CHAPTER ONE

D-Seq: Genome-wide detection of dihydrouridine modifications in RNA

Austin S. Draycott[a], **Cassandra Schaening-Burgos**[b], **Maria F. Rojas-Duran**[a], **and Wendy V. Gilbert**[a,*]

[a]Yale School of Medicine, Department of Molecular Biophysics & Biochemistry, New, Haven, CT, United States
[b]Massachusetts Institute of Technology, Department of Biology, Cambridge, MA, United States
*Corresponding author. e-mail address: wendy.gilbert@yale.edu

Contents

1. Introduction 4
   1.1 RNA modifications impact RNA function 4
   1.2 Methods to detect dihydrouridine 6
2. Step-by-step method details 8
   2.1 Sample preparation 8
3. D-Seq Library preparation 12
   3.1 RNA fragmentation 12
   3.2 Borohydride reduction 12
   3.3 3′ End healing 12
   3.4 RNA size selection 13
   3.5 3′ Adapter ligation 13
   3.6 Reverse transcription 14
   3.7 Size selection 15
   3.8 5′ Adapter ligation 15
   3.9 PCR amplification 16
4. D-seq data analysis 16
   4.1 Trimming of adapter sequences 17
   4.2 PCR duplicate collapsing 18
   4.3 UMI removal 18
   4.4 Read mapping 18
   4.5 Read end position gathering 18
   4.6 D-site identification 18
5. Experimental considerations 19
   5.1 Negative control selection 20
   5.2 Sequencing depth and biological replicates 20

Methods in Enzymology, Volume 692
ISSN 0076-6879, https://doi.org/10.1016/bs.mie.2023.09.001

6. Solutions and reagent recipes 20
6.1 Solutions 20
Acknowledgments 21
References 21

Abstract

In addition to A, C, G and U, RNA contains over 100 additional chemically distinct residues. An abundant modified base frequently found in tRNAs, dihydrouridine (D) has recently been mapped to over 100 positions in mRNAs in yeast and human cells. Multiple highly conserved dihydrouridine synthases associate with and modify mRNA, suggesting there are many D sites yet to be found. Because D alters RNA structure, installation of D in mRNA is likely to effect multiple steps in mRNA metabolism including processing, trafficking, translation, and degradation. Here, we introduce D-seq, a method to chart the D landscape at single nucleotide resolution. The included protocols start with RNA isolation and carry through D-seq library preparation and data analysis. While the protocols below are tailored to map Ds in mRNA, the D-seq method is generalizable to any RNA type of interest, including non-coding RNAs, which have also recently been identified as dihydrouridine synthase targets.

1. Introduction

1.1 RNA modifications impact RNA function

RNA is heavily modified with more than 100 distinct post-transcriptional modifications that affect RNA structure, stability, and function (Boccaletto et al., 2018). The recent development of protocols for transcriptome-wide mapping of some RNA modifications has revealed a complex and dynamic mRNA 'epitranscriptome' that includes N6-methyladenosine (m^6A), 5-methylcytidine (m^5C), inosine (I), pseudouridine (Ψ), 5-hydroxymethylcytidine (hm^5C), N1-methyladenosine (m^1A), N4-acetylcytidine and 2′-*O*-methylribose (Nm) (Arango et al., 2018; Gilbert, Bell, & Schaening, 2016; Li, Xiong, & Yi, 2016; Schaefer, Kapoor, & Jantsch, 2017). Emerging evidence suggests that these RNA modifications have multiple effects on mRNA, altering structure, trafficking, translation, splicing and degradation patterns of mRNAs (Wang et al., 2014). In addition to their functions in mRNA, RNA modifications in non-coding RNAs are known to impact RNA function. The ubiquitous tRNA modification dihydrouridine (D) is found at many positions in tRNAs from organisms of all domains of life, and profoundly impacts tRNA folding (Holley et al., 1965). Four eukaryotic DUS enzymes install D at positions within the D and

variable loops of tRNAs. (Rider, Ottosen, Gattis, & Palfey, 2009; Xing, Hiley, Hughes, & Phizicky, 2004). The recent development of transcriptome scale methods for D profiling revealed that at least one DUS enzyme modifies mRNA, and that D is a new component of the mRNA 'epitranscriptome' (Draycott et al., 2022; Finet et al., 2021).

Although the functions and consequences of dihydrouridylation in mRNA are still unclear, D could have multiple impacts on RNA metabolism. DUS catalyze a reduction of the C5-C6 double bond in uridine to form D (Fig. 1A), which has multiple effects on RNA structure. D subtly distorts the pyrimidine ring (Emerson & Sundaralingam, 1980) causing destacking of bases in a D-containing RNA (Dalluge, Hashizume, Sopchik, McCloskey, & Davis, 1996). These changes to the RNA backbone conformation disrupt normal A-form RNA helical geometry and allow for greater flexibility in RNAs. D nucleotides antagonize formation of RNA

