbes, J. (2012). Leukocyte telomere length in HIV-infected

- and HIV-exposed uninfected children: Shorter telomeres with uncontrolled HIV viremia. *PLoS ONE*, 7(7), e39266. <https://doi.org/10.1371/journal.pone.0039266>
- Evans, M. D., Dizdaroglu, M., & Cooke, M. S. (2004). Oxidative DNA damage and disease: Induction, repair and significance. *Mutation Research*, 567(1), 1–61. <https://doi.org/10.1016/j.mrrev.2003.11.001>
- Fox, E. J., Reid-Bayliss, K. S., Emond, M. J., & Loeb, L. A. (2014). Accuracy of next generation sequencing platforms. *Next Generation Sequencing & Applications*, 1, 1000106. <https://doi.org/10.4172/jngsa.1000106>
- Fraga, C. G., Shigenaga, M. K., Park, J. W., Degan, P., & Ames, B. N. (1990). Oxidative damage to DNA during aging: 8-hydroxy-2'-deoxyguanosine in rat organ DNA and urine. *Proceedings of the National Academy of Sciences*, 87(12), 4533–4537. <https://doi.org/10.1073/pnas.87.12.4533>
- Gerschman, R., Gilbert, D. L., Nye, S. W., Dwyer, P., & Fenn, W. O. (1954). Oxygen poisoning and X-irradiation: A mechanism in common. *Science*, 119(3097), 623–626. <https://doi.org/10.1126/science.119.3097.623>
- Guaraldi, G., Orlando, G., Zona, S., Menozzi, M., Carli, F., Garlassi, E., ... Palella, F. (2011). Premature age-related comorbidities among HIV-infected persons compared with the general population. *Clinical Infectious Diseases*, 53(11), 1120–1126. <https://doi.org/10.1093/cid/cir627>
- Harman, D. (1956). Aging: A theory based on free radical and radiation chemistry. *Journal of Gerontology*, 11(3), 298–300. <https://doi.org/10.1093/geronj/11.3.298>
- Hiatt, J. B., Patwardhan, R. P., Turner, E. H., Lee, C., & Shendure, J. (2010). Parallel, tag-directed assembly of locally derived short sequence reads. *Nature Methods*, 7(2), 119–122. <https://doi.org/10.1038/nmeth.1416>
- Hsue, P. Y., Lo, J. C., Franklin, A., Bolger, A. F., Martin, J. N., Deeks, S. G., & Waters, D. D. (2004). Progression of atherosclerosis as assessed by carotid intima-media thickness in patients with HIV infection. *Circulation*, 109(13), 1603–1608. <https://doi.org/10.1161/01.CIR.0000124480.32233.8A>
- Jabara, C. B., Jones, C. D., Roach, J., Anderson, J. A., & Swanstrom, R. (2011). Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proceedings of the National Academy of Sciences*, 108(50), 20166–20171. <https://doi.org/10.1073/pnas.1110064108>
- Jitratkosol, M. H. J., Sattha, B., Maan, E. J., Gadawski, I., Harrigan, P. R., Forbes, J. C., ... Côté, H. C. F. (2012). Blood mitochondrial DNA mutations in HIV-infected women and their infants exposed to HAART during pregnancy. *AIDS*, 26(6), 675–683. <https://doi.org/10.1097/QAD.0b013e32835142eb>
- Kennedy, S. R., Salk, J. J., Schmitt, M. W., & Loeb, L. A. (2013). Ultra-sensitive sequencing reveals an age-related increase in somatic mitochondrial mutations that are inconsistent with oxidative damage. *PLoS Genetics*, 9(9), e1003794. <https://doi.org/10.1371/journal.pgen.1003794>
- Kiyosawa, H., Suko, M., Okudaira, H., Murata, K., Miyamoto, E., Chung, M.-H., ... Nishimura, S. (1990). Cigarette smoking induces formation of 8-hydroxydeoxyguanosine, one of the oxidative DNA damages in human peripheral leukocytes. *Free Radical Research Communications*, 11(1–3), 23–27. <https://doi.org/10.3109/10715769009109664>
- Kreutzer, D. A., & Essigmann, J. M. (1998). Oxidized, deaminated cytosines are a source of C → T transitions in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 3578–3582.
- Lee, H. C., Lu, C. Y., Fahn, H. J., & Wei, Y. H. (1998). Aging- and smoking-associated alteration in the relative content of mitochondrial DNA in human lung. *FEBS Letters*, 441(2), 292–296. [https://doi.org/10.1016/S0014-5793\(98\)01564-6](https://doi.org/10.1016/S0014-5793(98)01564-6)
- Lewis, P. D., Fradley, S. R., Griffiths, A. P., Baxter, P. W., & Parry, J. M. (2002). Mitochondrial DNA mutations in the parotid gland of cigarette smokers and non-smokers. *Mutation Research*, 518(1), 47–54. [https://doi.org/10.1016/S1383-5718\(02\)00066-9](https://doi.org/10.1016/S1383-5718(02)00066-9)
