Video Article

Simultaneous Mapping and Quantitation of Ribonucleotides in Human Mitochondrial DNA

Katrin Kreisel[1], Martin K.M. Engqvist[1,2], Anders R. Clausen[1]

[1]Department for Medical Biochemistry and Cell Biology, University of Gothenburg

[2]Department of Biology and Biological Engineering, Chalmers University of Technology

Correspondence to: Anders R. Clausen at anders.clausen@gu.se

Abstract

Established approaches to estimate the number of ribonucleotides present in a genome are limited to the quantitation of incorporated ribonucleotides using short synthetic DNA fragments or plasmids as templates and then extrapolating the results to the whole genome. Alternatively, the number of ribonucleotides present in a genome may be estimated using alkaline gels or Southern blots. More recent *in vivo* approaches employ Next-generation sequencing allowing genome-wide mapping of ribonucleotides, providing the position and identity of embedded ribonucleotides. However, they do not allow quantitation of the number of ribonucleotides which are incorporated into a genome. Here we describe how to simultaneously map and quantitate the number of ribonucleotides which are incorporated into human mitochondrial DNA *in vivo* by Next-generation sequencing. We use highly intact DNA and introduce sequence specific double strand breaks by digesting it with an endonuclease, subsequently hydrolyzing incorporated ribonucleotides with alkali. The generated ends are ligated with adapters and these ends are sequenced on a Next-generation sequencing machine. The absolute number of ribonucleotides can be calculated as the number of reads outside the recognition site per average number of reads at the recognition site for the sequence specific endonuclease. This protocol may also be utilized to map and quantitate free nicks in DNA and allows adaption to map other DNA lesions that can be processed to 5´-OH ends or 5´-phosphate ends. Furthermore, this method can be applied to any organism, given that a suitable reference genome is available. This protocol therefore provides an important tool to study DNA replication, 5´-end processing, DNA damage, and DNA repair.

Video Link

The video component of this article can be found at https://www.jove.com/video/56551/

Introduction

In a eukaryotic cell, the concentration of ribonucleotides (rNTPs) is much higher than the concentration of deoxyribonucleotides (dNTPs)[1]. DNA polymerases discriminate against ribonucleotides, but this discrimination is not perfect and, as a consequence, ribonucleotides instead of deoxyribonucleotides may be incorporated into genomes during DNA replication. Ribonucleotides may be the most common non-canonical nucleotides incorporated into the genome[2]. Most of these ribonucleotides are removed during Okazaki fragment maturation by RNase H2 initiated ribonucleotide excision repair (RER) or by Topoisomerase 1 (reviewed in reference[3]). Ribonucleotides that cannot be removed stay stably incorporated in the DNA[2,4] and may affect it in both harmful and beneficial ways (reviewed in reviewed[5]). Besides being able to act as positive signals, for example in mating type switch in *Schizosaccharomyces pombe*[6] and marking the nascent DNA strand during mismatch repair (MMR)[7,8], ribonucleotides affect the structure[9] and stability of the surrounding DNA due to the 2´-hydroxyl group of their ribose[10], resulting in replicative stress and genome instability[11]. The abundance of ribonucleotides in genomic DNA (gDNA) and their relevance in replication and repair mechanisms, as well as the implications for genome stability, give reason to investigate their precise occurrence and frequency in a genome-wide manner.

RNase H2 activity has not been found in human mitochondria and ribonucleotides are therefore not efficiently removed in mitochondrial DNA (mtDNA). Several pathways are involved in the supply of nucleotides to human mitochondria and to investigate whether disturbances in the mitochondrial nucleotide pool cause an elevated number of ribonucleotides in human mtDNA, we developed a protocol to map and quantitate these ribonucleotides in human mtDNA isolated from fibroblasts, HeLa cells, and patient cell lines[12].

Most *in vitro* approaches (reviewed in reviewed[13]) to determine DNA polymerases' selectivity against rNTPs are based on single ribonucleotide insertion or primer extension experiments where competing rNTPs are included in the reaction mix, allowing the identification or relative quantitation of ribonucleotide incorporation in short DNA templates. Quantitative approaches on short sequences may not reflect dNTP and rNTP pools at cellular concentrations and therefore provide insight into polymerase selectivity but are of limited significance regarding whole genomes. It has been shown that the relative amount of ribonucleotides incorporated during the replication of a longer DNA template, such as a plasmid, can be visualized on a sequencing gel using radiolabeled dNTPs and hydrolyzing the DNA in an alkaline milieu[14]. Furthermore, gDNA has been analyzed on Southern blots following alkaline hydrolysis, allowing strand-specific probing and determination of absolute rates of ribonucleotide

incorporation *in vivo*[15]. These approaches allow a relative comparison of incorporation frequency but deliver no insight into the position or identity of the incorporated ribonucleotides. More recent approaches to analyze the ribonucleotide content in gDNA *in vivo*, like HydEn-Seq[16], Ribose-Seq[17], Pu-Seq[18], or emRiboSeq[19], take advantage of the embedded ribonucleotides' sensitivity to alkaline or RNase H2 treatment, respectively, and employ Next-generation sequencing to identify ribonucleotides genome-wide. These methods do not provide insight into the absolute incorporation frequency of the detected ribonucleotides. By adding the step of sequence specific enzymatic cleavage to the HydEn-seq protocol, the method we describe here conveniently extends the information gained from a sequencing approach, allowing simultaneous mapping and quantitation of embedded ribonucleotides[12]. This method is applicable to virtually any organism given that highly intact DNA extracts can be generated and a suitable reference genome is available. The method could be adapted to quantitate and determine the location of any lesion that can be digested by a nuclease and leaves a 5′-phosphate or a 5′-OH end.

