Characterizing mitochondrial cell-free DNA in plasma as a novel liquid biopsy biomarker for oral cancer detection

UCLA Dentistry

Chase McGill[2], Gauri Vanjari[1], Alice Yu[3], Irene Choi[1], Jordan Cheng[1], Akanksha Arora[1], Karolina Kaczor-Urbanowicz[1], David T.W. Wong[1]

[1]W Lab, School of Dentistry, UCLA, [2]Cornell University, [3]University of Central Florida

Abstract

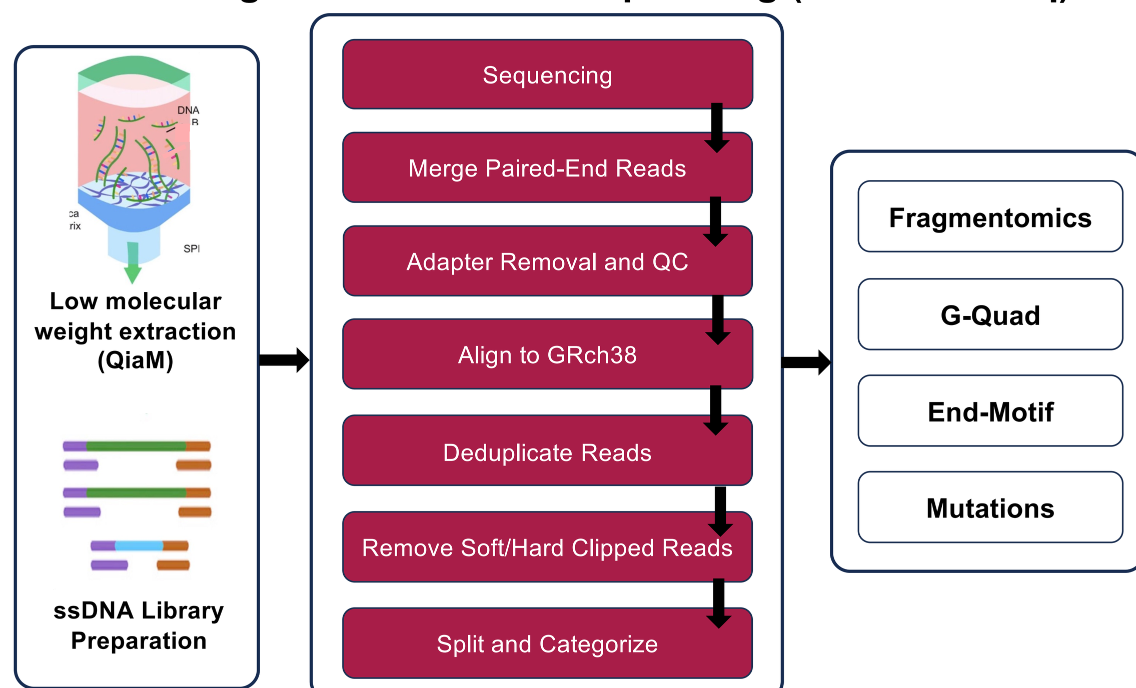
Plasma cell-free DNA (cfDNA) has been emerging as a promising source of biomarkers for oncological diseases. In addition to analyzing cfDNA from nuclear origin, reports have shown that mitochondria cell-free DNA (mitcfDNA) is also a viable biomarker for cancer detection. Previously, we demonstrated that using broad-range cell-free DNA Sequencing (BRcfDNA-Seq), which couples low-molecular weight extraction with single-stranded library preparation, reveals the presence of ultrashort single-stranded cell-free DNA (uscfDNA) in plasma in addition to mononucleosomal cell-free DNA (mncfDNA). We observed that BRcfDNA-Seq enhances the retention of mitcfDNA population, which contains characteristics to be a useful biomarker for cancer detection. Using strand-specific nucleases and library preparation modifications, we deduce that mitcfDNA is made up of a combination of ssDNA and dsDNA fragments. Based on end-motif analysis, we predict that the nucleases regulating mitcfDNA are different than those of nuclear origin. Additionally, we found that mitcfDNA has higher inherent coverage and G-Quad signatures. As a proof of concept, we analyzed 20 non-progressing and 20 progressing oral cancer samples using BRcfDNA-Seq, revealing that end-motif, mutations, and fragmentomic analysis of mitcfDNA are useful metrics for differentiating the two cohorts.

