A probe-based capture enrichment method for detection of A-to-I editing in low abundance transcripts

Emma Lamb[a], **Dyuti Pant**[b], **Boyoon Yang**[c], **and Heather A. Hundley**[b,*]

[a]Genome, Cell and Developmental Biology Graduate Program, Indiana University, Bloomington, Indiana, United States
[b]Department of Biology, Indiana University, Bloomington, Indiana, United States
[c]Biochemistry Graduate Program, Indiana University, Bloomington, Indiana, United States
*Corresponding author. e-mail address: hahundle@iu.edu

Contents

Abstract

Exactly two decades ago, the ability to use high-throughput RNA sequencing technology to identify sites of editing by ADARs was employed for the first time. Since that time, RNA sequencing has become a standard tool for researchers studying RNA biology and led to the discovery of RNA editing sites present in a multitude of organisms, across tissue types, and in disease. However, transcriptome-wide sequencing is not without limitations. Most notably, RNA sequencing depth of a given transcript is correlated with expression, and sequencing depth impacts the ability to robustly detect RNA editing events. This chapter focuses on a method for enrichment of low-abundance transcripts that can facilitate more efficient sequencing and

Methods in Enzymology
ISSN 0076-6879, https://doi.org/10.1016/bs.mie.2024.11.033

detection of RNA editing events. An important note is that while we describe aspects of the protocol important for capturing intron-containing transcripts, this probe-based enrichment method could be easily modified to assess editing within any low-abundance transcript. We also provide some perspectives on the current limitations as well as important future directions for expanding this technology to gain more insights into how RNA editing can impact transcript diversity.

1. Introduction

The use of high-throughput sequencing moved the ADAR field from a handful of known adenosine (A) to inosine (I) editing sites discovered mostly serendipitously to millions of sites identified across the human transcriptome (Eisenberg & Levanon, 2018). Since those first sequencing approaches (Athanasiadis, Rich, & Maas, 2004; Blow, Futreal, Wooster, & Stratton, 2004; Kim et al., 2004; Levanon et al., 2004), researchers have used high-throughput sequencing to explore editing in a multitude of organisms and across tissue-types and developmental timescales (Cuddleston et al., 2022; Rajendren et al., 2021; Tan et al., 2017; Zhang et al., 2023). These technologies have also been used to examine transcriptome-wide RNA editing levels of patients suffering from many types of cancer (Han et al., 2015; Paz et al., 2007; Paz-Yaacov et al., 2015) and neuropathological diseases (Hwang et al., 2016; Ma et al., 2021). With these catalogues of A-to-I changes, researchers have been able to get insight into the major impact of ADARs in generating transcriptome diversity and impacts of RNA editing on mRNA expression and various biological processes (Erdmann, Mahapatra, Mukherjee, Yang, & Hundley, 2021).

It is now well-established that RNA editing varies across tissue-types, development and disease, and the amount of editing observed does not directly correlate with expression levels of the ADAR editing enzymes (Deffit & Hundley, 2016). To begin to understand how RNA editing is regulated, researchers have performed RNA-sequencing on cell lines and mutant animals lacking potential regulators, often other RNA binding proteins (Freund et al., 2020; Quinones-Valdez et al., 2019; Sapiro et al., 2020; Washburn et al., 2014). While these types of studies have identified several regulators of RNA editing, there are limitations to these approaches. Most importantly, RNA-sequencing coverage is biased by the expression of a given transcript. Consequently, RNA editing detection and quantification in high-throughput sequencing datasets is limited by sequencing depth (Bazak et al., 2014). Therefore, using unbiased,

transcriptome-wide RNA sequencing to identify regulators of RNA editing is limited to identifying factors that regulate RNA editing of abundant RNAs.

Herein, we describe a protocol for enrichment of low abundance RNAs that can be coupled to high-throughput sequencing and analysis to allow for accurate quantification of RNA editing levels. Enrichment methods for sequencing RNA editing sites have been used previously to study ADAR target RNAs in the human transcriptome. The first methodology used padlock DNA probes to recognize predicted edited regions in complementary DNA (cDNA) (generated from reverse transcription of cellular RNA), which were then amplified to yield enrichment prior to high-throughput RNA-sequencing (Li et al., 2009). However, it was noted that this approach was difficult to achieve accurate quantification due to the wide range of expression of the targeted genes (Zhang et al., 2014). Furthermore, as ADAR target RNAs are double-stranded, and reverse transcriptases used for general conversion and amplification of the transcriptome have difficulties with structured regions (Verwilt, Mestdagh, & Vandesompele, 2023), it is possible that conversion to cDNA prior to enrichment reduces the ability to examine edited regions of transcripts.

