PLOS BIOLOGY

RESEARCH ARTICLE

Somatic nuclear mitochondrial DNA insertions are prevalent in the human brain and accumulate over time in fibroblasts

Weichen Zhou[1☯], Kalpita R. Karan[2☯], Wenjin Gu[1], Hans-Ulrich Klein[3,4], Gabriel Sturm[2,5], Philip L. De Jager[3,4], David A. Bennett[6], Michio Hirano[3], Martin Picard[2,7,8,9]*, Ryan E. Mills[1,10]*

1 Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, Michigan, United States of America, **2** Department of Psychiatry, Division of Behavioral Medicine, Columbia University Irving Medical Center, New York, New York, United States of America, **3** Center for Translational & Computational Neuroimmunology, Department of Neurology, Columbia University Irving Medical Center, New York, New York, United States of America, **4** Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Irving Medical Center, New York, New York, United States of America, **5** Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California, United States of America, **6** Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois, United States of America, **7** Department of Neurology, H. Houston Merritt Center, Columbia University Translational Neuroscience Initiative, Columbia University Irving Medical Center, New York, New York, United States of America, **8** New York State Psychiatric Institute, New York, New York, United States of America, **9** Robert N Butler Columbia Aging Center, Columbia University Mailman School of Public Health, New York, New York, United States of America, **10** Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan, United States of America

☯ These authors contributed equally to this work.
* martin.picard@columbia.edu (MP); remills@umich.edu (REM)

Check for updates

OPEN ACCESS

Citation: Zhou W, Karan KR, Gu W, Klein H-U, Sturm G, De Jager PL, et al. (2024) Somatic nuclear mitochondrial DNA insertions are prevalent in the human brain and accumulate over time in fibroblasts. PLoS Biol 22(8): e3002723. https://doi.org/10.1371/journal.pbio.3002723

Academic Editor: Thomas B.L. Kirkwood, University of Newcastle upon Tyne, UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND

Received: February 28, 2024

Accepted: June 26, 2024

Published: August 22, 2024

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pbio.3002723

Copyright: © 2024 Zhou et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The Whole genome sequence of ROSMAP can be obtained through the

Abstract

The transfer of mitochondrial DNA into the nuclear genomes of eukaryotes (Numts) has been linked to lifespan in nonhuman species and recently demonstrated to occur in rare instances from one human generation to the next. Here, we investigated numtogenesis dynamics in humans in 2 ways. First, we quantified Numts in 1,187 postmortem brain and blood samples from different individuals. Compared to circulating immune cells ($n = 389$), postmitotic brain tissue ($n = 798$) contained more Numts, consistent with their potential somatic accumulation. Within brain samples, we observed a 5.5-fold enrichment of somatic Numt insertions in the dorsolateral prefrontal cortex (DLPFC) compared to cerebellum samples, suggesting that brain Numts arose spontaneously during development or across the lifespan. Moreover, an increase in the number of brain Numts was linked to earlier mortality. The brains of individuals with no cognitive impairment (NCI) who died at younger ages carried approximately 2 more Numts per decade of life lost than those who lived longer. Second, we tested the dynamic transfer of Numts using a repeated-measures whole-genome sequencing design in a human fibroblast model that recapitulates several molecular hallmarks of aging. These longitudinal experiments revealed a gradual accumulation of 1 Numt every ~13 days. Numtogenesis was independent of large-scale genomic instability and unlikely driven by cell clonality. Targeted pharmacological perturbations including chronic glucocorticoid signaling or impairing mitochondrial oxidative phosphorylation (OxPhos) only modestly increased the rate of numtogenesis, whereas patient-derived *SURF1*-mutant cells

NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS, https://dss.niagads.org/) data set NG00067 or by contacting them directly at help@niagads.org. Whole genome sequences of aged primary human dermal fibroblasts can be found in a recent study at https://columbia-picard.shinyapps.io/shinyapp-Lifespan_Study/. Illumina-sequenced 2504 independent individuals from the 1000 Genomes Project Phase 3 can be found by contacting the project at info@1000genomes.org or downloading directly from ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/. The Numt callsets by Dinumt from 1000 Genomes Project (GRCh38) and from ROSMAP and Lifespan dataset (GRCh37), the SV callsets by DELLY, Manta, and Canvas from Lifespan dataset (GRCh37), and the MEI callset by MELT from Lifespan dataset (GRCh37) can be found at GitHub page: github.com/mills-lab/numts-and-aging-in-fibroblasts-and-brains-. The raw data for generating the plots in the figures and supplementary figures can be found in Supporting Information S1 Data. Dinumt: https://github.com/mills-lab/dinumt. The scripts and command lines in the project can be found on the GitHub page: github.com/mills-lab/numts-and-aging-in-fibroblasts-and-brains, and ZENODO: https://zenodo.org/records/11625510 with DOI https://doi.org/10.5281/zenodo.11625510.

Funding: This work was supported by NIA R01AG066828 (including salary support to K.R.K., G.S., M.H., and M.P.), 1R21HG011493-01 (including salary support to R.E.M. and W.Z.), and the BaszuckiBrain Research Fund. W.Z. was partially supported by and provided salary support from the NIH/NIA-funded Michigan Alzheimer's Disease Research Center (P30AG072931) and the University of Michigan Alzheimer's Disease Center Berger Endowment. ROSMAP is supported by P30AG10161, P30AG72975, R01AG15819, R01AG17917, U01AG46152, and U01AG61356. ROSMAP resources can be requested at https://www.radc.rush.edu. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: AC, anterior caudate; AD, Alzheimer's dementia; DLPFC, dorsolateral prefrontal cortex; MAP, Memory and Aging Project; MCI, mild cognitive impairment; MEI, mobile element insertion; mtDNA, mitochondrial DNA; NCI, no cognitive impairment; OxPhos, oxidative phosphorylation; PBMC, peripheral blood mononuclear cell; PCC, posterior cingulate cortex; ROS, Religious Orders Study; SNV, single-

exhibiting mtDNA instability accumulated Numts 4.7-fold faster than healthy donors. Combined, our data document spontaneous numtogenesis in human cells and demonstrate an association between brain cortical somatic Numts and human lifespan. These findings open the possibility that mito-nuclear horizontal gene transfer among human postmitotic tissues produces functionally relevant human Numts over timescales shorter than previously assumed.

Introduction

The incorporation of mitochondrial DNA into the nuclear genomes of organisms is an ongoing phenomenon [1–8]. These nuclear mitochondrial insertions, referred to as "Numts," have been observed in the germline of both human [6,8–11] and nonhuman [7,12–22] species. These insertions occur as part of a wider biological process termed numtogenesis [23,24], which has been defined as the occurrence of any mitochondrial DNA (mtDNA) components into the nucleus or nuclear genome. Once integrated, Numts are biparentally transmitted to future generations, like other types of genetic variation. While mostly benign, Numts have been implicated with cellular evolution and function [1,25], various cancers [23,24], and can confound studies of mitochondrial DNA heteroplasmy [26,27], maternal inheritance of mitochondria [10,28–31], and forensics [32–34].

Investigations of Numts have been conducted in numerous species, but yeast, in particular, has provided an excellent experimental platform as a model organism due to its smaller genome and fast replication timing. Mechanisms of Numt integration involve genome replication processes in several yeast species [1,2] and have further been linked to the yeast *YME1* (yeast mitochondrial escape 1) gene and double-stranded break repair [35–37]. Interestingly, Numts have also been associated with chronological aging in *Saccharomyces cerevisiae* [3], suggesting a model where the accumulation of somatic mutations with aging [38,39], particularly structural genomic changes, could provide an opportunistic environment for somatic numtogenesis.

In humans, neural progenitor cells and cortical neurons harbor extensive tissue-specific somatic mutations, including single-nucleotide variants (SNVs) [40–42], transposable elements [43–45], and larger structural variants [46–48]. However, to date, no studies have investigated the extent of Numts specific in human brain regions, though several studies have now explored somatic numtogenesis in various cancers [23,49]. Using blood as the source of DNA, rare events of germline numtogenesis leading to a new Numt absent from either parent are estimated to occur every 4,000 human births and to be more frequent in solid tumors but not hematological cancers [4]. Using the observations in yeast as a foundation, we hypothesized that the accumulation of somatic mutations with age in the human brain also could be associated with numtogenesis and an increase in the number of somatically acquired (i.e., de novo) Numts. Mechanistically, numtogenesis requires the release of mitochondrial fragments into the cytoplasm and nucleus [50,51], where they can be integrated into autosomal sequences. In this context, we note that neuroendocrine, energetic, and mitochondrial DNA maintenance stressors in human and mouse cells trigger mitochondrial DNA release into the cytoplasm [52] and even in the bloodstream [53,54]. Thus, intrinsic genetic perturbations to mitochondrial biology or environmentally induced stressors could increase numtogenesis across the lifespan.

nucleotide variant; SV, structural variation; VAF, variant allele frequency; WB, whole blood; WGS, whole genome sequencing.

We investigated these scenarios through a multifaceted approach using postmortem human brain tissue and blood from large cohorts of older individuals, as well as a longitudinal analysis of cultured primary human fibroblasts from healthy donors and patients deficient for *SURF1*, a gene associated with Leigh syndrome and cytochrome *c* oxidase deficiency [55] that alters oxidative phosphorylation (OxPhos). We further examined the potential role of environmental stress on numtogenesis through the treatment of these cells with oligomycin (OxPhos inhibitor) and dexamethasone (glucocorticoid receptor agonist).

Results

Somatic Numt integration differs by tissue, age, and cognitive status

We applied our Numt detection approach, *dinumt*, to whole genome sequencing (WGS) data generated in the ROSMAP cohort [56,57] comprising 466 dorsolateral prefrontal cortex (DLPFC), 260 cerebella, 68 posterior cingulate cortex (PCC), and 4 anterior caudate (AC) tissue samples as well as non-brain tissue from 366 whole blood (WB) and 23 peripheral blood mononuclear cells (PBMCs) samples (Methods, Fig 1A). We note that the sequencing coverage of these samples was too low (~45×, S1 Table) to confidently identify lower-level somatic mosaicism within any individual tissue sample, though some would still likely be detected by our approach. Our strategy instead was to examine whether there were any potential clonal mosaic events occurring predominantly in one or more tissues compared to others by filtering out all common germline Numts, with the expectation that any rare germline Numts that were not filtered out would be present at a constant rate across all tissues, age ranges, and cognitive categories and thus any differences we observe between them would be driven by somatic events.

We obtained 3,758 unique quality-pass Numt calls from ROSMAP 1,187 samples. Only 17 of these tissue samples were from the same individual, and thus typical somatic variant analysis using multiple tissues to distinguish germline and early somatic variation from tissue-specific events was prohibitive. To mitigate this, we cross-referenced all of our detected non-reference Numts against large population cohorts including the 1000 Genomes Project [6] and another recent study using the 100,000 Genomes Project in England [4]. This step identified 45 Numts shared between one of our samples and the population-level reference, which were filtered out (Methods). The remaining 3,713 Numts were then examined in aggregate across tissues to determine whether there were differences in Numt abundance between specific tissues. To identify and rectify any false positive calls that may have been triggered by potential bacterial mitochondria contamination [58], we further conducted an investigation into the length of paired-end read fragments mapped to the nuclear genome and supported Numt calls and found that in all cases (Methods), there were at least 150 bp of sequence anchored to human chromosomal sequences (S1 Fig), demonstrating that there is likely no microbiome interference in the data.

We identified a mean of 10.4 Numts per sample across all tissues, of which ~3 on average were found to be tissue-specific after filtering for germline Numt polymorphisms found in other samples, population-scale controls, or somatic Numts in other tissues, as described above. We observed no correlation between the number of Numts detected in each sample and its genomic sequence coverage ($r^2 = 0.003$, S1 Table), indicating that our results are robust across a range of sequence depths [59]. The detected Numts ranged from 22 bp to 8,172 bp in length, with a median of 73 bp, a mean of 1,169 bp, and s.d. 1,882 bp (S1 Table), consistent with previous results from population-scale data in blood DNA [4,6,7]. We observed an average of 3.52 Numts in each whole blood tissue sample, closely comparable to the 4.9 Numts in blood samples reported in Wei and colleagues [4].