To map and quantitate ribonucleotides in genomic DNA, the method combines cleavage by a sequence specific endonuclease and alkaline hydrolysis generating 5′-phosphate ends at sites where the specific recognition sequence for the endonuclease is located and 5′-OH ends at positions where ribonucleotides were located. Since the generated free ends are subsequently ligated with adapters and sequenced using Next-generation sequencing, it is of importance to use highly intact DNA and avoid random fragmentation during DNA extraction and library preparation. Assessing these reads normalized to the reads at the endonuclease cleavage sites allows a simultaneous quantitation and mapping of the detected ribonucleotides. Free 5′-ends are detected in control experiments where the alkaline hydrolysis of DNA is replaced by treatment with KCl. The acquired data provide insight into ribonucleotide location and quantity and allows analyses with respect to ribonucleotide content and incorporation frequency.

Protocol

This protocol is outlined in **Figure 1** and includes the isolation of gDNA, digestion with restriction enzymes to be able to quantitate the number of ribonucleotides, treatment with alkali to hydrolyze the phosphodiester bonds of ribonucleotides incorporated into the gDNA, phosphorylation of free 5′-OH ends, ssDNA ligation of adapters, second strand synthesis, and PCR amplification before sequencing.

1. Adapters and Index Primers

1. Obtain ARC49, ARC140 oligonucleotides, ARC76/77, adapter and ARC78-ARC107 index primers (see **Table 1**).
   NOTE: Oligonucleotides should be HPLC purified. ARC76/77 are ordered as duplex.
2. Prepare 100 µM stock solutions of each oligonucleotide in Tris-EDTA (TE) buffer (see the **Table of Materials**) and store at -20 °C.
3. Prepare 10 µM solutions of ARC67/77 and 2 µM solutions of ARC49 and index primers by diluting in elution buffer (EB; see the **Table of Materials**). Store at -20 °C.

2. Growth and Harvest of Cells

1. Grow HeLa cells in 70 mL Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum in a 250 mL Spinner flask at 37 °C.
2. Count the number of cells and collect 5×10^6 cells in a 50-mL tube, centrifuge for 5 min at 200 x g, and discard the supernatant.
3. Wash the cells with 20 mL 1x PBS, centrifuge for 5 min at 200 x g, and discard the supernatant.
4. Freeze the pellets at -20 °C or continue with DNA purification.

3. DNA Purification and Quantitation

1. **Purify gDNA using phenol-chloroform extraction as described below.**
   1. Resuspend the cells in 2 mL lysis buffer (see the **Table of Materials**) and incubate for 30 min at 42 °C on a heating block.
      CAUTION: Lysis buffer contains hazardous components. SDS solution is irritating, Proteinase K is sensitizing, irritating, and toxic. Wear protective clothes and gloves.
   2. Split the sample in two 2 mL tubes and add 1 volume (V) of phenol-chloroform-isoamyl alcohol (25:24:1).
      CAUTION: Phenol-chloroform-isoamyl alcohol is toxic, mutagenic, corrosive, and hazardous to aquatic environments. Use in a fume hood, wear protective clothes and gloves, and discard in special phenol-chloroform waste.
   3. Mix by inversion for 30-60 s and centrifuge for 5 min at 15,000 x g at room temperature.
      NOTE: Do not vortex DNA to avoid introducing random strand breaks, which would distort the results.
   4. Transfer the upper, aqueous phase to a new 2 mL tube and add 1 V of phenol-chloroform-isoamyl alcohol (25:24:1).
   5. Mix by inversion and centrifuge for 5 min at 15,000 x g at 4 °C.
   6. Transfer upper, aqueous phase to a new 2 mL tube and add 20 µL NaCl (5 M) and 1 V of cold isopropanol.
      CAUTION: Isopropanol is flammable, irritating, and toxic. Store it in a ventilated cabinet, wear protective clothes and gloves, and keep it away from flames.
   7. Mix by inversion and incubate for at least 1 h at -20 °C.
   8. Centrifuge for 20 min at 15,000 x g, 4 °C and discard supernatant.
   9. Wash DNA pellet with 200 µL cold 70% ethanol, centrifuge for 20 min at 15,000 x g at 4 °C and discard supernatant.
      CAUTION: 70% ethanol is flammable and irritating. Keep working solution at -20 °C, otherwise store in ventilated cabinet, wear protective clothes and gloves, and keep it away from flames.
   10. Dry the DNA pellet at room temperature for 20-25 min.
   11. Dissolve DNA pellets in 100 µL TE buffer and pool the samples in one tube.

2. Quantitate DNA concentration using a dsDNA quantitation reagent according to manufacturer's specifications (see the **Table of Materials**).
   NOTE: Use a dsDNA quantitation reagent, because spectrophotometric DNA quantitation can be affected by residual phenol.

