











RESEARCH ARTICLE

Accurate mapping of mitochondrial DNA deletions and duplications using deep sequencing

Swaraj Basu ^{1‡}, Xie Xie ^{1‡}, Jay P. Uhler¹, Carola Hedberg-Oldfors², Dusanka Milenkovic³, Olivier R. Baris ^{4,5}, Sammy Kimoloi⁴, Stanka Matic³, James B. Stewart ^{3,6}, Nils-Göran Larsson ⁷, Rudolf J. Wiesner ^{4,8}, Anders Oldfors ², Claes M. Gustafsson ¹, Maria Falkenberg ^{1‡*}, Erik Larsson ^{1‡*}

1 Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, **2** Department of Laboratory Medicine, Institute of Biomedicine, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, **3** Max Planck Institute for Biology of Ageing, Cologne, Germany, **4** Center for Physiology and Pathophysiology, Institute of Vegetative Physiology, University of Cologne, Cologne, Germany, **5** Equipe MitoLab, UMR CNRS 6015, INSERM U1083, Institut MitoVasc, SFR ICAT 4208, Université d'Angers, Angers, France, **6** Wellcome Centre for Mitochondrial Research, Newcastle University Biosciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle Upon Tyne, United Kingdom, **7** Division of Molecular Metabolism, Department of Medical Biochemistry and Biophysics (MBB), Karolinska Institute, Stockholm, Sweden, **8** Cologne Excellence Cluster on Cellular Stress Responses in Aging-associated Diseases (CECAD), University of Cologne, Cologne, Germany

‡ SB and XX share first authorship on this work. MF and EL are joint senior authors on this work.

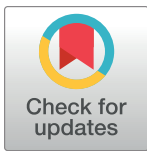
* maria.falkenberg@medkem.gu.se (MF); erik.larsson@gu.se (EL)

Abstract

Deletions and duplications in mitochondrial DNA (mtDNA) cause mitochondrial disease and accumulate in conditions such as cancer and age-related disorders, but validated high-throughput methodology that can readily detect and discriminate between these two types of events is lacking. Here we establish a computational method, MitoSAIt, for accurate identification, quantification and visualization of mtDNA deletions and duplications from genomic sequencing data. Our method was tested on simulated sequencing reads and human patient samples with single deletions and duplications to verify its accuracy. Application to mouse models of mtDNA maintenance disease demonstrated the ability to detect deletions and duplications even at low levels of heteroplasmy.

Author summary

Deletions in the mitochondrial genome cause a wide variety of rare disorders, but are also linked to more common conditions such as neurodegeneration, diabetes type 2, and the normal ageing process. There is also a growing awareness that mtDNA duplications, which are also relevant for human disease, may be more common than previously thought. Despite their clinical importance, our current knowledge about the abundance, characteristics and diversity of mtDNA deletions and duplications is fragmented, and based to large extent on a limited view provided by traditional low-throughput analyses.



OPEN ACCESS

Citation: Basu S, Xie X, Uhler JP, Hedberg-Oldfors C, Milenkovic D, Baris OR, et al. (2020) Accurate mapping of mitochondrial DNA deletions and duplications using deep sequencing. *PLoS Genet* 16(12): e1009242. <https://doi.org/10.1371/journal.pgen.1009242>

Editor: Ed Reznik, Memorial Sloan Kettering Cancer Center, UNITED STATES

Received: May 18, 2020

Accepted: November 2, 2020

Published: December 14, 2020

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.pgen.1009242>

Copyright: © 2020 Basu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The mouse sequencing data has been deposited in the European Nucleotide Archive (ENA) under accession PRJEB37552. The patient sequencing

data has been deposited in the European Genome-phenome Archive (EGA) under accession EGAS00001004380.

Funding: The work described here was supported by the Swedish Research Council (2018-02439 to M.F., 2017-01257 to C.M.G., and 2018-02852 to E.L.), the Swedish Cancer Foundation (2019-816 to M.F., 2017-631 to C.M.G., and 2018-747 to E.L.), the Knut and Alice Wallenberg Foundation (KAW 2017.0080 to M.F. and KAW 2015.0144 to E.L.), the European Research Council (683191 to M.F.) and grants from the Swedish state under the agreement between the Swedish government and the county councils, the ALF agreement (ALFGBG-727491 to M.F., and ALFGBG-728151 to C.M.G.). 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.

Here, we describe a bioinformatics method, MitoSAIt, that can accurately map and classify mtDNA deletions and duplications using high-throughput sequencing. Application of this methodology to mouse models of mitochondrial deficiencies revealed a large number of duplications, suggesting that these may previously have been underestimated.

Introduction

Mitochondria contain a separate genome which encodes essential subunits of the oxidative phosphorylation system and the RNA molecules (ribosomal and transfer RNA) needed for mitochondrial translation. Mitochondrial DNA (mtDNA) in humans is a small 16.6 kb circular molecule with only a few non-coding regions[1,2]. Thus, large deletions and duplications in mtDNA almost invariably lead to disruption of mitochondrial gene function. These types of structural alterations can be spontaneous or attributed to mutations affecting the nuclear-encoded mtDNA maintenance machinery, e.g. the mitochondrial DNA polymerase γ (POL γ) [3] or the replicative Twinkle helicase[4,5]. Deletions are a common cause of mitochondrial disorders[6–9] while also being linked to cancer[10–12], diabetes[13,14], neurodegenerative disorders[15,16], and the ageing process[16,17]. Duplications are less commonly described, but have for instance been described in patients with disease-causing mutations in *MGME1* [18,19] or in mice expressing a proof-reading-deficient version of Poly[20].

Despite the clinical significance of mtDNA structural alterations, our current knowledge about their abundance, diversity and exact localization is fragmented. A significant challenge is the multi-copy nature of mtDNA, with each cell containing hundreds to thousands of individual molecules. Most mtDNA alterations are heteroplasmic, meaning that wild-type mtDNA co-exists with mutant variants[21]. This complex DNA landscape makes the molecular characterization of mtDNA variants difficult, with low-level heteroplasmic variants being particularly hard to detect. The most commonly used detection methods, Southern blotting and long-range PCR[22], have limited resolution and cannot define all mtDNA variants in a given sample[23]. Even a variant present at high levels can remain undetected depending on the selection of primers, probes or restriction enzymes, and in the past, using these methods, duplications have wrongly been classified as deletions[18,19,24,25].

An attractive idea is therefore to use high-throughput sequencing to detect mtDNA deletions and duplications, as this potentially can provide more sensitive, less biased and more accurate mapping of these alterations. This would also dramatically simplify the workflow, and would enable exploration of mtDNA deletions and duplications in a large body of preexisting sequencing datasets. Due to the high copy number of mtDNA in cells ($n = 1,000\text{--}10,000$), mtDNA-derived reads are typically highly abundant in genomic sequencing data, in principle making the technology ideally suited for the purpose. The basic bioinformatics principles for determining structural alterations from short read sequencing are well-known, specifically identification of discordant paired-end reads or gapped/split alignment of individual reads to the reference genome. However, details in the implementation may have a large influence on performance, and tools for mapping structural changes in the nucleus show a surprising degree of discordance[26]. While the small size of the mitochondrial genome simplifies the problem, it is made harder by the fact that mitochondrial deletions commonly occur near repetitive sequences, and mapping of structural events on a circular genome presents additional challenges.