- Li, M., Foli, Y., Liu, Z., Wang, G., Hu, Y., Lu, Q., ... Paintsil, E. (2017). High frequency of mitochondrial DNA mutations in HIV-infected treatment-experienced individuals. *HIV Medicine*, 18(1), 45–55. <https://doi.org/10.1111/hiv.12390>
- Li, M., Schröder, R., Ni, S., Madea, B., & Stoneking, M. (2015). Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. *Proceedings of the National Academy of Sciences*, 112(8), 2491–2496. <https://doi.org/10.1073/pnas.1419651112>
- Loft, S., Vistisen, K., Ewertz, M., Tjønneland, A., Overvad, K., & Poulsen, H. E. (1992). Oxidative DNA damage estimated by 8-hydroxydeoxyguanosine excretion in humans: Influence of smoking, gender and body mass index. *Carcinogenesis*, 13(12), 2241–2247. <https://doi.org/10.1093/carcin/13.12.2241>
- Longley, M. J., Nguyen, D., Kunkel, T. A., & Copeland, W. C. (2001). The fidelity of human DNA polymerase gamma with and without exonucleolytic proofreading and the p55 accessory subunit. *The Journal of Biological Chemistry*, 276(42), 38555–38562. <https://doi.org/10.1074/jbc.M105230200>
- Marcelino, L. A., & Thilly, W. G. (1999). Mitochondrial mutagenesis in human cells and tissues. *Mutation Research*, 434(3), 177–203. [https://doi.org/10.1016/S0921-8777\(99\)00028-2](https://doi.org/10.1016/S0921-8777(99)00028-2)
- Martin, A. M., Hammond, E., Nolan, D., Pace, C., Den Boer, M., Taylor, L., ... Mallal, S. (2003). Accumulation of mitochondrial DNA mutations in human immunodeficiency virus-infected patients treated with nucleoside-analogue reverse-transcriptase inhibitors. *American Journal of Human Genetics*, 72(3), 549–560. <https://doi.org/10.1086/367849>
- Michikawa, Y., Mazzucchelli, F., Bresolin, N., Scarlato, G., & Attardi, G. (1999). Aging-dependent large accumulation of point mutations in the human mtDNA control region for replication. *Science*, 286(5440), 774–779. <https://doi.org/10.1126/science.286.5440.774>
- Miquel, J., Economos, A. C., Fleming, J., & Johnson, J. E. J. (1980). Mitochondrial role in cell aging. *Experimental Gerontology*, 15(6), 575–591. [https://doi.org/10.1016/0531-5565\(80\)90010-8](https://doi.org/10.1016/0531-5565(80)90010-8)
- Munsch, C., Müller-Höcker, J., & Kadenbach, B. (1993). Human aging is associated with various point mutations in tRNA genes of mitochondrial DNA. *Biological Chemistry Hoppe-Seyler*, 374(12), 1099–1104. <https://doi.org/10.1515/bchm3.1993.374.7-12.1099>
- Narayan, A., Carriero, N. J., Gettinger, S. N., Kluytenaar, J., Kozak, K. R., Yock, T. I., ... Patel, A. A. (2012). Ultrasensitive measurement of hotspot mutations in tumor DNA in blood using error-suppressed multiplexed deep sequencing. *Cancer Research*, 72(14), 3492–3498. <https://doi.org/10.1158/0008-5472.CAN-11-4037>
- Payne, B. A. I., Wilson, I. J., Hateley, C. A., Horvath, R., Santibanez-Koref, M., Samuels, D. C., ... Chinnery, P. F. (2011). Mitochondrial aging is accelerated by anti-retroviral therapy through the clonal expansion of mtDNA mutations. *Nature Genetics*, 43(8), 806–810. <https://doi.org/10.1038/ng.863>
- Petitti, D. B., & Kipp, H. (1986). The leukocyte count: Associations with intensity of smoking and persistence of effect after quitting. *American Journal of Epidemiology*, 123(1), 89–95. <https://doi.org/10.1093/oxfordjournals.aje.a114227>
- Polyak, K., Li, Y., Zhu, H., Lengauer, C., Willson, J. K. V., Markowitz, S. D., ... Vogelstein, B. (1998). Somatic mutations of the mitochondrial genome in human colorectal tumours. *Nature Genetics*, 20(3), 291–293. <https://doi.org/10.1038/3108>
- Schmitt, M. W., Kennedy, S. R., Salk, J. J., Fox, E. J., Hiatt, J. B., & Loeb, L. A. (2012). Detection of ultra-rare mutations by next-generation sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 109(36), 14508–14513. <https://doi.org/10.1073/pnas.1208715109>
- Schriner, S. E., Linford, N. J., Martin, G. M., Treuting, P., Ogburn, C. E., Emond, M., ... Rabinovitch, P. S. (2005). Extension of murine life span by overexpression of catalase targeted to mitochondria. *Science*, 308(5730), 1909–1911. <https://doi.org/10.1126/science.1106653>

