

Review

The Mighty NUMT: Mitochondrial DNA Flexing Its Code in the Nuclear Genome

Liyong Xue ^{1,†}, Jesse D. Moreira ^{2,†} , Karan K. Smith ¹ and Jessica L. Fetterman ^{1,*} 

¹ Evans Department of Medicine, Boston University Chobanian & Avedisian School of Medicine, Boston, MA 02118, USA

² Department of Health Sciences, Programs in Human Physiology, Boston University Sargent College, Boston, MA 02215, USA

* Correspondence: jefetter@bu.edu

† These authors contributed equally to this work.

Abstract: Nuclear-mitochondrial DNA segments (NUMTs) are mitochondrial DNA (mtDNA) fragments that have been inserted into the nuclear genome. Some NUMTs are common within the human population but most NUMTs are rare and specific to individuals. NUMTs range in size from 24 base pairs to encompassing nearly the entire mtDNA and are found throughout the nuclear genome. Emerging evidence suggests that the formation of NUMTs is an ongoing process in humans. NUMTs contaminate sequencing results of the mtDNA by introducing false positive variants, particularly heteroplasmic variants present at a low variant allele frequency (VAF). In our review, we discuss the prevalence of NUMTs in the human population, the potential mechanisms of de novo NUMT insertion via DNA repair mechanisms, and provide an overview of the existing approaches for minimizing NUMT contamination. Apart from filtering known NUMTs, both wet lab-based and computational methods can be used to minimize the contamination of NUMTs in analyses of human mtDNA. Current approaches include: (1) isolating mitochondria to enrich for mtDNA; (2) applying basic local alignment to identify NUMTs for subsequent filtering; (3) bioinformatic pipelines for NUMT detection; (4) k-mer-based NUMT detection; and (5) filtering candidate false positive variants by mtDNA copy number, VAF, or sequence quality score. Multiple approaches must be applied in order to effectively identify NUMTs in samples. Although next-generation sequencing is revolutionizing our understanding of heteroplasmic mtDNA, it also raises new challenges with the high prevalence and individual-specific NUMTs that need to be handled with care in studies of mitochondrial genetics.

Keywords: nuclear-mitochondrial DNA segments mitochondrial genome; mitochondrial DNA; bioinformatics



Citation: Xue, L.; Moreira, J.D.; Smith, K.K.; Fetterman, J.L. The Mighty NUMT: Mitochondrial DNA Flexing Its Code in the Nuclear Genome. *Biomolecules* **2023**, *13*, 753. <https://doi.org/10.3390/biom13050753>

Academic Editor: Peter Celec

Received: 15 March 2023

Revised: 7 April 2023

Accepted: 24 April 2023

Published: 27 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Nuclear-mitochondrial DNA segments (NUMTs) are fragments of the mitochondrial genome (mtDNA) that have been inserted into the nuclear genome of multiple organisms, including domestic cats, great apes, fruit flies, and humans [1–5]. NUMTs are likely due to multiple insertion events rather than duplications of nuclear DNA [6]. The concept of NUMTs was first suggested in 1967 [7] but experimental evidence of NUMTs only began appearing in the literature in the early 1980s [8,9]. Mitochondrial-like sequences in the nuclear genome have been identified in multiple species, including humans [10], yeast [11], mice [12], and others [13,14].

NUMTs have been used to define phylogenetic outgroups and elucidate mtDNA evolution [10,14]. Recent versus ancient nuclear insertions of mtDNA segments can be determined based on the degree of homology between NUMTs and the mtDNA equivalent within the individual. Once integrated into the nuclear genome, NUMTs undergo a slower mutation rate than the mtDNA, which is likely more similar to the slower mutation rate

of the nuclear genome compared to the mtDNA [6,15,16]. Therefore, NUMTs that are dissimilar to their mtDNA counterparts are often thought of as “nuclear fossils”, fragments derived from ancient mtDNA [3,17]. In contrast, NUMTs that are more similar to their homologous mtDNA sequences are likely more recent insertions into the nuclear genome.

NUMTs can originate from any part of the mtDNA and are found throughout the nuclear genome [1,18]. The number of NUMTs found in the human genome varies, with most individuals harboring several NUMTs [1,3,18,19]. Based upon the characterization of NUMTs in more than 67,000 human genomes, NUMTs are estimated to arise *de novo* once in every 10^4 births [1]. Comparisons of parent-child triads indicate that the incorporation of mtDNA fragments into the nuclear genome is an ongoing process [1]. NUMTs range in size from 24 bp to nearly the entire length of the mtDNA [1]. The majority of NUMTs are less than 500 bps in length and most frequently originate from the D-loop, a non-coding region containing many of the mtDNA regulatory elements [1].

NUMTs pose a challenge in studies evaluating mitochondrial genetic variation. Due to the high sequence similarity of NUMTs to the mtDNA within the individual, variants may be misidentified as being in the mtDNA rather than attributed to a NUMT. NUMTs interfere with the calculation of variant allele frequency (VAF), the number of copies of the mtDNA that contain a variant, used to define heteroplasmy [20]. Early studies reveal that NUMTs are often co-amplified with true mtDNAs, due to the use of non-specific primers in polymerase chain reaction (PCR) experiments or techniques that inadvertently select for nuclear DNA, creating additional challenges in sequencing techniques that involve the use of PCR [21–23].

NUMTs often do not encode functional proteins; however, some NUMTs are associated with disease [15]. One study mistook variants in NUMTs as missense variants in the mtDNA genes encoding subunits of cytochrome c oxidase and then associated those variants with Alzheimer’s disease [24]. The association of these NUMT-derived variants with Alzheimer’s disease has since been disproven, but this highlights the difficulties in attributing variants to the mtDNA versus NUMTs [23,25]. Further study of NUMTs is necessary to prevent the false identification of mtDNA variants. A number of methods are emerging for the improved detection of NUMTs. For this review, we will focus on the characterization of NUMTs and current methods for identifying NUMTs.

2. Transfer of mtDNA Segments into the Nuclear Genome

NUMTs are found throughout the entire human nuclear genome [1,15,26], but how mtDNA fragments get out of the mitochondrion and into the nucleus is not known. Currently, no consensus exists regarding the mechanism for the insertion of mtDNA segments into the nuclear genome. However, several studies have provided some insights into the potential mechanisms and non-random site selection of NUMT insertion.