Several methods have recently been developed specifically for identification of mtDNA deletions from high-throughput short read sequencing, including MitoDel[27], Splice-Break

[28], eKLIPse[29], MitoMut[30], and a PERL script provided in (Zambelli et al., 2017)[31]. These methods rely on gapped alignments to predict deletions, but fail to recognize that every such event can represent either a deletion or a duplication affecting the arc complementary to the deleted part; a consequence of the circularity of mtDNA. Duplications can form as a consequence of mutations in mitochondrial replication factors, and correct identification and classification of such alterations is therefore an important requirement for any bioinformatics method pertaining to analysis of mtDNA structural changes.

Here we present the first high-throughput computational pipeline, MitoSAlt (Mitochondrial Structural Alterations), for identification, quantification and visualization of both deletions and duplications in mtDNA. The performance of MitoSAlt was carefully established using simulated sequencing data, patient samples with single events, and mouse models of mtDNA maintenance disease. MitoSAlt also introduces a way of visualizing the results such that duplications and deletions, as well as start and end positions, are unambiguously indicated. Using MitoSAlt, we also demonstrate that disease-causing mutations affecting specific steps in mtDNA replication cause distinct structural alterations in mtDNA.

Results

Detection of deletions and duplications with MitoSAlt

MitoSAlt is designed to take single- or paired-end sequencing reads as input to generate a map of predicted deletions and duplications, visualized in a circular plot along with tab delimited tables detailing the breakpoint positions and heteroplasmy levels (Fig 1A, further detailed in Materials and Methods). The pipeline relies on an initial alignment of sequencing reads to the nuclear and mitochondrial (Mt) genome using HISAT2[32] to remove nuclear reads while retaining mtDNA-mapped and unmapped reads. This step accelerates the analysis, but may optionally be disabled when working with species having extensive nuclear mitochondrial DNA (NUMT) regions such as mouse[33] to avoid patch-wise reduced mtDNA read coverage. This is followed by alignment to mtDNA using LAST[34], processing of the LAST results to identify deletions and duplications based on split alignments, and classification of deletions and duplications along with plotting the results and generating final tables. Additionally, in the case of whole genome sequencing (WGS), when no mtDNA or nuclear enrichment has been

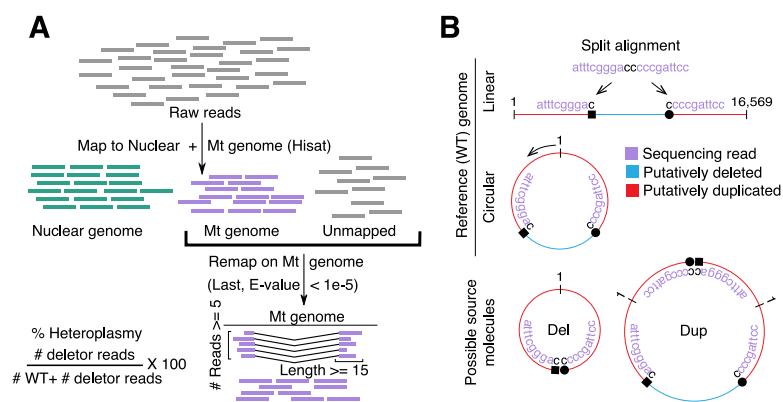


Fig 1. MitoSAlt pipeline overview. (A) Raw sequencing reads are mapped first to the nuclear and mitochondrial (Mt) genomes using a fast aligner, followed by precision alignment of unmapped and Mt mapped reads to the Mt genome to identify “split” reads informative of structural breakpoints. (B) Dual interpretation of split alignments: a split read can represent either a deletion or a complementary arc duplication, and these scenarios are indistinguishable using short-read sequencing.

<https://doi.org/10.1371/journal.pgen.1009242.g001>

performed, MitoSAlt can compare mitochondrial and nuclear read counts to estimate relative mtDNA levels, which are indicative of mtDNA copy number.

Similar to other methods, MitoSAlt relies on identification of reads aligning in a split/gapped fashion to the linear mitochondrial genome (Fig 1A). However, it is important to note that on a circular genome, every split read can represent either a deletion or, alternatively, a duplication of the mtDNA arc complementary to the deletion, and these two possibilities are indistinguishable when using short read sequencing (Fig 1B). MitoSAlt handles this by initially assuming that all events are deletions, followed by complementation and re-classification as a duplication in cases where the altered mtDNA molecule is deemed incapable of replicating due to loss of one or both origins (OriH or OriL; positions are user-definable). The favored interpretation is thus one where both origins are unaltered or, when this is not possible, none are deleted. The deletion/duplication classification is always non-ambiguous, since only one interpretation will satisfy these criteria while the other will violate them. Furthermore, the circularity of mtDNA implies that both deletions and duplications can produce alignments where the split segments map in reverse order to the linear reference, and care has been taken for MitoSAlt to handle and interpret this correctly (S1 Fig).

Evaluation of MitoSAlt on simulated sequencing data

We first evaluated the ability of the pipeline to accurately detect and classify both duplications and deletions based on a small set of simulated alterations present at high heteroplasmy levels. These were designed to cover the main classes of conceivable events that may still maintain mtDNA replicability. To this end, deletions (2,001–3,999 and the so-called common deletion [35] at 8,470–13,446; coordinates indicate start and end of the affected segment) and duplications (16,069–500, 2,500–3,500, 5,000–6,000, and 9,000–10,000) were introduced into the human reference mitochondrial genome (rCRS), each one at 16.7% heteroplasmy. These were combined with the nuclear genome to emulate a mitochondrial copy number of 6,000, and 10 million reads (5 million 2×126 bp) were generated using a model that emulates Illumina HiSeq characteristics [36]. Both alignment steps were performed (nuclear and mtDNA using HISAT2 followed by LAST on unmapped and Mt aligned reads). Eventually, 98.4% of mitochondrial reads ($n = 210,235$) were mapped to mtDNA, resulting in a mean coverage of $\sim 1,600\times$ (Fig 2A).

MitoSAlt accurately detected all events at single bp resolution and correctly classified them as deletions or duplications, with heteroplasmy estimates varying between 12.5% and 16.7% (Fig 2A). Between 148 and 213 reads were correctly aligned across each breakpoint (theoretical expectation 266 without any dropouts), while a smaller number of alignments (0–4 reads) supported breakpoints within 5 bp of the actual positions (Fig 2A). No other events were detected despite inclusion of nuclear chromosomes in the simulations. These results support that MitoSAlt can accurately identify and classify deletions and duplications without additional false positive detections.

