
©2013 The Authors

This is the final published version of the article accepted for publication in Cell Reports following peer review. The definitive published version (see citation above) is located on the article abstract page of the publisher, Elsevier.

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-No Derivative Works License, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

This version was made available in the UWA Research Repository on 3 October 2014 in compliance with the publisher’s policies on archiving in institutional repositories.
Mapping of Mitochondrial RNA-Protein Interactions by Digital RNase Footprinting

Ganqiang Liu,1,2 Timothy R. Mercer,1,2 Anne-Marie J. Shearwood,3 Stefan J. Siira,3 Moira E. Hibbs,3 John S. Mattick,1,4 Oliver Rackham,3 and Aleksandra Filipovska1,*

1Garvan Institute of Medical Research, Sydney NSW 2010, Australia
2Institute for Molecular Bioscience, The University of Queensland, Brisbane QLD 4072, Australia
3Western Australian Institute for Medical Research, Centre for Medical Research and School of Chemistry and Biochemistry, The University of Western Australia, Perth WA 6000, Australia
4St Vincent’s Clinical School and the School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney NSW 2052, Australia
*Correspondence: aleksandra.filipovska@uwa.edu.au
http://dx.doi.org/10.1016/j.celrep.2013.09.036

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-No Derivative Works License, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

SUMMARY

Human mitochondrial DNA is transcribed as long polycistronic transcripts that encompass each strand of the genome and are processed subsequently into mature mRNAs, tRNAs, and rRNAs, necessitating widespread posttranscriptional regulation. Here, we establish methods for massively parallel sequencing and analyses of RNase-accessible regions of human mitochondrial RNA and thereby identify specific regions within mitochondrial transcripts that are bound by proteins. This approach provides a range of insights into the contribution of RNA-binding proteins to the regulation of mitochondrial gene expression.

INTRODUCTION

Mitochondria appear to possess very few transcription factors, but rather have a wide range of RNA-binding proteins that control gene expression at the posttranscriptional level (Falkenberg et al., 2007; Rackham et al., 2012). The mitochondrial genome is transcribed as long polycistronic transcripts that are subsequently cleaved and processed into individual mRNAs (Brzezniak et al., 2011; Holzmann et al., 2008; Sanchez et al., 2011; Ojala et al., 1981). Recent sequencing of the human mitochondrial transcriptome has revealed previously unexpected complexities of mitochondrial RNAs and identified new transcripts (Mercer et al., 2011). The large variation observed in mature mitochondrial transcripts has provided evidence that regulation of RNA processing, maturation, stability, translation, and degradation orchestrated by mitochondrial RNA-binding proteins play a major role in mitochondrial gene expression. This indicates that posttranscriptional mechanisms have assumed a pre-eminent role in mammalian mitochondrial gene regulation and expression (Rackham et al., 2012) and that achieving a profile of their interactions with mitochondrial RNAs will provide regulatory insights akin to transcription factor binding in the nuclear genome (Hesselberth et al., 2009). Here, we develop a digital RNase footprinting method that provides a global profile of RNA-protein interactions in human mitochondria and permits the investigation of posttranscriptional regulation.

RESULTS

Identification of Footprints on Mitochondrial RNAs

To identify footprints of proteins bound to mitochondrial RNAs, we treated isolated mitochondrial preparations with three endoribonucleases with different cleavage specificities (Figure 1A): RNase A, which cleaves single-stranded RNA after pyrimidine nucleotides; RNase T1, which cleaves single-stranded RNA after guanine residues; and RNase V1, which is a sequence-independent endonuclease that cleaves double-stranded RNAs. A mock-digested mitochondrial preparation was used as an accompanying control. By comparison to matched control digestions of purified mitochondrial RNA, a RNA footprint detection algorithm identified significant contigous sites with a lower C score than their flanking regions. These represent footprints within mitochondrial transcripts protected from endonuclease cleavage by mitochondrial RNA-binding proteins (Figure S1A).