The theory of mitochondrial endosymbiosis posits an evolutionary benefit of full-length mitochondrial gene transfer to specific loci in the eukaryotic nucleus that results in the production of a protein product [27–29]. The endosymbiotic relationship between the primitive cell and the bacteria from which mitochondria originate is thought to have occurred at the root of eukaryotic evolution, approximately 2.5 billion years ago [30], at which time much of the mtDNA was transferred to the nuclear genome, with the mitochondria retaining the genes encoding core catalytic subunits of the OXPHOS enzymes. However, NUMTs do not appear to be inserted as a functional transfer of mitochondrial genes to the nuclear genome, despite NUMTs of all lengths, from short fragments to the entire length of the mtDNA, occurring across different species [1,22]. While the mtDNA has been shown to be transferred to the nuclear genome in its entirety as a single NUMT in some individuals, often only fragments of the mtDNA are typically transferred, which are likely to be non-functional [1,3,18,19]. NUMTs may be inserted in a way that is similar or distinct from the process of full mitochondrial gene transfer to the nuclear genome. While full gene transfer of mitochondrial genes to the nuclear genome may result in the

co-expression of both genomes to regulate mitochondrial function, NUMTs may serve a different function, or possibly no function at all.

NUMTs are found across all nuclear chromosomes as insertions; hence, a biased mechanism of insertion towards any given locus seems unlikely [1]. However, a bias for open chromosomal locations adjacent to A + T oligomers has been noted [31]. Nonetheless, NUMTs do not appear to be independent loci under the control of their own promoters/repressors. As such, NUMTs appear to be seemingly accidental inclusions of mtDNA in the nuclear genome without a functional presence.

NUMTs may be preferentially inserted at points of double-stranded DNA breaks. The insertion of mtDNA fragments at double-stranded DNA breaks in the nuclear genome is thought by some to be an intentional process occurring during non-homologous end joining, a major double-stranded DNA break repair process in eukaryotic cells. Non-homologous end joining, while an effective way to repair double-stranded DNA breaks, results in deletion errors, possibly resulting in disease-causing frameshifts [32]. Environmental stressors and events that create double-stranded DNA breaks, including ionizing radiation exposure [33] and age-related free radical generation [34], are associated with an increase in the insertion of NUMTs, termed numtogenesis. Interestingly, the inclusion of NUMTs at non-homologous end joining repair sites in the nuclear genome of yeast was associated with a 46% reduction in DNA deletions [35]. Hence, greater biological fitness may be conferred by the incorporation of NUMTs during non-homologous end joining compared to cells without insertions of NUMTs.

A key question in the field is how the mtDNA fragments come to be inside the nucleus and available for insertion into double-stranded DNA breaks. Some literature suggests the involvement of mitophagy in the creation of freely available mtDNA fragments for insertion as NUMTs [36]. Additionally, double-stranded mtDNA was found to be transported across both mitochondrial membranes through the voltage-dependent anion channel [37]. Whether or not this is the primary mechanism of mtDNA movement out of mitochondria to be available in the nucleus for insertion into the nuclear genome as a NUMT remains unclear.

3. Technical Approaches for Limiting NUMTs

In addition to the computational approaches available for detecting NUMTs, several technical approaches can be used to limit the NUMTs in a sample. The most straightforward technical approach involves the isolation of mitochondria, thereby removing the nuclei, and hence, the nuclear genome and NUMTs, from the sample [38,39]. Mitochondrial isolation can be achieved with relative ease in cell pellets or whole animal or human tissue through the use of commercially available kits [40–42]. Differential centrifugation techniques employed by these kits result in minimal nuclear contamination, which would not be expected to contribute a sufficient number of reads during next-generation sequencing to be problematic in heteroplasmic mtDNA variant detection (Figure 1).

Following mitochondrial isolation, a method called Mito-SiPE (sequence-independent, PCR-free mitochondrial DNA enrichment) can be employed to determine the true mtDNA sequences and heteroplasmic variants present in the mtDNA sample. After mitochondrial fraction enrichment from cells or tissues, ultra-deep sequencing is performed to gain $>80,000\times$ coverage of the mtDNA, thereby limiting the introduction of false heteroplasmic variants from PCR enrichment or NUMT contamination [38]. Several PCR-based techniques such as Mito-tiling and rolling circle amplification [43] also enrich for mtDNA sequences, particularly in samples with a low mtDNA copy number; however, such approaches may also co-amplify NUMTs due to high sequence similarity.