We further compared the performance of MitoSAlt with five published pipelines on the same simulated dataset (S1 Table). Two of the tools, eKLIPse and the PERL script provided in (Zambelli et al., 2017), identified all events, but with the duplications reported as complementary arc deletions (i.e. start and end coordinates in reverse order). eKLIPse and Zambelli et al identified at most 94 and 186 breakpoint-spanning reads, respectively, suggesting that eKLIPse in particular has reduced sensitivity compared to MitoSAlt. Zambelli et al was less accurate when breakpoints were flanked by repeats: the 8,470–13,446 common deletion (flanked by a 13 bp identical repeat) was reported at 8482–13,447, and the 2,500–3,500 duplication (flanked by a longer imperfect repeat) was reported as a deletion at 3,525–2,500. The remaining tools

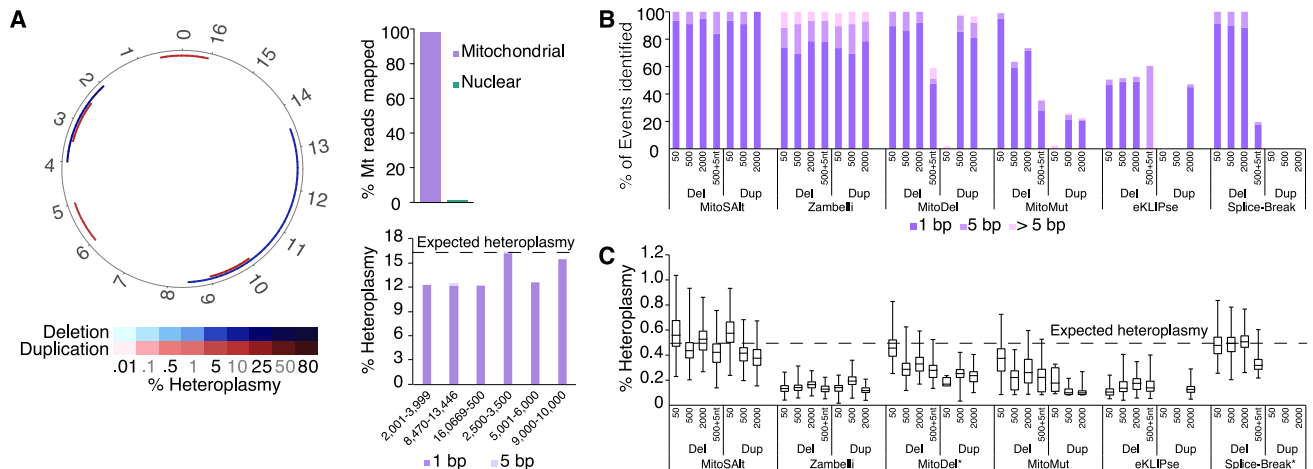


Fig 2. MitoSAlt pipeline performance on simulated data. (A) Evaluation on simulated sequencing data harboring two synthetic deletions and 4 duplications, each at 16.7% heteroplasmy (2×126 bp, 10,000,000 reads, resulting in $\sim 2,000\times$ mtDNA coverage). The circular plot shows deleted (blue) or duplicated (red) segments. The upper bar graph indicates the fraction of Mt reads mapped to the mitochondrial genome, while the lower shows heteroplasmy levels estimated by MitoSAlt for each event (events with 1 bp or 5 bp of the expected breakpoints are quantified separately). (B) Evaluation of sensitivity on simulated sequencing datasets containing large numbers of low heteroplasmy deletions and duplications of various sizes. Each data set contained 200 minor or major arc events, each at 0.5% heteroplasmy (2×126 bp, 50,000,000 reads, resulting in $\sim 6,000\times$ mtDNA coverage. “500+5nt” refers to 500 bp deletions with 5 bp non-template random insertions at the breakpoint. (C) Box and whisker plot of heteroplasmy levels estimated by different pipelines. The boxes show 25th to 75th percentiles, and whiskers show the minimum and maximum value. *, These tools do not directly report heteroplasmy levels, and estimates were instead made based on the reported number of reads supporting each event and the average mitochondrial read-depth.

<https://doi.org/10.1371/journal.pgen.1009242.g002>

identified 4 out of 6 events at best. Splice-Break specifically failed to identify duplications associated with inverse-order split alignments (S1 Fig), suggesting that the algorithm is not designed to handle this case. In addition to finding 4 out of the 6 true alterations, MitoMut identified 4 additional small deletions (S1 Table). Similar results were obtained when simulated reads were generated using an error model derived from empirical data (S1 Table).

Next, we generated simulated datasets containing large numbers of low heteroplasmy level (0.5%) deletions or duplications of various sizes (50, 500 and 2000 bp). Each dataset contained 200 events of a single type distributed across the major and minor arcs. Additionally, a dataset with 500 bp deletions with 5 bp random insertions was generated, to test the ability to handle non-template insertions at breakpoints. Mitochondrial number was set to 6,000, and 50 million reads (5 million 2×126 bp) were generated for each dataset, resulting in a mean coverage of $\sim 5,900\times$ on chrM. All events were detected by MitoSAlt and Zambelli et al, though the latter had lower accuracy with respect to exact determination of breakpoint coordinates (Fig 2B and S1 Table). Remaining tools all showed reduced or no sensitivity with respect to small duplications or duplications in general, as well as deletions with non-template insertions. Heteroplasmy estimates reported by MitoSAlt ranged from 0.38% to 0.57% on average in each dataset (Fig 2C). No events were detected by MitoSAlt or the other tools in a simulated wild type dataset of similar size. MitoSAlt thus compared favorably to other tools in terms of sensitivity and breakpoint coordinate accuracy, in addition to being the only method capable of differentiating between duplications and deletions.

Application to mitochondrial disease patients

We next tested the MitoSAlt pipeline on muscle biopsy DNA from mitochondrial disease patients with single high-heteroplasmy mtDNA deletions or duplications present at high levels as detected by long-range PCR (LX-PCR). Two patients carried a deletion while the third

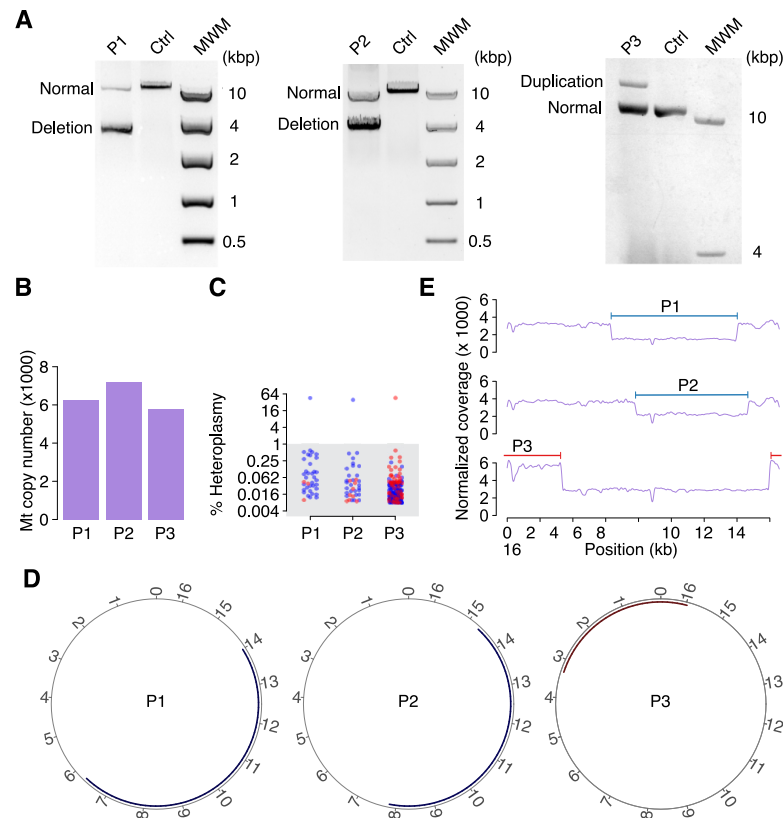


Fig 3. Assessment of MitoSAlt on patient samples with a single deletion or duplication. (A) Total DNA from patients (P1, P2 and P3) and controls were analyzed by LX-PCR using two different primer sets. A single deletion was detected in patients P1 and P2 using primers LX1 and LX2, while a single duplication was detected in P3 with primers LX3 and LX4. Amplicons from wild type mtDNA (denoted “normal”) were also detected in all patients. (B) Predicted mtDNA copy number in the patients. (C) Heteroplasmy levels for the identified deletions/duplications (marked in blue and red, respectively) in the patient samples. All cases have single events at heteroplasmy levels (> 35%), in addition to multiple low-heteroplasmy alterations (<1%, grey area). (D) Circular plots showing deletions/duplications at heteroplasmy >1%, all being consistent with the LX-PCR results. (E) Read coverage depth across the Mt genome for the human samples shows drastic changes in the regions identified as being deleted or duplicated (marked in blue and red respectively). MWM, molecular weight marker.