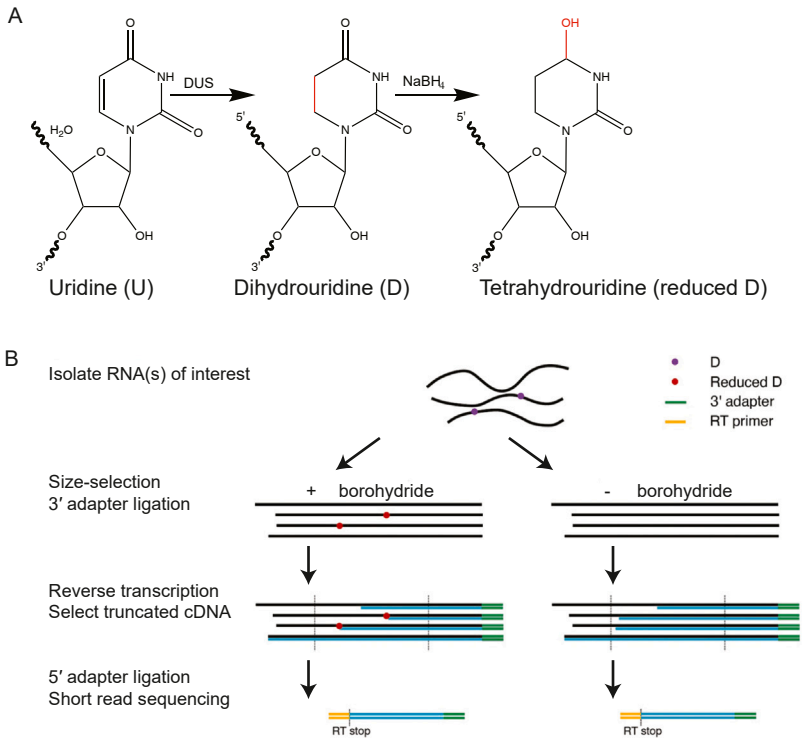

Fig. 1 (A) Structures of uridine, dihydrouridine and tetrahydrouridine. (B) Schematic of D-seq library construction.

duplexes, which are crucial for many aspects of mRNA metabolism. RNA secondary structures have been found to affect the efficiency and regulation of translation initiation, alternative splicing, RNA localization and RNA stability (Gilbert et al., 2016; Martinez & Gilbert, 2018; Roundtree, Evans, Pan, & He, 2017). D also substantially destabilizes the typical C3′ endo conformation of the ribose sugar, and instead favors the C2′ endo conformation in a D nucleotide (Dalluge et al., 1996). Because of this effect, D is expected to perturb binding of numerous regulatory RNA binding proteins by favoring the C2′ endo conformation that is preferentially bound by K homology (KH) domains and RNA recognition motifs (RRM) (Kligun & Mandel-Gutfreund, 2015). KH and RRM domains are responsible for sequence-specific binding of hundreds of human proteins that regulate all aspects of mRNA processing and function.

Installation of D in mRNA has broad potential to impact RNA metabolism. Emerging evidence suggests important roles for dysregulation of RNA modifications in stress response, cancer progression, and viral infection (Daniloski et al., 2021; Kato et al., 2005; Niu et al., 2021). The D-seq method presented here will enable future work to determine how alterations in the D-landscape impacts these diverse cellular processes.

1.2 Methods to detect dihydrouridine

D was initially discovered in yeast $tRNA^{Ala}$ in 1965 by unexpected migration of fragments of RNase T1 digests of yeast $tRNA^{Ala}$. The resolution and throughput of D mapping methods was subsequently substantially increased by Feng Xing and Eric Phizicky, who developed a primer extension method to map D (Xing et al., 2004). The Xing method works because dihydrouridine is susceptible to alkaline hydrolysis in basic conditions. Alkaline hydrolysis of D results in the cleavage of the 3–4 linkage of D ring. The product of the ring-opening reaction, β-ureidopropionic acid loses the capacity for base pairing. The damaged D site then causes a highly penetrant reverse transcriptase stop during library preparation. While alkaline hydrolysis triggers robust RT stops at D, it is not compatible with RNAs other than highly structured and stable RNAs such as tRNA. Instead of relying on alkaline hydrolysis, the D-seq method presented here relies on reduction of D with sodium borohydride to generate RT stops (Fig. 1B). Borohydride treatment does not require harsh pH or conditions and is compatible with any RNA type.

In this chapter, we provide instructions for performing D-seq, a method to find genome-wide (Draycott et al., 2022). D-seq uses borohydride reduction

of D to trigger RT-stops, which can then be mapped by high-throughput sequencing. In this protocol, we provide methods to generate D-seq data from any RNA type. While the protocols presented below are optimized for *Saccharomyces cerevisiae*, they can easily be adapted to other model systems.

Key resources table

Note that not all areas will be used in every protocol.

Reagent or Resource	Source	Identifier
Chemicals and reagents		
Oligo (dT) cellulose beads	NEB	S1408
Dynabeads oligo$(dT)_{25}$	Thermo Fisher	61002
SuperScript III	Invitrogen	18080093
Plastics		
MaXtract high density	Qiagen	129065
RNeasy MinElute columns	Qiagen	74204
Dynabeads MyOne silane	Thermo Fisher	37002D
Experimental models: Organisms/strains		
BY4741	Eric Phizicky	
Oligonucleotides		
Pre-adenylated 3′ adapter: 5′-adenyl-GGAATTCTCGGGTGCCAAGG-dideoxycytidine-3′	IDT	
RT primer, oWG915: 3′-CCTTAAGAGCCCACGGTTCCG-5′	IDT	
5′ Adapter, oWG920: 3′-C3 Spacer-GTGCAAGTCTCAAGATGTCAGGCT CGATCNNNNNNNNN-phosphate-5′	IDT	
Library PCR reverse primer: 5′-AATGATACGGCGACCACCG AGATCTACACGTTCAGAGTT CTACAGTCCGA-3′	IDT	
Barcoded Library PCR forward primer: 3′-ACCTTAAGAGCCCACGGTTCCTTGAG GTCAGTGXXXXXXTAGAGCATACGGC AGAAGACGAAC-5′ where XXXXXX = unique barcode sequence	IDT	

Software and algorithms	
BBmap	https://sourceforge.net/projects/bbmap/
bedtools	https://github.com/arq5x/bedtools2
Python	https://www.python.org/downloads/

2. Step-by-step method details

2.1 Sample preparation

The first step in D-seq library preparation is the purification of total RNA from a biological sample, followed by isolation of the RNA type of interest (rRNA, tRNA, mRNA or ncRNA). The methods employed to purify total RNA will be determined by the type and origin of the sample, and the type of RNA of interest. Here, we provide a protocol for the isolation of total RNA from exponentially growing cultures of S. cerevisiae using hot acid phenol (Collart & Oliviero, 2001). Subsequent to RNA isolation, we also provide protocols for isolation of polyadenylated RNAs using oligo (dT) cellulose (Carlile, Rojas-Duran, & Gilbert, 2015) or magnetic beads (Green & Sambrook, 2019), or tRNAs using size selection on silica columns.