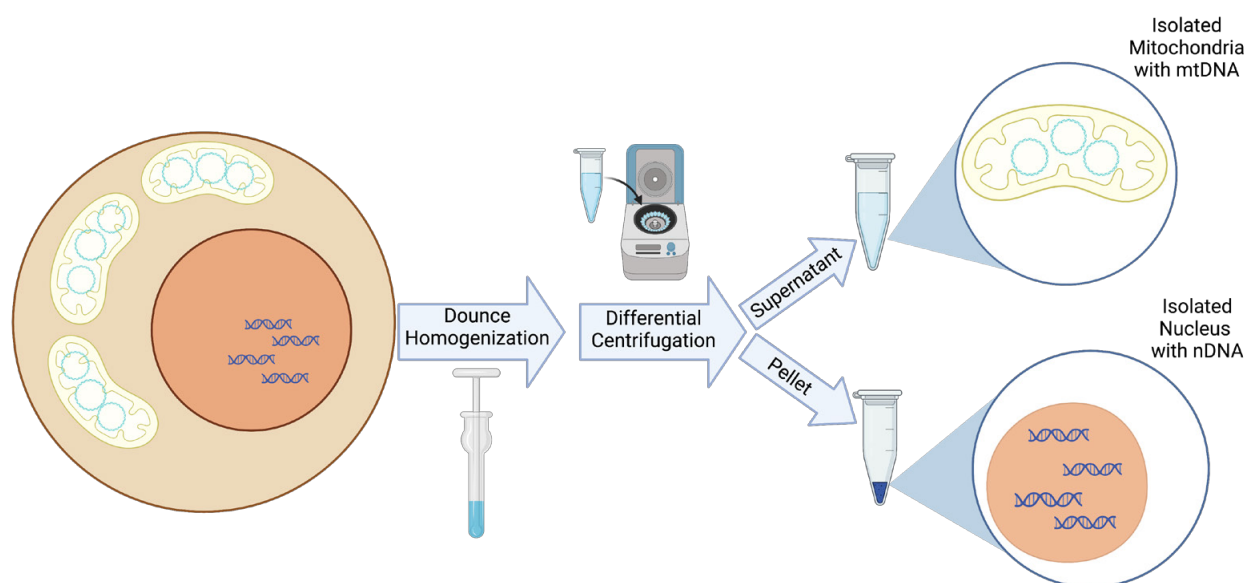


Figure 1. A schematic representation of mitochondrial and nuclear isolation. In brief, a membrane detergent is added to a pellet of whole cells or tissues. Next, dounce homogenization is performed with the specific number of dounce strokes dependent upon the cell or tissue type. Following homogenization, differential centrifugation over several steps first pellets nuclei and then mitochondria based on their different masses, allowing for the collection of both purified nuclear DNA (nDNA) and mtDNA from the same cells or tissue.

4. Computational Identification of NUMTs

To prevent the interference of NUMTs in studies of mitochondrial genetic variation, some studies simply remove loci from existing catalogs of NUMTs from their analyses; however, this approach is likely not sufficient. Apart from common NUMTs that are widely shared, most NUMTs are rare and specific to individuals [1]. The use of bioinformatic approaches is required to efficiently and effectively identify all NUMTs in next-generation sequencing datasets prior to performing analyses of mtDNA variants with outcomes and phenotypes of interest.

Before the advent of next-generation sequencing, NUMTs were primarily detected through a basic local alignment search (BLAST) of a PCR-amplified DNA sequence to reference genomes [20,44]. To discover novel NUMTs in a genome, exhaustive BLAST searches were also performed. By aligning the entire mtDNA sequence of multiple samples to the reference nuclear genome using a nucleotide BLAST, a database of common NUMTs in humans was built [45]. Similarly, an exhaustive BLAST search aligning a specific region of the mtDNA to the nuclear genome of different species provided additional insights into how NUMTs of certain mtDNA regions differ across species [46]. BLAST is still used to investigate or confirm whether annotated ancestral NUMTs are present in next-generation sequencing data, which can then be excluded prior to analyses of mtDNA variants with phenotypes of interest (Figure 2) [47]. However, more sophisticated tools are now available for NUMT detection in next-generation sequencing data and will be the focus of this section of the review.

A novel NUMT discovery pipeline for paired-end whole genome sequencing (WGS) data has been developed and applied in multiple studies (Figure 3) [1,48,49]. In the pipeline, discordant paired-reads, with one read mapped to the nuclear genome and the other read mapped to the mtDNA, are selected and clustered based on their location, orientation, and insertion size. Discordant read clusters with more than five paired reads are used to identify the breakpoint location to determine the NUMT location within the nuclear genome. WGS reads in regions of putative breakpoints are then reviewed again in search of split reads, where half of the read is mapped to the nuclear genome, and half to the

mitochondrial genome. With five or more split reads of good quality, the breakpoints can then be readily defined. To further locate the site of insertion of the NUMTs, split reads are realigned to both nuclear and mitochondrial reference genomes. Novel NUMTs in WGS data detected by this pipeline are further validated using long-read sequencing or through an additional examination on genome browsers.

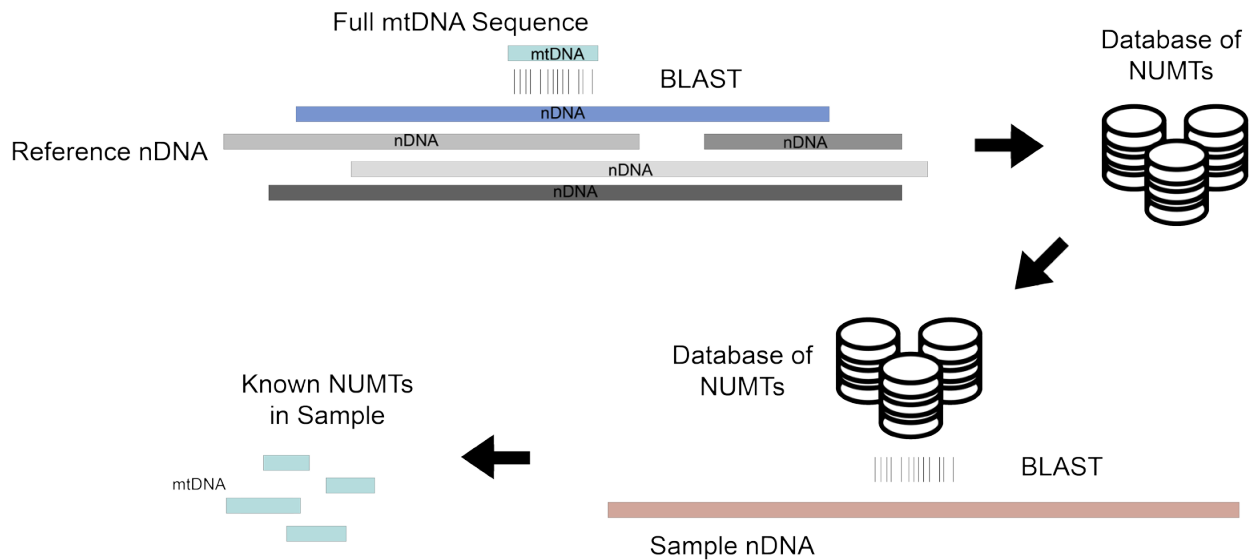


Figure 2. Using BLAST to identify NUMTs: The full mtDNA sequence from multiple individuals is aligned against the reference nuclear DNA (nDNA) with nucleotide BLAST in order to build a database of common NUMTs. Nucleotide BLAST searches performed between a sample nDNA and a database of common NUMTs can detect annotated ancestral NUMTs present in the sample nDNA, but cannot effectively detect polymorphic NUMTs specific to individuals, which are the majority of NUMTs found in humans. Hence, nucleotide BLAST will miss the vast majority of NUMTs, which are polymorphic NUMTs unique to individuals [1].