<https://doi.org/10.1371/journal.pgen.1009242.g003>

patient had a duplication (Fig 3A). WGS resulted in a coverage between 83,737 \times and 121,703 \times on chrM, and the estimated mtDNA levels, which can be used to predict mtDNA copy number, varied between 5,789 and 7,204 for all samples (Fig 3B). MitoSAlt detected a single high-level heteroplasmy (>50%) deletion or duplication in each patient as expected (Fig 3C). Additional low-level heteroplasmy (<1%) events often had breakpoints close to the main alterations, which may represent inaccurate alignments caused by sequencing errors (Fig 3C). The major breakpoints predicted by MitoSAlt (deletions at 6,330–13,993, 7,826–14,673 and a duplication spanning the D-loop at 15,973–3,326) were compatible with the LX-PCR results and corresponded closely to breakpoints estimated from chrM read depth (Fig 3D and 3E).

Additionally, we tested the MitoSAlt pipeline on whole genome sequencing data from three human tumors (deriving from liver, pancreas and skin), where read depth-based analysis previously suggested presence of large mtDNA duplications or deletions [37], and found that these events were confirmed by our approach (S2 Fig). These results provide further support that MitoSAlt can correctly identify breakpoints and classify events as deletions or duplications based on retention or loss of replication origins.

MitoSAlt detects large numbers of duplications in mouse models of mtDNA disease

Having validated the MitoSAlt pipeline on patients carrying single large-scale mtDNA duplications or deletions, we decided to extend our analysis to more complex DNA samples. To this end, we obtained DNA from mice previously shown to harbor multiple mtDNA structural alterations due to mutations in the gene for the Twinkle helicase (*Twink*^{K320E}; two different mice, M1 and M2), knockout of the mtDNA maintenance exonuclease Mgm1 (*Mgm1*^{-/-}; two different mice, M3 and M4), or mutations in the exonuclease domain of DNA polymerase gamma Poly (*Polg*^{D257A}; one mouse, M5). All three genes are important for mtDNA maintenance in mice and in humans [19,24,38–40]. The mutant mouse samples (M1-M5) and wild-type controls (C1-C5) were subjected to WGS (*Twink*^{K320E} and *Mgm1*^{-/-}) or sequencing following an mtDNA enrichment protocol (*Polg*^{D257A}), resulting in a coverage on chrM ranging from 35,913× to 150,182× (S3 Fig).

MtDNA level estimates for the *Twink*^{K320E} and *Mgm1*^{-/-} mutants were comparable to wild type control samples (Fig 4A and S1 Table), while the use of mtDNA enrichment precluded mtDNA level estimation in the *Polg*^{D257A} mutant sample. A large number of events were detected in all mutant samples (ranging from 95 to 4841), mostly duplications present at low

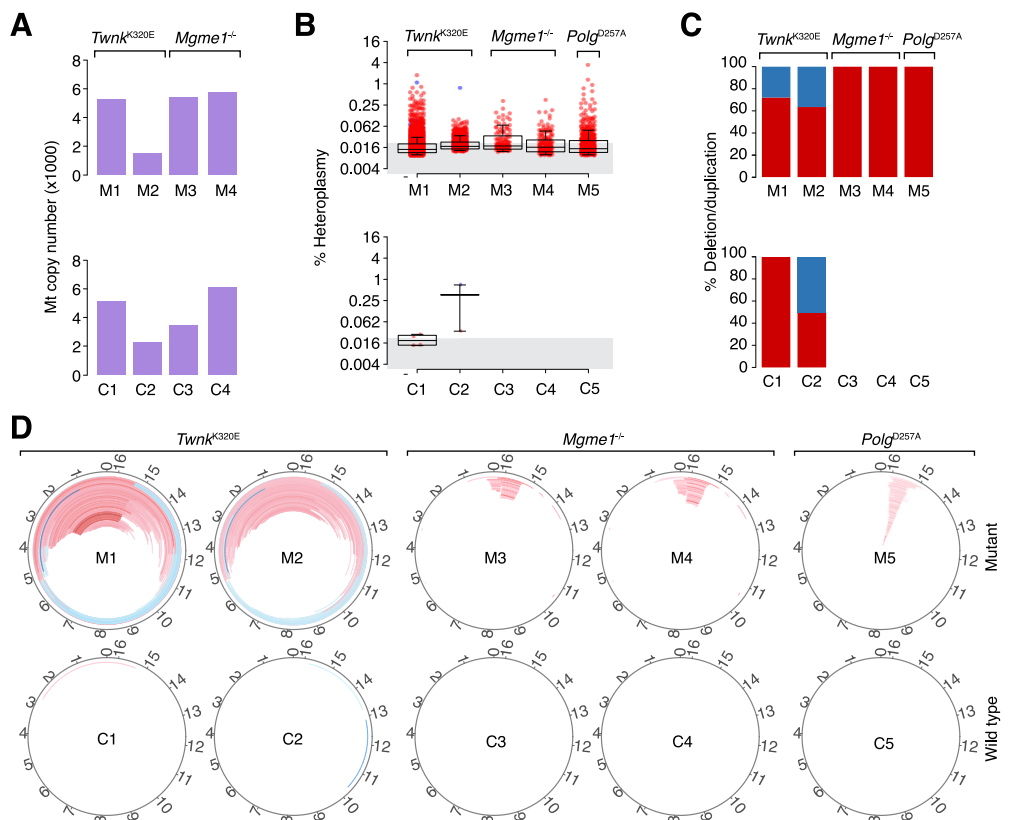


Fig 4. Identification of mtDNA structural alterations in wild-type and *Mgm1*, *Twink* or *Polg* mutant mice using the MitoSAlt pipeline. (A) Predicted copy number for the given mutant and wild-type samples (denoted M and C, respectively). Copy number could not be estimated for the *Polg* samples due to use of an mtDNA enrichment protocol. (B) Heteroplasmy levels for the deletions (blue) and duplications (red) identified in the mutant and wild-type samples. The grey area delineates low-heteroplasmy events (<0.02%). (C) Fraction deletions (blue) and duplications (red) in each sample. (D) Circular plots showing the deletions (blue) and duplications (red) identified in the mutant and wild-type samples. For visual clarity, a heteroplasmy cut-off of 0.02% was used for all samples.

<https://doi.org/10.1371/journal.pgen.1009242.g004>

heteroplasmy levels (maximum 3.47% and with the average per sample ranging from 0.023% to 0.038%; **Fig 4B and 4C**). In contrast, the negative control samples were essentially void of structural events (in total 5 events, all below 0.01%; **Fig 4B and 4C**).

Visualization of the events on the circular Mt genome revealed two distinct patterns, where *Mgme1*^{-/-} and *Polg*^{D257A} shared a common signature involving multiple, shorter duplications in the non-coding region (NCR), while *Twnk*^{K320E} instead was characterized by abundant longer duplications, spanning from a hotspot in the NCR to another hotspot in the middle of the minor arc (**Fig 4D**). These alteration signatures may reflect similarities and differences in the underlying molecular processes leading to breakpoint formation.