2.1.1 RNA isolation from S. cerevisiae

Inoculate a 10-mL starter culture in YPAD and grow overnight at 30 °C. Use this starter to inoculate 700 mL of YPAD in a baffled flask to an OD 600 of 0.05, and grow to mid-log phase. Harvest cells by centrifugation at 16,000×*g* for 5 min at 4 °C. Resuspend the cell pellet in 25 mL cold water and transfer to a 50-mL conical tube. Pellet cells by centrifugation at 3400×*g* for 5 min at 4 °C and remove supernatant. Cell pellets can be either used directly for RNA isolation or snap-frozen in liquid N2 and stored at −80 °C.

To purify total RNA, add 5 mL acid phenol (Sigma P4557) and 5 mL AES Buffer to the cell pellet. Incubate for 30 min in a 65 °C water bath, vortexing for 30 s every 5 min. Add 5 mL chloroform (Sigma C2432) and centrifuge at 3400×*g* for 5 min at room temperature. Transfer the upper,

aqueous phase to a fresh 15-mL conical tube, add 5 mL of acid phenol:-chloroform:isoamyl alcohol (Ambion AM9732), and vortex. Centrifuge and transfer aqueous phase to a new 15-mL conical tube as above. Repeat the extractions until the interface between the organic and aqueous phases is free of protein (typically two to three times). Perform a final extraction by transferring the aqueous phase to a new 15-mL MaXtract High Density tube (Qiagen 129065), adding 5 mL chloroform, and spinning as above. Transfer the aqueous phase to an Oakridge Tube (Thermo 3114-0030) and add 1/9th volume 3 M NaOAc, pH 5.3, and 1 vol isopropanol. Spin at 14,000×*g* for 30 min at 4 °C. Wash the RNA pellet twice with 10 mL ice-cold 70% ethanol, spinning at 14,000×*g* for 10 min at 4 °C. Air dry the pellet and Resuspend RNA in 1–2 mL H_2O. This protocol is adapter from (Carlile et al., 2015).

2.1.2 Poly(A) selection using oligo (dT) cellulose beads

Wash 1.5 mL of 50% slurry of oligo (dT) cellulose beads (NEB S1408) in 15 mL conical tubes. Pellet the beads at 3000×*g* for 30 s at room temperature and remove the supernatant. Wash three times with 3 mL water, followed by two washes with 3 mL of TES + NaCl, pelleting as above between washes. Prior to pelleting the final wash, distribute equal volumes to individual 15 mL conical tubes for. Bring 7.5–10 mg of total RNA up to a volume of 4.5 mL in TES (without NaCl) in a 15-mL conical tube. Denature at 65 °C for 15 min and then place on ice for 2 min. To bind poly (A) RNA to the beads, add 563 μL 5 M NaCl to the denatured RNA and transfer to the oligo (dT) cellulose pellet. Vortex and incubate at room temperature for 15 min. To perform a second round of binding, which increases mRNA yield, pellet the beads as above, transfer the supernatant to a 15-mL conical tube, denature at 65 °C for 10 min, and place on ice for 2 min. Then, add the supernatant back to the beads, vortex to mix, and incubate at room temperature for 15 min. Pellet the beads as above and discard the supernatant. Wash the beads three times with 5 mL TES + NaCl, incubating each wash at room temperature for 2 min with rotation. Wash once with 2 mL ice-cold water, vortex briefly, pellet as above, and discard the supernatant. To elute poly(A) RNA, add 2 mL 55 °C water to the beads and incubate at 55 °C for 5 min. Pellet as above and transfer supernatant to a 15-mL conical tube. Repeat elution as before and pool eluates. Wash the beads once in 5 mL water and once in 5 mL TES + NaCl for reuse in the second round of poly(A) selection. Bring the pooled eluates up to 5 mL total volume in TES (without NaCl) by adding

50 μL 1 M Tris, pH 7.6; 10 μL 0.5 M EDTA, pH 8.0; 25 μL 20% SDS; and water to 5 mL total. Denature RNA at 65 °C for 10 min and place on ice for 2 min. To bind poly(A) RNA, add 626 μL 5 M NaCl to the denatured RNA, transfer to the washed beads, and incubate at room temperature for 15 min. To perform a second round of denaturation, pellet the beads as above, transfer the supernatant to a 15-mL conical tube, denature at 65 °C for 5 min, and place on ice for 2 min. To rebind, add the supernatant back to the beads, vortex to mix, and incubate at room temperature for 15 min. Perform the washes and elutions as above, except elute twice in 1.8 mL 55 °C water, and pool. To remove residual oligo (dT) cellulose beads, pass the eluates through a 0.45-μm cellulose acetate filter (VWR 28145-481) with a syringe and transfer in 900 μL aliquots to 2 mL microcentrifuge tubes. To each aliquot, add 1/9th volume 3 M NaOAc, pH 5.3; 2 μL GlycoBlue (Invitrogen AM9516); and one volume isopropanol and precipitate at 20 °C for at least 30 min. Spin in a microcentrifuge at max speed at 4 °C for 30 min. Wash the pellet in 750 μL ice-cold 70% ethanol, spin at max speed at 4 °C for 10 min, and air dry for 2 min. Resuspend each pellet in 6 μL of water and pool into a single PCR tube. This protocol is adapter from (Carlile et al., 2015).