An alignment-free approach to detect NUMTs has also been developed (Figure 4). Local alignment assumes that the divergence between two sequences is small; hence, local alignments may fail to detect significant similarity between sequences with larger mutational dynamics, as is the case with NUMTs [50]. A k-mer-based NUMT detection method was developed that uses a moving window of 3000 bases. The algorithm moves a step of $1/8$ the window size, targeting all NUMTs within the range of 3000 base pairs of length. After applying moving windows to both mtDNA and nuclear DNA in a sample, the frequency of k-mers in each window is recorded, resulting in multiple k-mer frequency distributions. Jensen–Shannon divergence is then used to measure the similarity between the two k-mer frequency distributions (one for each genome). A window size and moving steps of $k = 7$ was found to be optimal at efficiently providing unique k-mer distributions for ancestral NUMT detection [50].

NUMTs within WGS datasets can also be identified using mtDNA copy number (Figure 5). MtDNA copy number negatively correlates with NUMT-derived heteroplasmic mtDNA variants [51]. Most false positive NUMT-derived heteroplasmic variants have a VAF of 0–5% in human blood samples [51]. Theoretically, a NUMT-derived heteroplasmic variant will have a VAF of approximately $1/(1 + \text{mtDNA copy number})$, where mtDNA coverage is twice the nuclear coverage. NUMTs can be identified by using Spearman’s correlation to identify variants with a VAF correlated with $1/(1 + \text{mtDNA copy number})$, after which, the reads containing the candidate NUMT-derived variants are remapped back to the human reference genome. This method also identifies NUMT-derived heteroplasmic mtDNA variants where at least two variants are derived from the same NUMT insertion.

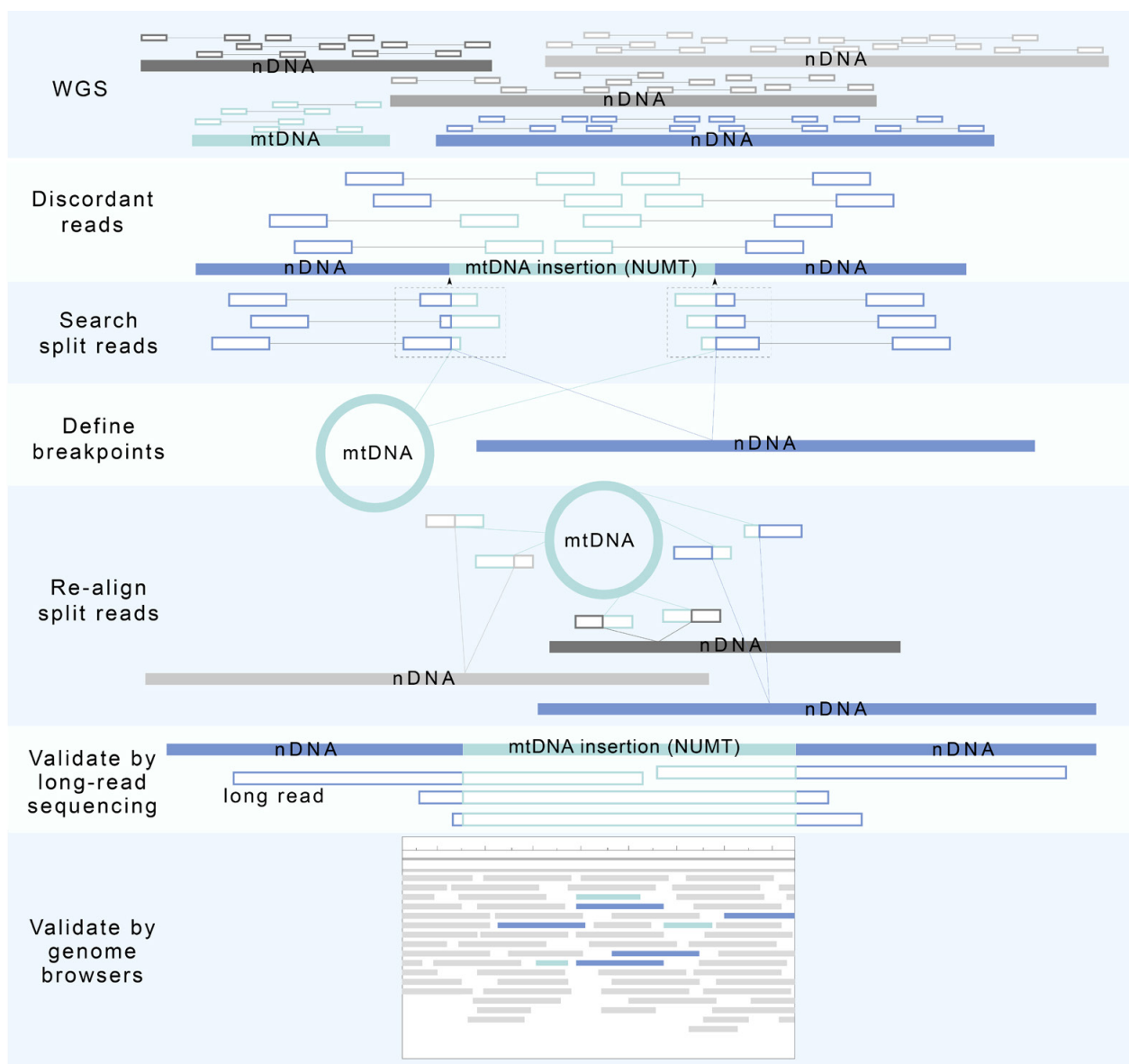


Figure 3. A novel NUMT detection pipeline for paired-end WGS data. In this pipeline, discordant read pairs, with one read in the pair aligning to the nuclear DNA (nDNA) and the other read aligning to the mtDNA, are first detected and clustered to locate a putative breakpoint. With multiple split reads retained, a breakpoint can be identified. Split reads are then re-aligned to the reference mtDNA and nDNA to locate the origin of NUMT insertion. Long-read sequencing and genome browsers can be used to further validate the location of detected NUMTs.

