A high-efficient capture-based NGS approach for comprehensive analysis of mitochondrial transcriptom

Zhenni Wang  
Air Force Medical University

Kaixiang Zhou  
Air Force Medical University

Qing Yuan  
Northwestern Polytechnical University

Dongbo Chen  
Northwestern Polytechnical University

Xi'e Hu  
Tangdu Hospital Fourth Military Medical University: Air Force Medical University Tangdu Hospital

Fanfan Xie  
Air Force Medical University

Yang Liu  
Tangdu Hospital Fourth Military Medical University: Air Force Medical University Tangdu Hospital

Jinliang Xing (xingjinliang@163.com)  
Air Force Medical University  https://orcid.org/0000-0002-7010-1822

Research Article

Keywords: Mitochondrial RNA, Capture-based sequencing, Transcriptome profiling

Posted Date: September 9th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3322535/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Background

The transcription of mitochondrial genome is pivotal for maintenance of mitochondrial functions, and deregulated mitochondrial transcriptome contributes to various pathological changes. Despite substantial progress has been achieved in uncovering the transcriptional complexity of the nuclear transcriptome, many unknowns and controversies remain for the mitochondrial transcriptome, partially owing to the lack of high-efficient mitochondrial RNA (mtRNA) sequencing and analysis approach.

Methods

Here, we first comprehensively evaluated the influence of essential experimental protocols, including strand-specific library construction, two RNA enrichment strategies and optimal rRNA depletion, on accurately profiling mitochondrial transcriptome in whole transcriptome sequencing (WTS) data.

Results

Based on these insights, we developed a high-efficient approach specifically suitable for targeted sequencing of whole mitochondrial transcriptome, termed capture-based mtRNA seq (CAP), in which strand-specific library construction and optimal rRNA depletion was applied. Compared with WTS, CAP has a great decrease of required data volume, without affecting the sensitivity and accuracy of detection. In addition, CAP also characterized the unannotated mt-tRNA transcripts whose expression level is below the detection limits of conventional WTS. As a proof-of-concept characterization of mtRNAs, the transcription initiation sites and mtRNA cleavage ratio were accurately identified in CAP data. Moreover, CAP had a very reliable performance in plasma and single-cell samples, highlighting its wide application.

Conclusions

All together, the present study has established a high-efficient pipeline for targeted sequencing of mtRNAs, which may pave the way toward functional annotation of mtRNAs and mtRNA-based diagnostic and therapeutic strategies in various diseases.

Background

As the key organelle in eukaryotic cells, mitochondria function vitally in regulating diverse biological processes, including oxidative phosphorylation, carbohydrate and fat metabolism, antioxidant defense and biosynthesis of hormones (Vyas et al., 2016). Human mitochondria contain a compact double-strand circular genome, known as mitochondrial DNA (mtDNA), which is 16.6 kilobases in length (Kopinski et al., 2021). Mitochondrial RNAs (mtRNAs) are transcribed as long polycistronic precursor transcripts initiated
from two heavy strand promoters (HSP1 and HSP2) and one light strand promoter (LSP), which necessitates extensive processing to release the individual mRNAs, rRNAs, and tRNAs (Rackham and Filipovska, 2022). In detail, the heavy (H) strand encodes 12 protein subunits of the oxidative phosphorylation system, 2 rRNAs, and 14 tRNAs, whereas the light (L) strand encodes 1 protein subunit and 8 tRNAs (Rackham and Filipovska, 2022). Deep understanding of the unique transcriptional system owned by mitochondria provides avenues for the treatment of diseases caused by impaired energy conversion, including neurodegenerative, cardiovascular, and neoplastic diseases (Rackham and Filipovska, 2022).

Advances in RNA sequencing (RNA-Seq) technologies permit the unbiased profile of the human transcriptome, providing a comprehensive annotation of the functional transcripts (Stark et al., 2019). These phenotype-related transcripts are in the form of not only protein-coding RNAs but also non-coding RNAs that participate in regulation of transcription (de Klerk and t Hoen, 2015). The recent advent of deep sequencing has provided a global profile of the nuclear transcriptome, uncovering the unexpected transcriptional complexity that includes prevalent post-transcriptional processing and numerous functional noncoding RNAs (Jacquier, 2009). However, little is known about the sophisticated features of the mitochondrial transcriptome, especially the regulation of mtRNA abundance, mtRNA processing and modification sites, as well as its maturation and degradation in various diseases.