In total, we identified 88 distinct footprints, 33 within mRNAs, 8 from rRNAs, and 7 within tRNAs. The majority of footprints are in specific regions of mitochondrial mRNAs (Figure 1B). We found that the remaining 40 footprints were in transcription regulatory sites within the D loop and in noncoding transcripts encoded by the light strand of the mitochondrial genome. The positioning of footprints within regions of noncoding transcripts indicate a role for proteins in their biogenesis or degradation (Rackham et al., 2011).
Figure 1. Mapping of Mitochondrial RNA-Protein Interactions and Small RNAs

(A) Schematic illustrating the method for mapping mitochondrial RNA-protein interactions by digital RNase footprinting.

(B) The identified mitochondrial RNA-binding protein footprints are shown in the mitochondrial transcriptome, including those of the TRMT10C protein. The outside and inner circles show mitochondrial genes on the heavy and light strand, respectively.

(legend continued on next page)
The digital RNase footprinting provides a global profile of protein interactions with the mitochondrial transcriptome. To provide insight into the identity of specific proteins interacting with the transcriptome, we performed small interfering RNA (siRNA) knockdown of the mitochondrial RNase P protein 1 (MRPP1), recently renamed tRNA methyltransferase 10C (TRMT10C), one of the components of the RNase P complex required for cleavage of the 5' ends of mitochondrial tRNAs (Holzmann et al., 2008), and determined its depletion in cells to be over 95% by quantitative RT-PCR and immunoblotting (Figure S1B).

To identify TRMT10C footprints, we compared the data sets where TRMT10C was knocked down to data sets where mitochondria were treated with control siRNAs. We identified 17 unique footprints (Table S1) that were specifically absent from TRMT10C-knockdown mitochondrial data sets. These footprints predominantly reside within mitochondrial tRNAs, such as that of the TRMT10C-knockdown mitochondrial data sets. These footprints were specifically absent from TRMT10C-knockdown mitochondrial data sets. These footprints predominantly reside within mitochondrial tRNAs, such as that of the TRMT10C-knockdown mitochondrial data sets. These footprints were specifically absent from the endonuclease-treated samples, finding only a small number of the identified siRNAs produced from the human mitochondrial transcriptome overlap with RNA-binding protein footprints (Figure 1H). This suggests that most siRNAs are free to take part in intermolecular RNA-RNA interactions and may have regulatory roles in mitochondrial gene expression.

**Identification of PPR Protein Footprints**

RNA-binding proteins of the PPR family have emerged as important regulators of organelle gene expression and consequently cell health (Rackham and Filipovska, 2012; Schmitz-Linneweber and Small, 2008). These proteins are similar to the PUF and TALE proteins in that they are predicted to have a sequence-specific mode of nucleic acid recognition (Boch et al., 2009; Filipovska and Rackham, 2012; Moscou and Bogdanove, 2009; Wang et al., 2002). Therefore, there is considerable interest in identifying their RNA targets to understand better their role in RNA metabolism or to use them in biotechnology for targeting specific RNAs of interest. Here, we used RNase digital footprinting to identify binding sites of the mammalian PPR protein, pentatricopeptide repeat domain protein 1 (PTCD1), previously found to affect mitochondrial RNA metabolism (Sanchez et al., 2011; Rackham et al., 2009). We identified five PTCD1-specific RNA footprints in samples from cells where this protein was knocked down using siRNAs, relative to controls. Examples of PTCD1 footprints in mtRNAth and mtRNAde are shown in Figures 2A and 2B. We investigated the interaction between the identified RNA targets with PTCD1 in vitro using an RNA electrophoretic mobility shift assay (EMSA) (Figure 2C). The binding of the identified RNA targets to PTCD1 compared to a scrambled RNA control validated the specificity of this protein for these targets. This suggests that the RNase digital footprinting method combined with specific siRNA knockdown can be used to effectively predict the targets for RNA-binding proteins of interest.