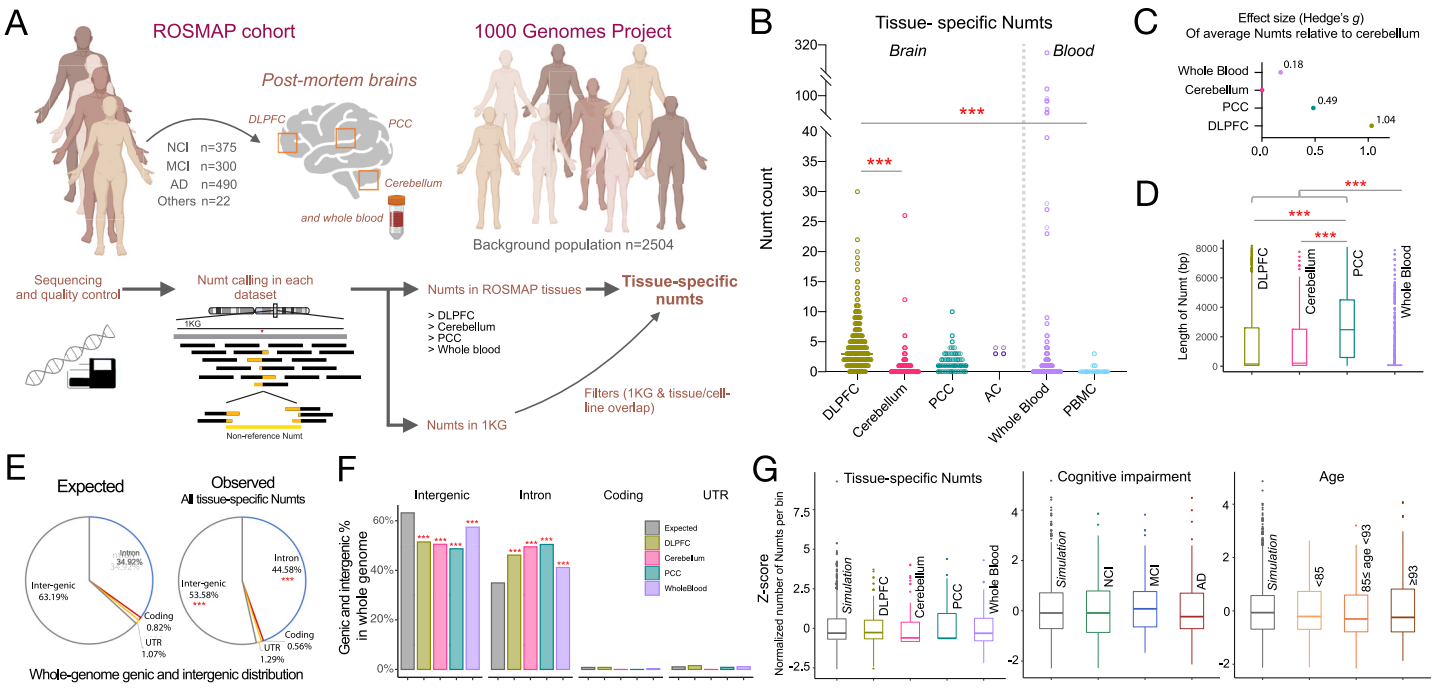

Fig 1. Characteristics of tissue-specific Numts in postmortem human brain regions and blood. (A) Overview of approach to identify tissue-specific Numts in ROSMAP cohort. (B) Abundance of tissue-specific Numts across brain regions and blood cells from ROSMAP participants. (C) Effect size (Hedge's *g*) of average tissue-specific Numts relative to cerebellum. (D) Length of tissue-specific Numts across brain regions and whole blood. (E) Genic and intergenic distribution of all tissue-specific Numts versus the expected distribution in the whole genome for all samples. (F) Percentage of genic and intergenic distribution of tissue-specific Numts delineated by brain regions and whole blood. (G) Random genomic distributions of tissue-specific Numts across tissues (left), cognitive impairment stratifications (middle), and age groups (right), based on comparison to simulation data. The age groups were defined as less than 85 (n = 295), between 85 and 93 (n = 545), and greater than or equal to 93 (n = 319). One-way ANOVA was used to test the significance in B and D. Pearson's chi-square test was used to test the significance in E. Fisher's exact test was employed instead of Pearson's chi-square test when sample sizes are small. ***, **, and * represent a significant p-value less than 0.001, 0.01, and 0.05, respectively. DLPFC, dorsolateral prefrontal cortex; PCC, posterior cingulate cortex. NCI, no cognitive impairment; MCI, mild cognitive impairment; AD, Alzheimer's dementia. Graphical artwork in Fig 1A were created with BioRender.com and are pursuant to BioRender's Academic License Terms. The data underlying this figure can be found in S1 Data.

https://doi.org/10.1371/journal.pbio.3002723.g001

Interestingly, we found the majority of tissue-specific Numts fell within DLPFC regions (mean = 4.13 per person), representing a 5.5-fold (p-value <0.001) higher frequency compared to the cerebellum (mean = 0.75), 2.4-fold higher than PCC (mean = 1.71, p-value <0.001), and 15.9-fold higher than PBMC (mean = 0.26, p-value <0.001) (Fig 1B). From the 17 individuals with multiple tissue samples comprising 9 pairs of the cerebellum and DLPFC samples, we consistently observed significant differences of tissue-specific Numts between the 2 tissues (S2 Fig, p-value = 0.033, Student's T test, paired, two-sided) which aligns with our larger analysis across the cohort. Relative to the cerebellum, DLPFC and AC also showed the highest effect sizes (Hedge's g = 1.03 and 1.38, respectively) in Numt abundance, followed by PCC (g = 0.49) (Fig 1C). In addition, we observed a significant correlation between the mtDNA copy number (mtDNAcn) and somatic Numts in DLPFC (r^2 = 0.146, $p < 0.001$), but not in other tissues (S3 Fig). mtDNAcn was higher in DLPFC than in other brain regions and tissues (median copies per cell: DLPFC = 4,047, versus cerebellum = 999, PCC = 4,042, WB = 172, and PBMCs = 125). Interestingly, we found that the length of tissue-specific Numts in 3 brain regions (DLPFC, cerebellum, and PCC) were significantly longer than those observed in whole blood (median = 63 bp, one-way ANOVA, p-value <0.001), with PCC-specific Numts themselves exhibiting larger lengths (median = 2,477 bp) than DLPFC and cerebellum (median = 152

bp and 210 bp, one-way ANOVA, p-value <0.001, Fig 1D). The absence of large Numts in blood immune cells could reflect negative selection against new Numts [4].

We next explored Numt integration using a gene-centric approach in the tissues with the largest number of samples (DLFPC, cerebellum, PCC, and WB). We focused on the Numts that were inserted in and around transcribed regions of the genome (introns, coding, UTR, and intergenic regions) and examined the proportion of our detected NUMTs that fell within each region. Surprisingly, we found that our somatic Numts integrated into introns at a significantly higher frequency compared to its overall expected proportion of the genome based on Ensembl gene annotations (44.58% v.s. 34.92%, p-value <0.001), while a negative enrichment was observed in intergenic regions (53.58% v.s. 63.92%, p-value <0.001, Methods, Fig 1E), though these could be the result of differences of sequence mappability within these regions precluding the detection of Numts. These significant differences in genomic distributions were further observed across the various tissues (Fig 1F).

We lastly hypothesized that the genomic distribution of somatic Numt integration sites might differ within individual tissues, cognitive status, or age groups from an expected random distribution throughout the genome. We tested this hypothesis by iteratively assigning random positions for each of our observed Numts 50,000 times across the human genome reference and assessing whether our observed integration sites differed significantly when compared within predefined 10 Mb windows across the genome as a permutation test. After the multiple test correction (Benjamini–Hochberg procedure), we observed no significant difference from random for any of the tested tissues (Fig 1G). We further stratified our results by age and cognitive status (Methods) and likewise observed no differences in genomic distribution. This is in agreement with previous studies that suggest Numt integration is a random occurrence [6,7,11], though we note that the paucity of Numts may lead to the permutation test being underpowered and thus preclude an accurate assessment at such broad regions across the genome.

DLPFC-specific Numts are negatively associated with age at death in persons without cognitive impairment

On the basis of potential adverse genomic effects of Numts [4,50,51] and the results above, we hypothesized that tissue-specific Numts were associated with mortality and age at death, though this correlation might differ between tissues or clinical diagnoses. We first examined NUMTs in aggregate across tissues and observed almost no correlation in Numt abundance with the age of death (Fig 2A). Given the existence of mitochondrial DNA defects alterations in the human brain with cognitive decline [60], we next stratified individuals with tissue-specific Numts by their cognitive status into no cognitive impairment (NCI), mild cognitive impairment (MCI), and Alzheimer's dementia (AD, COGDX score, Methods, S1 and S2 Tables). We found that in DLPFC tissues, NCI individuals that carried more Numts died earlier (r^2 = 0.094, p-value <0.001), with 2 additional Numt insertions observed per decade of life lost (Fig 2B). MCI individuals exhibited a similar but lower negative association (r^2 = 0.031, p-value <0.05). However, no correlation between Numts and age at death was observed in the AD group (r^2 = 0.009, p-value = 0.19). In the cerebellum, we observed similar patterns of correlation between Numts and age at death among cognitive groups, albeit with weaker correlations (NCI: r^2 = 0.044, p-value <0.05; MCI: r^2 = 0.007, p-value = 0.514; and AD: r^2 = 0.045, p-value <0.05, S4 Fig). These results indicate that Numts are negatively associated with age at death in certain brain regions of non-AD individuals, suggesting the possibility that brain numtogenesis is deleterious and that the pathogenicity of AD may be uncoupled from age-dependent Numt integration.