Discussion

MitoSAlt is the first pipeline explicitly designed to identify and correctly classify both deletions and duplications in mtDNA. MitoSAlt also provides a novel way of visualizing complex mtDNA alteration patterns, where deletions and duplications are unambiguously indicated along with their start and end positions and heteroplasmy levels. While primarily designed to be used on genomic sequencing data (whole genome or mtDNA enriched), MitoSAlt may in principle also be applicable to transcriptome or exome sequencing data, although the latter often exhibits limited mtDNA coverage. The performance of MitoSAlt was verified using simulated sequencing data, which showed that low heteroplasmy (0.5%) events are detectable with high sensitivity even at moderate sequencing depths. MitoSAlt was further applied to sequence data from human patients carrying single deletion/duplication events confirmed by LX-PCR, and mutant mice strains previously shown to harbor large numbers of mtDNA structural alterations[18,20,24,38].

Results from LX-PCR analysis of mice expressing *Twnk*^{K320E} (corresponding to the disease causing mutation *TWNK*^{K319E} in humans) have previously been interpreted as evidence for mtDNA deletions [39]. Interestingly, MitoSAlt instead predicted far more duplications (more than 85%) than deletions (less than 15%) in *Twnk*^{K320E} mice. Duplications also outnumbered deletions in mice with full-body knockout of *Mgme1*, recapitulating patients with homozygous nonsense mutations in *MGME1*[18], and in mice expressing exonuclease deficient Pol γ , *Polg*^{D257A}, which were previously proposed to harbor duplications in the same region [20]. Our results thus support that mtDNA duplications may be prevalent.

The mechanisms underlying mtDNA deletion formation have been carefully studied, leading to different models, including copy-choice recombination[41] and double-strand break repair[42]. How duplications are formed, and which enzymes are responsible, is still unclear, but the detailed data provided by MitoSAlt can be a useful resource for developing mechanistic hypotheses. For example, the similar alteration patterns seen in *Mgme1*^{-/-} and *Polg*^{D257A} mice (short duplications in the NCR) could indicate that these two enzymes are required for a common molecular function, a conclusion supported by previous studies, which have linked *Mgme1* and *Poly* to the formation of ligatable nicks during termination of mtDNA replication in the NCR[43–45].

MitoSAlt also estimates relative mtDNA levels, which are indicative of mtDNA copy number. However, for a more accurate mtDNA copy determination, the presence of large structural alterations in mtDNA must be considered. For example, long deletions present at high heteroplasmy will lead to a drop in mtDNA levels, even if the mtDNA copy number remains unchanged. In a related way, mtDNA copy number drops in the *Mgme1*^{-/-} mice [18], but mtDNA levels remain unchanged due to the constant production of long, linear mtDNA fragments that cannot be replicated or expressed.

In conclusion, MitoSAlt is carefully validated tool for precision mapping of mtDNA structural alterations, specifically designed to detect and discriminate between deletions and

duplications. MitoSAlt will facilitate further dissection of the mechanistic basis underlying the formation of these types of events, and will enable detailed analysis of samples from patients with mitochondrial diseases.

Materials and methods

Ethics statement

The transgenic mice studies were approved by the Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein–Westfalen (reference numbers 84–02.04.2015.A103, 84–02.05.50.15.004 and 2013-A165) and performed in accordance with the recommendations and guidelines of the Federation of European Laboratory Animal Science Associations (FELASA). Human patients gave informed consent for the investigations made and the study was approved by the Regional Ethics Committee at the University of Gothenburg, Sweden (number 390–07).

MitoSAlt

The MitoSAlt pipeline is comprised of three modules combined into a single pipeline: (1) alignment of sequencing reads (using PERL wrapper third party softwares), (2) parsing aligned reads to identify Mt breakpoints (PERL and R), and (3) plotting the results on the circular Mt genome and analysis of breakpoint repeats (R programming environment).

Alignment of sequencing reads

The raw sequencing reads are aligned to the source genome (Nuclear + Mitochondrial) using HISAT2[46]. HISAT2 is run with default parameters for RNA sequencing and specific parameters are used to customize it for DNA sequencing (`—no-temp-splicesite—no-spliced-alignment—max-intronlen 5000`). Following the first round of alignment the reads which remain unmapped or are mapped to the mitochondrial genome are extracted and converted to a concatenated FASTQ using Samtools. The FASTQ is realigned to the mitochondrial genome using the `lastal (-Q1 -e80)`, processed using `last-split` and converted from MAF to BAM and TAB format using `maf-convert`, where all the binaries are part of the LAST software package. The results in TAB format are parsed in PERL and R to classify the potential deletions and duplications. If the input sequencing data is enriched for mitochondrial DNA/RNA, then the pipeline skips the initial HISAT2 mapping and concatenates the FASTQ files using `reformat.sh` from BBMap software suite and maps the concatenated reads on the mitochondrial genome using LAST, where the downstream processing remains the same.

Parsing aligned reads to identify Mt breakpoints

The TAB formatted output is parsed in PERL to remove duplicated reads (both wildtype and mutant) and generate three output files a) BED format file with the list of split reads which may support a deletion or a duplication b) BREAKPOINT file with the list of breakpoints identified c) CLUSTER file, which groups the breakpoints at a given distance threshold and estimates the heteroplasmy at a given pair of clustered breakpoints as the ratio of reads supporting the breakpoints by the number of wildtype reads overlapping the breakpoints.

Final report and circular plots

The CLUSTER, and BREAKPOINT files are further used by an R script to generate a final table, classifying each cluster as a duplication or a deletion using the logic described in [S1 Fig](#). This report also contains information about direct repeat sequences overlapping with or

flanking the breakpoints. It should be noted that genomic coordinates in the final table refer to start and end positions of the deleted or duplicated segments, rather than junction coordinates. Finally, the breakpoint positions (at the cluster level) are plotted on a circular plot (size of the input mitochondrial genome) as arcs using the R plotrix package, where the individual arcs are colored to indicate whether they represent deletions or duplications, and where the estimated heteroplasmy is indicated by the intensity of the color.

Generation of simulated sequencing data

For the initial evaluation, involving a limited number of high heteroplasmy level events, six mutant mitochondrial reference genomes were generated, each containing a large deletion or duplication as detailed in Results. These were concatenated such that each would be present at a heteroplasmy of 16.7% and included in multiple copies together with the nuclear human chromosomes (hg19 assembly) to emulate an mtDNA copy number of 6000. Next we generated simulated reads using *InSilicoSeq*, a Python software package[36]. Two different error models were used: the default Illumina HiSeq model (10,000,000 2×126 bp paired-end reads) and an empirical error model base on NextSeq 500 generated whole genome sequencing data (6,000,000 2×76 bp paired-end reads). To evaluate the performance on a larger number of low heteroplasmy events, 6 separate datasets were generated, each containing 200 events as described in Results. These datasets were generated by concatenating mitochondrial genomes containing different deletions or duplications such that each would have a heteroplasmy level of 0.5%. These were combined with a nuclear human genome to emulate mtDNA copy number of 6,000. Simulated reads were generated using the Illumina HiSeq Model (50,000,000 2×126 bp paired-end reads).