Introduction

Oral cancer, which includes cancers of the mouth and throat, accounts for roughly 3% of cancers diagnosed annually. Unfortunately, most oral cancers are identified at late stages where treatment options is limited. The ability to detect or determine malignancy status of oral premalignant lesions (OPML) at the early stages would dramatically improve patient outcomes. Cell-free DNA (cfDNA), which is found in different biofluids is commonly derived from the death of cells. cfDNA derived from cancerous cells or altered behavior from immune cells has recently been the focus of a new wave of research called liquid biopsy. Using BRcfDNA-Seq reveals that the presence of mncfDNA, which is double stranded and ~167bp long, and ultrashort cfDNA, which is single stranded and ~50bp long. However, mitochondrial cfDNA (mitcfDNA) has remained relatively unstudied. Although within the mitochondria, mitDNA is known to be circular and double stranded, and composed of a guanine-rich H-chain and a cytosine rich L-chain, its presentation in circulation is unclear. Because of its relatively high abundance and high mutation rate, mitcfDNA holds great potential as a tumor biomarker. We hypothesize that BRcfDNA-Seq enriches the abundance of mitcfDNA compared to conventional pipeline. As a proof of concept, we hypothesize that mitcfDNA contains unique characteristics which can be be useful in differentiating progressing from non-progressing oral cancer in patients presenting with oral premalignant lesions.

Methods

Broad-Range Cell-Free DNA Sequencing (BRcfDNA-Seq)



Nucleic acids from 1 mL of plasma from 1 non-cancer and 18 non-progressing and 18 progressing subjects presenting with OPML were extracted using the miRNA protocol from QiAmp Circulating Nucleic Acids kit and single-stranded library preparation was performed using the Claret SRSLY PicoPlus kit. Following 2x150bp sequencing at ~40million reads/sample, raw data was aligned to the human genome (hg38) using the BWAMEM algorithm. The processed files were split into ultrashort (40-70bp), mononucleosomal (120-250bp), and mitochondrial regions. Fragmentomics analysis was conducted to assess the lengths of the reads from each regions. A fragmentomics score was normalized to the fragment length at which a peak was identified. The G-quadruplex (G4) structure was predicted and identified using the REGEXfinder tool using the following formula: *d(G3+N1-7G3+N1-7G3+N1-7G3+)*. For end-motif, the first four base pairs at the 5' end of each read were also analyzed. The frequency of the unique combination of the end-motif were quantified within the mitcfDNA sequences. Mutect2 was used to determine basepair substitutions.