In the past few years, several whole transcriptome-wide approaches have been employed to interrogate the mitochondrial transcriptome dynamics, which have revisited the genetic content of the mitochondrial genome and identified previously overlooked noncoding transcripts (Mercer et al., 2011b), (Rackham et al., 2016). A recently developed circularized RNA seq has enabled the discovery of unprocessed mtRNAs and revealed simultaneous processing and degradation of mtRNAs (Kuznetsova et al., 2017). Despite the diverse whole transcriptome sequencing (WTS)-based technologies that have been successfully employed to characterize the mitochondrial transcriptome in various contexts, they have several caveats: (i) the atypical features of mitochondrial transcription require special considerations in sequencing and analysis, including polycistronic transcription, lack of recognizable features such as poly-A tails, and the presence of multiple precursor/intermediate forms; (ii) the profiling of mtRNAs is limited by the sensitivity of sequencing methods to detect mitochondrial transcripts with low-abundance and may be partially improved by depletion of ribosomal RNA, and hence requiring higher data volume in delineating the biological relevance of low-level mtRNAs. Owing to these limitations, current studies have still not yet comprehensively elaborated the characteristics of mitochondrial transcriptome, fueling ongoing debates in describing mitochondria-specific mechanisms of RNA transcription and processing.

To reliably detect and quantify multiple forms of mtRNAs and thus uncover the complexity of the mitochondrial transcriptome, here we developed a highly efficient mtRNA targeted enrichment approach termed capture-based mtRNA seq. In this approach, oligonucleotide probes were utilized to hybridize mt-cDNAs and facilitate mtRNA purification, which dramatically increase the abundance of selected mitochondrial transcripts. Considering the unique feature of mitochondrial polycistronic transcription, the present study first systematically assessed the feasibility of strand-specific library construction, two RNA
enrichment strategies and optimal rRNA depletion in mitochondrial transcriptome profiling. Furthermore, we have provided the first comprehensive analysis of the human mitochondrial transcriptome based on capture-based mtRNA seq in a cost-effective and efficient manner. Importantly, the capture-based mtRNA seq could also be widely applied in various types of samples, providing a foundation for further broad clinical utility of mtRNA detection.

Methods

Sample processing and data collection

A total of 21 cell line samples, 15 tissue samples, 3 plasma samples, and 3 single cell samples were included in the present study. The miRNeasy tissue kit (Qiagen, Germany) and plasma/serum circulating and exosomal RNA purification kit (Norgen Biotek, Canada) were utilized for total RNA extraction from tissues or cell lines and plasma samples, respectively. Extracted RNA samples were stored at −80°C until use. In addition, the REPLI-g® WTA single cell kit (Qiagen, Germany) was employed to extract and amplify total RNA from single cell samples. RNA quantity was determined using a Qubit RNA BR Assay kit (Invitrogen, USA). The RNA degradation was assessed using 2100 BioAnalyzer (Agilent, USA), and only samples with RIN (RNA integrity number) > 7 were included for library construction. Public raw data from whole transcriptome sequencing (WTS) of three human osteosarcoma cell lines were downloaded from PRJNA251691. All sample and data information were summarized in Supplemental Table 1. This study was approved by the Ethics Committee of FMMU, and written consent was obtained from each individual.

