

Quantitation and Characterization of Cell-Free Mitochondrial DNA in Plasma by Deep Sequencing

Grant T. Daly

University of South Alabama

Viktor M. Pastukh

University of South Alabama

Yong B. Tan

University of South Alabama

C. Michael Francis

University of South Alabama

C. Zack Aggen

University of South Alabama

S. Chris Groark

University of South Alabama

Carson Edwards

University of South Alabama

Mohammad Hamo

University of South Alabama

Jon D. Simmons

University of South Alabama

Matthew E. Kutcher

University of Mississippi

Emily M. Hartsell

University of South Alabama

Darrell L. Dinwiddie

University of New Mexico

Zachary M. Turpin

Florida State University

Hank W. Bass

Florida State University

Mark N. Gillespie (✉ mgillesp@southalabama.edu)

University of South Alabama

Raymond J. Langlely

University of South Alabama

Research Article

Keywords: Quantitation, Characterization, Cell-free Mitochondrial DNA, Plasma, Deep Sequencing, molecular pattern (DAMP), multiple human disorders

Posted Date: July 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-668341/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Abstract

The presence in plasma of mitochondrial DNA (mtDNA) fragments, a proinflammatory damage-associated molecular pattern (DAMP), is positively associated with outcomes of multiple human disorders. Because mtDNA comprises only a small percentage of the total DNA in plasma, qPCR is typically employed as an analytic strategy. However, this method provides little insight into sequence origins or other characteristics of circulating mtDNA fragments. Here we found that target bait-capture applied to plasma mtDNA derived from severely injured patients enriched recovery by ≈ 1400 -fold, thus affording depth sufficient for NextGen sequence analysis. Our method excluded nuclear mitochondrial insertions (NUMTs) using a stringent alignment and filtering strategy which calls and quantifies both NUMT in the reference genome and polymorphic NUMTs. After enrichment, we sequenced 2,000–60,000x mean coverage over the mtDNA genome. Normalization of mtDNA abundance to NUMT coverage reduced batch variability. Two massively-transfused trauma patients and two non-transfused patients displayed time-dependent increases in mtDNA DAMP coverage and decreases in the mean fragment length, which were more pronounced in massively transfused patients. Finally, our approach enabled detection of low-frequency heteroplasmic variants. Collectively, these findings suggest that our target bait-capture, deep sequencing and attendant analytic protocols could provide unprecedented characterization of cell-free plasma mtDNA and NUMTs.

Introduction

Several broad lines of evidence implicate mitochondrial DNA (mtDNA) damage-associated molecular patterns (DAMPs) in the pathogenesis of multiple forms of chronic and acute illness and injury. Observational studies in human patients show that plasma levels of mtDNA DAMPs are associated with clinical outcomes in a wide range of disorders including cardiometabolic diseases, sepsis, ischemia-reperfusion injury, trauma, and others^{1–8}. The pathophysiologic involvement of mtDNA DAMPs is further implied by studies in laboratory animal and cell culture models showing that administration of exogenous mtDNA fragments recapitulates many aspects of disease, in part by propagating cellular dysfunction through activation of nucleic acid receptor-mediated innate immune responses^{9–13}. Pharmacologic inhibition of mtDNA DAMP formation using mitochondrial-targeted DNA repair enzyme constructs and degradation of mtDNA DAMPs with exogenous DNase1 also are protective in numerous experimental systems mimicking key aspects of human disorders^{4,10,14–19}.

Detection and analysis of plasma mtDNA DAMPs are critical to a comprehensive assessment of their pathophysiologic importance and utility as biomarkers of disease. Conventional approaches employ qPCR to quantify the presence of a few ≈ 100 –200 bp mtDNA regions preselected as a proxy for the 16.5 kb expanse of the mitochondrial genome. While sensitive, qPCR can be confounded by false-positive contributions from homologous nuclear mitochondrial (NUMTs) insertions co-amplify along with the MtDNA DAMPs^{20,21}. In addition, because the choice of primers and amplicon sizes²² predict that an unknown number of plasma mtDNA DAMP sequences likely go undetected. Apart from the mtDNA

sequence selected for amplification, the qPCR strategy is incapable of delineating the mitochondrial genomic origin of other mitochondrial fragments present in plasma but outside of the selected sequence. Finally, qPCR lacks information about sequence variants in mtDNA DAMPs, limiting opportunities to utilize the cell-free DNA as a diagnostic marker, as it done in some studies of cancer²³⁻²⁵.

Whole genome sequencing (WGS) of mtDNA DAMPs has the potential to overcome the limitations noted above²², but in practice the amount of mtDNA relative to the total plasma DNA pool is generally too low to permit sufficient sequencing depth required to accurately determine changes in mtDNA DAMP abundance, sequences of origin, or variants. The goals of this proof-of-concept study were to validate the use of a novel RNA target-bait capture approach to enrich mtDNA from the total plasma DNA pool in concert with WGS and a bioinformatics protocol to differentiate mtDNA DAMPs from NUMTs in cell-free plasma. Our aim was to provide a more complete characterization of plasma mtDNA as defined by accurate quantitation, definition of fragment length distributions, identification of mtDNA DAMP sequences of origin, and detection of heteroplasmic (low-frequency) variants and homoplasmic (high-frequency consensus) variants²². We performed these protocols using mtDNA enriched from plasma derived from four severely injured human patients each sampled at two time points, upon hospital admission and 72h later. Two of the four patients required massive transfusions between procurement of the plasma samples.

Methods

Study subjects: This was an observational study approved by the University of South Alabama Institutional Review Board which involved analyzing plasma samples of patients enrolled in the Surgical/Trauma Intensive Care Unit (STICU) at the University of South Alabama Health University Hospital Trauma Center. Blood draws were taken on arrival and 72h hours later. Informed written consent was obtained from the patient or legal representative prior to the samples being utilized for experimental purposes. Any samples collected which were not subsequently consented to were destroyed. All experimental protocols were approved by University Hospital, and the methods performed were in accordance with the relevant guidelines and regulations. Inclusion criteria were adults (≥ 18 years) that were admitted to the STICU with an Injury Severity Score (ISS) ≥ 15 . Plasma (8 ml) was collected immediately upon admission to the STICU and 72h later. Two patients were given the massive transfusion protocol ($> 15U$ of blood products) while two patients with similar ISS did not require transfusion products. Plasma also was obtained from a human biorepository, but in this instance, mtDNA was below the limits of qPCR detection or the mtDNA present was full-length (not fragmented) (Fig. 1A).

mtDNA Library preparation from extracellular DNA in plasma and deep DNA sequencing: Fresh plasma was centrifuged at 700g (4°C) for 5m. Supernatant was carefully removed by pipette to avoid aspirating the cell pellet. DNA was isolated from 200 μ l of plasma using the QIAamp DSP DNA Blood Mini Kit (Qiagen, Inc, Hilden, Germany) according to manufacturer's instructions and stored in 75 μ l elution buffer (EB) -80°C.

For DNA-seq library preparation, The KAPA Hyper Prep Kit was used following manufacturer's instructions. Briefly, end repair and adenine tailing were combined in a single step, followed by adaptor ligation. DNA was purified 1.5X AMPure XP Beads (Beckman Coulter). Library amplification (11 cycles) was performed. DNA was purified using AMPure beads and stored 30 µl EB. To enrich mtDNA, we used MyBaits Human Global Panel mtDNA kit (Arbor Biosciences) following manufacturer's instructions. Briefly, 7µl of library was hybridized with the RNA mtDNA baits for 40h at 55°C. The RNA-DNA hybridized baits were captured with Dynabeads MyOne Streptavidin C1 (Arbor Biosciences). After purification, the library was enriched with KAPA HIFI kit (KAPA Inc) for 8 cycles of PCR. Quantification was determined using the KAPA Library Quantification kit. Analysis of the libraries utilized the Bio-Rad Bioanalyzer (Experion BioRad Inc.).

Next Generation sequencing was performed for two patients by HudsonAlpha on a *HiSeq2000*. 2x50 paired reads, n=4 libraries $1.03 \times 10^8 \pm 7.46 \times 10^6$ reads (mean \pm StDev) The other two samples were sequenced at Florida State University on a *HiSeq2500* at 2x50 paired reads, n=7 libraries with $3.14 \times 10^7 \pm 8.70 \times 10^6$ reads (mean \pm StDev). Four samples were run prior to enrichment to determine enrichment efficiency of the MyBaits protocol. However, one of the libraries failed due to low amount of isolated cfDNA; after enrichment, however, the sample contained adequate mtDNA for sequencing.