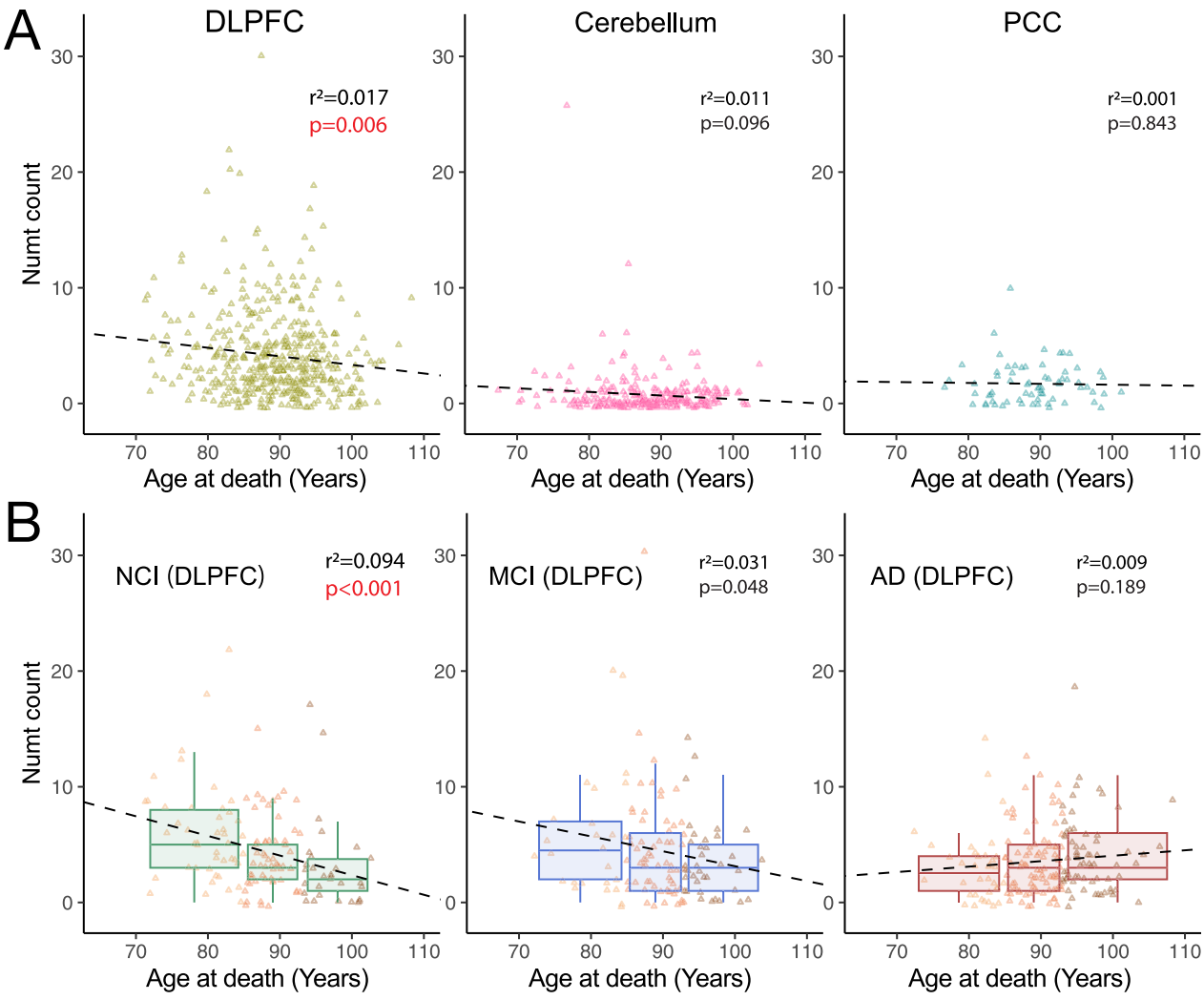

Fig 2. Numt association with age at death by tissue and cognitive impairment. (A) Correlation between age at death and abundance of DLPFC-specific, cerebellum-specific, and PCC-specific Numts, respectively. (B) DLPFC samples correlated with age at death, stratified by cognitive diagnosis status as NCI (n = 121, left), MCI (n = 112, middle), and AD (n = 176, right). Data points are colored by arbitrary age groups (see Methods) in light yellow, orange, and brown, respectively. r^2 and p-values are calculated using standard least-squares regression models. DLPFC, dorsolateral prefrontal cortex; PCC, posterior cingulate cortex. NCI, no cognitive impairment; MCI, mild cognitive impairment; AD, Alzheimer's dementia. The data underlying this figure can be found in S1 Data.

https://doi.org/10.1371/journal.pbio.3002723.g002

Somatic Numts accumulate over time in fibroblasts

The cross-tissue analysis of the ROSMAP cohort provided compelling results that Numts have both tissue and age-dependent characteristics of their integration in aggregate. However, the lack of relationships between individual tissue samples prohibits a direct measurement of Numt integration rates or numtogenesis. We therefore tested the dynamic transfer of Numts using a longitudinal, repeated-measures WGS study in primary human fibroblasts cultured in vitro under physiological conditions (Table 1) [5,61–63]. Over time, replicating cells exhibit conserved epigenomic (hypomethylation), telomeric (shortening), transcriptional (senescence-associated markers), and secretory (pro-inflammatory) features of human aging, representing a useful model to quantify the rate of dynamic age-related molecular processes in a human system [5]. We recently showed that primary mitochondrial bioenergetic defects

Table 1. Experimental design for Numt accumulation study in the in vitro fibroblast aging model.

Individual	Gender	Treatment	Collection day	# Data point
Control				
Donor 1	Male	None	7 ~ 211	8
		Dex	27 ~ 166	8
		Oligo	27 ~ 103	6
Donor 2	Female	None	8 ~ 189	8
		Dex	43 ~ 211	10
		Oligo	28 ~ 189	8
Donor 3	Male	None	3 ~ 189	10
		Dex	28 ~ 90	4
		Oligo	28 ~ 105	4
SURF1 defect				
Patient 1	Male	None	16 ~ 90	5
Patient 2	Male		8 ~ 152	8
Patient 3	Female		8 ~ 58	5

https://doi.org/10.1371/journal.pbio.3002723.t001

accelerate the rate of aging based on the telomere shortening per cell division, DNA methylation clocks, and age-related secreted proteins [63]. Therefore, using this model to monitor the accumulation of overall and donor-specific unique Numts absent in the general population, we analyzed cultured fibroblasts from 3 unrelated healthy donors, aged in culture under physiological conditions for up to 211 days [5] (Fig 3A). Instead of focusing on a single cell line tested in triplicates, we opted to include 3 separate donors, which provides a more robust test of our hypothesis.

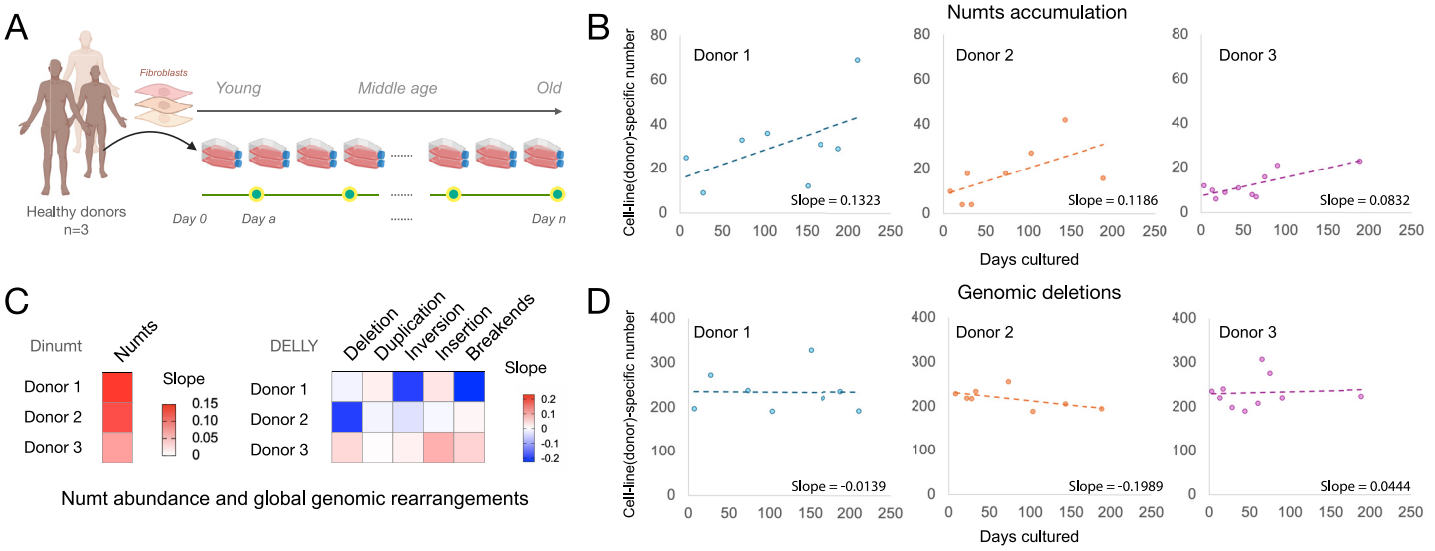

Fig 3. Numts accumulate in human fibroblasts during normal aging. (A) Study design of cellular aging model using primary fibroblasts from 3 healthy donors. (B) Cell line-specific Numts accumulate over time in aging fibroblasts obtained from 3 healthy donors cultured up to 211 days. (C) Heatmap of slopes based on the linear regression between days cultured and the cell line-specific calls, including Numts (from Dinumt, left) and structural variants (from DELLY, right). (D) Time-course of cell line-specific deletions in the 3 healthy donors. Graphical artwork in Fig 3A were created with BioRender.com and are pursuant to BioRender's Academic License Terms. The data underlying this figure can be found in S1 Data.

https://doi.org/10.1371/journal.pbio.3002723.g003

Across the 3 donors, we observed a positive correlation between time in culture and the number of unique Numts (r^2 = 0.30, 0.38, and 0.59, respectively) with a positive slope range of 0.08 to 0.13 Numts/day. Thus, on average, human fibroblasts accumulated a novel Numt every 12.6 days of culture (or 0.79 Numt per 10 days, 95% C.I. = 0.28 to 1.31). Numtogenesis was also evident when delineating our Numts per donor into total Numts observed at each time point (total Numts) and Numt insertions specific to individual cell lines with treatment within each donor (cell line-specific Numts) (Figs 3B and S5).

To determine if global genomic instability could account for this effect, we conducted the same analysis on multiple types of somatic structural variation (e.g., deletion, duplication, inversion, insertion, and breakends; Methods). The accumulation rates of these variant types were significantly lower compared to the rate of numtogenesis (Fig 3C). For example, although we observed a positive yet more moderate increase in deletion abundance compared to Numts when considering all such variants, we did not observe the same increase in the cell line-specific deletions as with the somatic Numts (Fig 3D). These results indicate that Numt insertions occur at a higher rate than autosomal deletions in this system and suggest a higher rate of age-related somatic numtogenesis rate than the other genetic variants (S6 Fig). We further observed no significant increase in cell line-specific genomic duplications over time, thus indicating that the increase in the total number of Numts over time is likely due to novel integration events and not duplications of preexisting copies.

We questioned whether the accumulation of these apparently somatic Numts could be driven by the simple clonal expansion of few Numts-containing cells. Even within a given donor line followed longitudinally, all observed Numts were unique in their length and sequence (average 562 bp, s.d. 1,400 bp) and showed no evidence of relatedness with one another. This lack of sequence overlap is most parsimoniously explained by the random nature of our sequencing coverage (sequencing depth: 25×; the total number of genomes in each experiment 2×10^6 diploid genomes) and a large number of new Numts accumulating over time. Thus, the unique identity of all observed Numts in these in vitro experiments argues against the clonal origin of these events.

Impact of environmental and genetic stress on somatic Numt integration rates

We next explored whether the cellular environment could impact somatic numtogenesis by testing if chronic exposure to a stress-mimetic or an inhibitor of mitochondrial OxPhos would alter the rate of Numt accumulation in otherwise healthy and aging fibroblasts. We analyzed human fibroblasts derived from the same 3 healthy donors described above that were treated with (a) the glucocorticoid receptor agonist dexamethasone (Dex, 100 nM); and (b) the ATP synthesis inhibitor oligomycin (Oligo, 100 nM). Similar to the untreated donors (see Methods, r^2 = 0.30, *p*-value = 0.004 linear regression), both treatment groups exhibited an accumulation of new Numts over time (see Methods, linear regression for Dex group r^2 = 0.59, *p*-value $<$0.001; for Oligo group r^2 = 0.22, *p*-value = 0.052). Compared to the untreated group (0.79 Numt per 10 days, 95% C.I. = 0.28 to 1.31), Dex and Oligo treatments tended to increase the rate of numtogenesis to 1.07 Numt (95% C.I. = 0.65 to 1.49) and 2.15 Numt (95% C.I. = −0.02 to 3.27) per 10 days, respectively (Fig 4A, 4B and 4D). Although these differences in effects did not reach statistical significance, the accumulation of Numts over time in these biologically independent experiments from those above further document Numtogenesis in aging human cells in vitro.