**Protein Footprints upon Stalled Mitochondrial Translation**

Protein translation within mitochondria has diverged significantly from translation in the cytoplasm and prokaryotes (Suzuki et al., 2001a, 2001b) and is still poorly understood. To examine the
translational regulation of mRNAs, we treated cells with chloramphenicol to stall mitochondrial translation and performed RNA footprinting analyses. We compared data sets from mitochondria treated in the presence and absence of chloramphenicol to identify 270 footprints that were protected from endonuclease cleavage as a result of inhibiting translation. We identified 124 protein footprints (from the total of 270) that were within mitochondrial mRNAs, of which 22 were 25–35 nt in length and were recognized as mitochondrial ribosome stalling sites or mitoribosome footprints (Figure 3A). An example of a mitochondrial ribosomal footprint in MT-CO1 mRNA during translational stalling is shown in Figure 3B. The ribosome footprints were not enriched at start codons of mitochondrial mRNAs, suggesting that ribosome stalling induced by chloramphenicol may be

Figure 2. Identification of PTCD1 Footprints on Mitochondrial RNAs
(A and B) Specific PTCD1 protein footprints are shown at the 3’ ends of tRNA^{His} (A) and tRNA^{Ile} (B).
(C) An RNA EMSA shows specific in vitro binding of purified PTCD1 protein with its RNA targets identified in (A) and (B) compared to control RNA.
attributed to the downstream secondary structure. The other 102 protected sites in mitochondrial mRNAs are likely RNA-binding protein footprints that are transiently associated with mRNAs between rounds of translation. The remaining 146 protein footprints from the total 270 identified were in noncoding RNAs: 36 in rRNAs, 10 in tRNAs, and 7 in the D loop, as well as 85 in transcripts encoded by the light strand and 8 in noncoding regions of the heavy strand. Interestingly, we identified 9 ribosomal footprints in transcripts encoded by the light strand, suggesting that mitochondrial ribosomes may indiscriminately scan or initiate translation of RNAs that have cryptic start codons.

To understand the cause for the ribosomal stalling at specific regions in mitochondrial mRNAs, we investigated the correlation between the 22 identified mitoribosome footprints and RNA secondary structure. Regions within mitochondrial transcripts that are immediately upstream and downstream of the ribosomal footprints have low R scores (see Experimental Procedures), suggesting that these regions are less structured, compared to the regions further away from the footprints that are more structured and consequently have higher R scores (Figure 3C). We confirmed that the single-strand RNA structure upstream and downstream of ribosomal footprints is significant using the Wilcoxon test (see Experimental Procedures), suggesting the ribosome prevents folding of RNA immediately upstream and downstream during translation. We also considered whether the 22 identified ribosomal footprints correlate with codon usage by the mitochondrial ribosome (Jia and Higgs, 2008). We selected the lowest frequency codons in each ribosomal footprint with an average frequency of ~0.008 (Figure 3D) and compared these to the frequency of rare codons in 22 randomly selected mRNAs (Experimental Procedures). The median p value is 0.338 (Figure 3E), which suggests that ribosome stalling is not strongly influenced by the frequency of rare codons.

**Digital Mapping of RNA Secondary Structures within Mitochondria**

RNase cleavage coupled with deep sequencing can be used as a high-throughput method for RNA structural mapping (Fourmy and Yoshizawa, 2012). Because RNase V1 preferentially cleaves double-stranded RNAs, we used it to analyze local secondary structures of mitochondrial transcripts in our control data sets. We found that the majority of mapped reads aligned to mitochondrial rRNA genes (92.5% in mitochondrial lysates and 89.2% in purified RNA), consistent with the extensive double-stranded regions within these transcripts. To investigate the regions of mitochondrial transcripts that would be susceptible to RNase V1 cleavage, we assigned an R score to each nucleotide across the entire mitochondrial transcriptome from all data sets and identified the regions with strong secondary structure (Figure 4A; Figure S3). The R score of the mitochondrial rRNAs correlates positively with secondary structure predictions by the Matthews correlation coefficient (MCC) (Table S2; Figures S4 and S5), validating the digital approach to secondary structure prediction.