If the mtDNA copy number is high enough (>500 copies) for a sample, as is the case for highly energetic tissues, NUMTs are likely not problematic for detecting heteroplasmic variants with a low VAF of 1–5%. In tissues with high mtDNA copy numbers, most NUMT-derived heteroplasmic variants will have a VAF below the commonly used VAF thresholds of 3–10% for defining heteroplasmic variants [51]. However, most population-level studies use whole blood or peripheral blood mononuclear cells for WGS, which do not have sufficiently high mtDNA copy numbers to allow for false positive heteroplasmic variants to be confidently removed at low VAF thresholds. Hence, studies using WGS from blood samples require the implementation of additional computational strategies for NUMT detection.

To distinguish low-level heteroplasmic variants due to mtDNA variants rather than variants found in NUMTs, software such as DREEP [52] has been developed to provide DQS scores, a Phred-like quality score. DQS scores are measures of the deviation of the observed minor allele count from the expected error count, derived from a reference panel to indicate whether the mtDNA heteroplasmic variant is the result of a sequencing error. When mtDNA variants have a VAF ≥ 0.015 and a ≥ 4 DQS score, no false positive variants due to NUMTs are identified, suggesting that the combined thresholds for minor allele frequency and DQS is sufficient for high-confidence mtDNA variant identification [52]. Using a reference panel can be extremely problematic, however, as multiple studies have shown that samples have very different mtDNA copy numbers, even when the same DNA extraction kits and sequencing protocols are used [51]. As mtDNA copy number is closely related to the probability of identifying false positive variants due to NUMT contamination [51], using a reference panel like DREEP may be problematic.

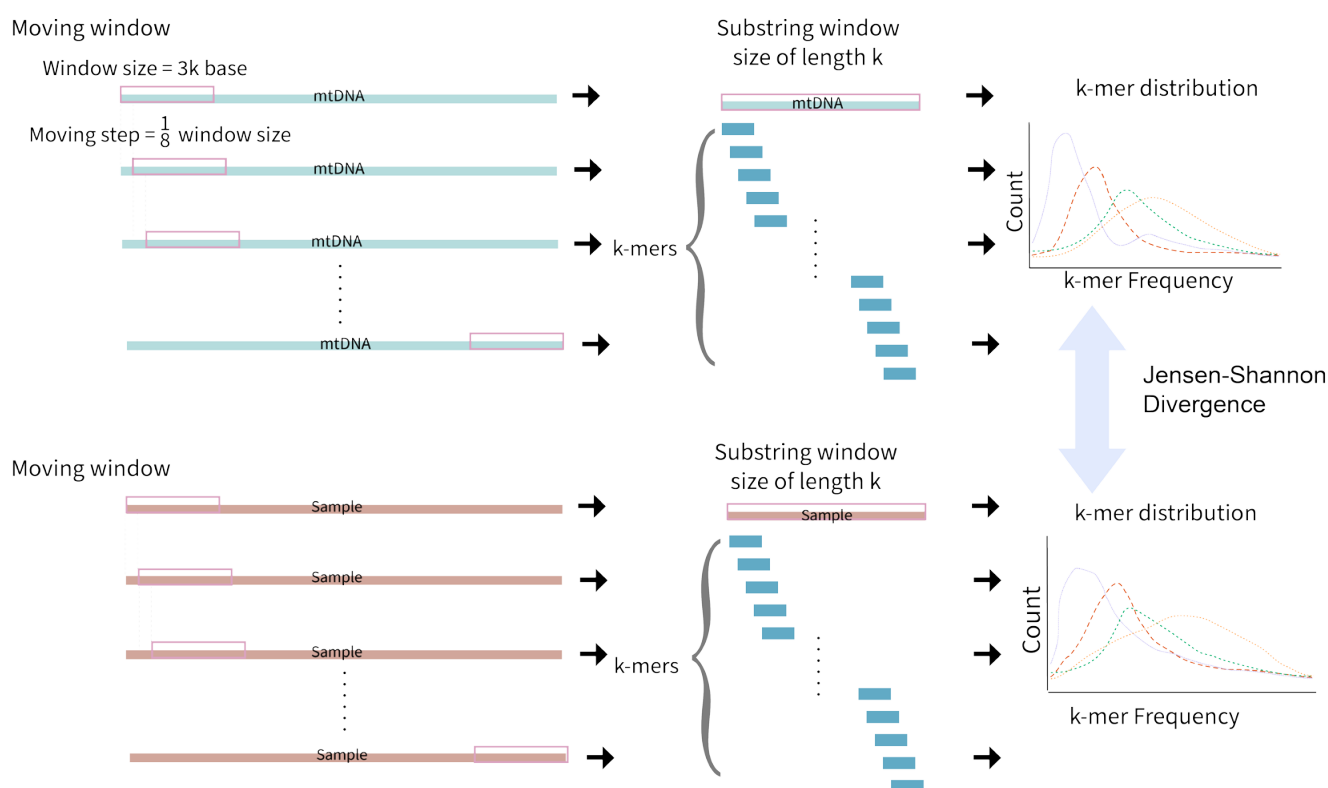


Figure 4. The k-mer-based NUMT detection method. A 3000 base pair moving window first takes a step of 1/8 window size on the full mtDNA sequence. Each window is sub-stringed with size k and the distribution of k -mers is recorded. The same moving window is applied to sample nuclear DNA (nDNA), and the k -mer distributions are compared with those from mtDNA using Jensen–Shannon Divergence. This approach cannot effectively detect polymorphic NUMTs specific to individuals when performed with reference genomes, and hence, may miss a number of NUMTs.

To validate that the detected variants are truly from mtDNA, some bioinformatic pipelines generate a de novo mtDNA reference for each individual sample and conduct a second iteration of variant calling [43,53]. Using a de novo consensus mtDNA reference specific to each individual reduces potential bias introduced by the commonly used revised Cambridge Reference sequence [54], as more reads are aligned with high confidence using the de novo consensus mtDNA reference. The increase in the coverage afforded by the consensus reference results in higher confidence in the identification of variants. Although only applying a consensus mtDNA as a reference may fall short of minimizing the contamination of long NUMTs, the majority of the NUMTs are short insertions [1] and

confirmation of the existence of long NUMT sequences necessitates the use of long-read sequencing. No studies have directly compared the mtDNA variant calling results between the revised Cambridge Reference sequence and the consensus reference sequence, but using a de novo consensus mtDNA reference for each sample will likely reduce the contamination of NUMTs, particularly in cell or tissue types with a low mtDNA copy number.