PolyA selection and rRNA depletion

The Poly(A)Purist™ MAG kit (Invitrogen, USA) was used for the PolyA selection from total RNA according to manufacturer’s protocols. In brief, the polyA-containing mRNA molecules were purified from 1 µg of total RNA using magnetic beads attached with polyT oligo. In addition, QIAseq FastSelect -rRNA HMR kit (Qiagen, Germany) was used to deplete rRNA from 1 µg of total RNA following the manufacturer instructions. Different quantities of rRNA probes (1×, 0.2×, 0.1×, 0.02×, 0×) were evaluated for the optimal rRNA depletion.

cDNA library construction and WTS

The 1 µg of total RNA samples treated with polyA selection or rRNA depletion were used for construction of WTS libraries using QIAseq stranded RNA library kits (Qiagen, Germany). Briefly, the processed RNA samples were first sonicated for fragmentation. Then, the synthesis of first-strand and second-strand cDNAs was performed using random hexamers and 5’ phosphorylated random primers, respectively. Moreover, the 5’ phosphorylated adapters were efficiently ligated to construct strand-specific libraries. The libraries were quantified using Qubit (Invitrogen, USA) and their quality were evaluated using 2100 BioAnalyzer (Agilent, USA). Finally, the libraries were sequenced by Illumina Novaseq 6000 (Illumina) platform using paired-end runs with 2 x 150 cycles (PE150).

Capture-based mtRNA sequencing
Capture-based mtRNA sequencing was performed as previously described (Zhou et al., 2020). In brief, 200 ng of WTS library was mixed with 10 ng of probe and hybridized at 65°C for 24 h. The captured mtRNA library was further amplified and sequenced on an Novaseq 6000 (Illumina) platform using PE 150.

Data processing and mapping

Raw sequencing data were trimmed using the fastp software (v 0.20.0), an ultra-fast all-in-one FASTQ preprocessor. In brief, the sequencing adaptors were first removed. A sliding window (4 bp in length) was then used to scan reads from front (5′) to tail (3′), and the bases with the average quality below Q30 in the window were discarded. The trimmed sequencing reads were mapped to the human nuclear reference genome grch38 and the revised Cambridge Reference Sequence (rCRS) using hisat2 software (v 2.2.1) and the reads that were not paired or with inconsistent directions were further removed. The mitochondrial reads in fastq format were then realigned to the rCRS based on no-mixed, no-discordant, and no-spliced alignment. Finally, the retained mitochondrial reads were extracted for further analysis according to their strand-of-origin with SAMtools (v 1.6).

Coverage and expression analysis

Considering the inconsistency of data volume among samples, the coverage of each sample was normalized as following: the normalized coverage = (the site depth/ the total depth) × 10^6. R package was used to calculate the gene expression as the fragments per kilobases of transcript per million mapped reads (FPKM). The formula for calculating FPKM was as follows: [number of fragments] / [(transcript length / 1,000) / (total reads) / 10^6)].

Identification of transcription initiation sites (TISs) and cleavage sites

TIS identification was carried out according to the method described previously (Blumberg et al., 2017) with some minor modifications. Briefly, a sliding window of 50 base pairs within the D-loop region was used to scan the specific segment with a sudden increase (which was the potential region of TISs in H and L strand) of mtRNA coverage. The site of highest peak within the identified region was identified as the TIS. The same criteria were used for the identification of transcription pausing site of L strand. In addition, the method previously described (Carbajosa et al., 2022) was used to analyze mtRNA cleavage sites and corresponding cleavage ratio. The cleavage ratio of each site was calculated as the proportion of reads starting or ending either side of site (processed reads) to total reads across the site (processed and unprocessed). Then, the site with highest cleavage ratio for each gene boundary region was identified as the cleavage site.

Results

Establishment of the capture-based mtRNA seq
To efficiently profile mitochondrial transcriptome with high resolution, a targeted sequencing strategy termed capture-based mtRNA seq was developed (Fig. 1). Briefly, several key procedures, including strand-specific library construction, polyA selection and rRNA depletion, were systematically evaluated and optimized. Then, mitochondrial cDNAs from the optimized library were further captured to provide robust coverage and permit the reliable detection of whole mitochondrial transcripts.