Next, we determined the average R score for mitochondrial mRNAs and found that their 5’ ends are significantly less structured (Figure 4B), corroborating previous observations (Jones et al., 2008; Montoya et al., 1981). This may facilitate the docking of the mitoribosomes onto mitochondrial mRNAs, as they lack methyl-guanosine caps or 5’ UTRs to recruit them. We investigated the structure of mitochondrial tRNAs, which differ from nuclear or bacterial tRNAs because they often miss semiconserved nucleotides and the size of the D and T loops can be reduced or absent (Giege et al., 2012; Suzuki et al., 2011). Our digital analyses confirm previous findings that the acceptor and anticodon stems are the most structured regions in tRNAs, such as those of tRNA^{Cys} and the anticodon loop is the least structured (Giege et al., 2012) (Figure 4C). Moreover, we have provided a method for investigating the unusual structures of mitochondrial tRNAs at a single-nucleotide level (Figure 4D), which has previously proven challenging.

**DISCUSSION**

The analyses of transcript structure and RNA-protein interactions with single-nucleotide resolution comprises a powerful approach to provide insight into the structures of mitochondrial RNAs and how they are regulated by RNA-binding proteins and recognized by the translation machinery. This is particularly valuable given that mammalian mitochondrial genomes cannot be investigated by targeted genetic manipulation. These methods provide valuable tools for future high-throughput secondary structure prediction of RNAs in cells and will enable the mitochondrial targets that are modified or bound by proteins to be deciphered. Furthermore, given the rapid improvements in next-generation sequencing approaches, the techniques and analyses developed here could be similarly applied to examine the roles of RNA-protein interactions in bacterial and cytoplasmic transcriptomes.

**EXPERIMENTAL PROCEDURES**

**Mitochondrial Isolation and Treatment**

Mitochondria or mitochondrial RNA were isolated from 143B cells, 143B cells treated with nontargeted control siRNAs, TRMT10C siRNAs, or PTCD1 siRNAs as described previously (Sanchez et al., 2011; Mercer et al., 2011). Knockdown of the proteins was confirmed by immunoblotting using TRMT10C, PTCD1, and porin antibodies (Figure S1B). To stall translation, cells were treated in the presence or absence of 100 μg·ml⁻¹ chloramphenicol for 30 min on ice followed by a mitochondrial isolation and purification as above. Purified mitochondrial (2 ng·ml⁻¹) from control cells or cells treated with siRNAs or chloramphenicol were lysed by addition of 200 μl of lysis buffer (100 mM Tris·HCl, 100 mM NaCl, 40 mM MnCl₂, 2 mM dithiothreitol [pH 7.5], 0.1% Triton X-100). For chloramphenicol-treated samples, 100 μg·ml⁻¹ chloramphenicol was included in all buffers prior to RNA isolation. The concentrations of the RNase A (10 U/μl), RNase T1 (0.1 U/μl), or RNase V1 (0.01 U/μl) were optimized and added to each mitochondrial lysate or purified RNA to generate 15–55 nt size fragments. All incubations were carried out at 37°C for up to 30 min and reactions were ended by addition of 700 μl Qiazol, followed by RNA isolation using the miRNeasy Mini Kit (Qiagen).

**Library Construction**

RNA concentration, purity, and integrity were confirmed by BioAnalyzer. The libraries were constructed using the Illumina TruSeq Small RNA Sample Prep Kit and deep sequencing of the mitochondrial small RNAs was performed by GeneWorks on an Illumina GAII according to the manufacturer’s instructions with one modification: sample isolation from the PAGE gel after adaptor ligation was performed with a modified set of size markers to facilitate capture of small RNAs between 15 and 55 nt.