DNA samples

For the LX-PCR and MitoSAlt analyses of human samples, total DNA was isolated from muscle biopsies from three patients with mitochondrial disease (Patient 1; age 9, Patient 2; age 16 and Patient 3; age 58) and age-matched control individuals using standard protocols. For MitoSAlt analysis of murine samples the following mice variants were used: *Twkn*^{K320E} transgenic mice expressing a dominant-negative mutant version of the *Twinkle* gene in skeletal muscle, which were generated by crossing R26-K320E-Twinkle^{loxP/+} mice[39] with *Mlc1f-cre* mice[47], *PolgA*^{D257A} mice carry a point mutation in the 3'-5' exonuclease domain of the replicative DNA polymerase POLG[38], and *Mgme1*^{-/-} knockout mice are deficient in the MGME1 exonuclease[24]. Total DNA was isolated from muscle for *Twkn*^{K320E} analysis, from heart for *Mgme1*^{-/-} analysis, and mtDNA was isolated from heart for PolgAD257A analysis using standard techniques.

LX-PCR

LX-PCR was performed on total DNA extracted from human muscle specimens to detect possible large scale mtDNA deletions and/or duplications using GoTaq Long PCR Master Mix according to the manufacturer's protocols (Promega, Madison WI, USA). The mtDNA was amplified with two sets of primers: set 1, LX1_m.5420-5447 (TGA ACA TAC AAA ACC CAC CCC ATT CCT C) and LX2_m.16232-16259 (GTG GCT TTG GAG TTG CAG TTG ATG TGT G) and set 2, LX3_m.8020-8000 (CGG GAG TAC TAC TCG ATT GTC) and LX4_m.13940-13972 (GCA CAA TCC CCT ATC TAG GCC TTC TTA CGA GCC) resulting in PCR products of size 10.8 kb and 10.6 kb, respectively, based on wild type mtDNA. PCR products were analysed by electrophoresis on 0.6% agarose gels.

Illumina sequencing

The patient samples were sequenced at Science for Life Laboratory in Stockholm, Sweden, using an Illumina NovaSeq 6000, resulting in 647.8–901.2 million 2×150 bp reads. The *Polg* mouse samples were sequenced at the Max Planck Genome Center in Cologne, Germany, using an Illumina HiSeq 2500, resulting in 100.7–101.4 million 2×250 bp reads. The *Twnk* and *Mgme1* mouse samples were sequenced at the Genomics Core Facility at the Sahlgrenska Academy in Gothenburg, Sweden, using an Illumina NovaSeq 6000, resulting in 602.7–691.5 million 2×150 bp reads.

Software availability

MitoSAlt is available through SourceForge at <https://sourceforge.net/projects/mitosalt>.

Supporting information

S1 Fig. Both deletions and duplications may give rise to forward or reverse order split alignments on a circular genome. The circularity of mtDNA presents special challenges when it comes to handling gapped/split alignments of short reads. Both deletions and duplications may give rise to split alignments where the split segments align in both forward or reverse order on the linear genome, depending on the type of alteration and its location relative to position 1, indicated here as O_H . Each gapped/split alignment, whether segments are in forward or reverse order, may represent either a deletion or a duplication, and these two possibilities are indistinguishable. Specifically, deletion of a specific segment A or duplication of the segment complementary to A (i.e. the remainder of the circular genome not covered by segment A) will produce identical split read alignments.

(EPS)

S2 Fig. Circular plots showing deletions/duplications in three cancer genomes. Estimated heteroplasmies are shown in the center of each circle.

(EPS)

S3 Fig. Read coverage on mouse chrM for the included samples.

(EPS)

S1 Table. Overview of sequenced DNA samples including basic statistics, performance of MitoSAlt compared to five published pipelines based on simulated sequencing data, and additional numerical data underlying figures.

(XLSX)

Acknowledgments

We thank Brith Leidvik for technical assistance. We would also like to acknowledge the Clinical Genomics Stockholm facility at Science for Life Laboratory and the Genomics Core Facility at the Sahlgrenska Academy for providing assistance in next generation sequencing. We acknowledge the contributions of the many clinical networks across ICGC and TCGA, enabling the analysis of whole genome sequencing data from human tumors in this study.

Author Contributions

Conceptualization: Claes M. Gustafsson, Maria Falkenberg, Erik Larsson.

Data curation: Swaraj Basu, Xie Xie.

Formal analysis: Swaraj Basu, Xie Xie.

Funding acquisition: Anders Oldfors, Claes M. Gustafsson, Maria Falkenberg, Erik Larsson.

Investigation: Jay P. Uhler, Carola Hedberg-Oldfors, Anders Oldfors, Maria Falkenberg.

Methodology: Swaraj Basu, Xie Xie, Maria Falkenberg, Erik Larsson.

Project administration: Maria Falkenberg, Erik Larsson.

Resources: Dusanka Milenkovic, Olivier R. Baris, Sammy Kimoloi, Stanka Matic, James B. Stewart, Rudolf J. Wiesner, Maria Falkenberg.

Software: Swaraj Basu, Xie Xie.

Supervision: Maria Falkenberg, Erik Larsson.

Validation: Nils-Göran Larsson.

Visualization: Swaraj Basu, Xie Xie, Jay P. Uhler.

Writing – original draft: Swaraj Basu, Xie Xie, Jay P. Uhler.

Writing – review & editing: Swaraj Basu, Xie Xie, Jay P. Uhler, Dusanka Milenkovic, James B. Stewart, Nils-Göran Larsson, Rudolf J. Wiesner, Anders Oldfors, Maria Falkenberg, Erik Larsson.

References

1. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. *Nature*. 1981; 290(5806):457–65. Epub 1981/04/09. <https://doi.org/10.1038/290457a0> PMID: 7219534.
2. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*. 1999; 23(2):147. Epub 1999/10/03. <https://doi.org/10.1038/13779> PMID: 10508508.
3. Van Goethem G, Dermaut B, Lofgren A, Martin JJ, Van Broeckhoven C. Mutation of POLG is associated with progressive external ophthalmoplegia characterized by mtDNA deletions. *Nat Genet*. 2001; 28(3):211–2. Epub 2001/06/30. <https://doi.org/10.1038/90034> PMID: 11431686.
4. Spelbrink JN, Li FY, Tiranti V, Nikali K, Yuan QP, Tariq M, et al. Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria. *Nat Genet*. 2001; 28(3):223–31. Epub 2001/06/30. <https://doi.org/10.1038/90058> PMID: 11431692.
5. Goffart S, Cooper HM, Tynismaa H, Wanrooij S, Suomalainen A, Spelbrink JN. Twinkle mutations associated with autosomal dominant progressive external ophthalmoplegia lead to impaired helicase function and in vivo mtDNA replication stalling. *Hum Mol Genet*. 2009; 18(2):328–40. Epub 2008/10/31. <https://doi.org/10.1093/hmg/ddn359> PMID: 18971204; PubMed Central PMCID: PMC2638771.
6. Holt IJ, Harding AE, Morgan-Hughes JA. Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature*. 1988; 331(6158):717–9. Epub 1988/02/25. <https://doi.org/10.1038/331717a0> PMID: 2830540.
7. Moraes CT, DiMauro S, Zeviani M, Lombes A, Shanske S, Miranda AF, et al. Mitochondrial DNA deletions in progressive external ophthalmoplegia and Kearns-Sayre syndrome. *N Engl J Med*. 1989; 320(20):1293–9. Epub 1989/05/18. <https://doi.org/10.1056/NEJM198905183202001> PMID: 2541333.
8. Poulton J, Deadman ME, Gardiner RM. Duplications of mitochondrial DNA in mitochondrial myopathy. *Lancet*. 1989; 1(8632):236–40. Epub 1989/02/04. [https://doi.org/10.1016/s0140-6736\(89\)91256-7](https://doi.org/10.1016/s0140-6736(89)91256-7) PMID: 2563411.
9. Hedberg-Oldfors C, Macao B, Basu S, Lindberg C, Peter B, Erdinc D, et al. Deep sequencing of mitochondrial DNA and characterization of a novel POLG mutation in a patient with arPEO. *Neurol Genet*. 2020; 6(1):e391. Epub 2020/02/12. <https://doi.org/10.1212/NXG.000000000000391> PMID: 32042919; PubMed Central PMCID: PMC6975171.
10. Durham SE, Krishnan KJ, Betts J, Birch-Machin MA. Mitochondrial DNA damage in non-melanoma skin cancer. *Br J Cancer*. 2003; 88(1):90–5. Epub 2003/01/31. <https://doi.org/10.1038/sj.bjc.6600773> PMID: 12556965; PubMed Central PMCID: PMC2376793.