**Stranded sequencing library construction for accurate identification of mitochondrial transcripts**

First, the two strategies of sequencing library construction (stranded and non-stranded), which were commonly used in RNA seq, were evaluated for accurate mitochondrial transcriptome profiling by using the paired public datasets from three human osteosarcoma cell lines. No significant difference of base quality value and mitochondrial genome (chrM) mapping rate was observed between non-stranded and stranded RNA seq data (Fig. S1a-b). As shown in Fig. 2a, the transcripts derived from mitochondrial H and L strands were accurately identified in stranded RNA seq data, with H strand transcripts of average 95.17% and L strand transcripts of average 4.83%, respectively. However, the strand origin of transcripts was undistinguishable in non-stranded RNA seq data, as the double-stranded cDNA sequencing reads lead to the loss of mtRNA origin information. Then, the expression level of mitochondrial genes was compared between both stranded and non-stranded RNA seq data. Our data showed a clearly increased expression level of 13 mitochondrial mRNAs (mt-mRNAs) in the non-stranded RNA seq data when compared to stranded RNA seq data, although the expressions were highly correlated between the two groups (Fig. 2b). The remarkable inconsistency of the expression levels of mitochondrial tRNAs (mt-tRNAs) was found between two groups (Fig. 2c). Furthermore, we found that H and L strand transcripts were overlapped in over 98.66% regions of mitochondrial genome (Fig. S1c), which accounted for the inaccurate expression levels of mt-RNA in the non-stranded RNA-seq data.

**Comparison of two transcriptome isolation approaches for mtRNA sequencing**

PolyA selection and rRNA depletion are two commonly used transcriptome isolation approaches. Moreover, we compared the applicability of both approaches for accurate profiling of mitochondrial transcriptome using paired WTS data from three liver cancer cell lines. As shown in Fig. 3a-b, the remarkable difference of the normalized coverage was observed between WTS data with polyA selection and rRNA depletion, although H and L strand transcripts were clearly identified in both approaches. Further analysis showed that the proportion of reads aligned to 16S mt-rRNA regions in WTS data with polyA selection was significantly higher than that in WTS data with rRNA depletion (6.45% vs. 0.85%, \(P<0.05\)). In contrast, no significant difference was observed in the proportion of reads aligned to 12S mt-rRNA regions between two groups (0.98% vs. 0.69%, \(P=0.3779\)) (Fig. 3c). Our results are in line with a recent experimental report indicating that 16S but not 12S mt-tRNAs are highly polyadenylated (Begik et al., 2023). Moreover, we evaluated the concordance of mt-RNA gene expression level between two groups
(Fig. 3d-e). Our analysis showed that the expression levels of a portion of mt-mRNAs including CO1 in the rRNA depletion group were clearly higher than those in polyA selection group (Fig. 3d), despite the overall high concordance between the two groups, which is possibly due to various degree of polyadenylation of mt-mRNAs. Owing to the post-transcriptional addition of CCA but not polyA tails at the mature mt-tRNA 3’ end, it is understandable to observe the more obvious discrepancies of mt-tRNA expressions between two groups, with higher expression levels of most mt-tRNAs in the rRNA depletion group (Fig. 3e). Collectively, all these results indicate that rRNA depletion is a more suitable approach for mitochondrial transcriptome profiling.

**Optimal rRNA depletion for efficient profiling of whole mitochondrial transcriptome**

To investigate the optimal rRNA depletion for efficient mitochondrial transcriptome profiling, we performed systematic evaluation against rRNA depletion probe quantity using three liver cancer cell line samples. The proportion of mt-rRNA reads in total reads were used to assess the efficiency of rRNA depletion. As shown in Fig. 4a, the mt-rRNA proportion was decreased in probe dosage-dependent manner. To maximize the proportion of non-rRNA transcripts while retain appropriate rRNA content for downstream analysis, the retained mt-rRNA proportion was set to be close to the ratio of rRNA gene length to total mitochondrial genome length (~15%). An average of 17.74% mt-rRNA (ranging from 12.46–22.1%) was retained when using 0.1X rRNA probe, which was hence selected as the candidate optimal rRNA depletion quantity (Fig. 4a). Then, the normalized coverage of different rRNA probe groups was depicted to evaluate whether different degree of rRNA depletion would alter the original portrait of mt-rRNA coverage (Fig. 4b). There were favorable correlations of rRNA coverage between the 0X and 0.02X groups (r = 0.8851), and between the 0X and 0.1X groups (r = 0.6965) (Fig. 4c). Taken together, the 0.1X rRNA depletion probe was selected as optimal condition for mitochondrial transcriptome profiling in tissue samples.