**Mapping Sequenced Reads to Human Genome**

Sequenced reads were initially trimmed of the 3’-adapter sequence (5’-TGGA ATTCTCGGGTGCCAAG-3’). After mapping to the reference human
A

B

C

D

E

(legend on next page)
C score (control) must be greater than the C score (experiment) and the F score model by shuffling the footprints’ loci. For the candidate footprint locus, the of F score between the experimental and control conditions. To estimate if the assigned it as F score (control); second, we divided the F score (control) by the shell, and R scripts to search for footprints. We randomly selected 22 control described previously (Chen et al., 2010). We employed in-house perl, awk

(A) Mitochondrial protein footprints and ribosome stalling sites are shown across the whole transcriptome when mitochondrial translation is inhibited with chloramphenicol. The histogram indicates the number of ribosome stalling sites found in mitochondrial mRNAs correlated to their length.

(B) A specific ribosomal protein footprint in the MT-C01 transcript is shown that occurs as a result of translation inhibition by chloramphenicol treatment. (C) The average R score of 50 nt upstream and downstream of 22 ribosomal stalling sites (left y-axis); the blue dotted line ~0.8444 is the average R score of all mitochondrial mRNA nucleotides in the chloramphenicol-treated data set. The red dotted line shows the average p value of 1,000 times Wilcoxon test (95% confidence level of the interval) to compare the ribosomal footprint R score to random sites and calculated an average p value from the 1,000 tests.

Northern Blotting

Northern blotting of 5S rRNAs was carried out as described before (Pall and Hamilton, 2008), and the 5S rRNAs were detected with biotinylated probes as described before (Rackham et al., 2009).

Identification of Footprints of Mitochondrial RNA-Binding Proteins

We transformed the mapped results to wig format, which represent the value of the 5’ end of sequence read coverage of each nucleotide of the mitochondrial genome. Then, we employed CPMM (counts per million mappers, reads per million in a library) to normalize different sequencing data sets. For each nucleotide, there are four normalized values from mock-treated control, RNase A, RNase T1, and RNase V1 sequencing data sets under the same experimental conditions, respectively. The RNase accessibility of each RNA nucleotide if protected by RNA-binding proteins was quantified according to the C score, the base i of C score is defined as:

\[
C_{\text{score}} = \log_{10} \left( \frac{\max(A_{i-1}, T_{i-1}, V_{i-1}) + 1}{\text{Untreated}_{i} + 1} \right),
\]

where \(A_{i-1}, T_{i-1}, V_{i-1}\) and \(\text{Untreated}_{i}\) are the number of times the nucleotide immediately downstream to the inspected nucleotide was mapped as the first base of a sequence read in RNase V1-, A-, and T1-treated and untreated samples, respectively.

The footprint detection algorithm (Figure S1A) searches continuous nucleotides (8–40 nt), whose C score is lower than the average of left and right flanking nucleotides (3 nt), and we defined as the central or core footprinting region. A central footprinting region is determined according to an F score:

\[
F_{\text{score}} = \frac{10^{C_{\text{max}}} + 10^{C_{\text{min}}}}{10^{C_{\text{max}}}},
\]

where C = the average C score of the central footprinting component, \(L_{i}\) = the average C score of the left flanking component, and \(R_{i}\) = the average C score of the right flanking component. For each footprint, we first calculated its F score\(_{\text{experiment}}\) and the F score for the equivalent control data at the region, assigned it as F score\(_{\text{control}}\). Second, we divided the F score\(_{\text{experiment}}\) by the F score\(_{\text{control}}\) and performed a log2 transformation to yield the fold change of F score between the experimental and control conditions. To estimate if the F score fold change is significant for a footprint, we built an empirical null model by shuffling the footprints’ loci. For the candidate footprint locus, the F score\(_{\text{control}}\) must be greater than the F score\(_{\text{experiment}}\) and the F score fold change is filtered to achieve an expected 5% false discovery rate (FDR) relative to a score obtained by random shuffling of footprints 1,000 times, as described previously (Chen et al., 2010). We employed in-house perl, awk shell, and R scripts to search for footprints. We randomly selected 22 control regions from mitochondrial mRNAs, unrelated to the identified footprints, and shuffled them 1,000 times. For each upstream or downstream nucleotide, we employed the Wilcoxon test (95% confidence level of the interval) to compare the ribosomal footprint R score to random sites and calculated an average p value from the 1,000 tests.