Using a genetic approach, we further tested whether defects in mitochondrial OxPhos associated with mtDNA instability are sufficient to alter the rate of numtogenesis. We analyzed

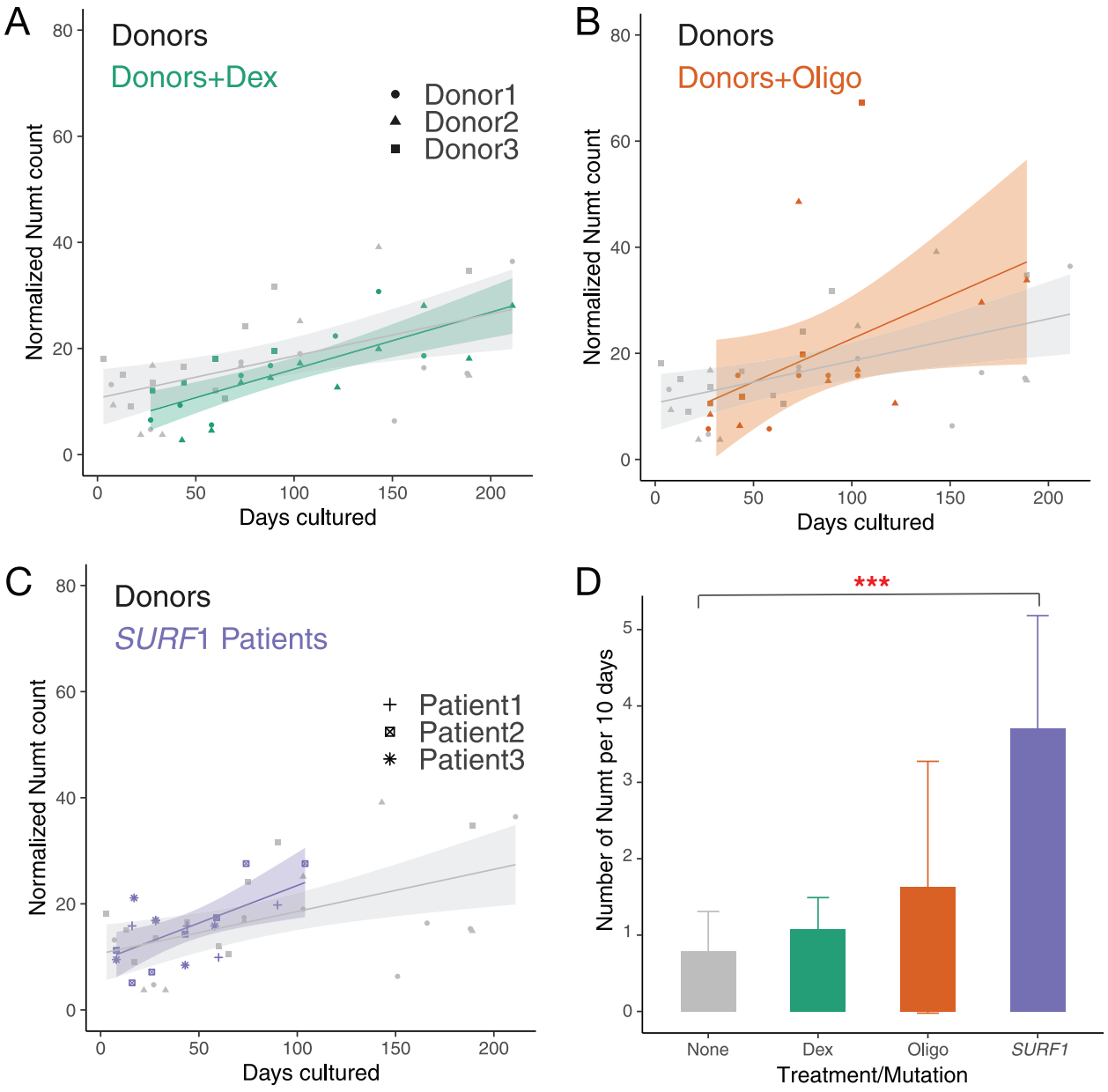

Fig 4. Effect of chronic pharmacological and genetic perturbation of mitochondrial function on Numt accumulation in fibroblasts. (A) Time course of Numt accumulation in healthy donors 1–3 and same cells cultured in dexamethasone (Dex) mimicking chronic Dex exposure. (B) Time course of Numt accumulation in healthy donors 1–3 and same cells cultured in oligomycin (Oligo). (C) Numt accumulation time course in the patient fibroblasts with *SURF1* gene defect (Patient 1–3) and the ones from 3 healthy donors (Donor 1–3). (D) Comparison of Slopes derived from all patients untreated, Dex-treated, oligo-treated, and *SURF1* gene defect. Numt counts for each group were normalized by the median value. A linear regression analysis was performed to derive the rate of the Numt accumulation and calculate the slopes in each group. ANOVA test was used to test the significance between the slopes of untreated donors and the ones of treated donors or patients in the hypothesis that pharmacological or genetic perturbation would increase the accumulation of Numts. ***, **, and * represent a significant *p*-value less than 0.001, 0.01, and 0.05, respectively, and ns represents non-significance. The data underlying this figure can be found in S1 Data.

https://doi.org/10.1371/journal.pbio.3002723.g004

data from a similar fibroblast culture system in 3 patient-derived fibroblasts with *SURF1* mutations (Patient 1–3) [5]. Mutations in *SURF1* represent one of the most frequent causes of cytochrome c oxidase and OxPhos deficiency in humans [64], and we recently showed that these *SURF1*-mutant fibroblasts accumulate large-scale mtDNA deletions over time in culture,

demonstrating mtDNA instability in this model. In these independent donors, we again observed the accumulation of new Numts over time (regression $r^2 = 0.64$, p-value <0.001). Strikingly, the rate of Numtogenesis in *SURF1*-mutant cells was 3.71 Numt per 10 days (95% C.I. = 2.24 to 5.18), in contrast to the rate of 0.79 Numt per 10 days in the healthy donors (4.7-fold of control, p-value = 8.20E-05, ANOVA test, Fig 4C and 4D). As in tumors [23,24] and yeast [1,2], this result further documents Numtogenesis as an inter-genomic event occurring in human cells over relatively short timescales, and establishes, using patient-derived cells with mtDNA instability, the modifiability of the rate of numtogenesis in vitro [63].

Discussion

The transfer of mitochondrial DNA into the nuclear genome of eukaryotes occurs in the germlines of various eukaryotes, suggesting that the endosymbiotic event initiated 1.5 billion years ago is still ongoing [4]. However, the extent and impact of somatic Numt insertions in specific human tissues have remained elusive outside of cancerous environments. Here, we provide some of the first evidence of the somatic nuclear accumulation of Numts in both healthy and impaired human brain tissue across different age ranges. We found that specific brain regions harbor more somatic insertions than others, in line with previous studies of other types of genomic variation [46–48], and that these rates do differ with the degree of cognitive impairment. We further extend these observations in a longitudinal study of primary human fibroblasts under various environmental and genetic conditions, documenting variable rates of ongoing numtogenesis in dividing human cells.

Our data provide new information concerning the rate of numtogenesis in humans. While the cross-sectional study of postmortem human brains does not allow us to draw conclusions about the rate of Numt transfer, the higher frequency of Numts in postmitotic brain tissue—relative to the commonly studied genomic material from blood—indicates a greater number of Numts. We further observed a significant correlation with mtDNA copy number in these samples, suggesting a potential biological mass action mechanism whereby higher mtDNAcn results in more potential transfers to the nuclear genome (S3 and S7 Figs). On the other hand, our longitudinal in vitro studies allow us to measure Numts among the same cell population (i.e., "individual") over time as cells accumulate age-related molecular hallmarks of aging. In this system, cells divide approximately every 40 h (1.7 days) when they are young (from days 0 to 80) and slow their replication rate dramatically towards the end of life when they undergo less than 1 replicative event per month (see growth curves in [65]). Although the temporal resolution of our trajectories is limited by the number of time points across the lifespan of each cell line (on average 7), our data suggests that the rate of numtogenesis is roughly linear across the wide range of replication rate, and therefore, more dependent on time rather than the rate of replication. This observation aligns with the results in cortical brain tissue (DLPFC), a postmitotic tissue where cell division is expected to be minimal to absent. This result is unlikely explained by clonality as all discovered Numts are unique. Consistent with the ROSMAP tissue data suggesting the accumulation of somatic Numts detected at very low variant allele frequency (VAF) in genomic material (median = 4.4%, mean = 5.3% ± 3.5%, with 99.7% of total number VAF <35%, S8 Fig), the VAF distribution for somatic Numts in fibroblasts was comparably low (median = 9.1%, mean = 13.1% ± 7.4%, with 98.3% of total number VAF <35%, S8 Fig). In cultured fibroblasts, each cell population contains between 1 and 5 million cells, meaning 2 to 10 million genomic copies. At 25× coverage, we therefore sample 0.00125% to 0.00025% of all copies. Our sampling is therefore essentially random. If the accumulation of Numts in aging fibroblasts was driven by clonality, we would see an increase in abundance of a

few Numts present at each passage in the same donor, rather than an increased frequency of unique Numts at each time point, as observed here.

At least 2 main findings from our results and that of others suggest that de novo Numts may be functionally meaningful. First, strong evidence that detectable Numts are excluded from coding DNA sequences and instead preferentially integrate within intergenic regions [4,6–8] suggest that they are under some functional constraint during development. In contrast, tumors frequently contain Numts within genes, and Numts may even contribute to oncogenesis, which in this case would drive positive selection in tumors [4]. Numts have also been implicated previously in several diverse disorders [66–71]. Second, Numts prevalence correlates with the somatic selection pressure for adverse mitochondrial genomic changes. In the replicating blood immune compartment of the bone marrow, high selection pressure occurs and eliminates mtDNA mutations over time [72,73], whereas de novo mtDNA defects accumulate at high levels in postmitotic tissues such as skeletal muscle, heart, and in the brain [74]. Similarly, it is possible that Numt insertions which negatively affect cell fitness in the lymphoid or myeloid lineages of the bone marrow are outcompeted and eliminated from the cell pool (or exist at low levels and are not sampled during blood draw), compared to somatic tissues where the same potentially deleterious Numt insertions in postmitotic cells cannot readily be outcompeted and eliminated without functionally compromising the tissue.

Some limitations of our study should be noted. While our human multi-tissue study includes >1,000 individuals, given the low number of Numts per person, we are likely underpowered to draw definitive conclusions around the cartography of Numts, such as potential Numts hotspots in the nuclear genome. The report by Wei and colleagues in 66,083 human genomes robustly addresses this point [4]. The absence of WGS data from multiple tissues in the same individuals also precludes direct comparisons of the specific rate of numtogenesis between tissues. This question could be addressed in other studies (e.g., GTEx) by sequencing dozens of tissues from the same individuals, but the current lack of such data set precludes a robust analysis of this kind. In relation to our longitudinal cellular studies with genetic and pharmacological mitochondrial OxPhos defects, the marginal increase in cell death when OxPhos is disrupted [63] or with chronic glucocorticoid stress [75] could offer a route of negative selection that eliminates deleterious de novo Numts; the most compromised cells die, cleansing the cell population of new variants. If this effect was strong, it would make numtogenesis undetectable to WGS, or reduce the observed effect size of the rate of numtogenesis in both our brain and fibroblast studies. This possibility could be addressed in future studies by systematically sequencing cellular debris or dead cells from the culture medium. Thus, while our positive results conclusively establish the dynamic nature of Numts transfer in healthy and stressed human cells, the magnitude reported across donors (range of 0.79 [controls] to 3.71 [*SURF1*-mutant] Numts every 10 days) may reflect a lower bound, and the rates of numtogenesis in cells under stress should be interpreted with caution. Validating the dynamic nature of numtogenesis across the lifespan in humans would require repeated measures, longitudinal WGS of postmitotic tissues (e.g., muscle biopsies), with the caveat that the same exact cells likely cannot be repeatedly biopsied, and therefore that new Numts would be missed. For the reasons mentioned above (negative selection in the immune compartment), repeated WGS in blood would likely underestimate the true rate of in vivo mito-nuclear genomic transfer.

We also considered if the increase in Numts over time could be driven by the clonal expansion of Numt-containing cells, which would suggest that there is no active transfer of mtDNA and numtogenesis in this model. Here, our observed increase in unique Numts over time suggests that these are not the product of clonal expansion, which also is consistent with the somatic Numts in ROSMAP samples, and in Wei and colleagues [4]. Several possible theories have been advanced for numtogenesis, with several studies showing environmental effects of

ionizing radiation [76,77] or mitochondrial reactive oxygen species [3,78] causing double-strand breaks associated with Numt accumulation. However, there have been few mechanistic studies in this area to conclusively determine the precise process. What is clear is that for such events to happen, whole or fragments of mtDNA must come in close proximity to the autosomal genes. There are now several reports of cytoplasmic mtDNA release where the mtDNA is released in "free form" and able to bind DNA-sensing receptors such as cGAS [50,79–81]. In fact, partial mitochondrial permeabilization that may lead to cytoplasmic mtDNA release triggers nuclear genomic instability [82], which could theoretically open the door ongoing to numtogenesis, possibly independent of cell division and genome replication.