Code Availability: Main code and a description of how code was run are contained at <https://github.com/GrantDaly/TRALI-2021>. Code for variant calling is contained at <https://github.com/GrantDaly/PyVCF/>.

Alignments: Adapter trimming was performed with *Cutadapt v2.10*. Paired-end reads were aligned to the GRCh38 reference assembly with the *Sentieon v201911.01* implementation of *BWA-MEM*. Reads were sorted and converted to the "bam" format with "sentieon util sort"^{27,33}.

Nuclear Mitochondrial Insertions: The previously annotated reference NUMT set available had originally been generated with the outdated *hg18* assembly, so we elected to generate a NUMT call-set for the current GRCh38 assembly. *NCBI Nucleotide Blast 2.6.0* was run on the *Alabama Supercomputer* with the query being the human mitochondrial sequence, database GRCh38, word size 9, reward 1, penalty -1, gapopen 2, and gapextend 2. This resulted in 1521 NUMT covering 1.03×10^5 bp. The intervals called are reported in Supplementary Table S5.

Polymorphic NUMTs are more recent insertions not found in the reference assembly, but are instead structural population variants. A "vcf" file retrieved from Dayama et al.²⁰ provided coordinates for the 95% CI of the insertion sites, and *Bedtools v2.26.0* "slop" added 200bp upstream and downstream of the insertion sites. GRCh37 coordinates were then converted to GRCh38 coordinates with the UCSC *Liftover* utility, with 138/141 sites successfully lifting over. Analysis intervals can be found in Supplementary Table S6.

Validation of Alignment Pipeline by Simulation: To evaluate the effect of NUMTs on the accuracy of our alignment pipeline, a read simulation method was employed. First, *Bedtools v2.26.0* “slop” added 100 bp flanking regions upstream and downstream to the previously described non-polymorphic NUMTs. Then *Bedtools* “merge” merged any of these intervals within 100 bp, which resulted in 699 remaining intervals. Using fragment size parameters derived from one of our samples, we used *Art Illumina Q Version 2.5.8* to simulate reads with the error model of a *HiSeq 2500* instrument, read length 50bp, insert size 100 ± 25 (mean \pm StDev). Reads were simulated for the mitochondrial sequence (with the “N” spacer character removed from position 3107), the merged NUMT sequences, and the known polymorphic NUMT sequences²⁰. Coverage was parameterized as 1,000 for the mitochondrial and NUMT reads. Simulated reads were processed through our standard alignment and coverage pipeline. Accounting for the fact that the mitochondria and NUMT were simulated at a known depth of coverage per locus, percent error metrics for mtDNA coverage were defined for each base of the mitochondrial genome. False negative % error was defined as the number of mitochondrial reads expected to be lost covering a site that was actually covered by 100 reads (equation (1)), while false positive percent error of a site was defined as the number of NUMT reads incorrectly aligned to a site in mitochondrial genome when the set of NUMTs was covered at 100-fold (equation (2)). It is important to note that the copy number of the nuclear and mitochondrial genomes are not necessarily equivalent, so the overall magnitude of erroneously aligned reads is dependent on the true copy number of the mitochondrial and nuclear genomes.

$$\text{False Negative \% Error} = \frac{(1000 - \text{Observed Coverage})}{10} \quad (1)$$

$$\text{False Positive \% Error} = \frac{\text{Observed Coverage}}{10} \quad (2)$$

Mitochondrial Coverage: Read coverages were calculated for the mitochondrial genome, reference NUMTs, and known polymorphic NUMT insertions sites. Coverages were generated with the pysam v0.15.3 library (<https://github.com/pysam-developers/pysam>) “count_coverage” method with quality threshold ≥ 20 ³⁴. Mean mtDNA coverage \div mean NUMT coverage for each sample was used to normalize mitochondrial coverage.

Variant Calling: Bcftools 1.11-19 was used to generate a preliminary vcf file for the mitochondrial genome with maximum depth 1000, minimum base quality 20, minimum mapping quality 20. The NCBI Variation Services API was accessed to add dbSNP identifiers to known alternate alleles. VCFs were normalized so only one alternate appeared per row and these data were processed by a script employed with a customized fork of *PyVCF* (<https://github.com/GrantDaly/PyVCF/blob/master/scripts/call-mito.py>). For sites with ≥ 500 -fold coverage of the reference and alternate, allele frequencies $\geq 1\%$ and $\leq 99\%$ were called heteroplasmies and allele frequencies $\geq 99\%$ were called homoplasmies. For loci with ≥ 10 and ≤ 500 -fold coverage heteroplasmies were allowed if ≥ 5 reads of the minor allele were identified, and were called homoplasmies if the allele fraction was $\geq 99\%$.

Statistics: Spearman's rank analysis was performed using *JMP Genomics* 8.0 (SAS inc., Cary NC, USA). Most figures were made using *Graphpad Prism* 7.0 (GraphPad Software, San Diego, CA, USA). Of particular importance were Python v3.7.9 (<http://www.python.org/>) executed in a Jupyter Notebook, Pandas and Numpy data analysis libraries, and Matplotlib and Seaborn plotting libraries to plot the fragment size distributions. *InscAPE* 1.0.2 was used for one diagram.

Results

We enrolled four trauma patients that were admitted to the University Hospital with an injury severity score (ISS) > 15. All four patients were males between the ages of 19 and 42 with blood type O+ (Supplementary Tables S1 and S2). The injuries were due to motor vehicle crash (two patients), a pedestrian hit by automobile, and a gunshot wound to the chest and abdomen. Two patients had injuries that required massive transfusion protocol, while transfusion was not necessary in the other two. Plasma was collected at the time of enrollment, prior to the administration of any blood products, and 72 hours thereafter, resulting in 8 samples altogether.

Experimental and Analytical Workflow:

Our goal in this study is to establish an analytical workflow that fully characterizes cell free plasma mtDNA DAMPs and in order to ultimately enable the use of mtDNA DAMPs as biomarkers of disease outcome (Fig. 1). In order of implementation, the initial challenge was to select a method that overcame the limitations of analyzing low abundance, highly fragmented mtDNA relative to the substantially larger nDNA pool. For this we used a commercially available RNA target-bait capture enrichment kit previously applied to intact mitochondrial genomes in concert with a deep sequencing protocol. Next, after sequencing, we developed alignment and filtering strategies to more fully quantify DAMP abundance over the entire mitochondrial genome, characterize fragment lengths, and identify mtDNA heteroplasmies.

Target-Bait Enrichment:

Utilizing standard isolation protocols, we recovered a range of total DNA (221 ± 193 ng/ μ l; mean \pm StDev) in 200 μ l of plasma from either healthy subjects or trauma patients. However, unlike DNA isolated from plasma taken from normal human subjects, which was detected as fragments exceeding 7000 bp, most of the DNA in trauma patient plasma was highly fragmented at ~ 150 bp or less (Fig. 2a).

In four samples from two patients, we performed whole genome sequencing (WGS) in non-enriched samples and compared them to target captured samples to calculate the efficiency of the protocol (Fig. 2b). Library preparation was successful in three of the four non-enriched samples and provided $2.32 \times 10^7 \pm 2.73 \times 10^6$ (mean \pm StDev) sequence reads from WGS, and $6.25 \times 10^7 \pm 4.07 \times 10^7$ (mean \pm StDev) reads from the bait-capture enriched samples. The paired-end sequence reads were initially aligned to the Revised Cambridge Reference Sequence. Total mtDNA sequence coverage was high in the enriched samples (Fig. 2b).

The Obfuscating Impact of NUMTs on mtDNA Quantification:

Careful inspection of read alignments for enriched mtDNA samples revealed a conspicuously large number of heteroplasmies (Fig. 2b). The source of heteroplasmies was ambiguous, originating from either true mtDNA variants or NUMTs, which occur at more than 1500 locations in the human reference genome. Given that many NUMTs have a high degree of sequence homology to the mitochondrial genome, we postulated that they, like authentic mtDNA, were pulled down by the target-bait capture, but in fact represent contaminants. To test this hypothesis, we first performed a simulation analysis of mitochondrial sequences and NUMT sequences to determine where the reads would preferentially align to the human genome reference GRCh38²⁷. The outcome of this approach, shown in Fig. 3a-b, revealed that certain regions of the mitochondrial genome share sufficient homology that the alignment algorithm was incapable of uniquely assigning them to the correct location. Specifically, we found 14 regions within the mitochondrial genome which had perfect homology to the nuclear genome, with 12 of the 14 regions being longer than 100bp (Fig. 3b). Conspicuously, the majority of these were near the same region of Chromosome 1, suggesting they originated from the same insertion site. Finally, the mitochondrial genomic region spanning approximately 9.5-16.5kb has much less NUMT contamination than the rest of the mitochondrial genome (Fig. 3a).