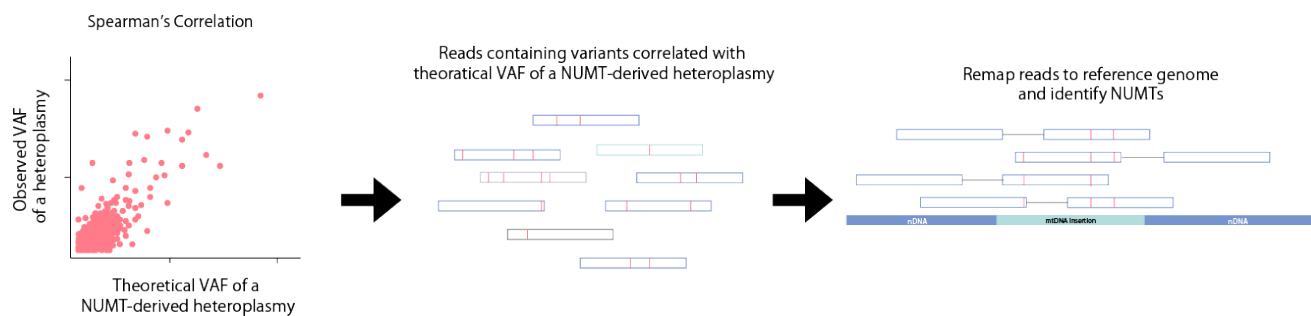


Figure 5. A NUMT detection method comparing the observed and theoretical VAFs of NUMT-derived heteroplasmic variants. Due to the relation of mtDNA coverage, nuclear DNA (nDNA) coverage, and mtDNA copy number, the theoretical VAF of a NUMT-derived heteroplasmic variant can be calculated. Heteroplasmic variants with a VAF that correlates with the theoretical VAF of a NUMT-derived heteroplasmic variant are likely false positives. The reads containing the false positive NUMT-derived heteroplasmic variant are then remapped to the reference genome to locate the NUMT insertion point.

5. Conclusions

NUMTs appear in all humans and across the entirety of the nuclear genome. Emerging evidence suggests that there may be a physiological and possible protective role of NUMT insertion into nuclear double-stranded break regions. However, false positive heteroplasmic variants due to the presence of NUMTs remains a challenge in confidently identifying low-level heteroplasmic variants in mtDNA sequencing datasets. Nonetheless, the methods for minimizing and identifying NUMTs, both technical and computational, are rapidly advancing. Mitochondrial isolation with subsequent sequencing represents an ideal scenario for removing the effect of NUMT contamination in mtDNA sequences. When mitochondrial isolation is not technically feasible, employing multiple computational methods can aid in identifying and removing NUMTs to facilitate studies of heteroplasmic variants with low VAFs. Ultimately, it is critical to run one or more of the aforementioned countermeasures to ensure accurate heteroplasmic mtDNA variant detection in mitochondrial genome sequencing.

Author Contributions: Conceptualization, all authors.; figure preparation, L.X., J.D.M. and J.L.F.; writing—original draft preparation, all authors.; writing—review and editing, all authors.; supervision, J.L.F.; funding acquisition, J.L.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Heart, Lung, and Blood Institute T32 HL007224-45 (J.D.M.) and K01 HL143142 (J.L.F.).