**Improved characterization of mitochondrial transcriptome by capture-based mtRNA seq**

Considering that mitochondrial RNAs only account for a small percentage of whole transcriptome in majority of human cell types, we evaluated the feasibility and cost-effectiveness of capture-based mtRNA seq (CAP) using 1 Gb of paired CAP and WTS data from six human liver tissue samples. As shown in Fig. 5a, very similar distribution pattern of coverage depth of both H and L strand transcripts was observed. Correlation analysis showed that the distribution of coverage depth was highly consistent between paired CAP and WTS groups for both H (r = 0.9062) and L (r = 0.9462) strands, suggesting the unbiased capture of mitochondrial transcriptome. Furthermore, we compared the expression level and relative proportion of 37 mtDNA-encoded genes between CAP and WTS groups. Our results showed significant correlation of proportion of three types of mtRNAs (r = 0.9923, P < 0.0001) and expression level of both H and L strand mtRNAs (r = 0.9630, P < 0.0001) between two groups (Fig. 5b), indicating that CAP
enables unbiased quantitative analysis of mtRNAs. Moreover, the coverage depth ratio of H and L strand mtRNAs also exhibited remarkable correlation between CAP and WTS groups ($r = 0.9411$, $P < 0.0001$; Fig. 5c), indicating an equivalent capture efficiency.

Except for the high-fidelity, CAP also holds several prominent performances compared to WTS. When the same data volume (1 Gb) was compared, CAP showed 36-fold higher average coverage depth for H strand transcripts (26492X vs. 728X, $P < 0.001$) and 56-fold higher average depth for L strand transcripts (2103X vs. 37X $P < 0.001$) than WTS (Fig. 5d). Similarly, the CAP group showed remarkable increase of the coverage depth of mt-tRNAs compared to WTS group (Fig. S2a). Besides, our analysis showed that significant fewer CAP data volume was required to achieve a series of mtRNA coverage depth for both H and L strand when compared with WTS (Fig. 5e). To further investigate the performance of CAP in detecting mt-tRNA, diverse precursors of ND1-tRNAs-ND2 were taken as examples (Fig. S2b). By benchmarking the CAP and WTS data to the same volume (1 Gb), our result showed that a greater diversity of mt-tRNAs were detected in CAP data when compared to WTS data (Fig. 5f). All these results indicate the feasibility and cost-effectiveness of CAP approach, which holds great superiority for sequencing large-scale samples and thus delineating the range, depth and complexity of mtRNAs far from fully characterized.

**Capture-based mtRNA seq enables comprehensive analysis of mitochondrial transcriptome**

Here, the mitochondrial transcriptome was comprehensively characterized by high-efficient capture-based mtRNA seq in three human liver tissue samples. The candidate transcription initiation sites (TISs) in the H and L strands were identified by screening the regions with a sudden increase of coverage. As shown in Fig. 6a, the TISs of H strand were identified within position 576 and 658, which were consistent with previous findings about HSP1 and HSP2 locations. The TIS of L strand was identified at position 412, which was also exactly consistent with the known LSP region (Fig. 6b). Moreover, the position 311 located in conserved sequence block 2 (CSB2), which was previously demonstrated to cause transcription pausing due to formation of a hybrid G-quadruplex, was hence identified as transcription pausing site (TPS). Next, the 5’ and 3’ cleavage sites of mitochondrial transcripts were systematically identified. By counting the number of 5’ and 3’ cleavage reads and then dividing this by the number of reads that fully overlap the site, we can generate a proxy ‘cleavage ratio’ for that position and infer putative mtRNA cleavage sites from peak cleavage ratio. Based on this strategy, the 5’ and 3’ cleave sites across H and L strands with calculated cleavage ratio were systematically delineated (Fig. 6c-f). Finally, both sense and antisense transcripts of mt-mRNAs and mt-tRNAs were identified and thus the expression level of each mtRNA was depicted in Fig. 6g-h and Fig. S3a. Taken together, our CAP data allow the comprehensive analysis of mitochondrial transcriptome and help to understand the unexpected complexity in mtRNAs.