RNA Electrophoretic Mobility Shift Assays

Purified PTCD1 (15 µM or 30 µM) was incubated with 40 nM fluorescein-labeled RNA oligonucleotides (Pharmaco) in 10 mM HEPES (pH 8.0), 1 mM EDTA, 50 mM KCl, 2 mM DTT, 0.1 mg/ml fatty acid-free BSA, and 0.02% Tween-20 at room temperature for 30 min. The following RNA sequences were used: rRNA\(^\text{5S}^\text{rRNA}\): 5’-CAGAGGCUAGCCACCCCUAUUUA-3’; rRNA\(^\text{5S}^\text{scrambled}\): 5’-CUCUAGAUCUAGAUUACCCCGA-3’; rRNA\(^\text{5S}^\text{scrambled}\): 5’-CCGUGCAUCUAGAUUACCCCGA-3’. RNA oligonucleotides alone or rRNA\(^\text{5S}^\text{scrambled}\) RNA oligonucleotides with PTCD1 were used as controls. At the end of the incubation, the reactions were analyzed by 10% PAGE in Tris-acetate-EDTA and fluorescence was detected using a Typhoon FLA 9500 biomolecular imager (GE).

Correlation of R Score and rRNA Secondary Structure

R score was defined to quantify the probability of each nucleotide to form a paired secondary structure. The base i is defined as

\[
R_{\text{score}} = \log_{2} \left( \frac{V_{i} + 1}{\max(A_{i-1}, T_{i-1}) + 1} \right),
\]

where \(V_{i}\) and \(A_{i-1}, T_{i-1}\) are the number of times the nucleotide immediately downstream to the inspected nucleotide was mapped as the first base of a sequence read in RNase V1-, A-, and T1-treated samples, respectively. We adopted the MCC to measure the correlation between the R score and RNA secondary structure. The secondary structure model of human mitochondrial 12S rRNA was derived from the CRW database (Cannone et al., 2002), including +Pseudoknots templates. The secondary structure model of human mitochondrial 16S rRNA was derived from previous studies (Mears et al., 2006; Seibel et al., 2008).

Unpaired predictive value = TP/(TP + FP)

Paired predictive value = TN/(TN + FN)
**Figure 4. In Vivo Analysis of Mitochondrial RNA Secondary Structure**

(A) Histogram of the R score of every nucleotide across the whole mitochondrial transcriptome from control mitochondria and its matched purified RNA data sets. From center, the outer two tracks show the R score of each nucleotide of transcripts on the mitochondrial heavy strand, while the inner two tracks correspond to the light strand. This figure was constructed using Circos (Krzywinski et al., 2009).

(B) R score across the coding sequences (CDS) of mitochondrial mRNA transcripts. Transcripts were aligned by their translational start and stop sites; the horizontal line denotes the average R score of mitochondrial CDS.

(C and D) R score across every nucleotide of tRNA\(^{\text{Val}}\) from control mitochondria (C) and its matched purified RNA data sets and their heat maps (D) showing the specific loops and stems of this tRNA.

**ACCESION NUMBERS**

The Sequence Read Archive (SRA) accession number for the data reported in this paper is SRA098892.

**SUPPLEMENTAL INFORMATION**

Supplemental Information includes five figures and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2013.09.036.
ACKNOWLEDGMENTS

The pETM30 plasmid was a kind gift from the EMBL Protein Expression and Purification Facility. This work was supported by the Australian Research Council (Future Fellowships to A.F. and O.R.) and the National Health and Medical Research Council of Australia (Australia Fellowship 631668 to J.S.M. and Project Grants to A.F. and O.R.).

Received: June 30, 2013
Revised: September 4, 2013
Accepted: September 25, 2013
Published: October 31, 2013

REFERENCES