11. Horton TM, Petros JA, Heddi A, Shoffner J, Kaufman AE, Graham SD Jr., et al. Novel mitochondrial DNA deletion found in a renal cell carcinoma. *Genes Chromosomes Cancer*. 1996; 15(2):95–101. Epub 1996/02/01. [https://doi.org/10.1002/\(SICI\)1098-2264\(199602\)15:2<95::AID-GCC3>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1098-2264(199602)15:2<95::AID-GCC3>3.0.CO;2-Z) PMID: 8834172.
12. Savre-Train I, Piatyszek MA, Shay JW. Transcription of deleted mitochondrial DNA in human colon adenocarcinoma cells. *Hum Mol Genet*. 1992; 1(3):203–4. Epub 1992/06/01. <https://doi.org/10.1093/hmg/1.3.203> PMID: 1339472.
13. Ballinger SW, Shoffner JM, Hedaya EV, Trounce I, Polak MA, Koontz DA, et al. Maternally transmitted diabetes and deafness associated with a 10.4 kb mitochondrial DNA deletion. *Nat Genet*. 1992; 1(1):11–5. Epub 1992/04/01. <https://doi.org/10.1038/ng0492-11> PMID: 1301992.
14. Rotig A, Bessis JL, Romero N, Cormier V, Saudubray JM, Narcy P, et al. Maternally inherited duplication of the mitochondrial genome in a syndrome of proximal tubulopathy, diabetes mellitus, and cerebellar ataxia. *Am J Hum Genet*. 1992; 50(2):364–70. Epub 1992/02/01. PMID: 1531167; PubMed Central PMCID: PMC1682469.
15. Horton TM, Graham BH, Corral-Debrinski M, Shoffner JM, Kaufman AE, Beal MF, et al. Marked increase in mitochondrial DNA deletion levels in the cerebral cortex of Huntington's disease patients. *Neurology*. 1995; 45(10):1879–83. Epub 1995/10/01. <https://doi.org/10.1212/wnl.45.10.1879> PMID: 7477986.
16. Ikebe S, Tanaka M, Ohno K, Sato W, Hattori K, Kondo T, et al. Increase of deleted mitochondrial DNA in the striatum in Parkinson's disease and senescence. *Biochem Biophys Res Commun*. 1990; 170(3):1044–8. Epub 1990/08/16. [https://doi.org/10.1016/0006-291x\(90\)90497-b](https://doi.org/10.1016/0006-291x(90)90497-b) PMID: 2390073.
17. Cortopassi GA, Arnheim N. Detection of a specific mitochondrial DNA deletion in tissues of older humans. *Nucleic Acids Res*. 1990; 18(23):6927–33. Epub 1990/12/11. <https://doi.org/10.1093/nar/18.23.6927> PMID: 2263455; PubMed Central PMCID: PMC332752.
18. Nicholls TJ, Zsurka G, Peeva V, Scholer S, Szczesny RJ, Cysewski D, et al. Linear mtDNA fragments and unusual mtDNA rearrangements associated with pathological deficiency of MGME1 exonuclease. *Hum Mol Genet*. 2014; 23(23):6147–62. Epub 2014/07/06. <https://doi.org/10.1093/hmg/ddu336> PMID: 24986917; PubMed Central PMCID: PMC4222359.
19. Kornblum C, Nicholls TJ, Haack TB, Scholer S, Peeva V, Danhauser K, et al. Loss-of-function mutations in MGME1 impair mtDNA replication and cause multisystemic mitochondrial disease. *Nat Genet*. 2013; 45(2):214–9. Epub 2013/01/15. <https://doi.org/10.1038/ng.2501> PMID: 23313956; PubMed Central PMCID: PMC3678843.
20. Williams SL, Huang J, Edwards YJ, Ulloa RH, Dillon LM, Prolla TA, et al. The mtDNA mutation spectrum of the progeroid Polg mutator mouse includes abundant control region multimers. *Cell Metab*. 2010; 12(6):675–82. Epub 2010/11/27. <https://doi.org/10.1016/j.cmet.2010.11.012> PMID: 21109200; PubMed Central PMCID: PMC3175596.
21. Moraes CT, Schon EA, DiMauro S, Miranda AF. Heteroplasmy of mitochondrial genomes in clonal cultures from patients with Kearns-Sayre syndrome. *Biochem Biophys Res Commun*. 1989; 160(2):765–71. Epub 1989/04/28. [https://doi.org/10.1016/0006-291x\(89\)92499-6](https://doi.org/10.1016/0006-291x(89)92499-6) PMID: 2541710.
22. Poulton J, Deadman ME, Turnbull DM, Lake B, Gardiner RM. Detection of mitochondrial DNA deletions in blood using the polymerase chain reaction: non-invasive diagnosis of mitochondrial myopathy. *Clin Genet*. 1991; 39(1):33–8. Epub 1991/01/01. <https://doi.org/10.1111/j.1399-0004.1991.tb02982.x> PMID: 1997213.
23. Moraes CT, Atencio DP, Oca-Cossio J, Diaz F. Techniques and pitfalls in the detection of pathogenic mitochondrial DNA mutations. *J Mol Diagn*. 2003; 5(4):197–208. Epub 2003/10/24. [https://doi.org/10.1016/S1525-1578\(10\)60474-6](https://doi.org/10.1016/S1525-1578(10)60474-6) PMID: 14573777; PubMed Central PMCID: PMC1907336.
24. Matic S, Jiang M, Nicholls TJ, Uhler JP, Dirksen-Schwanenland C, Polosa PL, et al. Mice lacking the mitochondrial exonuclease MGME1 accumulate mtDNA deletions without developing progeria. *Nat Commun*. 2018; 9(1):1202. Epub 2018/03/25. <https://doi.org/10.1038/s41467-018-03552-x> PMID: 29572490; PubMed Central PMCID: PMC5865154.
25. Poulton J, Morten KJ, Marchington D, Weber K, Brown GK, Rotig A, et al. Duplications of mitochondrial DNA in Kearns-Sayre syndrome. *Muscle Nerve Suppl*. 1995; 3:S154–8. Epub 1995/01/01. <https://doi.org/10.1002/mus.880181430> PMID: 7603518.
26. Kosugi S, Momozawa Y, Liu X, Terao C, Kubo M, Kamatani Y. Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. *Genome Biol*. 2019; 20(1):117. Epub 2019/06/05. <https://doi.org/10.1186/s13059-019-1720-5> PMID: 31159850; PubMed Central PMCID: PMC6547561.
27. Bosworth CM, Grandhi S, Gould MP, LaFramboise T. Detection and quantification of mitochondrial DNA deletions from next-generation sequence data. *BMC Bioinformatics*. 2017; 18(Suppl 12):407.