Results

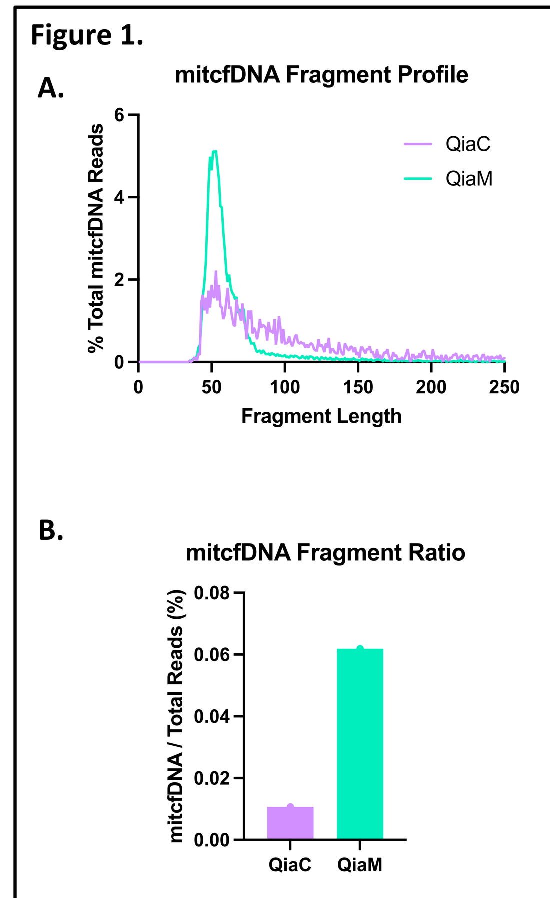


Figure 1. QiaM enriches short fragmented mitcfDNA compared to conventional extraction QiaC. A) QiaM shows a prominent peak for mitcfDNA at 54bp compared to the broad peak of QiaC. B) QiaM captures ~6 times more mitcfDNA than QiaC.