2.1.3 Poly(A) selection using oligo (dT) magnetic beads

Wash 400 μL Dynabeads oligo$(dT)_{25}$ (Thermo Fisher 61002) in 1.5 mL microcentrifuge tubes. Place tubes on magnetic support, magnetically separate beads from storage buffer and remove buffer. Remove tubes from magnetic support, and wash beads with 1 mL TES + NaCl. Separate buffer from beads, remove buffer and repeat wash three times. Bring 2 mg of total RNA up to a volume of 400 μL TES, and denature at 65 °C for 15 min. Add denatured RNA to beads, and incubate on ice for 2 min. Incubate the beads and RNA at room temperature for 10 min with rotation. Magnetically separate beads from total RNA, remove total RNA, and wash beads with 1 mL TES + NaCl. Wash again with 1 mL TES + NaCl, then repeat wash with TES. Magnetically separate beads and remove final wash. Elute polyA+ RNA in 180 μL TES by incubating beads at 65 °C with rotation for 5 min. Magnetically separate and transfer RNA to a fresh 1.5 mL microcentrifuge tube. Bring eluates up to 200 μL total volume in TES + NaCl by adding 20 μL 5 M NaCl. Wash the beads once 1 mL water and once in 1 mL TES + NaCl for reuse in the second round of poly (A) selection. Denature RNA at 65 °C for 15 min. Add denatured RNA to

beads, and incubate on ice for 2 min. Incubate the beads and RNA at room temperature for 10 min with rotation. Magnetically separate beads from total RNA, remove total RNA, and wash beads with 1 mL TES + NaCl. Wash again with 1 mL TES + NaCl, then repeat wash with TES. Magnetically separate beads and remove final wash. Elute polyA+ RNA in 180 μL TES by incubating beads at 65 °C with rotation for 5 min. Magnetically separate and transfer RNA to a fresh 1.5 mL microcentrifuge tube. To each tube, add 20 μL 3 M NaOAc, pH 5.3; 2 μL GlycoBlue and one 1 mL isopropanol and precipitate at −20 °C for at least 30 min. Spin in a microcentrifuge at max speed at 4 °C for 30 min. Wash the pellet in 750 μL ice-cold 70% ethanol, spin at max speed at 4 °C for 10 min, and air dry for 2 min. Resuspend each pellet in 6 μL of water and pool into a single PCR tube.

2.1.4 tRNA size selection using silica columns

Prepare smRNA enriched total RNA using RNeasy MinElute Columns (Qiagen 74204). Dilute 50 μg total RNA to 350 μL with RLT buffer in 1.5 mL microcentrifuge tubes. Add 350 μL 70% EtOH, and transfer sample to RNeasy mini spin column in 2 mL collection tube. Centrifuge at ≥8000×*g* for 15 s at RT. Transfer flow-through (containing smRNA) into fresh tube. Large RNA > 100 nt is still bound to column. Add 450 μL 100% EtOH to flow-through fraction and vortex.

Transfer sample to fresh RNeasy MinElute spin column on vacuum manifold. Run vacuum until all liquid has passed through column. Add 500 μL buffer RPE to column and run vacuum until all liquid has passed through column. Add 500 μL 80% EtOH to column and run vacuum until all liquid has passed through column. Remove RNeasy column from manifold and transfer to fresh 2 mL collection tube. Add 14 μL Rnase-free water onto membrane and centrifuge for 1 min at ≥8000×*g*. This is the smRNA enriched total RNA. To recover large RNA, add 700 μL RWT buffer into RNeasy mini column w/ large RNA bound from before. Transfer column vacuum manifold, and run vacuum until all liquid has passed through column Add 500 μL buffer RPE to column and run vacuum. Repeat wash with another 500 μL buffer RPE. Transfer column to fresh 2 mL collection tube. Centrifuge at ≥20,000×*g* for 1 min at room temperature. Place column in fresh 1.5 mL tube. Pipette 40 μL RNase-free water to column and spin at ≥8000×*g* for 1 min to elute large RNA.

3. D-Seq Library preparation

3.1 RNA fragmentation

Randomly fragmented RNA is needed to ensure even sequencing coverage of the RNA of interest. RNA can be randomly fragmented using divalent zinc cations. The cleavage and fragmentation efficiency in RNA fragmentation differs dramatically depending on the type and source of RNA being fragmented. It is critical to empirically determine the optimal fragmentation time and temperature may be necessary to obtain RNA fragments in the desired size range. For D-seq, fragments in the range of 100–200 nt range are suitable, though a 20 nt size distribution (e.g., 140–160 nt) should be used for a specific experiment. Fragment all of the RNA from Section 2. This should be approximately 2 μg of RNA. Dilute the RNA to a volume of 54 μL and place on ice. Add 6 μL of 100 mM $ZnCl_2$ (10 mM final) and fragment for 55 s at 94 °C. Quench the reaction by quickly placing on ice and adding 60 μL of 40 mM EDTA (20 mM final). Add 1/9th volume 3 M NaOAc and one volume isopropanol, and precipitate at 20 °C for at least 30 min. Spin and wash as described in Section 2.4. Resuspend fragmented RNA in 16.4 μL of water, and transfer to a PCR tube.

3.2 Borohydride reduction

Make a fresh solution of 100 mg/mL $NaBH_4$ (Sigma Aldrich 213462) in 10 mM KOH (Sigma Aldrich 484106). To each 16.4 μL of fragmented RNA, add 2 μL of 100 mg/mL $NaBH_4$ in 10 mM KOH, and 1.6 μL 500 mM Tris–HCl pH 7.5 (Sigma Aldrich 93362). Incubate at 0 °C for 1 h. Neutralize the reaction with 4 μL 6 M Acetic Acid. Add 1/9th volume 3 M NaOAc and one volume isopropanol, and precipitate at 20 °C for at least 30 min. Spin and wash as described in Section 2.4. Resuspend fragmented RNA in 6 μL of water, and transfer to a PCR tube.

3.3 3′ End healing

RNA fragmentation with divalent cations leaves a 2′,3′ cyclic phosphate, which must be removed to make the RNA fragments suitable substrates for subsequent adapter ligation. These cyclic phosphates can be converted to 3′ phosphates by T4 polynucleotide kinase (PNK), which can then be removed by the action of Calf Intestinal Alkaline Phosphatase (CIP). To 8 μL of RNA, add 0.5 μL of RNasin Plus (Promega N2615), 1.25 μL 10X PNK Buffer, 1.25 μL T4 PNK

(NEB M0201), and 1 μL CIP (NEB M0290) and incubate at 37 °C for 1–2 h. Add 12.5 μL of 2× RNA Loading Dye to prepare for size selection (Section 3.4).