Rigorous filtering and exclusion of NUMTs carries the risk of underestimating the total abundance of authentic mtDNA DAMPs. Conversely, inadequate filtering may overestimate abundance as well as lead to inaccurate identification of heteroplasmic variants. To illustrate the extremes of these filtering strategies, we plotted the percent error for mitochondrial and NUMT simulated reads (Fig. 3c) at Phred-scaled mapping quality scores of 0 and 20, with 0 corresponding to low confidence and 20 corresponding to $\sim 1/100$ chance of the read being erroneously aligned²⁸. These data indicate that rigorous filtering may increase false negative rate of alignment while reducing false positive alignments that can lead to false positive heteroplasmy calls.

Enrichment:

Based on the simulation data we realigned the sequence data to the total GRCh38 human reference genome²⁷. Consistent with our expectation, this enhanced alignment accuracy while diminishing read abundance (Fig. 4a). It is noteworthy that the total amount of NUMT sequences aligned to the mitochondrial genome exceeded the alignment of authentic mtDNA sequences. Therefore, we concluded that failure to correct for NUMT abundance can lead to overestimation, highly variable mtDNA coverage, and false assignments of heteroplasmies that actually originate from nuclear sequences of mitochondrial DNA insertions.

To assess the impact of sequence-capture enrichment, we performed Spearman's rank correlation of the mean abundance of 100 bp bins across the mtDNA genome for one patient at both time points between all enriched and non-enriched preparations. Correlations between paired enriched and non-enriched

samples were poor, whereas correlation of the enriched samples at the two time points were high (Fig. 4b). With confidence that the majority of reads were properly aligning either to the nDNA or the mtDNA, we then calculated mean coverage for total mtDNA, NUMTs and normalized coverage (mean mtDNA coverage ÷ mean NUMT coverage) (Supplemental Table S3). Normalized mtDNA coverage was additionally called over each base of the mitochondrial genome (Supplemental Table S4). Coverage statistics for reference NUMTs are included in Supplementary Table S5. The enrichment and alignment strategy led to a 1488 ± 1045 (mean \pm StDev) fold enrichment as compared to WGS (Fig. 4c).

Unique Challenges Imposed by Polymorphic NUMTs:

Our analyses were based on NUMTs identified in the human reference. However, a heterogeneous population of NUMTs not annotated in the human reference has emerged from the 1000 Genomes Project. Dayama and coworkers identified many of these polymorphic NUMTs by identifying spurious alignments of paired-end sequence reads in which one end uniquely aligned to the nuclear genome but the other end uniquely aligned to the mitochondrial genome²⁰. The prevalence of these insertions is highly variable within populations. A major problem with these NUMTs is the extremely high sequence homology to the human mtDNA reference (> 99% homology in many cases), thus making alignment accuracy exceedingly difficult to determine where these sequences originate. Here we designate this subset as “polymorphic NUMTs” and consider them population-level nuclear variants, distinct from genuine non-nuclear mitochondrial heteroplasmies. Because of these issues, we selected the human mitochondrial MyBaits designed for the “modern global human” global diversity panel of 197 sequences. These allow for maximal recovery of all human sequences – ideal for patient study samples and applicable world-wide.

Polymorphic NUMTs can be captured by RNA target-baits, as shown in Fig. 5a. Nuclear DNA sequences flanking these polymorphic NUMTs are also recovered. We hypothesized that polymorphic NUMTs could be delineated by sequence coverage of these flanking nuclear DNA insertion points. As depicted by the IGV snapshot shown in Fig. 5b, it is evident that sequence coverage was high within a NUMT found in Chromosome 4; in contrast it would not be detected in another patient indicating that this polymorphic NUMT is patient-specific.

Similar to the report by Dayama *et al.*, careful inspection showed that some read pairs aligned to both the nuclear and mitochondrial genomes (Fig. 5c). Consistent with our expectation, basic local alignment search tool (BLAST) analysis of regions flanking polymorphic NUMTs were not homologous to the mtDNA genome (Fig. 5D). Against this background, we then calculated mean coverage at 138 previously identified insertions to determine patient-specific associations (Supplementary Table S6). Our criteria for identifying these polymorphic NUMTs required ≥ 10 fold mean coverage in at least one of the samples at both time points, confirmation of the region by direct visualization on an IGV snapshot, and confirmation of paired reads mapping to both the mtDNA as well as the nDNA region of interest. Using this method we found 13 potential polymorphic NUMT sites, of which 3 NUMT genotypes were confirmed according to these criteria.

Accurate Delineation of mtDNA Coverage, Fragment Size, and Heteroplasmy

To illustrate the utility of the above-described workflow, we quantified mtDNA and NUMT coverage, fragment size and heteroplasmic signature for each of the four over time. In contrast to qPCR quantification of mtDNA DAMPs associated with a pre-specified mitochondrial genomic sequence, our method allows for quantitation over the entire mitochondrial genome. Additionally, this filtering strategy discriminates between authentic mtDNA fragments and contaminants and reference NUMTs, thus pointing to the prospect that mtDNA abundance could be normalized to the circulating nDNA pool. Towards these ends, mean NUMT coverage was relatively stable over time in this small sample of transfused and non-transfused patients (Fig. 6a). However, highly dynamic changes in mtDNA DAMPs mean coverage are evident in both transfused and non-transfused patients (Fig. 6b), with one patient displaying a conspicuously large increase after transfusion. In light of the stable NUMT abundance, mean mtDNA coverage was normalized to this nuclear DNA coverage shown in Fig. 6c. Inter-patient variability was notably reduced.

Fig. 7 displays histograms of the percentage of fragments observed of a given size. Notably, the massively transfused patients were enriched for small NUMT and mtDNA fragments at 72h. To analyze fragment size numerically we plotted the mean fragment size over time (Fig. 8). Mean fragment length of NUMTs and mtDNA DAMPs both increased moderately in non-transfused patients over time, while both NUMT and mtDNA fragment length decreased over time in massively transfused patients (Fig. 7a-d; Fig. 8; Supplementary Table S7). Finally, we calculated mtDNA heteroplasmy densities in each patient (Fig. 9; Supplementary Table S8 and S9). The criteria for designating a heteroplasmic variant is based on at least 500x read coverage and ≥ 1 percent of the reads calling the minor allele. One non-transfused patient maintained a steady level of heteroplasmy while the other increased heteroplasmy, and one massively transfused patient maintained similar levels of heteroplasmy levels while the other patient showed decreased heteroplasmy. This is an interesting observation, as we hypothesized that heteroplasmies would increase in transfused patients due to the exogenous addition of mtDNA from the transfusion products.

Discussion

In a range of human disorders, pro-inflammatory fragments of mtDNA DAMPs accumulate in the circulation where they propagate the initiating insult to distant sites¹⁻⁷ through proinflammatory immune responses²⁹⁻³². Central to obtaining a complete understanding of the pathophysiologic role of mtDNA DAMPs is the need to define the distribution of individual fragments relative to their origin within the mitochondrial genome and to measure the total fragment abundance. qPCR is the analytical method of choice, but although quantitatively sensitive, it is limited by primer and amplicon design choices which typically only sample small regions. Thus, the majority of the mtDNA DAMPs in plasma may escape detection by qPCR. In one of the few reports of its kind, Zhang and colleagues demonstrated through

whole genome sequencing that cell-free mtDNA in plasma is highly fragmented (average length ~ 60–200 bp) which makes accurate quantification by PCR very difficult²². Similar to Zhang et al., we were unable to detect highly fragmented mtDNA in plasma from normal subjects, so we focused our analysis exclusively on severely injured patients.

Against this background, the goal of this study was to develop an experimental protocol and analytical workflow that provide a more comprehensive assessment of plasma cell-free mitochondrial DNA fragments. Here we present new RNA target bait-capture, deep sequencing and bioinformatics protocols for more accurate quantitation and differentiation between mtDNA DAMPs and NUMTs in these samples. The high depth of coverage enabled by target-bait enrichment allows for interrogation of both NUMTs and authentic mtDNA at an unprecedented level of precision. Furthermore, the cost effectiveness of our enrichment protocol compared to WGS allows the procedure to be scaled with up to 50 patient samples per flow cell lane.