**Efficient application of capture-based mtRNA seq in plasma and single cell samples**
Finally, beyond tissue samples, we also assessed the application of the capture-based mtRNA seq in other specimens, including plasma and single cell samples. The normalized coverage of mtRNAs from three human plasma samples was depicted in Fig. 7a, indicating distinct proportions of mitochondrial transcripts from H and L strands. Although the plasma RNA samples were treated by the same quantity of rRNA depletion probes (0.1X) as tissue samples, the mt-rRNA transcripts were still kept at a high proportion of mt-rRNAs, which needs further optimization of rRNA depletion. Furthermore, the sense and antisense expressions of mt-mRNAs and mt-tRNAs were calculated and then shown in Fig. 7b-c and Fig. S3b. Next, the 5’ and 3’ cleavage sites of mitochondrial transcripts were successfully identified in plasma RNA samples (Fig. 7d-g). For single cell specimens, owing to the stranded RNA seq construction for low input samples were still inefficient to date, we hence constructed the non-stranded WTS library and investigated the application of capture-based mtRNA seq. As shown in Fig. S4, the coverage, expressions, and cleavage sites of mtRNAs were successfully characterized in single cell samples by capture-based approach, although H and L strand transcripts cannot be distinguished.

**Discussion**

To our knowledge, no approach for targeted sequencing of mitochondrial RNA is available until now. To address this demand, we have developed an optimized approach specifically suitable for targeted sequencing of whole mitochondrial transcriptome in a broad range of sample types. Our novel approach has several advantages: (i) it provides significant relevance to the selection of optimal experimental parameters for mtRNA sequencing by integrating the atypical features of mitochondrial transcription; (ii) it is competitive in economic terms, allowing comprehensively analysis of mitochondrial transcriptome with satisfactory resolution but reduced cost; (iii) it provides a powerful tool to investigate the hierarchy of mitochondrial transcription initiation and processing and thus enables a broad-spectrum characterization of mitochondrial transcriptome; (iv) it confers widespread applicability in the detection of different types of specimens, including plasma and single-cell samples.

Broader applications of RNA seq have greatly transformed our understanding on human nuclear transcriptome and revealed many essential phenotype-related features, including differential gene expression, alternative RNA splicing and modification (Bradley and Anczuków, 2023), (de Klerk and t Hoen, 2015). Compared to the nucleus, the mitochondria have evolved its unique features with respect to the organization and regulation of gene transcription (Jensen et al., 1988), whereas optimal RNA seq specifically for better characterization of mtRNAs is still lacking. In the present study, we comprehensively evaluated the influence of essential experimental protocols, including library construction, RNA enrichment, rRNA depletion, on accurately profiling mitochondrial transcriptome in WTS data. Our results indicate that stranded RNA-seq provides a more accurate identification of mitochondrial transcripts derived from either H or L strand when compared with non-stranded RNA-seq and is therefore recommended for mtRNA sequencing. In addition, the polyadenylation of mRNAs is critical for the stability and translation of transcripts (Gruber and Zavolan, 2019). A series of previous studies have reported that the polyadenylation is essential for completion of stop codons in 7 of the 13 H strand mt-mRNAs (ND1-4, CTYB, COX3 and ATP6) and the L strand mt-mRNA (ND6) is not polyadenylated.
(Rackham and Filipovska, 2022) (Pearce et al., 2017). Therefore, considering the different levels of mt-RNA polyadenylation, we evaluated the feasibility of the two commonly used RNA enrichment approaches, polyA selection and rRNA depletion, in mtRNA sequencing and found that the rRNA depletion was indeed a more suitable approach for the accurate profiling of mtRNAs. Furthermore, the 0.1X rRNA probe was demonstrated to retain a relative low level of mt-rRNAs while not alter the original mt-rRNA profile and was thus selected as optimal rRNA depletion in tissue samples, whereas further optimization is needed for other sample types, such as plasma. Based on these insights of performance evaluations, the highly efficient protocols can be selected for the capture-based mtRNA seq.