Using this protocol, we focus on enriching for low abundance RNAs prior to sequencing to facilitate more accurate quantification of RNA editing. Of particular interest are transcripts containing intronic regions of the genome, which may reside in newly synthesized RNA and/or transcripts with retained introns (Grabski et al., 2021). These areas are of particular interest as it is now well established that the majority of ADAR editing occurs co-transcriptionally (Hsiao et al., 2018; Rodriguez, Menet, & Rosbash, 2012), yet our knowledge of factors that influence RNA editing of intronic regions are limited to the speed of transcription (Saldi, Riemondy, Erickson, & Bentley, 2021; Szabo et al., 2024) and splicing (Licht et al., 2019).

2. Probe design

The ability to capture transcripts that contain editing sites of interest centers on functional probes that both efficiently hybridize to specific RNAs and allow for retrieval from a diverse population of nucleic acids. Specificity is largely governed by the interaction strength with the transcript of interest compared to potential off-target interactions. Interaction

strength of probes is influenced by both the type of nucleic acid and the length of the probe. For example, RNA probes are known to bind tighter to complementary RNA sequences than DNA probes (Farrell, 2005). The ability to retrieve the hybridized probe/transcript complex relies primarily on the use of modified nucleic acids. The most widely used modification of nucleic acids to allow for later retrieval is biotinylation, as the interaction of biotin with the protein Streptavidin is considered one of the strongest non-covalent interactions in nature (Chivers, Koner, Lowe, & Howarth, 2011). Biotin can be added to the ends of a nucleic acid molecule, added to RNA during in vitro transcription, or to DNA using polymerase chain reaction (PCR). There are advantages and disadvantages to the employment of each of these types of probes and modification methods in capture-based sequencing technologies. Commercially available biotinylated probes are readily available, but expensive. Internally labeled RNA probes require a template for RNA polymerase for each probe. In addition, short RNAs (< 50 nucleotides) are challenging to precisely generate via in vitro transcription, while longer RNA probes will require a very high annealing temperature for target RNA capture. Internally labeled DNA probes also have a similar requirement for a template; however, the presence of already generated genomic tools, such as the open reading frame (ORF) clones for the human genome, have been used in clever approaches to generate large numbers of biotinylated DNA probes (Sheynkman et al., 2020).

Here we describe a protocol for use of short (20 nucleotide) commercially synthesized DNA probes with 3′ biotinylated ends to enrich for transcripts that are known to contain editing sites within intronic regions of RNAs expressed in the model organism *Caenorhabditis elegans*. As these probes will be used to hybridize to transcripts that contain A-to-I editing events and ADARs target double-stranded RNA, it is important that both the location of editing sites and structure of the target transcript be considered when designing probes. Editing site information can be obtained from prior studies and then mapped onto genomic sequence information obtained from databases, such as UniProt (UniProt, 2023), or species-specific databases, such as Wormbase (Sternberg et al., 2024). As folding of spliced and intron-containing RNAs can vary, it is important annotate the intronic and exonic boundaries and then use a folding prediction software, such as mFold (Zuker, 2003), to predict the structures of various transcripts that contain the edited region.

When designing probes, it is important to bear in mind the region of the transcript that the editing events reside in and the possible different

RNA species that will contain these regions in the cell. For example, editing events within the 3′ untranslated region of a transcript will be present in species that also contain exonic sequences; thus, probes can be designed to anneal within the exons to pull down any isoforms of RNA with the 3′ UTR. To enable efficient retrieval of the intron-containing transcripts, the single-stranded DNA probes should be complementary to regions of the specific intron that contains editing (albeit not directly complementary to the edited region). Some studies have used a tiling approach, where multiple probes are designed to be either overlapping or slightly apart down the entire length of the transcript of interest (Theil, Imami, & Rajewsky, 2019). However, our initial tests with reporter RNAs indicated a similar enrichment with using a smaller number of probes (approximately 1 per 1000 nucleotides). Furthermore, probes for retrieving ADAR target RNAs should anneal to regions not within any predicted double-stranded region or where editing sites are known to occur (Fig. 1). The presence of stretches of complementary base-pairs (ex. AAAATTTT) within the probe sequence should also be avoided to prevent intramolecular pairing of the probe. Other items to consider include length, content of GC base-pairs, repetitive elements, off-targets in the genome, etc (Chu, Quinn, & Chang, 2012). Upon design of the probes, the sequence can be commercially ordered with a 3′ biotin modification. It is suggested that to

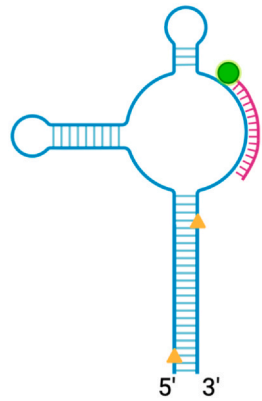

Fig. 1 Example of a predicted mRNA structure and probe design. The structure in blue represents the predicted mRNA structure obtained using software such as mFold. The structure in pink represents the 20 nt probe complementary to the single stranded region of the mRNA transcript, with the green circle representing biotinylation on the 3′ end of the probe. The yellow triangle represents editing sites present on the mRNA transcript. Figure was prepared using Biorender.

avoid hindrance issues between the probe and the magnetic Streptavidin-coated beads used for retrieval, a triethyleneglycol spacer can be placed between the DNA sequence and the biotin (Pritchard, 2023).