3. Store DNA at -20 °C or continue with HincII treatment.

4. HincII Treatment and Alkaline Hydrolysis

1. Digest 1 µg of DNA in a reaction mix containing 5 µL 10x buffer 3.1, 1 µL (10 U) HincII, and nuclease-free H_2O to a final volume of 50 µL.
   NOTE: To achieve optimal conditions for ligation, second strand synthesis and PCR amplification, it may be necessary to increase the amount of input DNA if it is expected that the DNA contains a very low number of ribonucleotides. Similarly, it may be necessary to decrease input DNA if the number of ribonucleotides is very high.
2. Incubate for 30 min at 37 °C.
3. **Purify HincII treated DNA with paramagnetic beads.**
   NOTE: Keep the tube lids open in the following steps to not disturb pellets by opening the tubes.
   1. Add 1.8 V of paramagnetic beads to each sample, carefully mix by pipetting, and incubate at room temperature for 10 min.
   2. Use a magnetic rack to pellet the beads for 5 min, then remove and discard supernatant.
   3. Wash the pellet with 150 µL of 70% ethanol (room temperature) for about 30 s then remove and discard the supernatant.
   4. Wash the pellet with 200 µL of 70% ethanol (room temperature) for about 30 s then remove and discard the supernatant.
      NOTE: Residual ethanol can be removed with a 10 µL pipette. Droplets can be spun down briefly beforehand.
   5. Dry the samples at room temperature for around 15-20 min.
      NOTE: The exact time depends on the volume of beads and the shape of the pellet, therefore the pellets should be checked visually.
   6. Remove tubes from the magnetic rack and elute pellet in 45 µL EB, mix by pipetting carefully.
   7. Incubate for 5 min then pellet the beads on the magnetic rack and use 45 µL of purified DNA in step 4.4.
4. Add 5 µL of KOH (3 M) or KCl (3 M) to the DNA creating a total volume of 50 µL.
   CAUTION: 3 M KOH solution is corrosive. Wear protective clothes and gloves.
5. Incubate for 2 h at 55 °C in a hybridization oven followed by 5 min on ice.
   NOTE: It is recommended to perform the KOH treatment in an oven rather than a heating block to maintain a uniform heating of the tube and prevent condensation at the lid.
6. Precipitate DNA by adding 10 µL sodium acetate (3 M, pH = 5.2) and 125 µL cold 100% ethanol. Incubate on ice for 5 min.
   CAUTION: 100% ethanol is flammable and irritating. Store in a ventilated cabinet, wear protective clothes and gloves, and keep away from flames.
7. Pellet gDNA by centrifuging at 21,000 x g, 4 °C for 5 min and discard the supernatant.
8. Wash DNA pellet with 250 µL cold 70% EtOH, centrifuge at 21,000 x g, 4 °C for 5 min, and discard the supernatant.
   NOTE: To remove droplets, the tube can be spun down briefly again and supernatant can be removed with a 10 µl pipette.
9. Let the pellet dry in an open tube for about 5-10 min until any visible fluid has evaporated.
10. Let DNA pellet dissolve in 20 µL EB for 30 min at room temperature.

5. 5´ End Phosphorylation

1. Prepare the reaction mix for each sample in advance consisting of 2.5 µL 10x T4 polynucleotide kinase reaction buffer, 1 µL (10 U) 3´-phosphatase-minus T4 polynucleotide kinase, and 2.5 µL ATP (10 mM).
2. Transfer 19 µL of each DNA sample into a new 200 µL tube and denature for 3 min at 85 °C in a thermo-cycler.
3. Cool DNA samples on ice and add 6 µL of reaction mix to each sample.
4. Incubate reaction mixes at 37 °C for 30 min and stop the reaction by incubating the samples at 65 °C for 20 min.
5. Purify DNA as described in 4.3, using 1.8 V of paramagnetic beads but elute in 14 µL EB.

6. ssDNA Ligation

1. Prepare the reaction mix for each sample in advance consisting of 0.5 µL ATP (2 mM), 5 µL 10x T4 RNA ligase reaction buffer, 5 µL $CoCl_3(NH_3)_6$ (10 mM), 0.5 µL ARC140 (100 µM), and 25 µL 50% PEG 8000. Mix well by pipetting.
   CAUTION: $CoCl_3(NH_3)_6$ is carcinogenic, sensitizing, and hazardous to aquatic environment. Wear protective clothes and gloves.
2. Transfer 13 µL of purified DNA from step 5.5 to a new 200 µL tube and denature for 3 min at 85 °C in a thermo-cycler.
3. Cool the DNA on ice and add 36 µL of reaction mix to each sample, mix by pipetting, and spin down briefly.
4. Add 1 µL (10 U) of T4 RNA Ligase to each reaction, mix by pipetting, and spin down briefly.
5. Incubate the samples at room temperature in the dark overnight.

7. Second-strand Synthesis

1. Purify ligated DNA as described in 4.3, but use 0.8 V of paramagnetic beads, pellet the beads for 10 min and elute in 20 µL EB.
   NOTE: Due to the higher viscosity of the ligation reaction mix, the first pelleting step is prolonged.
2. Transfer 20 µL of DNA sample to a new 200 µL PCR tube. Repeat the purification step using 0.8 V of paramagnetic beads following the manufacturer's specifications and elute in 14 µL EB.
3. Prepare the reaction mix for each sample in advance consisting of 2 µL of 10x T7 DNA polymerase reaction buffer, 2 µL ARC76/77 (2 µM), 2 µL dNTPs (2 mM), and 0.8 µL BSA (1 mg/mL).
4. Transfer 12.8 µL purified DNA to a new 200 µL tube, denature for 3 min at 85 °C in a thermo-cycler.
5. Cool the DNA on ice and add 6.8 µL of reaction mix to each sample, mix by pipetting, spin down briefly, and incubate for 5 min at room temperature.
6. Add 0.4 µL (4 U) T7 DNA polymerase to each reaction and incubate for 5 min at room temperature.
7. Purify DNA as described in 4.3, using 0.8 V of paramagnetic beads and elute in 11 µL EB.