3.4 RNA size selection

After Borohydride treatment and 3′ end healing, RNA fragment sizes must be selected. Selection of a narrow range of fragment sizes, spanning 20 nt, at this step allows reliable separation of truncated from full-length cDNAs after RT. RNA fragments are size selected by excising and eluting the desired range of RNA fragment sizes from a denaturing PAGE gel. Prepare 8% TBE/Urea/Polyacrylamide mini-gels (8 × 10 cm), and prerun for 20 min at 200 V. While the gel is prerunning, prepare RNA fragments and 20 bp oligo ladder for loading. For each gel, prepare 20 μL of 20 bp oligo ladder (0.5 μL Ladder, 9.5 μL water, 10 μL 2× RNA Loading Dye). Denature the RNA fragments and ladder at 65 °C for 5 min, and then place on ice until loading. Load the gel and run for 45 min at 200 V. Remove the gel from the cassette, and stain for 5 min with SYBR Gold (Invitrogen S-11494) diluted 1:10,000 in 0.5X TBE, and then visualize by blue light transillumination. Excise several ranges of RNA fragment sizes (40–60, 60–80, 80–100, Fig. 2) guided by the oligo ladder. Proceed with one range of fragment sizes and keep the other size ranges for backup. Backup RNA fragments can be stored as gel slices at 80 °C or can be eluted from the gel slices, and the precipitating eluates can be stored in isopropanol at 20 °C. Elute the RNA fragments from the gel slices by incubating gel slices in 400 μL RNA Elution Buffer overnight at 4 °C with shaking. The next day, transfer to 1.5 mL microcentrifuge tubes, add 1 mL isopropanol, and 2 μL GlycoBlue. Precipitate at −20 °C for at least 30 min and spin as described in Section 2.2. Resuspend the pellet in 5.5 μL of water.

3.5 3′ Adapter ligation

To provide a uniform primer binding site for cDNA synthesis from the RNA fragments, ligation of a universal 3′ DNA adapter is required. To ligate a primer binding site, an adenylated oligonucleotide is required. To the 5.5 μL of size selected RNA fragments, add 0.5 μL of pre-adenylated adapter, 1.2 μL 10× T4 RNA ligase buffer, 1 μL RNAsin Plus, 1 μL T4 RNA ligase (NEB M0204S), and 3 μL of 50% PEG-8000. Do not add ATP. Incubate at 22 °C for 3 h. To precipitate, add 30 μL 3 M NaOAc, pH 5.3, 260 μL water, 2 μL GlycoBlue, and 300 μL isopropanol. Precipitate at 20 °C for at least 30 min and spin as described in Section 2.2. Resuspend the pellet in 7 μL water.

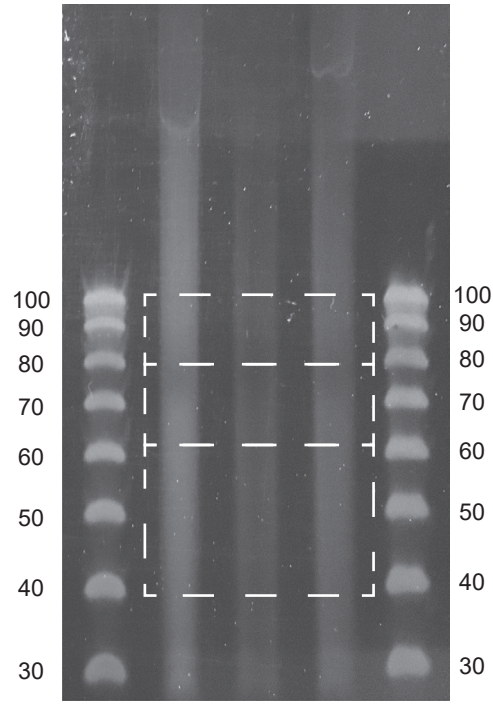

Fig. 2 RNA fragmentation gel.

3.6 Reverse transcription

To enrich for RT stops, we size select for truncate cDNAs. Multiple RTs terminate at reduced dihydrouridine, including AMV RT (Promega M5108), SSII (Invitrogen 18064022), SSIII (Invitrogen 18080093). Using SSIII, reverse transcribe RNA by adding 1 μL of 25 μM oWG915 RT primer and 1 μL of 10 mM dNTPs to each RNA sample. Also set up a no RNA control RT reaction. Heat to 65 °C and then slowly cool to 42 °C thermocycler changing temperature by 10 °C every 2 min and put on ice. Next, add 3 μL 5× First Strand Buffer, 1 μL DTT, 1 μL RNasin Plus, 1 μL SSII enzyme. Briefly centrifuge the tubes to collect any liquid and then place on ice. Incubate at 42 °C for 1 h Remove the RNA by adding 1.5 μL 1 M NaOH, and incubating in a thermocycler at 95 °C for 5 min. Neutralize the pH by adding 1.5 μL 1 M HCl. Prepare the samples for size selection by adding 17 μL of 2× RNA Loading Dye.

3.7 Size selection

Prepare 8% TBE/Urea/Polyacrylamide mini-gels (8 × 10 cm), and prerun for 20 min at 200 V. While the gel is prerunning, prepare cDNAs and 20 bp oligo ladder for loading. For each gel, prepare 20 μL of 20 bp oligo ladder (0.5 μL Ladder, 9.5 μL water, 10 μL 2× RNA Loading Dye). Denature the cDNAs and ladder at 65 °C for 5 min, and then place on ice until loading. Load the gel and run for 45 min at 200 V. Remove the gel from the cassette, and stain for 5 min with SYBR Gold (Invitrogen S-11494) diluted 1:10,000 in 0.5× TBE, and then visualize by blue light transillumination. Excise gel slices that correspond to truncated cDNAs (extended 25 nt beyond the primer, and 25 nt shorter than the corresponding full-length cDNA). Incubate gel slices in 400 μL RNA Elution Buffer overnight at 4 °C with shaking. The next day, transfer to 1.5 mL microcentrifuge tubes, add 1 mL isopropanol, and 2 μL GlycoBlue. Precipitate at −20 °C for at least 30 min and spin as described in Section 2.2. Resuspend the pellet in 5 μL water.