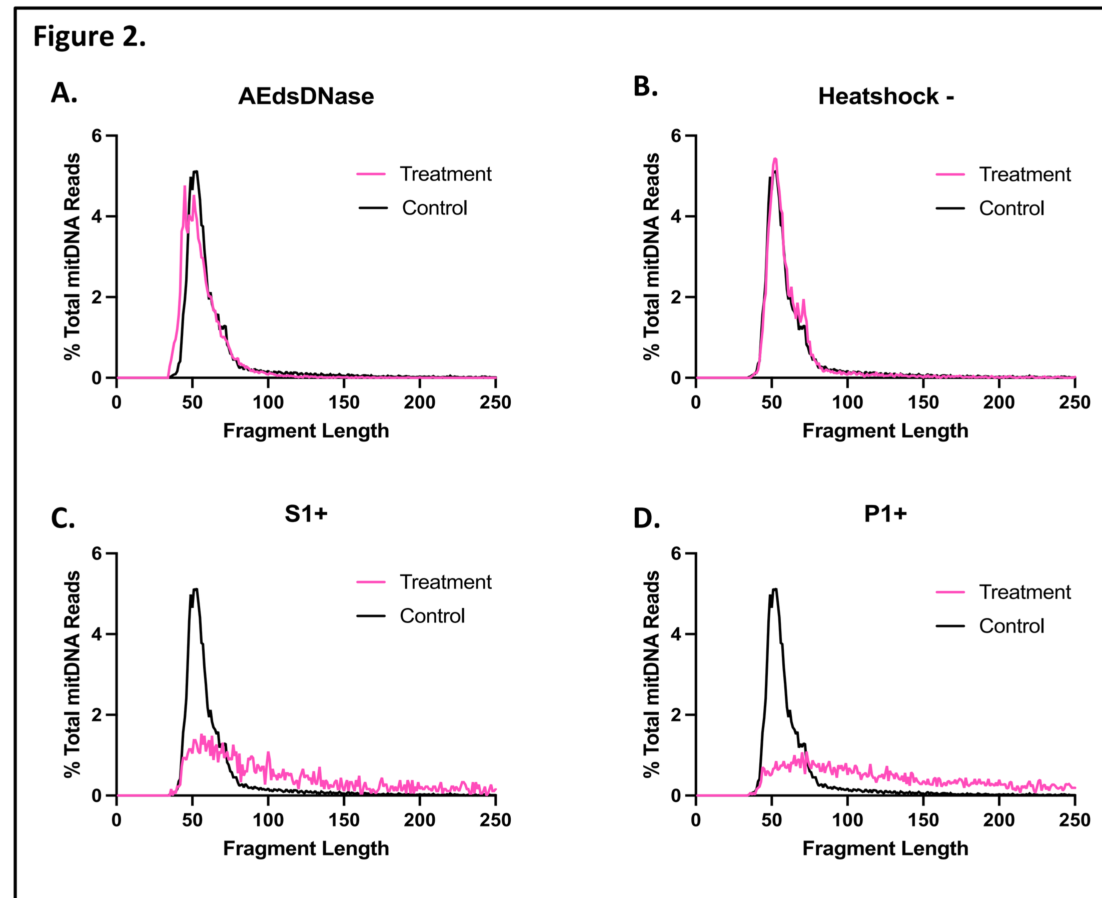


Figure 2. Digestion of mitcfDNA reveals information about its strandedness. A,B) After treatment with two procedures that remove dsDNA, a clear peak is still present, indicating that a large portion of mitcfDNA is single stranded. C,D) These two digestions removed all ssDNA, and the fact that there is still a significant number of reads indicates that a portion of mitcfDNA population is double stranded.

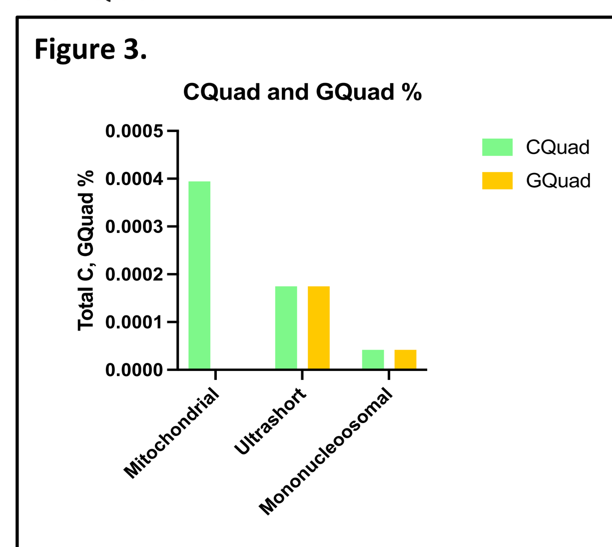


Figure 3. mitcfDNA has much higher percentages of CQuad positive sequences but is absent in Gquad positive sequences, while mncfDNA and uscfDNA have roughly equal percentages of the two.

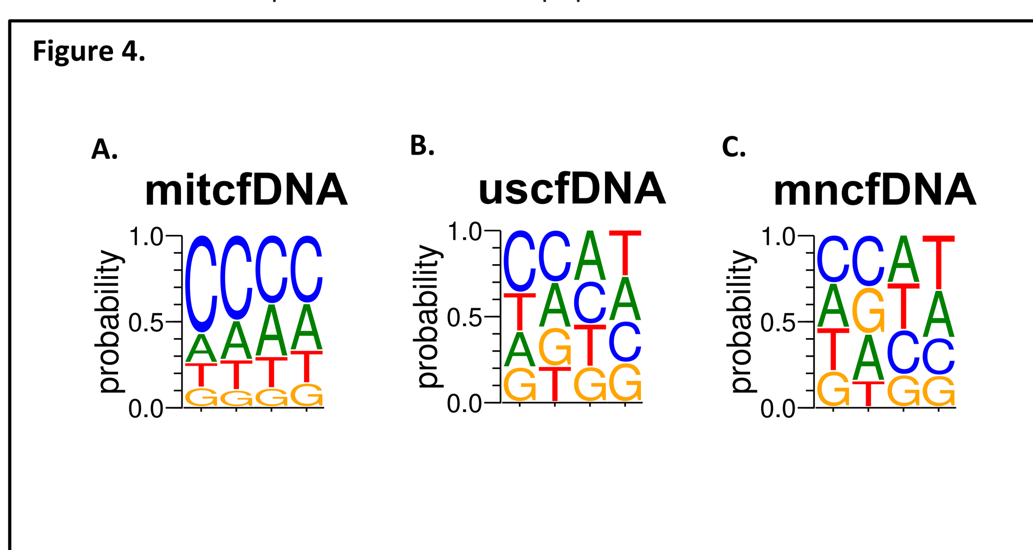


Figure 4. End motif analysis shows that mitcfDNA tends to have significantly more C's at the first 4 bps while mncfDNA and uscfDNA tend to have more variability.