- Epub 2017/10/27. <https://doi.org/10.1186/s12859-017-1821-7> PMID: 29072135; PubMed Central PMCID: PMC5657046.
28. Hjelm BE, Rollins B, Morgan L, Sequeira A, Mamdani F, Pereira F, et al. Splice-Break: exploiting an RNA-seq splice junction algorithm to discover mitochondrial DNA deletion breakpoints and analyses of psychiatric disorders. *Nucleic Acids Res.* 2019; 47(10):e59. Epub 2019/03/15. <https://doi.org/10.1093/nar/gkz164> PMID: 30869147; PubMed Central PMCID: PMC6547454.
 29. Goudenege D, Bris C, Hoffmann V, Desquiret-Dumas V, Jardel C, Rucheton B, et al. eKLIPse: a sensitive tool for the detection and quantification of mitochondrial DNA deletions from next-generation sequencing data. *Genet Med.* 2019; 21(6):1407–16. Epub 2018/11/06. <https://doi.org/10.1038/s41436-018-0350-8> PMID: 30393377.
 30. Elder CS, Welsh CE, editors. MitoMut: An Efficient Approach to Detecting Mitochondrial DNA Deletions from Paired-end Next-generation Sequencing Data. *BCB '19: Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*; 2019 September 2019; Niagara Falls NY USA.
 31. Zambelli F, Vancampenhout K, Daneels D, Brown D, Mertens J, Van Dooren S, et al. Accurate and comprehensive analysis of single nucleotide variants and large deletions of the human mitochondrial genome in DNA and single cells. *Eur J Hum Genet.* 2017; 25(11):1229–36. Epub 2017/08/24. <https://doi.org/10.1038/ejhg.2017.129> PMID: 28832570; PubMed Central PMCID: PMC5643964.
 32. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019; 37(8):907–15. Epub 2019/08/04. <https://doi.org/10.1038/s41587-019-0201-4> PMID: 31375807.
 33. Calabrese FM, Simone D, Attimonelli M. Primates and mouse NumtS in the UCSC Genome Browser. *BMC Bioinformatics.* 2012; 13 Suppl 4:S15. Epub 2012/05/02. <https://doi.org/10.1186/1471-2105-13-S4-S15> PMID: 22536961; PubMed Central PMCID: PMC3314570.
 34. Kielbasa SM, Wan R, Sato K, Horton P, Frith MC. Adaptive seeds tame genomic sequence comparison. *Genome Res.* 2011; 21(3):487–93. Epub 2011/01/07. <https://doi.org/10.1101/gr.113985.110> PMID: 21209072; PubMed Central PMCID: PMC3044862.
 35. Schon EA, Rizzuto R, Moraes CT, Nakase H, Zeviani M, DiMauro S. A direct repeat is a hotspot for large-scale deletion of human mitochondrial DNA. *Science.* 1989; 244(4902):346–9. Epub 1989/04/21. <https://doi.org/10.1126/science.2711184> PMID: 2711184.
 36. Gourle H, Karlsson-Lindsjo O, Hayer J, Bongcam-Rudloff E. Simulating Illumina metagenomic data with InSilicoSeq. *Bioinformatics.* 2019; 35(3):521–2. Epub 2018/07/18. <https://doi.org/10.1093/bioinformatics/bty630> PMID: 30016412; PubMed Central PMCID: PMC6361232.
 37. Yuan Y, Ju YS, Kim Y, Li J, Wang Y, Yoon CJ, et al. Comprehensive molecular characterization of mitochondrial genomes in human cancers. *Nat Genet.* 2020; 52(3):342–52. Epub 2020/02/07. <https://doi.org/10.1038/s41588-019-0557-x> PMID: 32024997; PubMed Central PMCID: PMC7058535.
 38. Trifunovic A, Wredenberg A, Falkenberg M, Spelbrink JN, Rovio AT, Bruder CE, et al. Premature ageing in mice expressing defective mitochondrial DNA polymerase. *Nature.* 2004; 429(6990):417–23. Epub 2004/05/28. <https://doi.org/10.1038/nature02517> PMID: 15164064.
 39. Baris OR, Ederer S, Neuhaus JF, von Kleist-Retzow JC, Wunderlich CM, Pal M, et al. Mosaic Deficiency in Mitochondrial Oxidative Metabolism Promotes Cardiac Arrhythmia during Aging. *Cell Metab.* 2015; 21(5):667–77. Epub 2015/05/09. <https://doi.org/10.1016/j.cmet.2015.04.005> PMID: 25955204.
 40. Hudson G, Deschauer M, Busse K, Zierz S, Chinnery PF. Sensory ataxic neuropathy due to a novel C100rf2 mutation with probable germline mosaicism. *Neurology.* 2005; 64(2):371–3. Epub 2005/01/26. <https://doi.org/10.1212/01.WNL.0000149767.51152.83> PMID: 15668446.
 41. Persson O, Muthukumar Y, Basu S, Jenninger L, Uhler JP, Berglund AK, et al. Copy-choice recombination during mitochondrial L-strand synthesis causes DNA deletions. *Nat Commun.* 2019; 10(1):759. Epub 2019/02/17. <https://doi.org/10.1038/s41467-019-08673-5> PMID: 30770810; PubMed Central PMCID: PMC6377680.
 42. Phillips AF, Millet AR, Tigano M, Dubois SM, Crimmins H, Babin L, et al. Single-Molecule Analysis of mtDNA Replication Uncovers the Basis of the Common Deletion. *Mol Cell.* 2017; 65(3):527–38 e6. Epub 2017/01/24. <https://doi.org/10.1016/j.molcel.2016.12.014> PMID: 28111015.
 43. Macao B, Uhler JP, Siibak T, Zhu X, Shi Y, Sheng W, et al. The exonuclease activity of DNA polymerase gamma is required for ligation during mitochondrial DNA replication. *Nat Commun.* 2015; 6:7303. Epub 2015/06/23. <https://doi.org/10.1038/ncomms8303> PMID: 26095671; PubMed Central PMCID: PMC4557304.
 44. Torregrosa-Munumer R, Hargas A, Goffart S, Blei D, Zsurka G, Griffith J, et al. Replication fork rescue in mammalian mitochondria. *Sci Rep.* 2019; 9(1):8785. Epub 2019/06/21. <https://doi.org/10.1038/s41598-019-45244-6> PMID: 31217442; PubMed Central PMCID: PMC6584726.

45. Uhler JP, Thorn C, Nicholls TJ, Matic S, Milenkovic D, Gustafsson CM, et al. MGME1 processes flaps into ligatable nicks in concert with DNA polymerase gamma during mtDNA replication. *Nucleic Acids Res.* 2016; 44(12):5861–71. Epub 2016/05/26. <https://doi.org/10.1093/nar/gkw468> PMID: 27220468; PubMed Central PMCID: PMC4937333.
46. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015; 12(4):357–60. <https://doi.org/10.1038/nmeth.3317> PMID: 25751142.
47. Bothe GW, Haspel JA, Smith CL, Wiener HH, Burden SJ. Selective expression of Cre recombinase in skeletal muscle fibers. *Genesis.* 2000; 26(2):165–6. Epub 2000/03/21. PMID: 10686620.