In this study, we have demonstrated a strategy to disentangle the overlapping signals from mtDNA and NUMTs²⁰. There are two varieties of NUMTs, those that have been included in the reference genome, and a second variety that is polymorphic, meaning they are found sporadically in the population²⁰. About 1500 reference NUMTs have been enumerated, spanning ~ 100Kbp of the nuclear genome. Polymorphic NUMTs are far more challenging to detect than the enumerated NUMT population as they often share greater than 99% homology, are not contained in the reference assembly, and can only be discovered by matching paired reads that align to the two genomes. The method described herein accounts for both polymorphic and reference NUMTs using a computational strategy that takes advantage of sequence variants and sequence coverage regions flanking NUMTs. Using this, we found that polymorphic NUMTs were a small and stable component of the presumed total mtDNA DAMP pool. Because of the apparent low frequency of polymorphic NUMTs within the population, the inclusion of this type of NUMT is unlikely to lead to overestimation of the authentic mtDNA DAMP abundance. However, our analysis found that failure to judiciously exclude annotated NUMTs was associated with spuriously elevated mtDNA DAMP abundances. We recommend, therefore, that our method be used to survey for the presence of polymorphic NUMTs in order to more accurately quantify de facto mtDNA DAMPs in plasma, removing false positives and yielding more robust heteroplasmy classification.

Our deep sequencing protocol allows determination of two additional characteristics of plasma mtDNA DAMPs: fragment length distribution, and heteroplasmic signatures. Even with a limited number of samples, we were able to detect changes in mean fragment size as a function of patient and time after hospital admission. With a larger number of patients, DNA fragment size distributions may be further characterized for a more detailed quantitative comparisons between and within patients. Notably, we also detected multiple heteroplasmic variants that changed over time, and some of the detected heteroplasmies are predicted to be coding variants (Supplemental Table S9). Other work demonstrates that certain heteroplasmies derive from specific tissues or cells. Accordingly, the current observations point to the prospect that heteroplasmic signatures in cell-free plasma mtDNA represent promising potential for the development of new markers of organ dysfunction.

The biological implications of this study are limited by the small patient number. Nevertheless, conspicuous differences between transfused and non-transfused patients were noted, including significant variability in the mean mtDNA abundance as a function of the transfusion protocol and time after hospital admission. We also noted that NUMT coverage was relatively consistent between patients irrespective of injury severity, transfusion, or time. This permitted the novel use of NUMT coverage as a factor to normalize mtDNA DAMP abundance. Normalized DAMP abundance was dynamic with time dependent changes noted in all patients.

In summary, here we have presented an experimental and analytical workflow that permits detailed evaluation of plasma cell-free mtDNA DAMP abundance and characteristics with increased accuracy and sensitivity over contemporary methods. Future studies can take advantage of this strategy to determine how mtDNA changes occur over time in patients with trauma, sepsis, and multiple other diseases that have been associated with mtDNA DAMPs.

Declarations

DATA AVAILABILITY

Raw sequencing data are available from the Sequence Read Archive accession PRJNA727684. Prior to public release, the data may be accessed through the following reviewer link:

<https://dataview.ncbi.nlm.nih.gov/object/PRJNA727684?reviewer=3qv5984f9oaeke91t1svrfaiju>

ACKNOWLEDGEMENTS

This work was supported by American Heart Association Grant # 830166 / Grant T. Daly / 2021. This work was supported by NIH grants K25 HL136869, 1K08GM109113-01, NR019338-01A1, 5KL2TR003097, R01HL113614, R01GM127823, and UL1TR003096.

This work was made possible in part by a grant of high performance computing resources and technical support from the Alabama Supercomputer Authority.

AUTHOR CONTRIBUTIONS

RJL, MNG, DLD developed the concept, secured direct funding, and designed experiments. VMP, YBT, ZMP, HWB processed samples and created libraries. Bioinformatics pipelines were designed and implemented by GTD, CMF, and MH. Downstream data analysis and figure design was performed by GTD, CZA, SCG, CE, and EMH. Clinical data was provided by YBT, JDS, and MEK. The manuscript was written by MNG, RJL, GTD, VMP, and EMH. All authors reviewed the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

References

1. Kung, C. T. *et al.* Plasma nuclear and mitochondrial DNA levels as predictors of outcome in severe sepsis patients in the emergency room. *Journal of translational medicine***10**, 130, doi:10.1186/1479-5876-10-130 (2012).
2. Nakahira, K. *et al.* Circulating mitochondrial DNA in patients in the ICU as a marker of mortality: derivation and validation. *PLoS Med***10**, e1001577; discussion e1001577, doi:10.1371/journal.pmed.1001577 (2013).
3. Harrington, J. S., Choi, A. M. K. & Nakahira, K. Mitochondrial DNA in Sepsis. *Curr Opin Crit Care***23**, 284-290, doi:10.1097/MCC.0000000000000427 (2017).
4. Simmons, J. D. *et al.* Mitochondrial DNA damage associated molecular patterns in ventilator-associated pneumonia: Prevention and reversal by intratracheal DNase I. *J Trauma Acute Care Surg***82**, 120-125, doi:10.1097/TA.0000000000001269 (2017).
5. Nakayama, H. & Otsu, K. Mitochondrial DNA as an inflammatory mediator in cardiovascular diseases. *Biochem J***475**, 839-852, doi:10.1042/BCJ20170714 (2018).
6. Harrington, J. S. *et al.* Circulating Mitochondrial DNA as Predictor of Mortality in Critically Ill Patients: A Systematic Review of Clinical Studies. *Chest***156**, 1120-1136, doi:10.1016/j.chest.2019.07.014 (2019).
7. Simmons, J. D. *et al.* Potential contribution of mitochondrial DNA damage associated molecular patterns in transfusion products to the development of acute respiratory distress syndrome after multiple transfusions. *J Trauma Acute Care Surg***82**, 1023-1029, doi:10.1097/TA.0000000000001421 (2017).
8. Tsuji, N. *et al.* Role of Mitochondrial DNA in Septic AKI via Toll-Like Receptor 9. *Journal of the American Society of Nephrology : JASN***27**, 2009-2020, doi:10.1681/ASN.2015040376 (2016).
9. Zhang, J. Z., Liu, Z., Liu, J., Ren, J. X. & Sun, T. S. Mitochondrial DNA induces inflammation and increases TLR9/NF-kappaB expression in lung tissue. *Int J Mol Med***33**, 817-824, doi:10.3892/ijmm.2014.1650 (2014).
10. Kuck, J. L. *et al.* Mitochondrial DNA damage-associated molecular patterns mediate a feed-forward cycle of bacteria-induced vascular injury in perfused rat lungs. *Am J Physiol Lung Cell Mol Physiol***308**, L1078-1085, doi:10.1152/ajplung.00015.2015 (2015).
11. Zhang, Q. *et al.* Circulating mitochondrial DAMPs cause inflammatory responses to injury. *Nature***464**, 104-107, doi:10.1038/nature08780 (2010).
12. Bliksoen, M. *et al.* Extracellular mtDNA activates NF-kappaB via toll-like receptor 9 and induces cell death in cardiomyocytes. *Basic Res Cardio***111**, 42, doi:10.1007/s00395-016-0553-6 (2016).