- Serrano, S., Marinoso, M. L., Soriano, J. C., Rubies-Prat, J., Aubia, J., Coll, J., ... Goday, A. (1995). Bone remodelling in human immunodeficiency virus-1-infected patients. A histomorphometric study. *Bone*, *16*(2), 185–191.
- Shiels, M. S., Pfeiffer, R. M., & Engels, E. A. (2010). Age at cancer diagnosis among persons with AIDS in the United States. *Annals of Internal Medicine*, *153*(7), 452–460. <https://doi.org/10.7326/0003-4819-153-7-201010050-00008>
- Spelbrink, J. N., Toivonen, J. M., Hakkaart, G. A., Kurkela, J. M., Cooper, H. M., Lehtinen, S. K., ... Jacobs, H. T. (2000). In vivo functional analysis of the human mitochondrial DNA polymerase POLG expressed in cultured human cells. *The Journal of Biological Chemistry*, *275*(32), 24818–24828. <https://doi.org/10.1074/jbc.M000559200>
- Tan, D., Goerlitz, D. S., Dumitrescu, R. G., Han, D., Seillier-Moiseiwitsch, F., Spornak, S. M., ... Shields, P. G. (2008). Associations between cigarette smoking and mitochondrial DNA abnormalities in buccal cells. *Carcinogenesis*, *29*(6), 1170–1177. <https://doi.org/10.1093/carcin/bgn034>
- Trifunovic, A., Wredenberg, A., Falkenberg, M., Spelbrink, J. N., Rovio, A. T., Bruder, C. E., ... Larsson, N.-G. (2004). Premature ageing in mice expressing defective mitochondrial DNA polymerase. *Nature*, *429*(6990), 417–423. <https://doi.org/10.1038/nature02517>
- Wallace, D. C. (1999). Mitochondrial diseases in man and mouse. *Science*, *283*(5407), 1482–1488. <https://doi.org/10.1126/science.283.5407.1482>
- Yasui, M., Kanemaru, Y., Kamoshita, N., Suzuki, T., Arakawa, T., & Honma, M. (2014). Tracing the fates of site-specifically introduced DNA adducts in the human genome. *DNA Repair*, *15*, 11–20. <https://doi.org/10.1016/j.dnarep.2014.01.003>
- Zanet, D. A. L., Thorne, A., Singer, J., Maan, E. J., Sathya, B., Le Campion, A., ... Côté, H. C. F. (2014). Association between short leukocyte telomere length and HIV infection in a cohort study: No evidence of a

relationship with antiretroviral therapy. *Clinical Infectious Diseases*, *58*(9), 1322–1332. <https://doi.org/10.1093/cid/ciu051>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Ziada AS, Lu MY, Ignas-Menzies J, et al. Mitochondrial DNA somatic mutation burden and heteroplasmy are associated with chronological age, smoking, and HIV infection. *Aging Cell*. 2019;18:e13018. <https://doi.org/10.1111/ace1.13018>

APPENDIX A

Other investigators of the CIHR Team in Cellular Aging and HIV Comorbidities in Women and Children (CARMA) include: Neora Pick, MD; Melanie Murray, MD; Deborah Money, MD; Hugo Soudeyans, PhD; Fatima Kakkar, MD; Ari Bitnun, MD; Jason Brophy, MD; Patricia Janssen, PhD; Joel Singer, PhD; Normand Lapointe, MD; Jerilynn Prior, MD; Michael Silverman, MD; Mary Lou Smith, PhD; Heather Macdonald, PhD.