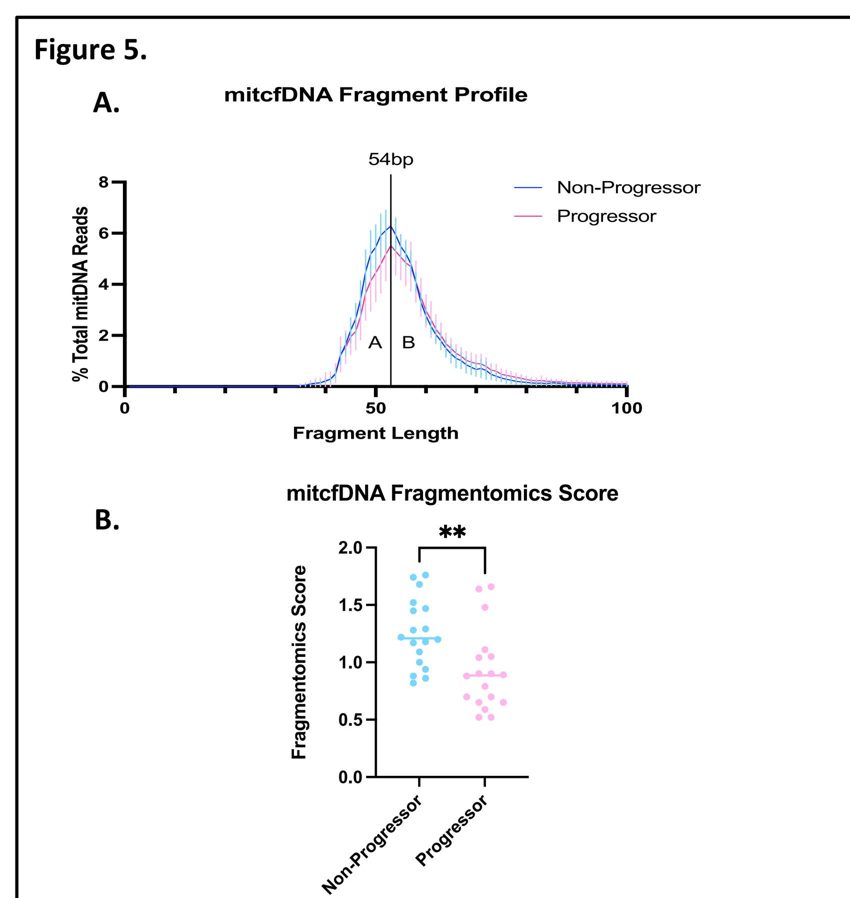


Figure 5. A) Non-progressor samples have a higher peak (more mitcfDNA reads) than progressor samples. B) Non-progressor samples have a higher mean fragmentomics score (percentage of reads from 35-54bp / percentage of reads from 55-90bp). Student t-test was performed, **p-value < 0.01.

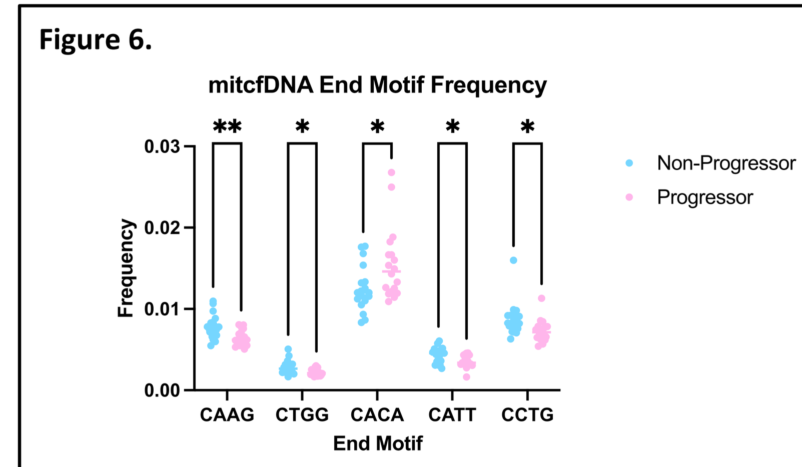


Figure 6. Candidate altered end-motif frequency between non-progressor and progressor groups. Multiple t-test was performed, unadjusted, *p-value < 0.05, **p-value < 0.01.