13. Alekseeva, A. Y., Kameneva, L. V., Kostyuk, S. V. & Veiko, N. N. Multiple Ways of cfDNA Reception and Following ROS Production in Endothelial Cells. *Adv Exp Med Bio***924**, 127-131, doi:10.1007/978-3-319-42044-8_25 (2016).
14. Tan, Y. B. *et al.* Enhanced Mitochondrial DNA Repair Resuscitates Transplantable Lungs Donated After Circulatory Death. *J Surg Res***245**, 273-280, doi:10.1016/j.jss.2019.07.057 (2019).
15. Yang, X. M. *et al.* Mitochondrially targeted Endonuclease III has a powerful anti-infarct effect in an in vivo rat model of myocardial ischemia/reperfusion. *Basic Res Cardio***110**, 459, doi:10.1007/s00395-014-0459-0 (2015).
16. Li, G. *et al.* DNaseI protects against Paraquat-induced acute lung injury and pulmonary fibrosis mediated by mitochondrial DNA. *Biomed Res Int***2015**, 386952, doi:10.1155/2015/386952 (2015).
17. Schuliga, M. *et al.* Self DNA perpetuates IPF lung fibroblast senescence in a cGAS-dependent manner. *Clin Sci (Lond)***134**, 889-905, doi:10.1042/CS20191160 (2020).
18. Mallavia, B. *et al.* Mitochondrial DNA Stimulates TLR9-Dependent Neutrophil Extracellular Trap Formation in Primary Graft Dysfunction. *Am J Respir Cell Mol Biol***62**, 364-372, doi:10.1165/rcmb.2019-01400C (2020).
19. Szczesny, B. *et al.* Mitochondrial DNA damage and subsequent activation of Z-DNA binding protein 1 links oxidative stress to inflammation in epithelial cells. *Sci Rep***8**, 914, doi:10.1038/s41598-018-19216-1 (2018).
20. Dayama, G., Emery, S. B., Kidd, J. M. & Mills, R. E. The genomic landscape of polymorphic human nuclear mitochondrial insertions. *Nucleic Acids Res***42**, 12640-12649, doi:10.1093/nar/gku1038 (2014).
21. Triant, D. A. & DeWoody, J. A. Extensive mitochondrial DNA transfer in a rapidly evolving rodent has been mediated by independent insertion events and by duplications. *Gene***401**, 61-70, doi:10.1016/j.gene.2007.07.003 (2007).
22. Zhang, R., Nakahira, K., Guo, X., Choi, A. M. & Gu, Z. Very Short Mitochondrial DNA Fragments and Heteroplasmy in Human Plasma. *Sci Rep***6**, 36097, doi:10.1038/srep36097 (2016).
23. Zhou, Q. *et al.* Circulating Cell Free DNA as the Diagnostic Marker for Ovarian Cancer: A Systematic Review and Meta-Analysis. *PLoS One***11**, e0155495, doi:10.1371/journal.pone.0155495 (2016).
24. Petit, J. *et al.* Cell-Free DNA as a Diagnostic Blood-Based Biomarker for Colorectal Cancer: A Systematic Review. *J Surg Res***236**, 184-197, doi:10.1016/j.jss.2018.11.029 (2019).
25. Chen, C., Chen, C. & Sadeghi, M. Evaluation of cell-free DNA accuracy as diagnostic biomarker for prostate cancer: A systematic review and meta-analysis. *Biotechnol Appl Biochem*, doi:10.1002/bab.2149 (2021).

26. Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics***23**, 147-147, doi:10.1038/13779 (1999).
27. Schneider, V. A. *et al.* Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res***27**, 849-864, doi:10.1101/gr.213611.116 (2017).
28. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res***8**, 175-185, doi:10.1101/gr.8.3.175 (1998).
29. West, A. P. & Shadel, G. S. Mitochondrial DNA in innate immune responses and inflammatory pathology. *Nat Rev Immunol***17**, 363-375, doi:10.1038/nri.2017.21 (2017).
30. Shimada, K. *et al.* Oxidized mitochondrial DNA activates the NLRP3 inflammasome during apoptosis. *Immunity***36**, 401-414, doi:10.1016/j.immuni.2012.01.009 (2012).
31. Grazioli, S. & Pugin, J. Mitochondrial Damage-Associated Molecular Patterns: From Inflammatory Signaling to Human Diseases. *Front Immunol***9**, 832, doi:10.3389/fimmu.2018.00832 (2018).
32. Julian, M. W., Shao, G., Vangundy, Z. C., Papenfuss, T. L. & Crouser, E. D. Mitochondrial transcription factor A, an endogenous danger signal, promotes TNFalpha release via RAGE- and TLR9-responsive plasmacytoid dendritic cells. *PLoS One***8**, e72354, doi:10.1371/journal.pone.0072354 (2013).
33. Freed, D., Aldana, R., Weber, J. A. & Edwards, J. S. The Sentieon Genomics Tools - A fast and accurate solution to variant calling from next-generation sequence data. *bioRxiv*, 115717, doi:10.1101/115717 (2017).
34. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics***25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

Figures

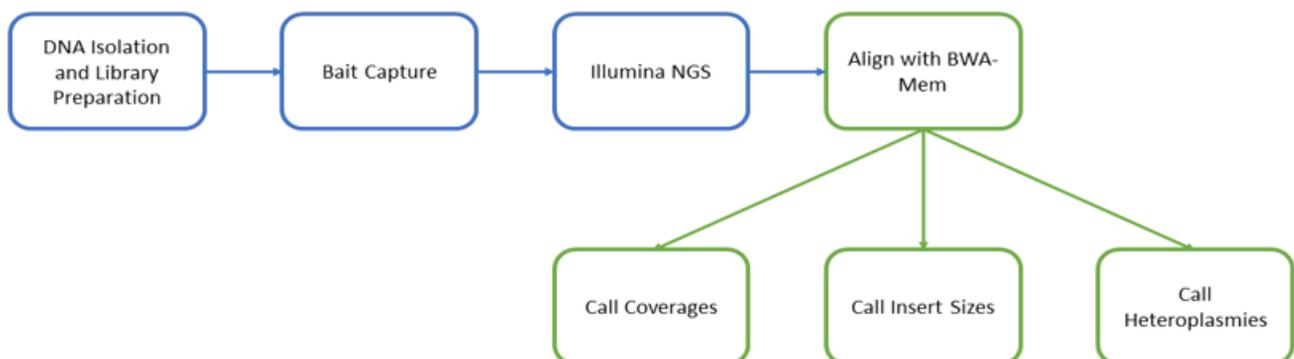


Figure 1

Experimental design and computational workflow for mtDNA DAMP enrichment, sequencing and analysis. Experimental design and computational workflow are depicted as blue boxes and green boxes, respectively. Figure prepared in MS Powerpoint.

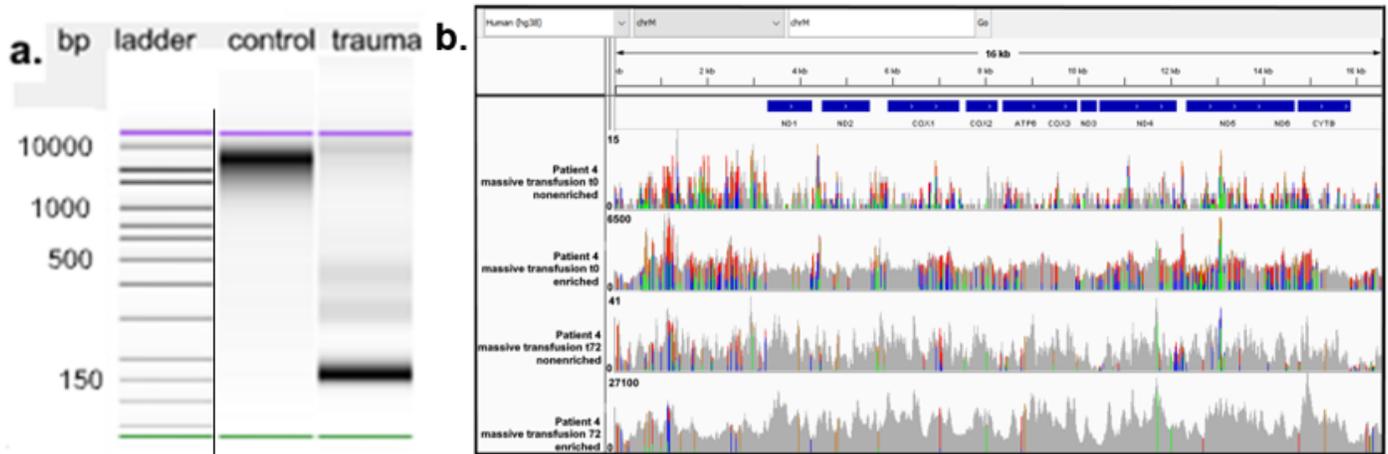
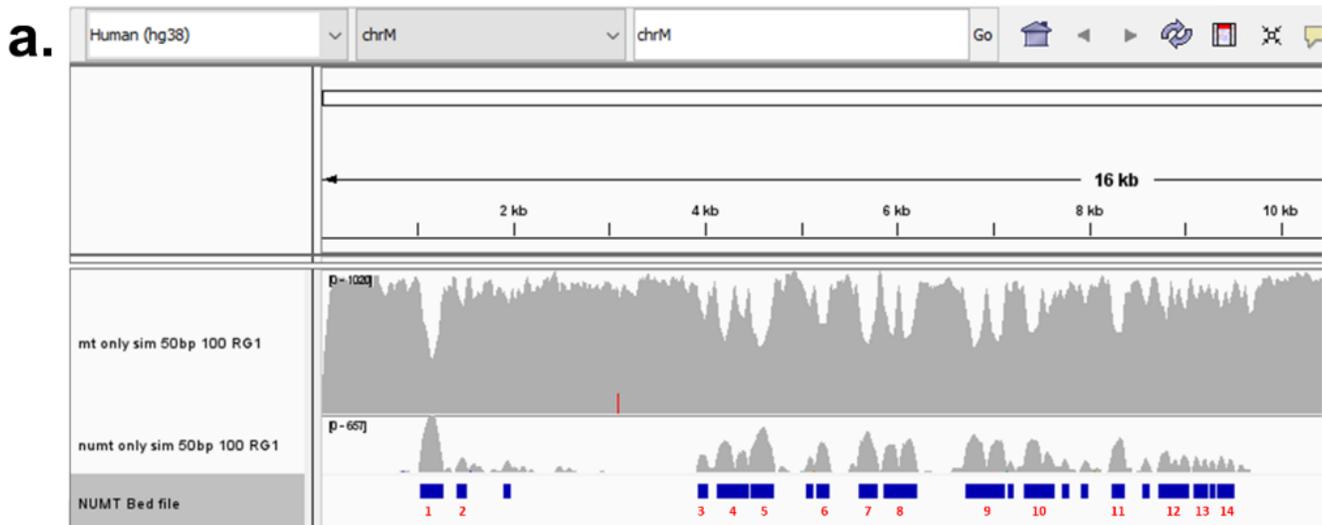