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: Not Applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wei, W.; Schon, K.R.; Elgar, G.; Orioli, A.; Tanguy, M.; Giess, A.; Tischkowitz, M.; Caulfield, M.J.; Chinnery, P.F. Nuclear-embedded mitochondrial DNA sequences in 66,083 human genomes. *Nature* **2022**, *611*, 105–114. [[CrossRef](#)] [[PubMed](#)]
2. Lopez, J.V.; Yuhki, N.; Masuda, R.; Modi, W.; O'Brien, S.J. 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. [[CrossRef](#)] [[PubMed](#)]
3. Tourmen, Y.; Baris, O.; Dessen, P.; Jacques, C.; Malthiery, Y.; Reynier, P. Structure and Chromosomal Distribution of Human Mitochondrial Pseudogenes. *Genomics* **2002**, *80*, 71–77. [[CrossRef](#)]
4. Thalmann, O.; Hebler, J.; Poinar, H.N.; Pääbo, S.; Vigilant, L. Unreliable mtDNA data due to nuclear insertions: A cautionary tale from analysis of humans and other great apes. *Mol. Ecol.* **2004**, *13*, 321–335. [[CrossRef](#)]
5. Parakatselaki, M.E.; Zhu, C.T.; Rand, D.; Ladoukakis, E.D. NUMTs Can Imitate Biparental Transmission of mtDNA—A Case in *Drosophila melanogaster*. *Genes* **2022**, *13*, 1023. [[CrossRef](#)]
6. Bensasson, D.; Feldman, M.W.; Petrov, D. Rates of DNA Duplication and Mitochondrial DNA Insertion in the Human Genome. *J. Mol. Evol.* **2003**, *57*, 343–354. [[CrossRef](#)] [[PubMed](#)]
7. Du Buy, H.G.; Riley, F.L. Hybridization between the nuclear and kinetoplast DNA'S of leishmania enriettii and between nuclear and mitochondrial DNA'S of mouse liver. *Proc. Natl. Acad. Sci. USA* **1967**, *57*, 790–797. [[CrossRef](#)]
8. Lewin, R. Promiscuous DNA Leaps All Barriers. *Science* **1983**, *219*, 478–479. [[CrossRef](#)]
9. Zullo, S.; Sieu, L.C.; Slightom, J.L.; Hadler, H.I.; Eisenstadt, J.M. Mitochondrial D-loop sequences are integrated in the rat nuclear genome. *J. Mol. Biol.* **1991**, *221*, 1223–1235.
10. Fukuda, M.; Wakasugi, S.; Tsuzuki, T.; Nomiya, H.; Shimada, K.; Miyata, T. Mitochondrial DNA-like sequences in the human nuclear genome: Characterization and implications in the evolution of mitochondrial DNA. *J. Mol. Biol.* **1985**, *186*, 257–266. [[CrossRef](#)]
11. Farrelly, F.; Butow, R.A. Rearranged mitochondrial genes in the yeast nuclear genome. *Nature* **1983**, *301*, 296–301. [[CrossRef](#)] [[PubMed](#)]
12. Hadler, H.I.; Dimitrijevic, B.; Mahalingam, R. Mitochondrial DNA and nuclear DNA from normal rat liver have a common sequence. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 6495–6499. [[CrossRef](#)]
13. Boogaart, P.V.D.; Samallo, J.; Agsteribbe, E. Similar genes for a mitochondrial ATPase subunit in the nuclear and mitochondrial genomes of *Neurospora crassa*. *Nature* **1982**, *298*, 187–189. [[CrossRef](#)]
14. Gellissen, G.; Bradfield, J.Y.; White, B.N.; Wyatt, G.R. Mitochondrial DNA sequences in the nuclear genome of a locust. *Nature* **1983**, *301*, 631–634. [[CrossRef](#)] [[PubMed](#)]
15. Hazkani-Covo, E.; Zeller, R.M.; Martin, W. Molecular Poltergeists: Mitochondrial DNA Copies (numts) in Sequenced Nuclear Genomes. *PLoS Genet.* **2010**, *6*, e1000834. [[CrossRef](#)] [[PubMed](#)]
16. Brown, W.M.; Prager, E.M.; Wang, A.; Wilson, A.C. Mitochondrial DNA sequences of primates: Tempo and mode of evolution. *J. Mol. Evol.* **1982**, *18*, 225–239. [[CrossRef](#)]
17. Zischler, H.; Geisert, H.; Von Haeseler, A.; Pääbo, S. A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. *Nature* **1995**, *378*, 489–492. [[CrossRef](#)]
18. Mourier, T.; Hansen, A.J.; Willerslev, E.; Arctander, P. The Human Genome Project Reveals a Continuous Transfer of Large Mitochondrial Fragments to the Nucleus. *Mol. Biol. Evol.* **2001**, *18*, 1833–1837. [[CrossRef](#)]
19. Woischnik, M.; Moraes, C.T. Pattern of Organization of Human Mitochondrial Pseudogenes in the Nuclear Genome. *Genome Res.* **2002**, *12*, 885–893. [[CrossRef](#)]
20. Parr, R.L.; Maki, J.; Reguly, B.; Dakubo, G.D.; Aguirre, A.; Wittrock, R.; Robinson, K.; Jakupciak, J.P.; E Thayer, R. The pseudo-mitochondrial genome influences mistakes in heteroplasmy interpretation. *BMC Genom.* **2006**, *7*, 185. [[CrossRef](#)]
21. Parfait, B.; Rustin, P.; Munnich, A.; Rötig, A. Co-amplification of nuclear pseudogenes and assessment of heteroplasmy of mitochondrial DNA mutations. *Biochem. Biophys. Res. Commun.* **1998**, *247*, 57–59. [[CrossRef](#)]
22. Zhang, D.-X.; Hewitt, G.M. Nuclear integrations: Challenges for mitochondrial DNA markers. *Trends Ecol. Evol.* **1996**, *11*, 247–251. [[CrossRef](#)] [[PubMed](#)]
23. Wallace, D.C.; Stugard, C.; Murdock, D.; Schurr, T.; Brown, M.D. Ancient mtDNA sequences in the human nuclear genome: A potential source of errors in identifying pathogenic mutations. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 14900–14905. [[CrossRef](#)] [[PubMed](#)]
24. Davis, R.E.; Miller, S.; Herrstadt, C.; Ghosh, S.S.; Fahy, E.; Shinobu, L.A.; Galasko, D.; Thal, L.J.; Beal, M.F.; Howell, N.; et al. Mutations in mitochondrial cytochrome c oxidase genes segregate with late-onset Alzheimer disease. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 4526–4531. [[CrossRef](#)] [[PubMed](#)]
25. Davis, J.N., 2nd; Parker, W.D., Jr. Evidence that two reports of mtDNA cytochrome c oxidase "mutations" in Alzheimer's disease are based on nDNA pseudogenes of recent evolutionary origin. *Biochem. Biophys. Res. Commun.* **1998**, *244*, 877–883. [[CrossRef](#)]
26. Maude, H.; Davidson, M.; Charitakis, N.; Diaz, L.; Bowers, W.H.T.; Gradovich, E.; Andrew, T.; Huntley, D. NUMT Confounding Biases Mitochondrial Heteroplasmy Calls in Favor of the Reference Allele. *Front. Cell Dev. Biol.* **2019**, *7*, 201. [[CrossRef](#)]