8. PCR Amplification and Library Quantitation

1. Prepare the reaction mix for each sample in a new 200 µL tube in advance consisting of 7.5 µL ARC49 (2 µM), 7.5 µL index primer (2 µM, unique for each sample), and 25 µL 2x hot start ready mix.
2. Add 10 µL of DNA sample to each reaction. Amplify the library using the following conditions: denature at 95 °C for 45 s, followed by 18 cycles of 98 °C for 15 s, 65 °C for 30 s, 72 °C for 30 s, ending with a final elongation at 72 °C for 2 min. Hold samples at 4 °C after amplification.
3. Purify libraries as described in 4.3, using 0.8 V of paramagnetic beads, and elute in 20 µL TE buffer.
4. Quantitate libraries using a dsDNA quantitation reagent, according to the manufacturer's specifications (see the **Table of Materials**).
5. Store samples at -20 °C or continue with library analysis.

9. Library Analysis and Pooling

1. Determine the quality of each library and estimate the average fragment size using a digital electrophoresis system.
   NOTE: The average fragment size is assessed by estimating where the area under the curve of the electropherogram is halved, disregarding peaks from markers. Representative results of suitable library profiles after KOH or KCl treatment are given in **Figure 2A**.
2. Calculate the concentration (nM) of the libraries as:
   $(c/10^3)/(p*650)]*10^9$
   where c is the concentration of the library in ng/µL and p is the average fragment size in bp, as estimated in 9.1.
3. **Pool equal molar amounts of up to 24 libraries amplified with different index primers for sequencing. Add TE buffer to a final volume of 25 µL and concentration of 10 nM.**
   NOTE: Depending on the number of libraries to be pooled, the amount of DNA from each library is adjusted. If primer dimers were detected in step 9.1 as a distinct peak of about 130 bp, the final volume of the library pool can exceed 25 µL, because the purification step is repeated as described in 4.3, using 0.8 V of paramagnetic beads, and DNA is eluted in 25 µL TE buffer.
   1. Determine the new library pool concentration using a dsDNA quantitation reagent according to manufacturer's specifications and the average peak size as described above. Proceed to sequencing and data analysis (sections 10 and 11).

10. Sequencing

1. Perform 75-base paired-end sequencing on pooled libraries[12].

11. Data Analysis

1. Trim all reads to remove adapter sequences, filter for quality and read length.
   NOTE: This can be done using cutadapt 1.2.1[20] with the command `cutadapt -f fastq --match-read-wildcards --quiet -m 15 -q 10 -a NNNNNNN <FILE>`, where NNNNNNN is replaced with the actual adapter sequence and <FILE> is replaced with the fastq file name.
2. Remove mates of reads that were discarded in the previous step using custom scripts.
3. Align Mate 1 of remaining pairs to an index containing the sequence of all oligonucleotides used in the library preparation (*e.g.,* using Bowtie 0.12.8[21] and the command line options -m1 -v2). Discard all pairs with successful alignments.
4. Align remaining pairs to the organism reference genome using Bowtie with the command line options -v2 -X10000--best.
5. Map reads that span between the mitochondrial molecule beginning and end by aligning Mate 1 of all unaligned pairs (using Bowtie with the command line options -v2).
6. Determine the count of 5´-ends for all single and paired end alignments. Shift the position of these by one base upstream to the position where the hydrolyzed ribonucleotides were.
7. Export data from the bowtie file format to a bedgraph file format using custom scripts for visualization in common genome browsers. Normalize the reads for each strand to reads per million.
8. Using the position and counts from the bedgraph file, reference the organism genome sequence to determine the identity of incorporated ribonucleotides.
   NOTE: For the human mitochondrial genome reads from the regions 16,200-300 and 5,747-5,847 for each strand should be excluded since these regions contain many free 5´-ends unrelated to ribonucleotide incorporation by DNA polymerase γ.
9. Divide the total reads, not including reads at the eleven HincII sites, with the mean number of reads per HincII site to get the number of ribonucleotides per single strand break, (*i.e.* the number of ribonucleotides per mitochondrial molecule).

Representative Results

Illustrating the methodology described above, representative data were generated analyzing human mitochondrial DNA from HeLa cells[12].
Figure 2B shows the summarized reads at all HincII sites in heavy (HS) and light strand (LS) of human mtDNA after KCl treatment (left panels).
Around 70% of all detected 5′-ends localize to the cut-sites, demonstrating the high efficiency of the HincII digestion. Treating libraries with KOH
to hydrolyze the DNA at embedded ribonucleotides decreases the number of reads at HincII sites to about 40% (**Figure 2B**, right panels). This is
expected since large numbers of 5′-ends are generated at the sites of ribonucleotide incorporation, and is indicative of a sufficient library quality.
Figure 2C illustrates the localization and frequency of 5′-ends (green) after KCl treatment and reads generated by HydEn-seq (magenta) after
KOH treatment, detecting both free 5′-ends and ends generated at ribonucleotides by alkaline hydrolysis. Free 5′-ends and ribonucleotides
localizing to the HS of human mtDNA are shown in the left panel and those localizing to the LS are shown in the right panel. The relative
numbers of raw reads at ribonucleotides (**Figure 2D**, upper panel) or HincII sites (lower panel) on HS and LS of mtDNA show, respectively, a 14-
fold or 31-fold stronger coverage of the LS relative to the HS, while a similar bias was not observed for nuclear DNA. This strand bias may be
explained by the distinct difference in base composition of the two strands and illustrates the importance of the normalization to reads at HincII
sites.