In conclusion, our results demonstrating the high prevalence of non-germline Numts in hundreds of human brains and their negative association with age at death suggests that numtogenesis occurs across the human lifespan and that they may have deleterious health effects. Using a longitudinal in vitro human system, we establish that primary human fibroblasts accumulate Numts over time and that numtogenesis may be accelerated by some stressors, in particular, *SURF1* defects associated with mtDNA instability. These findings build and extend previous evidence that numtogenesis is active in the human germline and can have deleterious genomic, cellular, and health effects on the host organism. The active transfer of mtDNA sequences to the nuclear genome adds to the vast repertoire of mito-nuclear communication mechanisms [83] that shape human health.

Materials and methods

Ethics statement

The ROS and MAP studies were approved by the Institutional Review Board of Rush University Medical Center, protocols #L91020181 (ROS), L86121802 (MAP), and #L99032481 (RADC repository). All participants signed an informed consent, Anatomical Gift Act, and a repository consent to share data and biospecimens. Patient-derived fibroblasts were utilized from the previous study and were approved through the Columbia University Irving Medical Center IRB #AAAB0483 in a previous publication [5]. All human studies were conducted according to the principles expressed in the Declaration of Helsinki.

ROSMAP cohort

Study participants. The Rush Memory and Aging Project (MAP) and the Religious Orders Study (ROS) [56,57] are 2 ongoing cohort studies of older persons, collectively referred to as ROSMAP. The ROS study enrolls older Catholic nuns, priests, and brothers, from more than 40 groups across the United States. The MAP study enrolls participants primarily from retirement communities throughout northeastern Illinois. Participants in both cohorts were without known dementia at study enrolment and agreed to annual evaluations and brain donation on death.

The clinical diagnosis of AD proximate to death was based on the review of the annual clinical diagnosis of dementia and its causes by the study neurologist blinded for postmortem data. Postmortem Alzheimer's disease pathology was assessed as described previously [84,85] and Alzheimer's disease classification was defined based on the National Institutes of Ageing-Reagan criteria [86]. Dementia status was coded as NCI, MCI, or AD from the final clinical diagnosis of dementia and the NIA Reagan criteria as previously described [87–89].

Data processing. We obtained and processed WGS samples from these cohorts through the NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) data set NG00067. In brief, we obtained sequence data for 1,187 tissue samples comprising 466 DLPFC, 260 cerebella, 68 PCC, 4 AC, 366 WB samples, and 23 PBMCs. We obtained these 1,187 samples from

1,170 individuals, where only 17 participants in the data set contributed 2 tissue different samples to the sequencing data (S1 and S3 Tables). All sequencing data were provided in CRAM format and aligned to the human genome reference (GRCh37) with an average read depth of 45×. Based on the clinical information of age, we also stratified all 1,187 samples into 3 age groups with a roughly similar sample size based on their percentiles (30%ile, 40%ile, and 30% ile) when fit to a Gaussian distribution: (a) samples died at an age younger than 85; (b) age at death older than or equal to 85, and less than 93; and (c) age at death older than or equal to 93 (S2 Table). All the alignment statistics are presented in S1 Table, along with the clinical characteristics of the study participants.

In vitro fibroblast aging model

Fibroblast collection and passaging. We further made use of processed WGS generated from a recent study of aged primary human dermal fibroblasts [5]. In brief, primary human dermal fibroblasts were obtained from distributors or our local clinic from 3 healthy and 3 *SURF1*-patient donors. Fibroblasts were isolated from biopsy tissue using standard procedures. Cells were passaged approximately every 5 days (+/− 1 day). Study measurements and treatment began after 15-day culture to allow for adjustment to the in vitro environment. Treatment conditions for healthy controls include the chronic addition of 1 nM oligomycin (oligo) to inhibit the OxPhos FoF1 ATP synthase and 100 nM dexamethasone (Dex) to stimulate the glucocorticoid receptor as a model of chronic stress [75,90]. Time points collected vary by assay, with an average sampling frequency of 15 days and 4 to 10 time points for each cell line and condition. Individual cell lines were terminated after exhibiting less than 1 population doubling over a 30-day period, as described in [5].

Whole-genome sequencing and processing. Whole-genome sequencing data were performed in the lifespan samples at each time point (overall 85 time points). Paired-end reads were aligned to the human genome (GRCh37) using Isaac (Isaac-04.17.06.15) [91]. Samtools (Ver1.2) [92] and Picard Toolkit (https://broadinstitute.github.io/picard/) were further used to process the aligned bam files and mark duplicates. The average read depth from the WGS and other alignment statistics used in this study can be found in S4 Table.

Population-scale WGS control data

We leveraged 2,504 independent individuals from the 1000 Genomes Project Phase 3 to serve as population-level controls. Samples were sequenced by 30× Illumina NovaSeq (https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/) [93], and the data were archived in CRAM format with the GRCh38 reference (https://ftp-trace.ncbi.nih.gov/1000genomes/ftp/technical/reference/GRCh38_reference_genome/). We also used Numts reported in the 100,000 Genomes Project in England [4] as population-level controls.

Detection of non-reference Numts

We applied an updated version of Dinumt [6,7,94] to identify non-reference Numts across the different sequenced cohorts. Dinumt is an established software that was first used in the 1000 Genomes Project and validated by orthogonal methods, including PCR and Sanger sequencing [6], and long-read sequencing [7]. Briefly, Dinumt identifies aberrant/discordant reads aligning with either the mtDNA or the reference Numts on one end and map elsewhere in the genome on the other end, read orientation, and various other filters to define insertion breakpoints. Reads are discarded if they do not align uniquely to the nuclear genome and have a mapping quality (MAPQ) of less than 10. Identified insertions are then filtered for quality using a Phred scale (≥50), a cutoff of supporting reads (≥4), and a cutoff of read depth (≥5×)

around the insertion point. We built the first set of Numts as populational and polymorphic controls from the 2,504 individuals of the 1KG Project recently re-sequenced to high coverage [93]. Individual non-reference Numt callsets were resolved and merged into a single VCF file using the merging module of Dinumt. All "PASS" Numts are lifted over [95,96] from GRCh38 to GRCh37 for the downstream analysis. Dinumt was used to identify non-reference Numts across the individual sequences from 1,187 samples in ROSMAP or 85 cell line genomes in the lifespan model. The same criteria were conducted in the pipelines (Fig 1A). The VAF of each Numt call was calculated based on the number of supporting reads reported by Dinumt divided by the number of overall read coverage in the sequenced genome. In addition, the length distribution of paired-end read fragments mapped to the nuclear genome and supported Numt calls were calculated to investigate the potential bacterial mitochondria contamination.

We then cross-referenced all of our detected non-reference Numts against large population cohorts including the 1000 Genomes Project and Numts reported from 66,083 genomes in the 100,000 Genomes Project in England [4]. In each case, we considered a Numt detected in our analysis as a germline polymorphic insertion if it fell within +/− 50 bp of a Numt reported in either of these studies.

We identified tissue/cell line-specific Numt insertions using the identified non-reference Numts. Tissue-specific Numts were derived from the ROSMAP callset, and cell line-specific calls were derived from the Numt callset of fibroblast lifespan data. Numts from each sample were first merged into an aggregated set for these 2 callsets. We then extracted all non-reference Numts that were found in only 1 specific tissue across all samples or cell line, respectively (S1 and S4 Tables). All analysis pipelines and the command lines for running Dinumt can be found at https://github.com/mills-lab/numts-and-aging-in-fibroblasts-and-brains [97].

Statistical analysis

Cell line-specific Numts from fibroblast lifespan data were grouped by both donor/patient and treatment status. There are 4 treatment statuses: no treatment donors, donors/cells cultured in dexamethasone (Dex), donors/cells cultured in oligomycin (Oligo), and patients' fibroblasts with *SURF1* gene mutation. To increase the statistical power of data points in each group, we normalized the Numt count, merged the data points from individual samples, and then conducted the linear regression. Normalized Numts counts were derived by normalizing by the median of Numt counts in each group as defined above for various donor/patient and treatment status combinations. A linear regression model was constructed for each category respectively as below:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

$$Y: \textit{Normalized Numts Count}; X: \textit{Days cultured}$$

Slopes (β_1) were compared by ANOVA test separately between 2 categories. All statistical analyses were performed in R 4.0.5.

Genomic analyses for non-reference Numts

Numt hotspots across chromosomes. We delineated the entire nuclear genome into 10 Mbp bins, resulting in an average of 10 detected Numts per bin. The frequency for tissue-specific Numts from ROSMAP in each bin was calculated. We performed a permutation analysis by randomly shuffling the genomic positions of each observed Numt 50,000 times to determine any hotspots across genome bins compared to the real data. An empirical *p*-value was

calculated for all the bins based on the frequency of Numt ranking in simulation data. A multiple test correction (Benjamini–Hochberg) was further conducted to decrease the false discovery rate. A bin with a *p*-value less than 0.05 after the adjustment was defined as a significant hotspot. A Z-score is calculated for normalizing Numt count, which measures the deviation of the Numt count in each 10 Mbp bin from the genome-wide average across all the bins. We stratified the tissue-specific Numts into different tissues, cognitive impairment levels, and age groups to perform the hotspot analysis separately.

Genomic content analysis and functional annotation. We conducted the genomic content analysis for non-reference Numt insertions. We calculated the genic distribution for the tissue-specific Numts from ROSMAP. Gene track (GRCh37) was obtained from Ensembl Genome Browser (https://grch37.ensembl.org/). Parameters for protein-coding regions, transcriptomes, and exons were calculated based on a previous report [98]. Pearson's chi-square test and Fisher's exact test were used to assess the statistical significance of the discrepancies between the observed and expected distributions of Numts within genic or intergenic regions. We compared 2 × 2 categorical variables, specifically the proportion of Numts within a particular genomic region against the one in the rest of the genome, with respect to either the observed or expected data. GC content and repeat sequence analyses were carried out both in the set of cell line-specific Numt insertions from the lifespan model and polymorphic Numts from the 1000 Genomes project. GC content and repeat sequence were downloaded from the GC content table and RepeatMasker track in the UCSC Genome Browser (https://genome.ucsc.edu/). Gene mapping was carried out by AnnotSV (https://lbgi.fr/AnnotSV/) [99] to determine the genes that were potentially affected by the tissue-specific Numts from ROSMAP (S5 Table) or cell line-specific Numts from the lifespan model (S6 Table).

Detection of structural variation (SV)

Background structural variations (SVs) were detected in the data of the 1000 Genomes Project, ROSMAP, and lifespan model. We used an integrated non-reference SV callset from the 1000 Genomes Project as the control in the project to filter out potential non-somatic SVs at the population level. It was derived from 13 callers and can be obtained from http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage_SV/working/20210104_JAX_Integration_13callers/. Delly2 (Version 0.8.5) [100] was applied to resolve non-reference SVs (including deletions, duplications, insertions, inversions, and translocations), and MELT (Version 2.1.4) [101] was used to identify a specific type of non-reference SVs, mobile element insertions (MEIs, including Alus, LINE-1s, and SVAs), in the sequenced genomes of lifespan experiments. Manta [102] and Canvas (Version 1.28.0) [103] were also applied to resolve non-reference SVs in the sequenced genomes of lifespan experiments. The same pipeline used in Numts was implemented to identify tissue-specific or cell line-specific SVs/MEIs among the ROSMAP and lifespan samples (S1 and S4 Tables).

Mitochondrial DNA copy number

We calculated the mtDNAcn using autosomal coverage with the following formula: mtDNAcn = $(cov_{MT} / cov_{autosomal}) \times 2$ [60] in both the ROSMAP data set and lifespan study. The median sequence coverages of the autosomal chromosomes *covnuc* and of the mitochondrial genome *covmt* were calculated using R/Bioconductor (packages GenomicAlignments and GenomicRanges). We filtered out the reads with MAPQ = 0 in the analysis. Ambiguous regions were excluded using the intra-contig ambiguity mask from the BSgenome package. The mtDNAcn was z-standardized within each brain region and DNA extraction kit and then logarithmized. The normalization facilitated the combined analysis of the 2 different kits used for the DLPFC

and resulted in approximately normal mtDNAcn measures [60]. To note, using either the median or mode will not have a noticeable impact on the mtDNA copy number estimates and downstream analyses (S9 Fig). R and shell scripts used for mtDNA analysis are deposited at GitHub: https://github.com/cu-ctcn/mtDNA [60].