27. Brandvain, Y.; Wade, M.J. The Functional Transfer of Genes from the Mitochondria to the Nucleus: The Effects of Selection, Mutation, Population Size and Rate of Self-Fertilization. *Genetics* **2009**, *182*, 1129–1139. [[CrossRef](#)] [[PubMed](#)]
28. Andersson, S.G.; Kurland, C.G. Origins of mitochondria and hydrogenosomes. *Curr. Opin. Microbiol.* **1999**, *2*, 535–541. [[CrossRef](#)]
29. Margulis, L. Origin of Mitochondria and Hydrogenosomes. *Hist. Philos. Life Sci.* **2008**, *30*, 473–477.
30. Hedges, S.B.; Chen, H.; Kumar, S.; Wang, D.Y.; Thompson, A.S.; Watanabe, H. A genomic timescale for the origin of eukaryotes. *BMC Evol. Biol.* **2001**, *1*, 4. [[CrossRef](#)]
31. Tsuji, J.; Frith, M.C.; Tomii, K.; Horton, P. Mammalian NUMT insertion is non-random. *Nucleic Acids Res.* **2012**, *40*, 9073–9088. [[CrossRef](#)] [[PubMed](#)]
32. Davis, A.J.; Chen, D.J. DNA double strand break repair via non-homologous end-joining. *Transl. Cancer Res.* **2013**, *2*, 130–143. [[CrossRef](#)]
33. Gaziev, A.I.; Shaikhaev, G.O. Ionizing radiation can activate the insertion of mitochondrial DNA fragments in the nuclear genome. *Radiats Biol. Radioecol.* **2007**, *47*, 673–683. [[PubMed](#)]
34. Cheng, X.; Ivessa, A.S. The migration of mitochondrial DNA fragments to the nucleus affects the chronological aging process of *Saccharomyces cerevisiae*. *Aging Cell* **2010**, *9*, 919–923. [[CrossRef](#)]
35. Hazkani-Covo, E.; Covo, S. Numt-Mediated Double-Strand Break Repair Mitigates Deletions during Primate Genome Evolution. *PLoS Genet.* **2008**, *4*, e1000237. [[CrossRef](#)] [[PubMed](#)]
36. Singh, K.K.; Choudhury, A.R.; Tiwari, H.K. Numtogenesis as a mechanism for development of cancer. *Semin. Cancer Biol.* **2017**, *47*, 101–109. [[CrossRef](#)]
37. Szabò, I.; Bãthori, G.; Tombola, F.; Coppola, A.; Schmehl, I.; Brini, M.; Ghazi, A.; De Pinto, V.; Zoratti, M. Double-stranded DNA can be translocated across a planar membrane containing purified mitochondrial porin. *FASEB J.* **1998**, *12*, 495–502. [[CrossRef](#)]
38. Walsh, D.J.; Bernard, D.J.; Pangilinan, F.; Esposito, M.; Harold, D.; Parle-McDermott, A.; Brody, L.C. Mito-SiPE is a sequence-independent and PCR-free mtDNA enrichment method for accurate ultra-deep mitochondrial sequencing. *Commun. Biol.* **2022**, *5*, 1269. [[CrossRef](#)]
39. Gould, M.P.; Bosworth, C.; McMahon, S.; Grandhi, S.; Grimerg, B.T.; LaFramboise, T. PCR-Free Enrichment of Mitochondrial DNA from Human Blood and Cell Lines for High Quality Next-Generation DNA Sequencing. *PLoS ONE* **2015**, *10*, e0139253. [[CrossRef](#)]
40. Louwagie, E.J.; Larsen, T.D.; Wachal, A.L.; Gandy, T.C.; Baack, M.L. Mitochondrial Transfer Improves Cardiomyocyte Bioenergetics and Viability in Male Rats Exposed to Pregestational Diabetes. *Int. J. Mol. Sci.* **2021**, *22*, 2382. [[CrossRef](#)]
41. Bhupana, J.N.; Huang, B.-T.; Liou, G.-G.; Calkins, M.J.; Lin-Chao, S. Gas7 knockout affects PINK1 expression and mitochondrial dynamics in mouse cortical neurons. *FASEB BioAdv.* **2020**, *2*, 166–181. [[CrossRef](#)]
42. Zeng, X.; Zhao, L.; Chen, S.; Li, X. Inhibition of mitochondrial and cytosolic calpain attenuates atrophy in myotubes co-cultured with colon carcinoma cells. *Oncol. Lett.* **2020**, *21*, 124. [[CrossRef](#)] [[PubMed](#)]
43. Ring, J.D.; Sturk-Andreaggi, K.; Peck, M.A.; Marshall, C. Bioinformatic removal of NUMT-associated variants in mitotiling next-generation sequencing data from whole blood samples. *Electrophoresis* **2018**, *39*, 2785–2797. [[CrossRef](#)]
44. Qu, H.; Ma, F.; Li, Q. Comparative analysis of mitochondrial fragments transferred to the nucleus in vertebrate. *J. Genet. Genom.* **2008**, *35*, 485–490. [[CrossRef](#)]
45. Simone, D.; Calabrese, F.M.; Lang, M.; Gasparre, G.; Attimonelli, M. The reference human nuclear mitochondrial sequences compilation validated and implemented on the UCSC genome browser. *BMC Genom.* **2011**, *12*, 517. [[CrossRef](#)] [[PubMed](#)]
46. Soto-Calderón, I.D.; Lee, E.J.; Jensen-Seaman, M.I.; Anthony, N.M. Factors Affecting the Relative Abundance of Nuclear Copies of Mitochondrial DNA (Numts) in Hominoids. *J. Mol. Evol.* **2012**, *75*, 102–111. [[CrossRef](#)] [[PubMed](#)]
47. Bintz, B.J.; Dixon, G.B.; Wilson, M.R. 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. [[CrossRef](#)]
48. Dayama, G.; Emery, S.B.; Kidd, J.M.; Mills, R.E. The genomic landscape of polymorphic human nuclear mitochondrial insertions. *Nucleic Acids Res.* **2014**, *42*, 12640–12649. [[CrossRef](#)] [[PubMed](#)]
49. Wei, W.; Pagnamenta, A.T.; Gleadall, N.; Sanchis-Juan, A.; Stephens, J.; Broxholme, J.; Tuna, S.; Odhams, C.A. Genomics England Research Consortium; BioResource; et al. Nuclear-mitochondrial DNA segments resemble paternally inherited mitochondrial DNA in humans. *Nat. Commun.* **2020**, *11*, 1740. [[CrossRef](#)]
50. Li, W.; Freudenberg, J.; Freudenberg, J. Alignment-free approaches for predicting novel Nuclear Mitochondrial Segments (NUMTs) in the human genome. *Gene* **2019**, *691*, 141–152. [[CrossRef](#)]
51. Laricchia, K.M.; Lake, N.J.; Watts, N.A.; Shand, M.; Haessly, A.; Gauthier, L.; Benjamin, D.; Banks, E.; Soto, J.; Garimella, K.; et al. Mitochondrial DNA variation across 56,434 individuals in gnomAD. *Genome Res.* **2022**, *32*, 569–582. [[CrossRef](#)] [[PubMed](#)]
52. Li, M.; Schroeder, R.; Ko, A.M.-S.; Stoneking, M. Fidelity of capture-enrichment for mtDNA genome sequencing: Influence of NUMTs. *Nucleic Acids Res.* **2012**, *40*, e137. [[CrossRef](#)] [[PubMed](#)]

53. Battle, S.L.; Puiu, D.; Verlouw, J.; Broer, L.; Boerwinkle, E.; Taylor, K.D.; I Rotter, J.; Rich, S.S.; Grove, M.L.; Pankratz, N.; et al. A bioinformatics pipeline for estimating mitochondrial DNA copy number and heteroplasmy levels from whole genome sequencing data. *NAR Genom. Bioinform.* **2022**, *4*, lqac034. [[CrossRef](#)] [[PubMed](#)]
54. Andrews, R.M.; Kubacka, I.; Chinnery, P.F.; Lightowlers, R.N.; Turnbull, D.M.; Howell, N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **1999**, *23*, 147. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.