Normalizing read counts to HincII gives a quantitative measure of the number of ribonucleotides per mitochondrial genome (**Figure 3A**). As
illustrated in **Figure 3B**, the reads after KOH treatment for each ribonucleotide normalized to the sequence composition of each strand show a
ratio different than 1, indicating a non-random distribution of reads suggesting a distinct ribonucleotide pattern and a high library quality. That
ratio is unaffected by previous digestion with HincII, verifying the enzyme's cleavage specificity. Normalizing the reads at the sites of embedded
ribonucleotides to those at HincII cleavage sites, as well as to the genome nucleotide content, generates a quantitative measure of how many of
each ribonucleotide are incorporated per 1,000 complementary bases (**Figure 3C**).
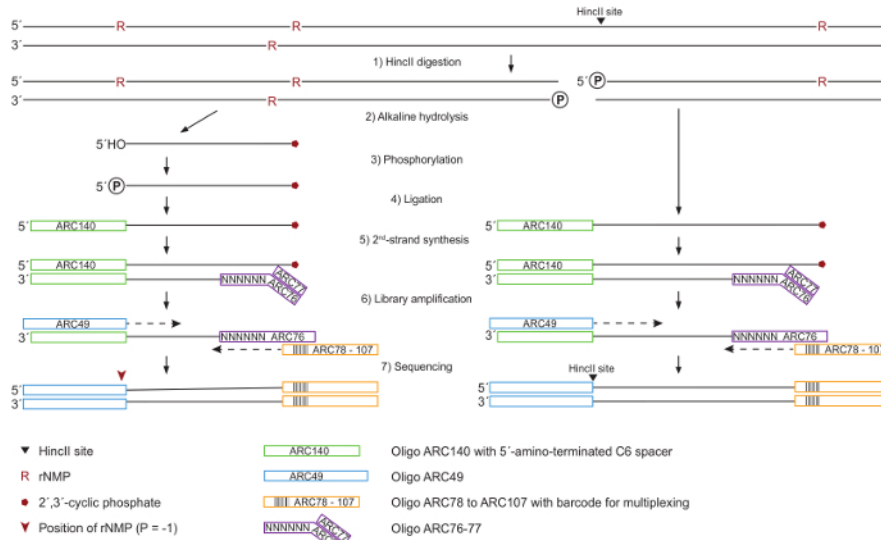

Figure 1: Schematic for DNA Processing and Library Preparation. (1) Whole genomic DNA is cleaved by HincII for normalization in
the subsequent quantitation of ribonucleotides, generating blunt ends at HincII sites (black arrowhead). (2) The DNA is treated with KOH
to hydrolyze at ribonucleotide sites, leading to 2′,3′-cyclic phosphate (red pentagon) at 3′-ends and free 5′-OH ends. (3) 5′-OH ends are
phosphorylated by T4 Polynucleotide Kinase 3′-phosphatase-minus. (4) All 5′-ends carrying a phosphate group are ligated to the ARC140
oligonucleotide by T4 RNA ligase. (5) The second strand is synthesized using T7 DNA Polymerase and the ARC76-77 oligonucleotides
containing random N_6 sequences. (6) The library is amplified by a high-fidelity DNA Polymerase using ARC49 and one of the ARC78 to ARC107
index primers containing a unique barcode for multiplexing. (7) 5′-ends are located by paired-end sequencing. Please click here to view a larger
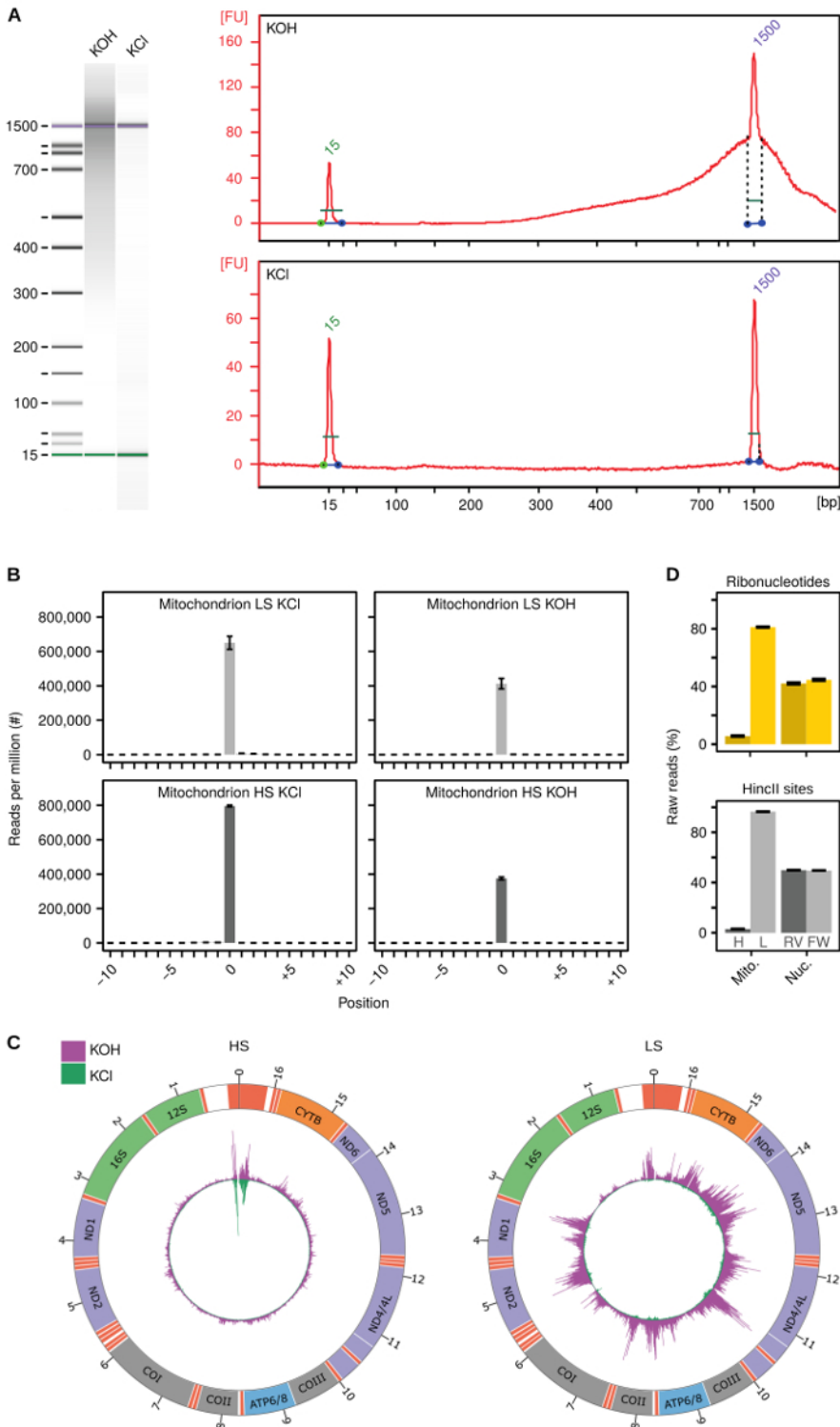version of this figure.