Although emerging WTS-based studies have raised great promise for diagnosis and treatment of various diseases, low-level mtRNAs may be unannotated under current protocols due to the insufficient coverage of mtRNAs. Recent studies have demonstrated the feasibility of capture-based approach in obtaining the enriched read coverage and accurate quantitation for nuclear RNA (Bussotti et al., 2016), (Mercer et al., 2011a), whereas the approach for mtRNA target enrichment and sequencing is still lacking till now. Hence, in the present study, we attempted to develop a highly efficient approach to interrogate mtRNA in a manner with lower costs and unprecedented resolution. Compared with WTS-based approach, capture-based NGS approach has a great decrease in required data volume, without affecting the sensitivity and accuracy of detection. In addition, the capture-based approach also characterized the unannotated mt-tRNA transcripts whose expression level is below the detection limits of conventional WTS. As a proof-of-concept characterization of mtRNAs, the transcription initiation and pausing sites were identified based on capture-based mtRNA seq. Additionally, due to the polycistronic nature of mitochondrial transcription, post-transcriptional events are pivotal in determining downstream events. We hence identified putative mtRNA cleavage sites and ratios by capture-based mtRNA seq, which may contribute to elaborate key mtRNA processing events. Moreover, the capture-based mtRNA seq also had a very reliable performance in plasma and single-cell samples, highlighting its wide application. Together, the present study has established a highly efficient mtRNA targeted sequencing approach, which enables unbiased mtRNA detection, showing evenly captured transcripts across mitochondrial transcriptome and therefore maintaining the original features of transcription.

Despite the above advantages, the capture-based mtRNA seq still needs to be optimized in several respects. First, the designed capture probes were fragmented products of mtDNA PCR amplification, which may be insufficient when studying novel or rare mtRNA intermediates. Nevertheless, the unnecessarily exact complementary matches between probes and target cDNAs together with the high depth coverage may serve as a versatile solution to comprehensively detect various intermediates and single-nucleotide polymorphisms in mtRNAs. Second, several technical defects may limit the completeness of capture-based NGS approach, including the shorter sequencing reads than many transcripts, incomplete reverse transcription of RNA templates, and RNA degradation before reverse transcription. These issues should be taken into consideration in the improvements of future protocols. The combination of third-generation sequencing and template-switching approaches would be appreciated.
To the best of our knowledge, this is the first study to provide a highly efficient capture-based mtRNA sequencing approach, holding great promise for the comprehensive analysis of human mitochondrial transcriptome that is far from fully characterized. We anticipate that based on the established approach in our study, the previously unknown features of mitochondrial transcription (such as initiation, processing, maturation and degradation of mtRNAs) would be efficiently delineated, perhaps adding more complementary views beyond previous WTS-based studies. Thus, our approach may pave the path toward functional mtRNA annotations, and helps to explore more extraordinary complexity of the mitochondrial transcriptional landscape.

Conclusions

In this study, we have established a high-efficient pipeline for targeted sequencing of mtRNAs, which may pave the way toward functional annotation of mtRNAs and mtRNA-based diagnostic and therapeutic strategies in various diseases.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The raw sequencing data underlying this article are available upon request.

Competing Interests

The authors declare that they have no competing interests.

Funding

National Natural Science Foundation of China [grants 82020108023, 81830070], and Autonomous Project of State Key Laboratory of Cancer Biology, China [grants CBSKL2019ZZ06, CBSKL2019ZZ27].

Authors’ Contribution

Z.N.W., K.X.Z., and Q.Y. carried out the sample collection, performed the data analysis, and drafted the manuscript. Q.Y., and F.F.X., collected the public data and participated in the bioinformatics analyses. C.D.B., performed the laboratory experiments. X.E.H., and Y.L., performed the statistical analysis. J.L.X.,
conceived of the study, and participated in its design and coordination and helped to revise the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This study has been supported by National Natural Science Foundation of China [grants 82020108023, 81830070], and Autonomous Project of State Key Laboratory of Cancer Biology, China [grants CBSKL2019ZZ06, CBSKL2019ZZ27]

References


**Figures**
Figure 1

Establishment of the capture-based mtRNA seq. Experimental workflow and optimization of capture-based mtRNA seq.
Stranded sequencing library construction for accurate identification of mitochondrial transcripts. (a) Normalized coverage of mitochondrial H and L strand transcripts from stranded and non-stranded whole transcriptome sequencing (WTS) data. top, stranded WTS; bottom, non-stranded WTS. Functional regions of mitochondrial genome were depicted by H and L strands. (b-c) Spearman correlation of the
mitochondrial gene expressions between non-stranded (vertical axis) and stranded (horizontal axis) for (b) mt-mRNAs, and (c) mt-tRNAs. Red dot, H strand transcripts; Blue dot, L strand transcripts.