Cellular mtDNA content was also quantified by qPCR on the same genomic material used for other DNA-based measurements in the ROSMAP and lifespan studies. Duplex qPCR reactions were performed to simultaneously quantify mitochondrial (mtDNA, ND1) and nuclear (nDNA, B2M) amplicons, details of which can be found in the previous studies [5,104]. We observed a significant correlation between qPCR and WGS method in terms of mtDNAcn. Spearman correlation analysis between mtDNAcn measures from qPCR versus WGS is presented across cell lines by treatment (S10 Fig).

Supporting information

S1 Table. Meta table for ROSMAP sequencing data, including alignment statistics, clinical information, and variant numbers.
(XLSX)

S2 Table. Sample counts in different age groups (age at death) and cognitive status across main tissues in ROSMAP.
(XLSX)

S3 Table. The list of 17 participants contributing 2 tissue samples in the ROSMAP with sample ID, information, and numbers of detected somatic Numt.
(XLSX)

S4 Table. Meta table for lifespan sequencing data, including alignment statistics, experimental information, and variant numbers.
(XLSX)

S5 Table. Meta table for gene annotation of tissue-specific Numts and Numt callset from ROSMAP.
(XLSX)

S6 Table. Meta table for gene annotation for cell line-specific Numts and Numt callset from the lifespan model.
(XLSX)

S1 Fig. Length distribution of the fragments in paired-end reads mapped to the nuclear genome. Left, 2 × 151 bp in ROSMAP and right, 2 × 149 bp in the lifespan WGS. The data underlying this figure can be found in S1 Data.
(PDF)

S2 Fig. Boxplots of Numt count in matched cerebellum and DLPFC samples from $n = 9$ individuals. P-value = 0.033, Student's T test, paired, two-sided. Samples are shown in jittered points. The data underlying this figure can be found in S1 Data.
(PDF)

S3 Fig. MtDNA copy number association with Numt count and age at death by tissue and cognitive impairment in ROSMAP cohort. Correlation between mtDNA copy number and Numt count in all ROSMAP samples (A), DLPFC (B), cerebellum (C), PCC (D), whole blood (E), and PBMC (F), respectively. (G) Correlation between mtDNA copy number and age at death in all ROSMAP samples, cerebellum, PCC, whole blood, and PBMC, respectively. (H) Correlation between mtDNA copy number and age at death in DLPFC and 3 cognitive groups

in DLPFC, respectively. r^2 and *p*-values are calculated using standard least-squares regression models. The data underlying this figure can be found in S1 Data.
(PDF)

S4 Fig. Cerebellum-, PCC-, and whole-blood-specific Numts are not associated with the age of death or cognitive status. (A) Cerebellum samples correlated with age at death, stratified by cognitive diagnosis status. (B) PCC samples correlated with age at death, stratified by cognitive diagnosis status. (C) Whole-blood samples correlated with age at death, stratified by cognitive diagnosis status. Data points are colored by arbitrary age groups (see Methods) in light yellow, orange, and brown, respectively. r^2 and *p*-values are calculated using standard least-squares regression models.
(PDF)

S5 Fig. Common fibroblast Numts are not abundant across the lifespan and are not associated with age. (A) Numts shared between cell lines (donors) are not significantly correlated with aging. (B) Slopes from cell line-specific Numts and shared Numts in the lifespan model. The data underlying this figure can be found in S1 Data.
(PDF)

S6 Fig. Background somatic SVs and MEIs during aging in primary human fibroblasts. (A) Heatmap of slopes based on the linear regression between days cultured and the cell line-specific Numts (from Dinumt). (B) Heatmap of slopes based on the linear regression between days cultured and the cell line-specific MEIs (from MELT). (C) Heatmap of slopes based on the linear regression between days cultured and the cell line-specific SVs (from DELLY).
(PDF)

S7 Fig. MtDNA copy number (mtDNAcn) association with Numt count and days cultured in lifespan model. (A) Correlation between mtDNA copy number and Numt count in 3 treatment groups and *SURF1* defect group, respectively. (B) Correlation between mtDNA copy number and the number of days cells were cultured (Days cultured) in 3 treatment groups and *SURF1* defect group, respectively. r^2 and *p*-values are calculated using standard least-squares regression models. The data underlying this figure can be found in S1 Data.
(PDF)

S8 Fig. Variant allele frequency of non-reference Numts in ROSMAP and lifespan study. The Numts were categorized into germline ones (red), which overlapped with 1KG and Wei and colleagues callset and shared between tissues (ROSMAP, left) or cell lines (lifespan study, right), and potential somatic ones (green), which are tissue-specific (ROSMAP, left) or cell line-specific (lifespan study, right). The data underlying this figure can be found in S1 Data.
(PDF)

S9 Fig. Effect of using the mean, median, or mode of the coverage on mtDNAcn estimates. Each dot represents one of the 455 ROSMAP samples from the dorsolateral prefrontal cortex (DLPFC). The black line indicates the diagonal and the blue line represents a linear regression line. The Pearson correlation is displayed in the top left corner. Scatterplots (A–C) depict (A) autosomal coverage, (B) MT coverage, and (C) mtDNAcn using the median (x-axis) versus the mean (y-axis). Scatterplots (D–F) depict (D) autosomal coverage, (E) MT coverage, and (F) mtDNAcn using the median (x-axis) versus the mode (y-axis). The data underlying this figure can be found in S1 Data.
(PDF)

S10 Fig. mtDNA copy number measures by qPCR and WGS are comparable in control and stressed fibroblasts during lifespan. Three donors in each group are merged for analysis. R-squared values and *p*-values are calculated using standard linear regression models. The data underlying this figure can be found in S1 Data.
(PDF)

S1 Data. Data underlying Figs 1B–1F, 2A, 2B, 2G, 3B, 3D and 4A–4D, S1, S2, S3, S4A–S4C, S5, S7, S8, S9A–S9F and S10A–S10D.
(XLSX)

Author Contributions

Conceptualization: Weichen Zhou, Kalpita R. Karan, Martin Picard, Ryan E. Mills.

Data curation: Weichen Zhou, Kalpita R. Karan, Wenjin Gu, Hans-Ulrich Klein, Gabriel Sturm, Ryan E. Mills.

Formal analysis: Weichen Zhou, Kalpita R. Karan, Wenjin Gu, Hans-Ulrich Klein, Gabriel Sturm.

Funding acquisition: Weichen Zhou, Philip L. De Jager, David A. Bennett, Michio Hirano, Martin Picard, Ryan E. Mills.

Investigation: Weichen Zhou, Martin Picard, Ryan E. Mills.

Methodology: Weichen Zhou, Kalpita R. Karan, Wenjin Gu, Hans-Ulrich Klein, Gabriel Sturm, Ryan E. Mills.

Project administration: Weichen Zhou, Martin Picard, Ryan E. Mills.

Resources: Weichen Zhou, Kalpita R. Karan, Hans-Ulrich Klein, Gabriel Sturm, Philip L. De Jager, David A. Bennett, Michio Hirano, Martin Picard, Ryan E. Mills.

Software: Weichen Zhou, Kalpita R. Karan, Ryan E. Mills.

Supervision: Martin Picard, Ryan E. Mills.

Validation: Weichen Zhou, Kalpita R. Karan, Hans-Ulrich Klein, Gabriel Sturm, Martin Picard.

Visualization: Weichen Zhou, Kalpita R. Karan, Wenjin Gu.

Writing – original draft: Weichen Zhou, Kalpita R. Karan, Ryan E. Mills.

Writing – review & editing: Weichen Zhou, Kalpita R. Karan, Wenjin Gu, Hans-Ulrich Klein, Gabriel Sturm, Philip L. De Jager, David A. Bennett, Michio Hirano, Martin Picard, Ryan E. Mills.

References

1. Chatre L, Ricchetti M. Nuclear mitochondrial DNA activates replication in Saccharomyces cerevisiae. PLoS One. 2011; 6:e17235. https://doi.org/10.1371/journal.pone.0017235 PMID: 21408151
2. Lenglez S, Hermand D, Decottignies A. Genome-wide mapping of nuclear mitochondrial DNA sequences links DNA replication origins to chromosomal double-strand break formation in Schizosaccharomyces pombe. Genome Res. 2010; 20:1250–1261. https://doi.org/10.1101/gr.104513.109 PMID: 20688779
3. Cheng X, Ivessa AS. The migration of mitochondrial DNA fragments to the nucleus affects the chronological aging process of Saccharomyces cerevisiae. Aging Cell. 2010; 9:919–923. https://doi.org/10.1111/j.1474-9726.2010.00607.x PMID: 20626726

4. Wei W, Schon KR, Elgar G, Orioli A, Tanguy M, Giess A, et al. Nuclear-embedded mitochondrial DNA sequences in 66,083 human genomes. Nature. 2022; 611:105–114. https://doi.org/10.1038/s41586-022-05288-7 PMID: 36198798
5. Sturm G, Monzel AS, Karan KR, Michelson J, Ware SA, Cardenas A, et al. A multi-omics longitudinal aging dataset in primary human fibroblasts with mitochondrial perturbations. Sci Data. 2022; 9:751. https://doi.org/10.1038/s41597-022-01852-y PMID: 36463290
6. Dayama G, Emery SB, Kidd JM, Mills RE. The genomic landscape of polymorphic human nuclear mitochondrial insertions. Nucleic Acids Res. 2014; 42:12640–12649. https://doi.org/10.1093/nar/gku1038 PMID: 25348406
7. Dayama G, Zhou W, Prado-Martinez J, Marques-Bonet T, Mills RE. Characterization of nuclear mitochondrial insertions in the whole genomes of primates. NAR Genom Bioinform. 2020; 2:lqaa089. https://doi.org/10.1093/nargab/lqaa089 PMID: 33575633
8. Hazkani-Covo E, Zeller RM, Martin W. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet. 2010; 6:e1000834. https://doi.org/10.1371/journal.pgen.1000834 PMID: 20168995
9. Ricchetti M, Tekaia F, Dujon B. Continued colonization of the human genome by mitochondrial DNA. PLoS Biol. 2004; 2:E273. https://doi.org/10.1371/journal.pbio.0020273 PMID: 15361937
10. Lutz-Bonengel S, Niederstätter H, Naue J, Koziel R, Yang F, Sänger T, et al. Evidence for multi-copy Mega-NUMTs in the human genome. Nucleic Acids Res. 2021; 49:1517–1531. https://doi.org/10.1093/nar/gkaa1271 PMID: 33450006
11. Lang M, Sazzini M, Calabrese FM, Simone D, Boattini A, Romeo G, et al. Polymorphic NumtS trace human population relationships. Hum Genet. 2012; 131:757–771. https://doi.org/10.1007/s00439-011-1125-3 PMID: 22160368
12. Calabrese FM, Simone D, Attimonelli M. Primates and mouse NumtS in the UCSC Genome Browser. BMC Bioinformatics. 2012. https://doi.org/10.1186/1471-2105-13-S4-S15 PMID: 22536961
13. Nacer DF, Raposo do Amaral F. Striking pseudogenization in avian phylogenetics: Numts are large and common in falcons. Mol Phylogenet Evol. 2017; 115:1–6. https://doi.org/10.1016/j.ympev.2017.07.002 PMID: 28690127
14. Calabrese FM, Balacco DL, Preste R, Diroma MA, Forino R, Ventura M, et al. NumtS colonization in mammalian genomes. Sci Rep. 2017; 7:16357. https://doi.org/10.1038/s41598-017-16750-2 PMID: 29180746
15. Richly E. NUMTs in Sequenced Eukaryotic Genomes. Mol Biol Evol. 2004:1081–1084. https://doi.org/10.1093/molbev/msh110 PMID: 15014143
16. Soto-Calderón ID, Clark NJ, Wildschutte JVH, DiMattio K, Jensen-Seaman MI, Anthony NM. Identification of species-specific nuclear insertions of mitochondrial DNA (numts) in gorillas and their potential as population genetic markers. Mol Phylogenet Evol. 2014; 81:61–70. https://doi.org/10.1016/j.ympev.2014.08.018 PMID: 25194325
17. Hazkani-Covo E. Mitochondrial insertions into primate nuclear genomes suggest the use of numts as a tool for phylogeny. Mol Biol Evol. 2009; 26:2175–2179. https://doi.org/10.1093/molbev/msp131 PMID: 19578158
18. Patterson EC, Lall GM, Neumann R, Ottolini B, Sacchini F, Foster AP, et al. Defining cat mitogenome variation and accounting for numts via multiplex amplification and Nanopore sequencing. Forensic Sci Int Genet. 2023; 67:102944. https://doi.org/10.1016/j.fsigen.2023.102944 PMID: 37820546
19. Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJ. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J Mol Evol. 1994; 39:174–190. https://doi.org/10.1007/BF00163806 PMID: 7932781
20. Antunes A, Pontius J, Ramos MJ, O'Brien SJ, Johnson WE. Mitochondrial introgressions into the nuclear genome of the domestic cat. J Hered. 2007; 98:414–420. https://doi.org/10.1093/jhered/esm062 PMID: 17660503
21. Pamilo P, Viljakainen L, Vihavainen A. Exceptionally high density of NUMTs in the honeybee genome. Mol Biol Evol. 2007; 24:1340–1346. https://doi.org/10.1093/molbev/msm055 PMID: 17383971
22. Hazkani-Covo E. A burst of numt insertion in the Dasyuridae family during marsupial evolution. Front Ecol Evol. 2022: 10. https://doi.org/10.3389/fevo.2022.844443
23. Singh KK, Choudhury AR, Tiwari HK. Numtogenesis as a mechanism for development of cancer. Semin Cancer Biol. 2017; 47:101–109. https://doi.org/10.1016/j.semcancer.2017.05.003 PMID: 28511886
24. Srinivasainagendra V, Sandel MW, Singh B, Sundaresan A, Mooga VP, Bajpai P, et al. Migration of mitochondrial DNA in the nuclear genome of colorectal adenocarcinoma. Genome Med. 2017. https://doi.org/10.1186/s13073-017-0420-6 PMID: 28356157