3.8 5′ Adapter ligation

To provide a uniform primer binding site for PCR amplification of the cDNAs, ligation of a universal 5′ DNA adapter is required. To the 5 μL of gel purified truncated, add 0.8 μL 80 μM oWG920, 1 μL 100% DMSO, incubate at 75 °C for 2 min, then place on ice for 2 min. Add 2 μL 10× RNA Ligase Buffer, 0.2 μL 100 mM ATP, 6.5 μL 50% PEG-8000, 3.6 μL water, and 0.5 μL 1 μL T4 RNA ligase (NEB M0204S). Mix thoroughly by pipetting up and down a few times. Incubate at 22 °C overnight.

Before PCR amplification, it is necessary to remove the excess unligated 5′ adapter using a magnetic bead based cleanup. Prepare Dynabeads MyOne Silane (Thermo Fisher 37002D), for clean up by pipetting 10 μL of beads into a 1.5 mL microcentrifuge tube. Magnetically separate beads, remove buffer, remove from magnet and resuspend in 500 μL RLT Buffer. Magnetically separate again, and remove RLT buffer. Resuspend beads in 35 μL RLT, add linker ligated cDNA, and add 45 μL 100% isopropanol. Mix by pipetting up and down twice, and incubate at room temperature for 5 min. Magnetically separate, remove supernatant, and wash beads with 200 μL 70% ethanol. Move the beads to a fresh 1.5 mL microcentrifuge tube, repeat wash step again, and air dry beads for 5 min on magnet. Remove from magnet, resuspend beads in 10 μL water. Magnetically separate again, and transfer water to a new tube.

3.9 PCR amplification

Linker ligated cDNAs can then be PCR amplified, and gel purified to yield a library suitable for sequencing. Before final library generation, the correct number of PCR cycles required to generate libraries with optimal amplification must be experimentally determined. Avoiding over amplification of final libraries is important to ensure that composition of library accurately reflects the RNA fragment/cDNA pool, and that the libraries are not dominated by certain sequences after PCR. Over amplification and unequal representation can be done mitigated by avoiding saturated PCR products, or those with higher molecular weight bands. To perform diagnostic PCRs, prepare a PCR Master Mix for each library in a 1.5 mL microcentrifuge tube. Per library add 15 μL HF Buffer, 1.5 μL 10 mM dNTPs, 3.78 μL Forward PCR Primer, 3.78 μL Barcoded Reverse PCR Primer, 52.6 μL H_2O, 0.75 μL Phusion High-Fidelity DNA Polymerase (NEB M0530L), 4.5 μL linker ligated cDNA. Transfer 16.7 μL of the master mix to 4 PCR tubes, and perform PCR with the following program: (1) 98 °C 30 s, (2) 98 °C 10 s, (3) 67 °C 20 s, (4) 72 °C 40 s. Repeat steps 2–4 for 12, 14, 16, or 18 cycles. Add 3.4 μL of DNA Loading Dye to each PCR reaction. Prepare an aliquot of NEB ssRNA low range ladder for each gel by mixing NEB ssRNA low range ladder (NEB N0364S), 15.7 μL H_2O, and 3.3 μL 6× DNA Loading Dye. Load the reactions on an 8% TBE/Polyacrylamide mini-gel and run at 200 V for 40 min. Disassemble, stain, and visualize the PCR gel as described in Section 3.4. Once the optimal PCR amplification cycle is determined, scale up the final to 50 μL final volume. Load and run the final PCRs on an 8% TBE/Polyacrylamide mini-gel as above, and excise the library band (Fig. 3). Elute the library from the gel slices as described in Section 3.4, and submit libraries for QC and sequencing.

4. D-seq data analysis

Dihydrouridine sites can be identified through analysis of D-seq libraries. The 5′ ends of D-seq reads are generated from RT stops. Reproducible 5′ read ends (RT stop sites) in libraries with borohydride and DUS activity that are not detected without borohydride or DUS activity represent D sites. Analysis of D-seq data is performed using a mix of publicly available software packages designed for the analysis of Illumina sequencing data, and custom python scripts.

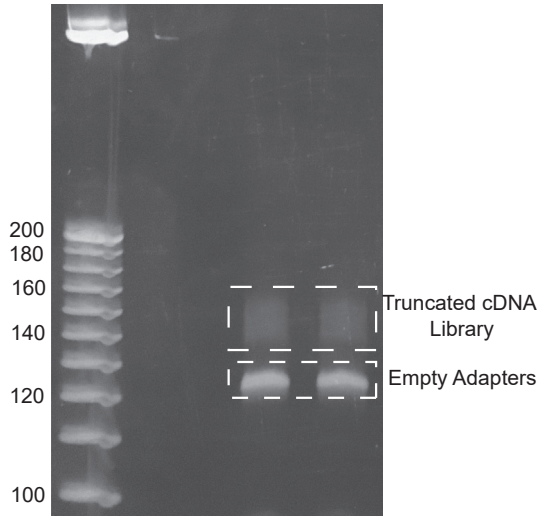

Fig. 3 Final PCR library purification gel.

Before analysis of D-seq data can be performed, several pre-processing steps must be completed. First, generation of a genome index file for the genome of interest is required. Next, processing of the D-seq reads is performed to remove adapters and index sequences. Because the D-seq library architecture includes a unique molecular identifier (UMI) in the 5′ adaptor, PCR duplicates can be identified by collapsing reads with identical UMIs. After duplicate collapsing, the UMIs can be trimmed with another round of sequence trimming. Once the reads are pre-processed, they can then be aligned to the target genome, and 5′ ends positions can be determined. Reproducible signals can then be identified and scored for analysis.