Figure 3

Comparison of two transcriptome isolation approaches for mtRNA sequencing. (a-b) Normalized coverage of mitochondrial H and L strand transcripts from WTS data with (a) polyA selection and (b)
rRNA depletion, respectively. Red, H strand transcripts; Blue, L strand transcripts. Functional regions of mitochondrial genome were depicted by H and L strands. (c) Proportion of 12s and 16s rRNA reads in WTS data with polyA selection and rRNA depletion, respectively. Yellow region, 16s rRNA; Orange region, 12s rRNA. (d-e) Spearman correlation analysis of the mitochondrial gene expressions between rRNA depletion (vertical axis) and polyA selection (horizontal axis) groups for (d) mt-mRNAs, and (e) mt-tRNAs. Red dot, H-strand transcripts; Blue dot, L-strand transcripts

Figure 4

Optimal rRNA depletion for efficient profiling of mitochondrial transcriptome. (a) Proportion of mt-rRNA reads among groups with different concentration of rRNA depletion probes. Mean + SD, n=3. (b) Normalized coverage of mt-rRNA transcripts among groups with different concentration of rRNA depletion probes. Functional regions of 12S mt-rRNA and 16S mt-rRNA were positioned (c) Spearman correlation of mt-rRNA coverage between each rRNA depletion group and un-depleted(0X) group.
Figure 5

**Improved characterization of mitochondrial transcriptome by capture-based mtRNA seq.** (a) Coverage depth of mitochondrial H (left) and L (right) strand transcripts per Gb data from CAP and WTS groups. Red, WTS; Blue, CAP. (b) Spearman correlation of proportions (left) and expressions (right) of 37 mtDNA-encoded transcripts between CAP and WTS groups. Left, Pink dot, mt-mRNAs; Blue square, mt-rRNAs; Green triangle, mt-tRNAs. Right, Red dot, H-strand transcripts; Blue dot, L-strand transcripts. (c) Spearman correlation of coverage depth ratio of H and L strand mtRNAs between CAP (vertical axis) and WTS (horizontal axis) groups. (d) Comparison of coverage depth for H strand transcripts (left) and L strand transcripts (right) per Gb data between WTS and CAP groups. (e) The required data volume for CAP- and WTS-based approach at a series of mtRNA coverage depth in both H strand (left) and L strand (right), respectively. (f) The number of precursor types for TI-TQ-TM detected in per Gb data from CAP and WTS groups. **, $P<0.01$; ***, $P<0.001$. Data were compared using paired $t$-test.
Figure 6

**Capture-based mtRNA seq enables comprehensive analysis of mitochondrial transcriptome.** (a-b) The transcription initiation sites identified in H (a) and L strand (b) of three human liver tissue samples. The candidate transcription pausing site of L strand was also labeled. (c-d) The distribution of 5’cleavage sites for (c) H and (d) L strands in the tissue samples. (e-f) The distribution of 3’cleavage sites for (e) H and (f) L strands in the tissue samples. (g-h) Expressions of mt-mRNA (g) sense transcripts and (h) anti-sense mt-
mRNA transcripts from both H and L strand in the tissue samples. Yellow region, transcripts from H strand; Green region, transcripts from L strand.

Figure 7

Efficient application of capture-based mtRNA seq in plasma samples. (a) Normalized coverage of mitochondrial H and L strand transcripts from three plasma samples. (b-c) Expressions of (b) sense
transcripts and (c) anti-sense mt-mRNA transcripts from both H and L strand. (d-e) The distribution of 5' cleavage sites for (d) H and (e) L strands in the plasma samples. (f-g) The distribution of 3' cleavage sites for (f) H and (g) L strands in the plasma samples.

**Supplementary Files**

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

- 3.supplementaryfiles.pdf