Figure 2: Method validation. (**A**) Representative electropherograms generated using an automated electrophoresis system to determine the quality of generated libraries treated with KOH or KCl. (**B**) Summarized signal at HincII sites in heavy (HS) and light strand (LS) human mtDNA after KCl (left panels) or KOH (right panels) treatment. (**C**) Circos figure of free 5′-ends (green) and from HydEn-Seq (free 5′-ends and ribonucleotides, magenta) in HS (left panel) and LS (right panel) human mtDNA. Peaks are normalized to per million reads and the maximum peak is adjusted to the maximum number of reads of the HydEn-seq library. (**D**) Summarized raw reads at ribonucleotides (upper panel) and HincII sites (lower panel) in heavy (H) and light (L) strand in human mtDNA (Mito.) or in reverse (RV) or forward (FW) strand in nuclear (Nuc.) DNA. Figures B, C and D are adapted from reference[12]. Error bars represent the standard error of the mean. Please click here to view a larger version of this figure.
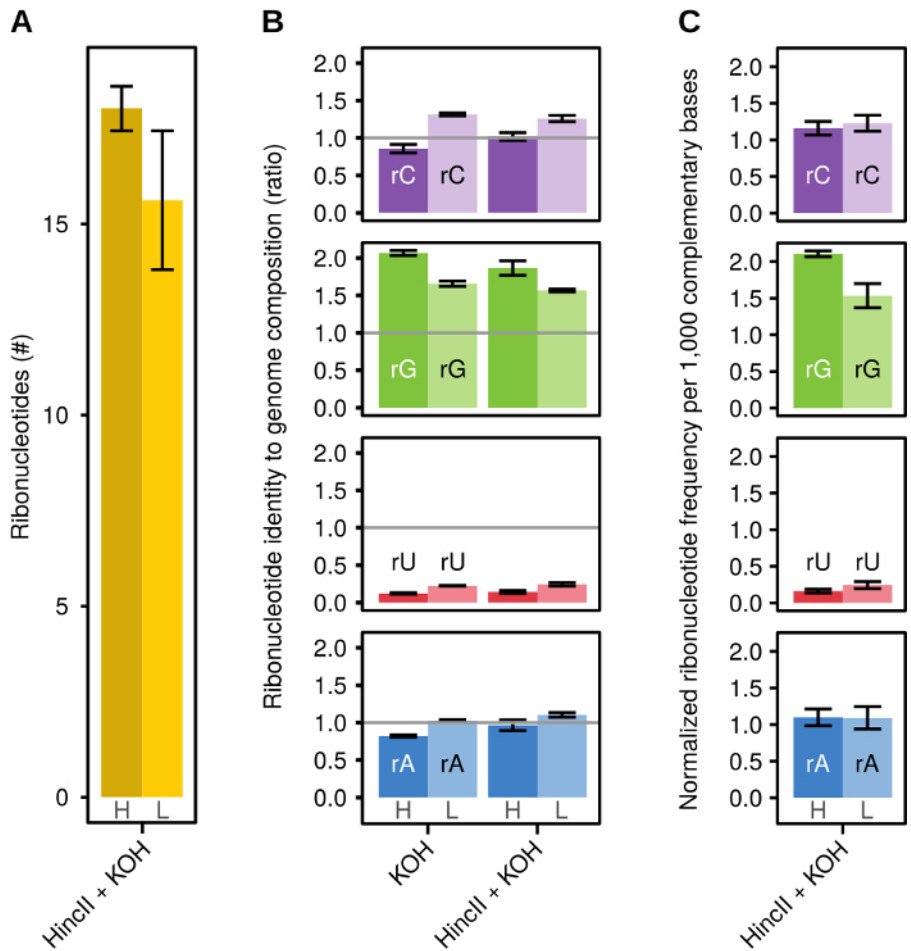
Figure 3: Representative Results. (**A**) The relative number of ribonucleotides normalized to reads at HincII sites for KOH treated libraries on the heavy (H) or light (L) strand. (**B**) Ratio of ribonucleotide identity to mtDNA genome composition for KOH treated (KOH) and HincII cleaved with KOH treated (HincII+KOH) libraries on the heavy (H) or light (L) strand of mtDNA. (**C**) Ribonucleotide frequency normalized to 1,000 complementary bases for HincII and KOH treated libraries on the heavy (H) or light (L) strand of mtDNA. Figures are adapted from reference[12]. Error bars represent the standard error of the mean. Please click here to view a larger version of this figure.