4.1 Trimming of adapter sequences

Accurate mapping of 5′ read end position is essential for the D-seq method. To ensure accurate read mapping, illumina adapters and index sequences must be trimmed. A fasta formatted file of Illumina adapter sequences can be obtained here: https://github.com/BioInfoTools/BBMap/blob/master/resources/adapters.fa. Multiple software packages can perform this step, but the bbduk.sh component of the BBMap toolkit (Bushnell, n.d.) is simple and efficient. It can be called using the following command:

```
>bbduk.sh in1=reads.fq out1=reads.trimmed ref=adapters.fa ktrim=r k=23 mink=11 hdist=1 tpe tbo.
```

4.2 PCR duplicate collapsing

Removal of sequences duplicated during library PCR amplification is crucial for accurate quantification of D-seq peak heights. Clumpify.sh from the BBMap toolkit can be used to collapse based on UMI using the following command:

```
>clumpify.sh in1=reads.trimmed out1=reads.collapsed
dedupe subs=0.
```

4.3 UMI removal

Next, using the following command, bbduk.sh will trim the UMI sequence from the collapsed sequencing reads:

```
>bbduk.sh in1=reads.collapsed out1=reads.clean ftl=10.
```

4.4 Read mapping

Adapter and UMI trimmed and PCR duplicate collapsed reads can be aligned to a reference genome using bbmap.sh with the following command:

```
>bbmap.sh ref=reference_genome.fa in=reads.clean
out=reads.bam slow k=12 ambig=toss.
```

4.5 Read end position gathering

5′ read ends positions can be determined using bedtools (Quinlan & Hall, 2010) and the following command:

```
>bedtools genomecov -d -5 -strand - -ibam reads.bam >
ends.minus.txt.
>bedtools genomecov -d -5 -strand + -ibam reads.bam >
ends.plus.txt.
```

5′ read ends positions can then be combined into a matrix across strands and replicates using custom python scripts.

4.6 D-site identification

Identification of RT stop positions (putative D-sites) can be determined by calculating the deviation from mean number of RT stops at each nucleotide for each U position in the transcriptome. In mRNA, this can be calculated with a 100 nt scanning window and a modified Z-score, in

which the position of interest is excluded when calculating the mean and standard deviation. The windowed Z-score calculation is as follows:

$$Z_{pos} = \frac{ends_{pos} - mean\,(ends_{window})}{stdev\,(ends_{window})}$$

Equation 1: Z-score calculation.

In structured and heavily modified RNAs, such as tRNAs and snRNAs, the Z-score will report false negatives, and a different statistical test is required. Analyzing the absolute deviation around the median (MAD), which scores sites relative to the median rather than the mean decreases the false positive rate, as the MAD-score is not dampened by multiple signals in a 100 nt window. The windowed MAD-score calculation is as follows:

$$M_{pos} = \frac{ends_{pos} - median\,(ends_{window})}{MAD\,(ends_{window})}$$

And

$$MAD = median\,(|\,ends_{pos} - median\,(ends_{window})\,|)$$

Equation 2: MAD-score calculation.

Once individual replicate peak scores are calculated, D-sites can be identified both by requiring reproducibility across replicates, and by empirical determination of peak score cutoffs. To determine peak score cutoffs, the distribution of peak scores in each library can be plotted as an inverse CDF. Cutoffs for D peaks can be determined by identifying the score at which the distributions of with scores from libraries with borohydride treatment and DUS activity separate from the distribution of scores from libraries made without borohydride treatment or DUS activity.

5. Experimental considerations

When planning a D-seq experiment, several factors need to be considered. First, selection of negative control conditions (either omission of borohydride treatment or ablation of DUS activity) is crucial for proper D-site identification. Additionally, adequate sequencing depth and number of biological replicates required must be determined before the experiment can be completed.

5.1 Negative control selection

Proper use of a negative control where RT-stops at D are not expected is essential for D-seq data analysis (Section 4.6). Two types of controls are possible: First, omission of borohydride treatment allows RT-readthrough of D positions, and borohydride omitted libraries can be used to compare D-site scores from borohydride treated libraries to. Alternatively, depending on the model system, ablation of DUS activity can be used as a negative control library. This has the added benefit of allowing assignment of a D-site to a specific DUS enzyme.

5.2 Sequencing depth and biological replicates

The sequencing depth required for D-mapping is dependent both on the expression level of the transcript of interest and the genome size of the experimental system. For budding yeast, a sequencing depth of 15 million reads per library permitted inspection of ~1% of all U nucleotides in the transcriptome for D. This can be scaled up or down depending on the expression level of the target RNA and genome size.

RT-stops can be generated by multiple RNA features and structures other than reduced D. Because of this, multiple biological replicates are required to accurately determine map D residues. We have performed D-seq with 3 replicates of each condition and were able to map Ds in highly expressed transcripts in yeast. If the target RNA to be inspected for D is lowly abundant, more replicates will be required to eliminate false positives.

6. Solutions and reagent recipes

Recipes for required solutions and reagents are below:

6.1 Solutions

YPAD: 1% (w/v) Bacto Yeast Extract, 2% (w/v) Bacto Peptone, 2% (w/v) Glucose, 0.004% (w/v) Adenine Sulfate.
AES Buffer: 50 mM NaOAc (pH 5.3), 10 mM EDTA (pH 8.0), 1% (w/v) SDS.
TES Buffer: 10 mM Tris (pH 7.6), 1 mM EDTA (pH 8.0), 0.1% (w/v) SDS.
TES+NaCl Buffer: 0.5 M NaCl, 10 mM Tris (pH 7.6), 1 mM EDTA (pH 8.0), 0.1% (w/v) SDS.

5× First Strand Buffer: 250 mM Tris–HCl, 375 mM KCl, 15 mM $MgCl_2$

2× RNA Loading Dye: 95% formamide, 5 mM EDTA (pH 8.0), 0.025% (w/v) SDS, 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol FF.