Figure 2

Differences between normal human subjects and trauma patients before and after enrichment. a) DNA fragment size was determined in total cell free DNA isolated from normal human and trauma patient plasma by bioanalyzer. Note that plasma from normal subjects contain relatively long fragment sizes, while plasma from trauma patients contained considerably smaller DNA fragments. Image is a Bioanalyzer output where the vertical black line shows where the output was cropped for clarity. The full output is presented as Supplemental Image 1. b) IGV screenshot depiction of plasma DNA fragments aligned to the mitochondrial genome before and after enrichment. The alignment strategy utilized the revised Cambridge reference sequence²⁶. We noted numerous heteroplasmic variants as depicted by the red, blue, green and orange bars.



b.

NUMT Peaks with 100% homology to mtDNA genome		
Peak #	Length	Location
1	209	chr13:109424125-109424333
2	104	chr11:10509340-10509443
3	106	chr1:629097-629202
4	191	chr1:629291-629481
5	250	chr1:629644-629893
6	137	chr1:630332-630468
7	202	chr1:630775-630976
8	163	chr1:631030-631192
9	222	chr1:631884-632105
10	189	chr1:632503-632691
11	146	chr1:633403-633548
12	128	chr1:633901-634028
13	75	chr1:634262-634336
14	91	chr1:634512-634602

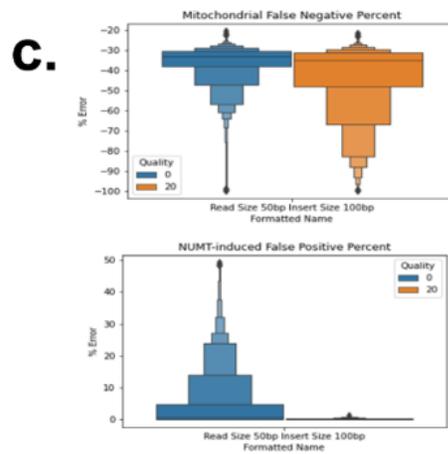


Figure 3

Simulated mtDNA and NUMT sequence reads identify numerous regions that align non-uniquely to the mitochondrial genome. a) IGV screenshot depicts simulated mtDNA and NUMT reads that align to the mitochondrial genome. There were 14 sites that the aligner could not differentiate between the mitochondrial genome and the nuclear genome. Figure prepared with MS Powerpoint b) Segments of highly homologous overlapping regions of NUMT and mitochondrial genome sequences. c) The false negative and false positive percentages as described in methods were displayed as “boxen” plots to display the quantiles of mitochondrial indices with a specific error. Data were colored to reflect the minimum mapping quality as quantified in Phred scale28 Figure from Seaborn plotting library.

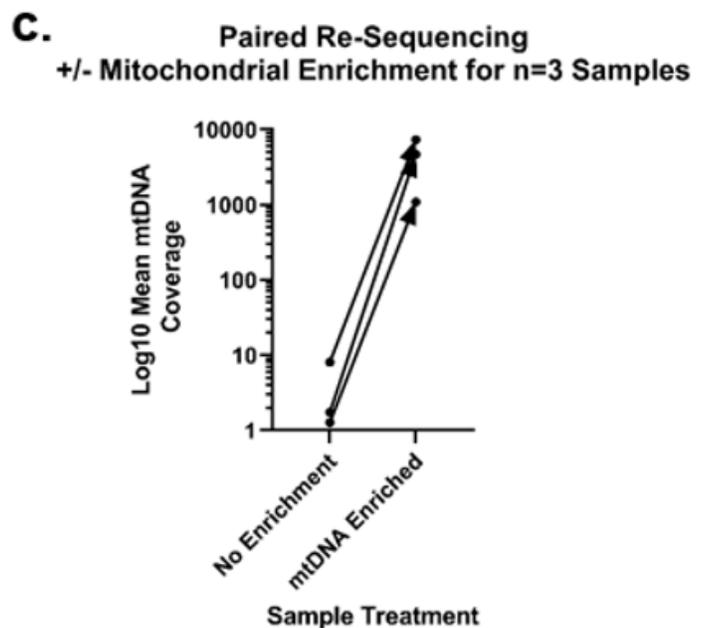
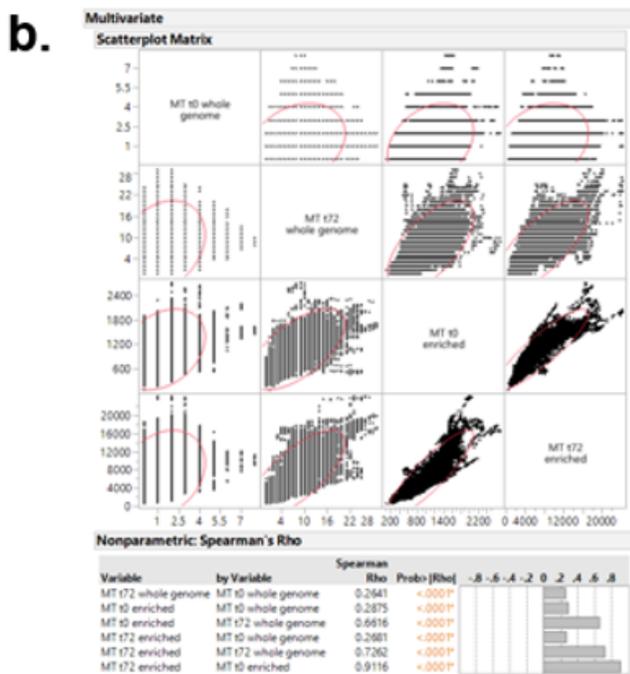
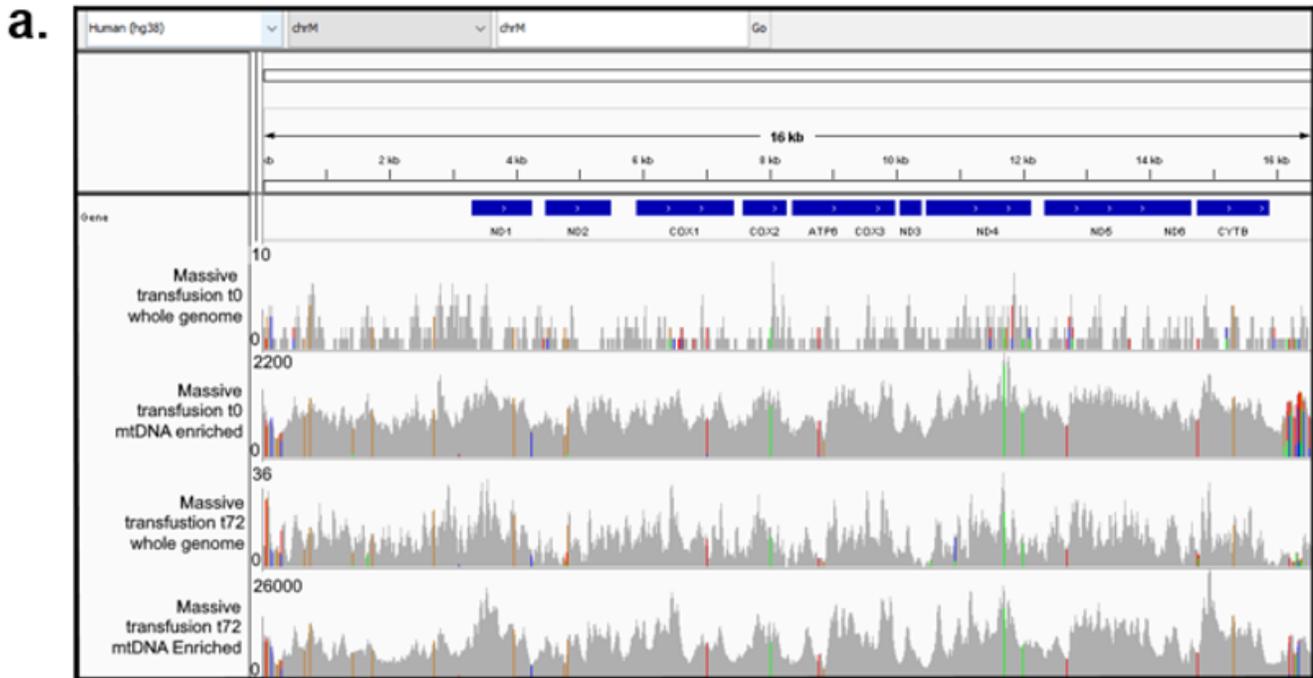