Name	Sequence
ARC49	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
ARC76	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNN*N*N**
ARC77	AGATCGGAAGAGCACACGTCTGAACTCCAGTC*A*C
ARC78	CAAGCAGAAGACGGCATACGAGAT**CGTGAT**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC84	CAAGCAGAAGACGGCATACGAGAT**ACATCG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC85	CAAGCAGAAGACGGCATACGAGAT**GCCTAA**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC86	CAAGCAGAAGACGGCATACGAGAT**TGGTCA**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC87	CAAGCAGAAGACGGCATACGAGAT**CACTGT**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC88	CAAGCAGAAGACGGCATACGAGAT**ATTGGC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC89	CAAGCAGAAGACGGCATACGAGAT**GATCTG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC90	CAAGCAGAAGACGGCATACGAGAT**TCAAGT**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC91	CAAGCAGAAGACGGCATACGAGAT**CTGATC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC93	CAAGCAGAAGACGGCATACGAGAT**AAGCTA**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC94	CAAGCAGAAGACGGCATACGAGAT**GTAGCC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC95	CAAGCAGAAGACGGCATACGAGAT**TACAAG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC96	CAAGCAGAAGACGGCATACGAGAT**TGTTGACT**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC97	CAAGCAGAAGACGGCATACGAGAT**ACGGAACT**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC98	CAAGCAGAAGACGGCATACGAGAT**TCTGACAT**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC99	CAAGCAGAAGACGGCATACGAGAT**CGGGACGG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC100	CAAGCAGAAGACGGCATACGAGAT**GTGCGGAC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC101	CAAGCAGAAGACGGCATACGAGAT**CGTTTCAC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC102	CAAGCAGAAGACGGCATACGAGAT**AAGGCCAC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC103	CAAGCAGAAGACGGCATACGAGAT**TCCGAAAC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC104	CAAGCAGAAGACGGCATACGAGAT**TACGTACG**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC105	CAAGCAGAAGACGGCATACGAGAT**ATCCACTC**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC106	CAAGCAGAAGACGGCATACGAGAT**ATATCAGT**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC107	CAAGCAGAAGACGGCATACGAGAT**AAAGGAAT**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
ARC140	/5AmMC6/ACACTCTTTCCCTACACGACGCTCTTCCGATCT

Table 1: Oligonucleotides. Listed are the oligonucleotides used for HydEn-Seq. Bold face indicates indexing. * indicates a phosphorothioate bond. ARC140 contains a 5´-amino group instead of a 5´-OH group, in combination with a C6 linker. This modification reduces formation of ARC140 concatemers during ligation.

Discussion

Here we present a technique to simultaneously map and quantify ribonucleotides in gDNA, and mtDNA in particular, by the simple introduction of DNA cleavage at sequence specific sites in the genome as an addition to the established HydEn-seq protocol. While this study focuses on human mtDNA, originally the HydEn-seq method was developed in *Saccharomyces cerevisiae*, illustrating the method's translation to other organisms[12,16].

For reliable results obtained from this approach, some critical steps should be noted: (A) Since sequencing adapters ligate to all available 5´-ends, it is crucial to work with highly intact DNA. DNA should be isolated and libraries should be made preferably immediately after DNA isolation, or the DNA can be stored at -20 °C. It is not recommended to store DNA in the fridge for a long time or to repeatedly freeze and thaw it. (B) To generate suitable libraries with this method, it is crucial to perform the KOH treatment of the DNA in an incubation oven, rather than a heating block, assuring homogenous heating of the whole sample and quantitative hydrolysis. (C) Furthermore, it is critical to control the quality of libraries before pooling and sequencing. The DNA should be quantified and analyzed using an automated electrophoresis system to ensure adequate amounts of library DNA, confirm appropriate fragment sizes, and check for primer dimers.

For a meaningful data analysis, it is also important to note that the informative value of this method is dependent on appropriate controls to assess background counts and sequence or strand biases. We routinely achieve a mapping efficiency in KCl samples of close to 70% when only digesting with the sequence specific endonuclease (**Figure 2B**, left panels). In addition, it is important to confirm that the endonuclease treatment is not affecting the overall detection of incorporated ribonucleotides by comparing HincII treated and untreated samples (**Figure 3B**). In these experiments, we have used HincII to introduce site specific cuts, though other high-fidelity restriction enzymes could also be used.

The protocol could be adapted to study other types of DNA lesions that can be processed to 5′-phosphate or 5′-OH ends. The accuracy of the results is dependent on the specificity of processing and requires suitable controls (*e.g.,* wild type or untreated) for verification. Moreover, when adapting this method to other applications or for use with other organisms, one should consider that the method in its current setup requires about 1 µg of DNA which is processed to a library. Since the number of ends is dependent on the number of embedded ribonucleotides, which varies depending on the organism or mutant, samples including a lower number of ribonucleotides would require more input DNA to generate a sufficient number of ends in the subsequent library construction. Similarly, if DNA samples have a much higher number of ribonucleotides, it would also require using less input DNA to obtain optimal conditions for ligation, second strand synthesis, and PCR amplification. It is noteworthy that the library construction as described in this protocol also generated data covering the nuclear genome (as displayed in **Figure 2D**) and only the data analysis was focused on mtDNA. This illustrates that larger genomes with moderately lower ribonucleotide frequencies are also captured by this method.