6× DNA Loading Dye: 30% (v/v) Glycerol, 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol FF.

RNA Elution Buffer: 300 mM NaOAc (pH 5.3), 1 mM EDTA (pH 8.0).

DNA Elution Buffer: 300 mM NaCl, 10 mM Tris (pH 8.0).

Acknowledgments

We thank members of the Gilbert lab for discussions and critical feedback on this protocol. Development of D-seq was supported by NIEHS 1R21ES031525 and NCI 5R21CA246118 to W. Gilbert. A. Draycott was supported by NCI 1F31CA254339 and a Gruber foundation fellowship.

References

Arango, D., Sturgill, D., Alhusaini, N., Dillman, A. A., Sweet, T. J., Hanson, G., ... Oberdoerffer, S. (2018). Acetylation of cytidine in mRNA promotes translation efficiency. *Cell, 175*, 1872–1886.e24.

Boccaletto, P., Machnicka, M. A., Purta, E., Piątkowski, P., Bagiński, B., Wirecki, T. K., ... Bujnicki, J. M. (2018). MODOMICS: A database of RNA modification pathways. 2017 update. *Nucleic Acids Research, 46*, D303–D307.

Bushnell, B. (n.d.). BBMap [WWW Document]. SourceForge. URL ⟨https://sourceforge.net/projects/bbmap/⟩ (accessed 5.23.21).

Carlile, T. M., Rojas-Duran, M. F., & Gilbert, W. V. (2015). Pseudo-Seq: Genome-wide detection of pseudouridine modifications in RNA. *Methods in Enzymology, 560*, 219–245.

Collart, M. A., & Oliviero, S. (2001). Preparation of yeast RNA. *Current Protocols in Molecular Biology, Chapter 13*, Unit13.12.

Dalluge, J. J., Hashizume, T., Sopchik, A. E., McCloskey, J. A., & Davis, D. R. (1996). Conformational flexibility in RNA: The role of dihydrouridine. *Nucleic Acids Research, 24*, 1073–1079.

Daniloski, Z., Jordan, T. X., Wessels, H.-H., Hoagland, D. A., Kasela, S., Legut, M., ... Sanjana, N. E. (2021). Identification of required host factors for SARS-CoV-2 infection in human cells. *Cell, 184*, 92–105.e16.

Draycott, A. S., Schaening-Burgos, C., Rojas-Duran, M. F., Wilson, L., Schärfen, L., Neugebauer, K. M., ... Gilbert, W. V. (2022). Transcriptome-wide mapping reveals a diverse dihydrouridine landscape including mRNA. *PLoS Biology, 20*, e3001622.

Emerson, J., & Sundaralingam, M. (1980). Structure of the potassium salt of the modified nucleotide dihydrouridine 3′-monophosphate hemihydrate: Correlation between the base pucker and sugar pucker and models for metal interactions with ribonucleic acid loops. *Acta Crystallographica, Section B, 36*, 537–543.

Finet, O., Yague-Sanz, C., Krüger, L. K., Tran, P., Migeot, V., Louski, M., ... Hermand, D. (2021). Transcription-wide mapping of dihydrouridine reveals that mRNA dihydrouridylation is required for meiotic chromosome segregation. *Molecular Cell, 82*, 404–419.e9.

Gilbert, W. V., Bell, T. A., & Schaening, C. (2016). Messenger RNA modifications: Form, distribution, and function. *Science (New York, N. Y.), 352*, 1408–1412.

Green, M. R., & Sambrook, J. (2019). Isolation of poly(A)+ messenger RNA using magnetic oligo(dT) beads. *Cold Spring Harbor Protocols, 2019*. https://doi.org/10.1101/pdb.prot101733.

Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., ... Zamir, A. (1965). Structure of a ribonucleic acid. *Science (New York, N. Y.), 147*, 1462–1465.

Kato, T., Daigo, Y., Hayama, S., Ishikawa, N., Yamabuki, T., Ito, T., ... Nakamura, Y. (2005). A novel human tRNA-dihydrouridine synthase involved in pulmonary carcinogenesis. *Cancer Research, 65*, 5638–5646.

Kligun, E., & Mandel-Gutfreund, Y. (2015). The role of RNA conformation in RNA-protein recognition. *RNA Biology, 12*, 720–727.

Li, X., Xiong, X., & Yi, C. (2016). Epitranscriptome sequencing technologies: Decoding RNA modifications. *Nature Methods, 14*, 23–31.

Martinez, N. M., & Gilbert, W. V. (2018). Pre-mRNA modifications and their role in nuclear processing. *Quantitative Biology, 6*, 210–227.

Niu, Y., Lin, Z., Wan, A., Sun, L., Yan, S., Liang, H., ... Wan, G. (2021). Loss-of-function genetic screening identifies aldolase A as an essential driver for liver cancer cell growth under hypoxia. *Hepatology (Baltimore, Md.), 74*, 1461–1479.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England), 26*, 841–842.

Rider, L. W., Ottosen, M. B., Gattis, S. G., & Palfey, B. A. (2009). Mechanism of dihydrouridine synthase 2 from yeast and the importance of modifications for efficient tRNA reduction. *The Journal of Biological Chemistry, 284*, 10324–10333.

Roundtree, I. A., Evans, M. E., Pan, T., & He, C. (2017). Dynamic RNA modifications in gene expression regulation. *Cell, 169*, 1187–1200.

Schaefer, M., Kapoor, U., & Jantsch, M. F. (2017). Understanding RNA modifications: The promises and technological bottlenecks of the "epitranscriptome". *Open Biology, 7*, 170077.

Wang, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., ... He, C. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature, 505*, 117–120.

Xing, F., Hiley, S. L., Hughes, T. R., & Phizicky, E. M. (2004). The specificities of four yeast dihydrouridine synthases for cytoplasmic tRNAs. *The Journal of Biological Chemistry, 279*, 17850–17860.