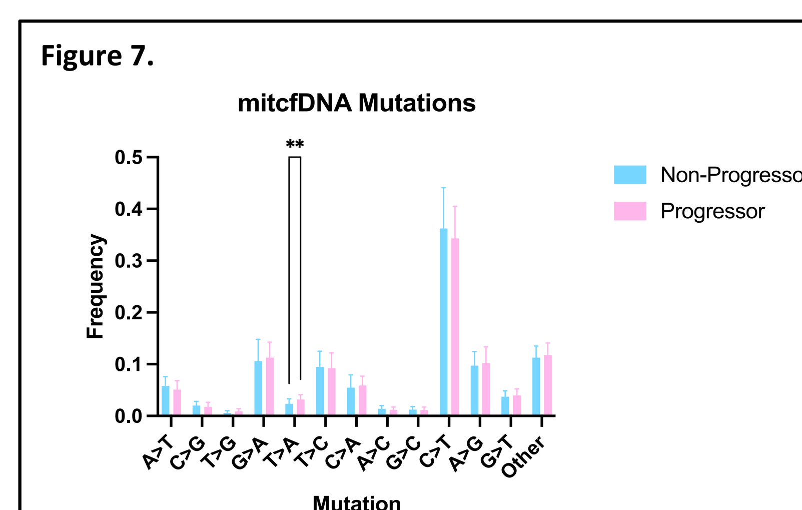


Figure 7. Higher T > A substitution mutations in mitcfDNA may be a signature in the progression of oral cancer. Multiple t-test with Welch's correction was performed, unadjusted **p-value < 0.01.

Discussion

- Combination of single-stranded and enhanced low molecular weight extraction reveals of QiaM an ultrashort population of mitcfDNA (Figure 1A, 1B)
- Mechanistically, QiaM protocol uses an increased ratio of isopropanol that is improves the ability to purify low molecular weight DNA
- Using different digestions and preparation methods, we report that mitcfDNA appears to be a combination of single and double stranded DNA
- The digestion and altered library preparation treatments that eliminated all dsDNA (Figure 2A, 2B) left a prominent peak, indicating that a large portion of mitcfDNA is single stranded
- After digestion with enzymes that eliminated ssDNA (Figure 2C, 2D), there was still a small but noticeable peak, suggesting a proportion of mitcfDNA is double stranded
- Single-stranded DNA has the ability to form stable secondary structures such as G-Quad in Guanine-rich sequences
- Interestingly, although G-Quad positive sequences were not found, the reverse complement C-Quad positives were present in high quantities in the mitcfDNA samples (Figure 3)
- This contrasted the equal number of G-Quad and C-Quad sequences in mncfDNA and uscfDNA (Figure 3)
- The C-rich end-motif signature (Figure 4A) in mitcfDNA most likely indicates that the L - chain is present in the biofluid of the mitcfDNA
- The uscfDNA and mncfDNA (Figure 4B, 4C) show a similar end-motif makeup that is different from mitcfDNA (Figure 4A)
- Fragmentomics analysis of cfDNA is a promising biomarker and yields information about properties such as fragment endpoints, specific motifs, and nucleosome positioning
- A lower peak in the progressor group suggests that the two cohorts may have different fragment profiles (Figure 5A)
- Those in the non-progressor group (did not advance to cancer) displayed a higher mean fragmentomics score than those in the progressor group (advanced to cancer) (Fig 5B), indicating that non-progressor samples have a higher proportion of short mitDNA and a smaller fragment size
- Fragmentomics score = percentage of reads from 35-54bp / percentage of reads from 55-90bp
- The cfDNA end-motifs are defined using the first 4-nucleotide sequence of each 5' end of aligned reads
- Aberrations in 4-mer end motfs in plasma DNA can serve as another biomarker
- The difference in the scores of candidate motifs between non-progressor and progressor samples are significant (Figure 6)
- The increase in these motifs (eg. CACA) in progressor samples can be indicative of a cancer diagnosis
- Using Mutect2, mutations and SNPs were identified
- There exists a significant T to A mutation in the progressor group (Figure 7)
- This may pose to be another biomarker to study in differentiating between cancer and non-cancers