25. Puertas MJ, González-Sánchez M. Insertions of mitochondrial DNA into the nucleus-effects and role in cell evolution. Genome. 2020; 63:365–374. https://doi.org/10.1139/gen-2019-0151 PMID: 32396758

26. Maude H, Davidson M, Charitakis N, Diaz L, Bowers WHT, Gradovich E, et al. NUMT Confounding Biases Mitochondrial Heteroplasmy Calls in Favor of the Reference Allele. Front Cell Dev Biol. 2019; 7:201. https://doi.org/10.3389/fcell.2019.00201 PMID: 31612134

27. Chow S, Yanagimoto T, Takeyama H. Author Correction: Detection of heteroplasmy and nuclear mitochondrial pseudogenes in the Japanese spiny lobster Panulirus japonicus. Sci Rep. 2022; 12:4435. https://doi.org/10.1038/s41598-022-08715-x PMID: 35292714

28. Balciuniene J, Balciunas D. A Nuclear mtDNA Concatemer (Mega-NUMT) Could Mimic Paternal Inheritance of Mitochondrial Genome. Front Genet. 2019; 10:518. https://doi.org/10.3389/fgene.2019.00518 PMID: 31244882

29. Pagnamenta AT, Wei W, Rahman S, Chinnery PF. Biparental inheritance of mitochondrial DNA revisited. Nat Rev Genet. 2021; 22:477–478. https://doi.org/10.1038/s41576-021-00380-6 PMID: 34031572

30. Wei W, Pagnamenta AT, Gleadall N, Sanchis-Juan A, Stephens J, Broxholme J, et al. Nuclear-mitochondrial DNA segments resemble paternally inherited mitochondrial DNA in humans. Nat Commun. 2020; 11:1740. https://doi.org/10.1038/s41467-020-15336-3 PMID: 32269217

31. Bai R, Cui H, Devaney JM, Allis KM, Balog AM, Liu X, et al. Interference of nuclear mitochondrial DNA segments in mitochondrial DNA testing resembles biparental transmission of mitochondrial DNA in humans. Genet Med. 2021; 23:1514–1521. https://doi.org/10.1038/s41436-021-01166-1 PMID: 33846581

32. Marshall C, Parson W. Interpreting NUMTs in forensic genetics: Seeing the forest for the trees. Forensic Sci Int Genet. 2021; 53:102497. https://doi.org/10.1016/j.fsigen.2021.102497 PMID: 33740708

33. Bintz BJ, Dixon GB, Wilson MR. Simultaneous detection of human mitochondrial DNA and nuclear-inserted mitochondrial-origin sequences (NumtS) using forensic mtDNA amplification strategies and pyrosequencing technology. J Forensic Sci. 2014; 59:1064–1073. https://doi.org/10.1111/1556-4029.12441 PMID: 24738853

34. Goios A, Amorim A, Pereira L. Mitochondrial DNA pseudogenes in the nuclear genome as possible sources of contamination. Int Congr Ser. 2006; 1288:697–699.

35. Thorsness PE, Fox TD. Escape of DNA from mitochondria to the nucleus in Saccharomyces cerevisiae. Nature. 1990; 346:376–379. https://doi.org/10.1038/346376a0 PMID: 2165219

36. Ricchetti M, Fairhead C, Dujon B. Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. Nature. 1999; 402:96–100. https://doi.org/10.1038/47076 PMID: 10573425

37. Yu X, Gabriel A. Patching broken chromosomes with extranuclear cellular DNA. Mol Cell. 1999; 4:873–881. https://doi.org/10.1016/s1097-2765(00)80397-4 PMID: 10619034

38. Lee MB, Dowsett IT, Carr DT, Wasko BM, Stanton SG, Chung MS, et al. Defining the impact of mutation accumulation on replicative lifespan in yeast using cancer-associated mutator phenotypes. Proc Natl Acad Sci U S A. 2019; 116:3062–3071. https://doi.org/10.1073/pnas.1815966116 PMID: 30718408

39. Kaya A, Lobanov AV, Gladyshev VN. Evidence that mutation accumulation does not cause aging in Saccharomyces cerevisiae. Aging Cell. 2015; 14:366–371. https://doi.org/10.1111/acel.12290 PMID: 25702753

40. Breuss MW, Yang X, Schlachetzki JCM, Antaki D, Lana AJ, Xu X, et al. Somatic mosaicism reveals clonal distributions of neocortical development. Nature. 2022; 604:689–696. https://doi.org/10.1038/s41586-022-04602-7 PMID: 35444276

41. Fasching L, Jang Y, Tomasi S, Schreiner J, Tomasini L, Brady MV, et al. Early developmental asymmetries in cell lineage trees in living individuals. Science. 2021; 371:1245–1248. https://doi.org/10.1126/science.abe0981 PMID: 33737484

42. Wang Y, Bae T, Thorpe J, Sherman MA, Jones AG, Cho S, et al. Comprehensive identification of somatic nucleotide variants in human brain tissue. Cold Spring Harbor Laboratory. 2020. p. 2020.10.10.332213. https://doi.org/10.1101/2020.10.10.332213

43. Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, et al. L1 retrotransposition in human neural progenitor cells. Nature. 2009; 460:1127–1131. https://doi.org/10.1038/nature08248 PMID: 19657334

44. Muotri AR, Chu VT, Marchetto MCN, Deng W, Moran JV, Gage FH. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature. 2005; 435:903–910. https://doi.org/10.1038/nature03663 PMID: 15959507

45. Zhu X, Zhou B, Pattni R, Gleason K, Tan C, Kalinowski A, et al. Machine learning reveals bilateral distribution of somatic L1 insertions in human neurons and glia. Nat Neurosci. 2021; 24:186–196. https://doi.org/10.1038/s41593-020-00767-4 PMID: 33432196

46. McConnell MJ, Lindberg MR, Brennand KJ, Piper JC, Voet T, Cowing-Zitron C, et al. Mosaic copy number variation in human neurons. Science. 2013; 342:632–637. https://doi.org/10.1126/science.1243472 PMID: 24179226

47. Wierman MB, Burbulis IE, Chronister WD, Bekiranov S, McConnell MJ. Single-Cell CNV Detection in Human Neuronal Nuclei. In: Frade JM, Gage FH, editors. Genomic Mosaicism in Neurons and Other Cell Types. New York, NY: Springer New York; 2017. p. 109–131.

48. Chronister WD, Burbulis IE, Wierman MB, Wolpert MJ, Haakenson MF, Smith ACB, et al. Neurons with Complex Karyotypes Are Rare in Aged Human Neocortex. Cell Rep. 2019; 26:825–835.e7. https://doi.org/10.1016/j.celrep.2018.12.107 PMID: 30673605

49. Ju YS, Alexandrov LB, Gerstung M, Martincorena I, Nik-Zainal S, Ramakrishna M, et al. Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer. elife. 2014: 3. https://doi.org/10.7554/eLife.02935 PMID: 25271376

50. McArthur K, Whitehead LW, Heddleston JM, Li L, Padman BS, Oorschot V, et al. BAK/BAX macropores facilitate mitochondrial herniation and mtDNA efflux during apoptosis. Science. 2018:359. https://doi.org/10.1126/science.aao6047 PMID: 29472455

51. Xian H, Watari K, Sanchez-Lopez E, Offenberger J, Onyuru J, Sampath H, et al. Oxidized DNA fragments exit mitochondria via mPTP- and VDAC-dependent channels to activate NLRP3 inflammasome and interferon signaling. Immunity. 2022; 55:1370–1385.e8. https://doi.org/10.1016/j.immuni.2022.06.007 PMID: 35835107

52. Trumpff C, Michelson J, Lagranha CJ, Taleon V, Karan KR, Sturm G, et al. Stress and circulating cell-free mitochondrial DNA: A systematic review of human studies, physiological considerations, and technical recommendations. Mitochondrion. 2021; 59:225–245. https://doi.org/10.1016/j.mito.2021.04.002 PMID: 33839318

53. Trumpff C, Marsland AL, Basualto-Alarcón C, Martin JL, Carroll JE, Sturm G, et al. Acute psychological stress increases serum circulating cell-free mitochondrial DNA. Psychoneuroendocrinology. 2019; 106:268–276. https://doi.org/10.1016/j.psyneuen.2019.03.026 PMID: 31029929

54. Maresca A, Del Dotto V, Romagnoli M, La Morgia C, Di Vito L, Capristo M, et al. Expanding and validating the biomarkers for mitochondrial diseases. J Mol Med. 2020; 98:1467–1478. https://doi.org/10.1007/s00109-020-01967-y PMID: 32851462

55. Péquignot MO, Dey R, Zeviani M, Tiranti V, Godinot C, Poyau A, et al. Mutations in the SURF1 gene associated with Leigh syndrome and cytochrome C oxidase deficiency. Hum Mutat. 2001; 17:374–381. https://doi.org/10.1002/humu.1112 PMID: 11317352

56. Bennett DA, Schneider JA, Buchman AS, Barnes LL, Boyle PA, Wilson RS. Overview and findings from the rush Memory and Aging Project. Curr Alzheimer Res. 2012; 9:646–663. https://doi.org/10.2174/156720512801322663 PMID: 22471867

57. Bennett DA, Buchman AS, Boyle PA, Barnes LL, Wilson RS, Schneider JA. Religious Orders Study and Rush Memory and Aging Project. J Alzheimers Dis. 2018; 64:S161–S189. https://doi.org/10.3233/JAD-179939 PMID: 29865057

58. Tan CCS, Ko KKK, Chen H, Liu J, Loh M, SG10K_Health Consortium, et al. No evidence for a common blood microbiome based on a population study of 9,770 healthy humans. Nat Microbiol. 2023; 8:973–985.