When considering this method, certain limitations should be taken into account: Although this method should, in theory, be applicable to virtually any organism, a suitable reference genome is necessary for the alignment of reads. Furthermore, the results obtained from our protocol represent the reads from a large number of cells. Specific ribonucleotide incorporation patterns of a subset of cells cannot be identified by this approach. If ribonucleotides are mapped in larger genomes with a very low number of ribonucleotides, it may be challenging to discriminate ribonucleotides from random nicks and appropriate controls are therefore needed.

The method we describe here, extends the available *in vivo* techniques such as HydEn-Seq[16], Ribose-Seq[17], Pu-Seq[18], or emRiboSeq[19]. These approaches take advantage of the embedded ribonucleotides' sensitivity to alkaline or RNase H2 treatment, respectively, employing Next-generation sequencing to identify ribonucleotides genome-wide, which allows their mapping and the comparison of relative incorporation. By cleaving the DNA sequence specifically, as described above, in addition to alkaline hydrolysis at embedded ribonucleotides, the reads for ribonucleotides can be normalized to those cleavage sites, allowing not only the identification and mapping of ribonucleotides, but also their quantitation for each DNA molecule. The application of our technique in the context of diseases related to DNA replication, DNA repair, and TLS could provide a deeper understanding of the role of ribonucleotides in underlying molecular mechanisms and genome integrity in general.

Disclosures

Acknowledgements

References

1. Traut, T. W. Physiological Concentrations of Purines and Pyrimidines. *Mol. Cell. Biochem.* **140**, 1-22 (1994).
2. McElhinny, S. A. N. *et al.* Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc. Natl. Acad. Sci. USA.* **107**, 4949-4954 (2010).
3. Williams, J. S., Lujan, S. A., & Kunkel, T. A. Processing ribonucleotides incorporated during eukaryotic DNA replication. *Nat. Rev. Mol. Cell Biol.* **17**, 350-363 (2016).
4. Clausen, A. R., Zhang, S., Burgers, P. M., Lee, M. Y., & Kunkel, T. A. Ribonucleotide incorporation, proofreading and bypass by human DNA polymerase delta. *DNA Repair.* **12**, 121-127 (2013).
5. Potenski, C. J., & Klein, H. L. How the misincorporation of ribonucleotides into genomic DNA can be both harmful and helpful to cells. *Nucleic Acids Res.* **42**, 10226-U10798 (2014).
6. Vengrova, S., & Dalgaard, J. Z. RNase-sensitive DNA modification(s) initiates S. pombe mating-type switching. *Gene. Dev.* **18**, 794-804 (2004).
7. Lujan, S. A., Williams, J. S., Clausen, A. R., Clark, A. B., & Kunkel, T. A. Ribonucleotides Are Signals for Mismatch Repair of Leading-Strand Replication Errors. *Mol. Cell.* **50**, 437-443 (2013).
8. Ghodgaonkar, M. M. *et al.* Ribonucleotides Misincorporated into DNA Act as Strand-Discrimination Signals in Eukaryotic Mismatch Repair. *Mol. Cell.* **50**, 323-332 (2013).
9. DeRose, E. F., Perera, L., Murray, M. S., Kunkel, T. A., & London, R. E. Solution Structure of the Dickerson DNA Dodecamer Containing a Single Ribonucleotide. *Biochemistry.* **51**, 2407-2416 (2012).
10. Li, Y. F., & Breaker, R. R. Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2 '-hydroxyl group. *J. Am. Chem. Soc.* **121**, 5364-5372 (1999).
11. McElhinny, S. A. N. *et al.* Genome instability due to ribonucleotide incorporation into DNA. *Nat. Chem. Biol.* **6**, 774-781 (2010).
12. Berglund, A. K. *et al.* Nucleotide pools dictate the identity and frequency of ribonucleotide incorporation in mitochondrial DNA. *Plos Genet.* **13** (2017).
13. Brown, J. A., & Suo, Z. C. Unlocking the Sugar "Steric Gate" of DNA Polymerases. *Biochemistry.* **50**, 1135-1142 (2011).
14. Sparks, J. L. *et al.* RNase H2-Initiated Ribonucleotide Excision Repair. *Mol. Cell.* **47**, 980-986 (2012).
15. Miyabe, I., Kunkel, T. A., & Carr, A. M. The Major Roles of DNA Polymerases Epsilon and Delta at the Eukaryotic Replication Fork Are Evolutionarily Conserved. *Plos Genet.* **7** (2011).
16. Clausen, A. R. *et al.* Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation. *Nat. Struct. Mol. Biol.* **22**, 185-191 (2015).
17. Koh, K. D., Balachander, S., Hesselberth, J. R., & Storici, F. Ribose-seq: global mapping of ribonucleotides embedded in genomic DNA. *Nat. Methods.* **12**, 251 (2015).

18. Keszthelyi, A., Daigaku, Y., Ptasinska, K., Miyabe, I., & Carr, A. M. Mapping ribonucleotides in genomic DNA and exploring replication dynamics by polymerase usage sequencing (Pu-seq). *Nat. Protoc.* **10**, 1786-1801 (2015).
19. Ding, J., Taylor, M. S., Jackson, A. P., & Reijns, M. A. M. Genome-wide mapping of embedded ribonucleotides and other noncanonical nucleotides using emRiboSeq and EndoSeq. *Nat. Protoc.* **10**, 1433-1444 (2015).
20. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal.* **17**, 10-12 (2011).
21. Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10** (2009).