Figure 4

Target-bait capture enrichment in combination with judicious exclusion of NUMTs leads to more precise quantitation of mtDNA DAMPs as compared to WGS. a) IGV screenshot histograms depict enrichment of mtDNA fragments after application of bioinformatics protocol. Note decreased heteroplasmy attributable to exclusion of NUMTs. b) Non-parametric Spearman's rank analysis of fragment base pair coverage along the mitochondrial genome in a single massively transfused patient before and after transfusion in enriched and non-enriched samples. Higher correlations are indicated by more tightly clustered data.

Figure prepared with JMP Genomics. c) Enrichment efficiency for whole genome sequenced and samples sequenced after enrichment in three individual patients. Figure from Graphpad Prism.

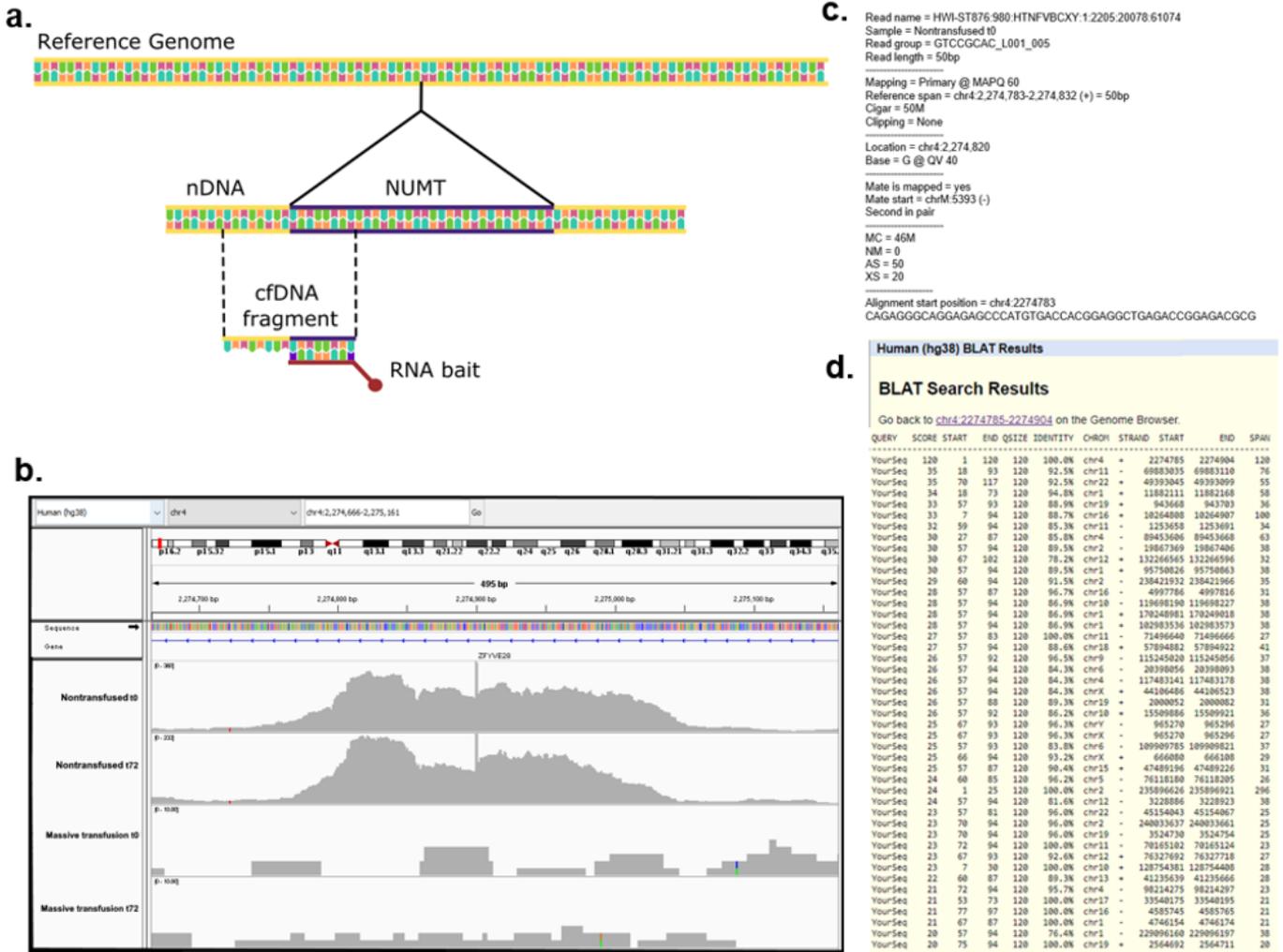


Figure 5

Target-bait capture enrichment protocol identifies insertion points of polymorphic NUMTs in the nuclear genome. a) Schematic depiction of how target-bait capture also leads to sequencing of flanking regions of polymorphic NUMTs. Figure prepared by EMH in Inkscape b) IGV histograms depicting a specific polymorphic NUMT in a massively transfused patient whereas the second patient lacks this insertion. c) Read meta-data taken from IGV demonstrates part of this read aligns to the nuclear genome flanking an annotated polymorphic NUMT, and the rest of the read maps to the mitochondria, consistent with an insertion event. d) BLAT analysis of the flanking region of the polymorphic NUMT depicts that the nuclear flanking region has no sequence homology to the mitochondrial genome.

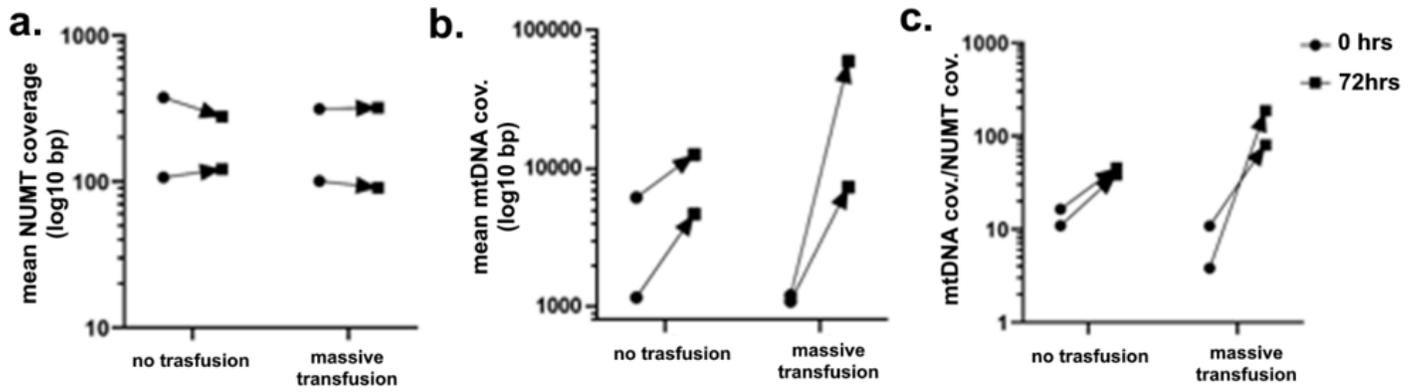


Figure 6

Sequence coverage of NUMTs and mtDNA as a function of patients over time. Mean coverage of NUMTs (a) and mtDNA DAMP (b) coverage depicts nDNA coverage remains relatively stable regardless of patient injury or transfusion status, while mtDNA DAMPs increase. c) Normalized coverage (mtDNA mean coverage ÷ NUMT mean coverage) in non-transfused and massively transfused patients. All charts from Graphpad Prism.