59. Kishikawa T, Momozawa Y, Ozeki T, Mushiroda T, Inohara H, Kamatani Y, et al. Empirical evaluation of variant calling accuracy using ultra-deep whole-genome sequencing data. Sci Rep. 2019; 9:1784. https://doi.org/10.1038/s41598-018-38346-0 PMID: 30741997

60. Klein H-U, Trumpff C, Yang H-S, Lee AJ, Picard M, Bennett DA, et al. Characterization of mitochondrial DNA quantity and quality in the human aged and Alzheimer's disease brain. Mol Neurodegener. 2021; 16:75. https://doi.org/10.1186/s13024-021-00495-8 PMID: 34742335

61. Sturm G, Cardenas A, Bind M-A, Horvath S, Wang S, Wang Y, et al. Human aging DNA methylation signatures are conserved but accelerated in cultured fibroblasts. Epigenetics. 2019; 14:961–976. https://doi.org/10.1080/15592294.2019.1626651 PMID: 31156022

62. Horvath S, Haghani A, Macoretta N, Ablaeva J, Zoller JA, Li CZ, et al. DNA methylation clocks tick in naked mole rats but queens age more slowly than nonbreeders. Nat Aging. 2022; 2:46–59. https://doi.org/10.1038/s43587-021-00152-1 PMID: 35368774

63. Sturm G, Karan KR, Monzel AS, Santhanam B, Taivassalo T, Bris C, et al. OxPhos defects cause hypermetabolism and reduce lifespan in cells and in patients with mitochondrial diseases. Commun Biol. 2023; 6:22. https://doi.org/10.1038/s42003-022-04303-x PMID: 36635485

64. Zhu Z, Yao J, Johns T, Fu K, De Bie I, Macmillan C, et al. SURF1, encoding a factor involved in the biogenesis of cytochrome c oxidase, is mutated in Leigh syndrome. Nat Genet. 1998; 20:337–343. https://doi.org/10.1038/3804 PMID: 9843204

65. Sturm G, Bobba-Alves N, Tumasian RA, Michelson J, Ferrucci L, Kempes CP, et al. Accelerating the clock: Interconnected speedup of energetic and molecular dynamics during aging in cultured human cells. bioRxiv. 2022:p. 2022.05.10.491392. https://doi.org/10.1101/2022.05.10.491392

66. Chen J-M, Chuzhanova N, Stenson PD, Férec C, Cooper DN. Meta-analysis of gross insertions causing human genetic disease: Novel mutational mechanisms and the role of replication slippage. Hum Mutat. 2005; 25:318–318. https://doi.org/10.1002/humu.20133 PMID: 15643617

67. Willett-Brozick JE, Savul SA, Richey LE, Baysal BE. Germ line insertion of mtDNA at the breakpoint junction of a reciprocal constitutional translocation. Hum Genet. 2001; 109:216–223. https://doi.org/10.1007/s004390100564 PMID: 11511928

68. Borensztajn K, Chafa O, Alhenc-Gelas M, Salha S, Reghis A, Fischer A-M, et al. Characterization of two novel splice site mutations in human factor VII gene causing severe plasma factor VII deficiency and bleeding diathesis. Br J Haematol. 2002; 117:168–171. https://doi.org/10.1046/j.1365-2141.2002.03397.x PMID: 11918550

69. Turner C, Killoran C, Thomas NST, Rosenberg M, Chuzhanova NA, Johnston J, et al. Human genetic disease caused by de novo mitochondrial-nuclear DNA transfer. Hum Genet. 2003; 112:303–309. https://doi.org/10.1007/s00439-002-0892-2 PMID: 12545275

70. Goldin E, Stahl S, Cooney AM, Kaneski CR, Gupta S, Brady RO, et al. Transfer of a mitochondrial DNA fragment to MCOLN1 causes an inherited case of mucolipidosis IV. Hum Mutat. 2004; 24:460–465. https://doi.org/10.1002/humu.20094 PMID: 15523648

71. Ahmed ZM, Smith TN, Riazuddin S, Makishima T, Ghosh M, Bokhari S, et al. Nonsyndromic recessive deafness DFNB18 and Usher syndrome type IC are allelic mutations of USHIC. Hum Genet. 2002; 110:527–531. https://doi.org/10.1007/s00439-002-0732-4 PMID: 12107438

72. Tang Z, Lu Z, Chen B, Zhang W, Chang HY, Hu Z, et al. A Genetic Bottleneck of Mitochondrial DNA During Human Lymphocyte Development. Mol Biol Evol. 2022: 39. https://doi.org/10.1093/molbev/msac090 PMID: 35482398

73. Franklin IG, Milne P, Childs J, Boggan RM, Barrow I, Lawless C, et al. T cell differentiation drives the negative selection of pathogenic mitochondrial DNA variants. Life Sci Alliance. 2023: 6. https://doi.org/10.26508/lsa.202302271 PMID: 37652671

74. Carelli V, Chan DC. Mitochondrial DNA: impacting central and peripheral nervous systems. Neuron. 2014; 84:1126–1142. https://doi.org/10.1016/j.neuron.2014.11.022 PMID: 25521375

75. Bobba-Alves N, Sturm G, Lin J, Ware SA, Karan KR, Monzel AS, et al. Chronic Glucocorticoid Stress Reveals Increased Energy Expenditure and Accelerated Aging as Cellular Features of Allostatic Load. bioRxiv. 2022:p. 2022.02.22.481548. https://doi.org/10.1101/2022.02.22.481548

76. Gaziev AI, Shaĭkhaev GO. [Ionizing radiation can activate the insertion of mitochondrial DNA fragments in the nuclear genome]. Radiats Biol Radioecol. 2007; 47:673–683. PMID: 18380326

77. Abdullaev SA, Fomenko LA, Kuznetsova EA, Gaziev AI. [Experimental detection of integration of mTDNA in the nuclear genome induced by ionizing radiation]. Radiats Biol Radioecol. 2013; 53:380–388. PMID: 25427370

78. Caro P, Gómez J, Arduini A, González-Sánchez M, González-García M, Borrás C, et al. Mitochondrial DNA sequences are present inside nuclear DNA in rat tissues and increase with age. Mitochondrion. 2010; 10:479–486. https://doi.org/10.1016/j.mito.2010.05.004 PMID: 20546951

79. Victorelli S, Salmonowicz H, Chapman J, Martini H, Vizioli MG, Riley JS, et al. Apoptotic stress causes mtDNA release during senescence and drives the SASP. Nature. 2023; 622:627–636. https://doi.org/10.1038/s41586-023-06621-4 PMID: 37821702

80. Zhang W, Li G, Luo R, Lei J, Song Y, Wang B, et al. Cytosolic escape of mitochondrial DNA triggers cGAS-STING-NLRP3 axis-dependent nucleus pulposus cell pyroptosis. Exp Mol Med. 2022; 54:129–142. https://doi.org/10.1038/s12276-022-00729-9 PMID: 35145201

81. Riley JS, Quarato G, Cloix C, Lopez J, O'Prey J, Pearson M, et al. Mitochondrial inner membrane permeabilisation enables mtDNA release during apoptosis. EMBO J. 2018:37. https://doi.org/10.15252/embj.201899238 PMID: 30049712

82. Ichim G, Lopez J, Ahmed SU, Muthalagu N, Giampazolias E, Delgado ME, et al. Limited mitochondrial permeabilization causes DNA damage and genomic instability in the absence of cell death. Mol Cell. 2015; 57:860–872. https://doi.org/10.1016/j.molcel.2015.01.018 PMID: 25702873

83. Picard M, Shirihai OS. Mitochondrial signal transduction. Cell Metab. 2022; 34:1620–1653. https://doi.org/10.1016/j.cmet.2022.10.008 PMID: 36323233

84. Bennett DA, Wilson RS, Schneider JA, Evans DA, Mendes de Leon CF, Arnold SE, et al. Education modifies the relation of AD pathology to level of cognitive function in older persons. Neurology. 2003; 60:1909–1915. https://doi.org/10.1212/01.wnl.0000069923.64550.9f PMID: 12821732

85. Bennett DA, Schneider JA, Tang Y, Arnold SE, Wilson RS. The effect of social networks on the relation between Alzheimer's disease pathology and level of cognitive function in old people: a longitudinal cohort study. Lancet Neurol. 2006; 5:406–412. https://doi.org/10.1016/S1474-4422(06)70417-3 PMID: 16632311

86. Hyman BT, Trojanowski JQ. Consensus recommendations for the postmortem diagnosis of Alzheimer disease from the National Institute on Aging and the Reagan Institute Working Group on diagnostic criteria for the neuropathological assessment of Alzheimer disease. J Neuropathol Exp Neurol. 1997; 56:1095–1097. https://doi.org/10.1097/00005072-199710000-00002 PMID: 9329452

87. Bennett DA, Wilson RS, Schneider JA, Evans DA, Beckett LA, Aggarwal NT, et al. Natural history of mild cognitive impairment in older persons. Neurology. 2002; 59:198–205. https://doi.org/10.1212/wnl.59.2.198 PMID: 12136057

88. Bennett DA, Schneider JA, Arvanitakis Z, Kelly JF, Aggarwal NT, Shah RC, et al. Neuropathology of older persons without cognitive impairment from two community-based studies. Neurology. 2006; 66:1837–1844. https://doi.org/10.1212/01.wnl.0000219668.47116.e6 PMID: 16801647

89. Bennett DA, Schneider JA, Aggarwal NT, Arvanitakis Z, Shah RC, Kelly JF, et al. Decision rules guiding the clinical diagnosis of Alzheimer's disease in two community-based cohort studies compared to standard practice in a clinic-based cohort study. Neuroepidemiology. 2006; 27:169–176. https://doi.org/10.1159/000096129 PMID: 17035694

90. Leung CS, Kosyk O, Welter EM, Dietrich N, Archer TK, Zannas AS. Chronic stress-driven glucocorticoid receptor activation programs key cell phenotypes and functional epigenomic patterns in human fibroblasts. iScience. 2022; 25:104960. https://doi.org/10.1016/j.isci.2022.104960 PMID: 36065188

91. Raczy C, Petrovski R, Saunders CT, Chorny I, Kruglyak S, Margulies EH, et al. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. Bioinformatics. 2013; 29:2041–2043. https://doi.org/10.1093/bioinformatics/btt314 PMID: 23736529

92. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

93. Byrska-Bishop M, Evani US, Zhao X, Basile AO, Abel HJ, Regier AA, et al. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. Cell. 2022; 185:3426–3440.e19. https://doi.org/10.1016/j.cell.2022.08.004 PMID: 36055201

94. Mills R, Zhou W (arthur). mills-lab/dinumt: Version for "Associations of nuclear mitochondrial DNA insertions and human lifespan in aging fibroblasts and in the human brain." 2021.

95. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, et al. The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. 2006; 34:D590–D598. https://doi.org/10.1093/nar/gkj144 PMID: 16381938

96. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002; 12:996–1006. https://doi.org/10.1101/gr.229102 PMID: 12045153

97. Zhou W (arthur). Mills-lab/numts-and-aging-in-fibroblasts-and-brains: Release.Jun.12th.2024. Zenodo. 2024.

98. Piovesan A, Antonaros F, Vitale L, Strippoli P, Pelleri MC, Caracausi M. Human protein-coding genes and gene feature statistics in 2019. BMC Res Notes. 2019; 12:315. https://doi.org/10.1186/s13104-019-4343-8 PMID: 31164174

99. Geoffroy V, Herenger Y, Kress A, Stoetzel C, Piton A, Dollfus H, et al. AnnotSV: an integrated tool for structural variations annotation. Bioinformatics. 2018; 34:3572–3574. https://doi.org/10.1093/bioinformatics/bty304 PMID: 29669011

100. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012; 28:i333–i339. https://doi.org/10.1093/bioinformatics/bts378 PMID: 22962449

101. Gardner EJ, Lam VK, Harris DN, Chuang NT, Scott EC, Pittard WS, et al. The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. Genome Res. 2017; 27:1916–1929. https://doi.org/10.1101/gr.218032.116 PMID: 28855259

102. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016:1220–1222. https://doi.org/10.1093/bioinformatics/btv710 PMID: 26647377

103. Roller E, Ivakhno S, Lee S, Royce T, Tanner S. Canvas: versatile and scalable detection of copy number variants. Bioinformatics. 2016; 32:2375–2377. https://doi.org/10.1093/bioinformatics/btw163 PMID: 27153601

104. Picard M, Zhang J, Hancock S, Derbeneva O, Golhar R, Golik P, et al. Progressive increase in mtDNA 3243A>G heteroplasmy causes abrupt transcriptional reprogramming. Proc Natl Acad Sci U S A. 2014; 111:E4033–E4042.