In conclusion, our study examines the properties of mitochondrial cell-free DNA (mitcfDNA), revealing its unique properties and clinical potential. Analysis of the nature of this DNA reveals a mixture of single- and double-stranded mitcfDNA, exclusively exhibiting "C-Quadruplex" structures. The fragmentomics score emerges as a promising cancer biomarker, with higher fragment scores correlated to non-progression. Using a mutation finder tool such as Mutect2 may allow for further discrimination of progression to cancer. Our findings deepen the understanding of mitcfDNA, providing insights that can contribute to future cancer diagnostic and therapeutic advances.

References

Burnham P, Kim MS, Agbor-Enoh S, Luikart H, Valantine HA, Khush KK, De Vlaminck I. Single-stranded DNA library preparation uncovers the origin and diversity of ultrashort cell-free DNA in plasma. Sci Rep. 2016 Jun 14;6:27859. doi: 10.1038/srep27859. PMID: 27297799; PMCID: PMC4906518.

Sahayasheela VJ, Yu Z, Hidaka T, Pandian GN, Sugiyama H. Mitochondria and G-quadruplex evolution: an intertwined relationship. Trends Genet. 2023 Jan;39(1):15-30. doi: 10.1016/j.tig.2022.10.006. Epub 2022 Nov 19. PMID: 36414480; PMCID: PMC9772288.

Liu Y, Zhou K, Guo S, Wang Y, Ji X, Yuan Q, Su L, Guo X, Gu X, Xing J. NGS-based accurate and efficient detection of circulating cell-free mitochondrial DNA in cancer patients. Mol Ther Nucleic Acids. 2021 Jan 1;23:657-666. doi: 10.1016/j.omtn.2020.12.017. PMID: 33575112; PMCID: PMC7851424.

An Q, Hu Y, Li Q, Chen X, Huang J, Pellegrini M, Zhou XJ, Rettig M, Fan G. The size of cell-free mitochondrial DNA in blood is inversely correlated with tumor burden in cancer patients. Precis Clin Med. 2019 Sep;2(3):131-139. doi: 10.1093/pcmedi/pbz014. Epub 2019 Oct 1. PMID: 31598384; PMCID: PMC6770274.

Haupts A, Vogel A, Foersch S, Hartmann M, Maderer A, Wachter N, Huber T, Kneist W, Roth W, Lang H, Moehler M, Hartmann N. Comparative analysis of nuclear and mitochondrial DNA from tissue and liquid biopsies of colorectal cancer patients. Sci Rep. 2021 Aug 18;11(1):16745. doi: 10.1038/s41598-021-95006-6. PMID: 34408162; PMCID: PMC8373949.

Zhang R, Nakahira K, Guo X, Choi AM, Gu Z. Very Short Mitochondrial DNA Fragments and Heteroplasmy in Human Plasma. Sci Rep. 2016 Nov 4;6:36097. doi: 10.1038/srep36097. PMID: 27811968; PMCID: PMC5095883.

Malkin EZ, De Michino S, Lambie M, Gill R, Zhao Z, Rostami A, Arruda A, Minden MD, Bratman SV. Cell-free DNA topology depends on its subcellular and cellular origins in cancer. JCI Insight. 2022 Oct 24;7(20):e159590. doi: 10.1172/jci.insight.159590. PMID: 36125881; PMCID: PMC9714790.