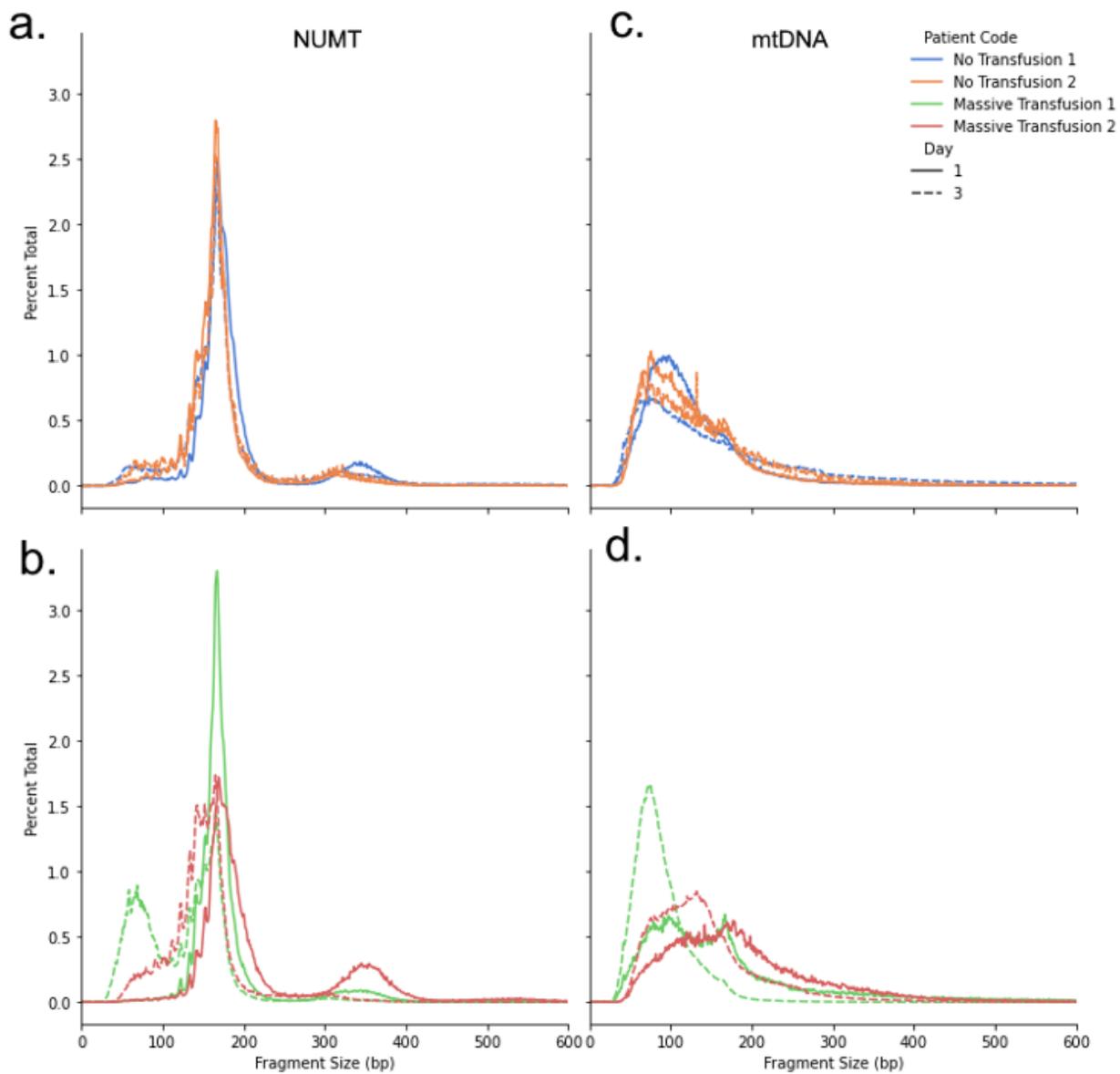


Figure 7

Fragment length characteristic histograms of mtDNA and NUMTs. Fragment length distributions of NUMTs (a,b) and mtDNA (c,d) demonstrate degradation of DNA fragments over time in the massively transfused patients (b,d). Figure prepared with the Seaborn plotting library.

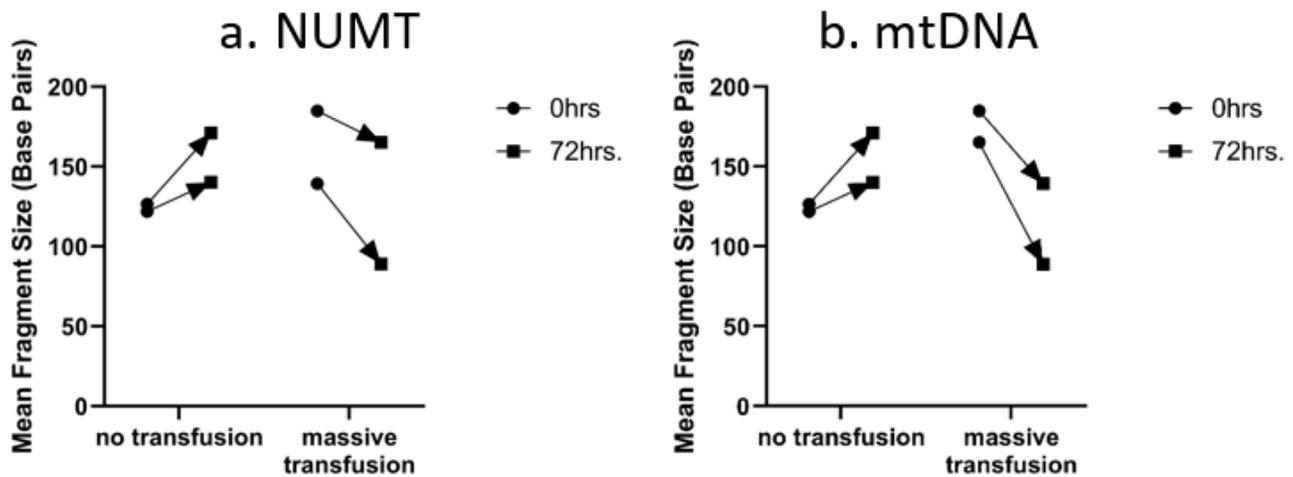


Figure 8

Mean fragment size with time and treatment. Mean fragment sizes plotted over time. a) NUMT fragment sizes increased over time for non-transfused patients while transfused patients had decreased NUMT fragment sizes. b) Mitochondrial mean fragment sizes similarly increased over time for the transfused patients and decreased over time in massively transfused patients. Both charts from Graphpad Prism.

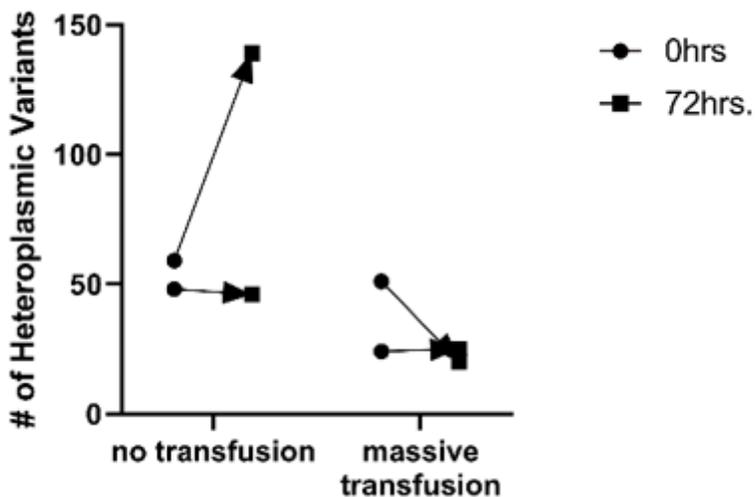


Figure 9

Changes in heteroplasmic variants with time and treatment. One non-transfused patient had a large increase in measured heteroplasmies while the second patient showed little change. One massively transfused patient showed a decrease in heteroplasmy over time while the other patient showed a similar level. Chart from Graphpad Prism.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalImage1.png](#)
- [SupplementalTables.xlsx](#)