3. RNA preparation

High quality, intact RNA is essential for the probes to bind and retrieve the transcripts of interest. As RNA is prone to degradation, caution should be taken to remove nucleases and other contaminants from all workspaces and instruments before working with samples throughout this entire protocol (Rio, Jr, Hannon, & Nilsen, 2010). Typical contamination removal includes wiping all surfaces with water, followed by a 70 % ethanol solution and then a commercial RNase removal cleaner, such as RNaseZap (Sigma). Samples should be manipulated with RNA-safe pipet tips (with barriers) and tubes, which can be purchased as low-retention/siliconized and free from RNases. In addition, all solutions that are used for the RNA samples should be made with RNase/nuclease free chemicals and stored in glassware free of contamination (baked at 200 °C overnight).

RNA is isolated from biological samples through homogenization and phase separation. Homogenization is typically achieved by treatment with ionic detergents or chaotropic salts, while extraction typically includes the use of organic solvents to deproteinize and separate RNA from DNA. A routine RNA extraction method uses the commercial phenol/guanidine isothiocyanate solution, TRIzol (Invitrogen). To yield high quality RNA, we couple the standard TRIzol extraction (see manufacturer's instructions) to a downstream DNase removal step. This step can occur after RNA precipitation and solubilization and with the addition of a commercial DNase enzyme, such as Turbo DNase (Ambion) followed by a column purification, such as the RNeasy columns (Qiagen). Alternatively, the aqueous phase after phase separation can be applied to a column, such as the Zymo Direct-zol RNA miniprep columns (Zymo Research), allowing for both RNA isolation and on column DNase removal.

Another consideration for the RNA samples of interest is whether it is important to perform removal of ribosomal RNA (rRNA) prior to target capture. As rRNA can constitute as much as 85 % of the RNA present in isolated total RNA, researchers have designed methods to efficiently degrade rRNA (Morlan, Qu, & Sinicropi, 2012). These methodologies can significantly improve read coverage of mRNAs for high-throughput

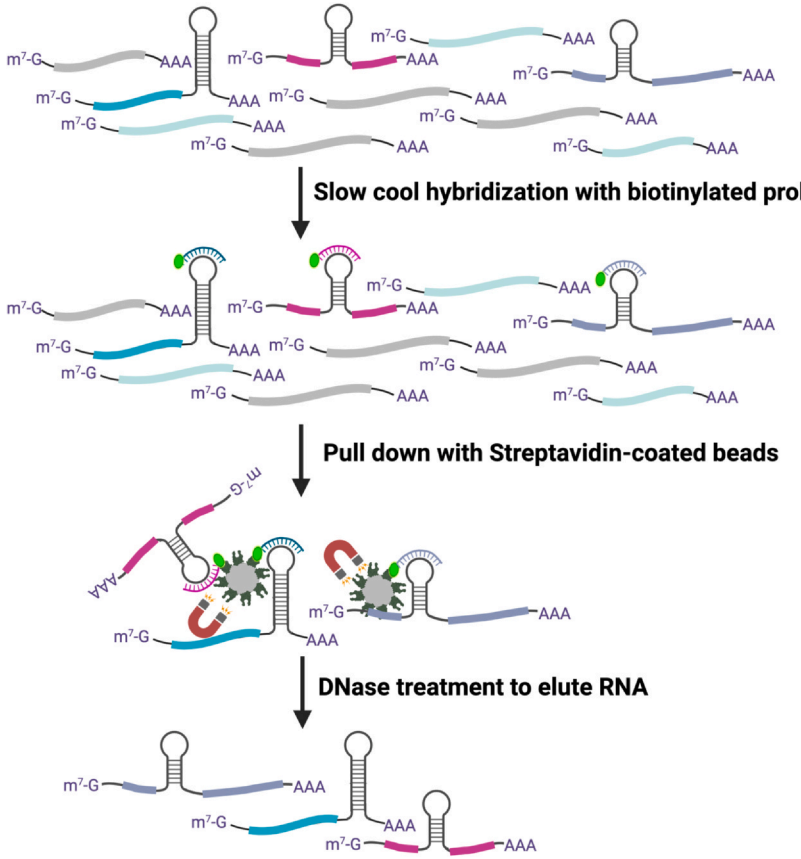

Fig. 2 Illustration of the capture-enrichment step-by-step protocol. The RNA species present in a sample are shown by lines and rectangles. The lines represent non-coding regions and the rectangles represent exonic regions. The m7G text represents the 5′ cap on mRNA, while the AAA text represents the poly-adenosine tail present at the 3′ end of mRNA. The biotinylated probes are depicted as in Fig. 1 (green circle connected to a line structure). The gray circle with green shapes attached represents the Streptavidin-coated magnetic beads, which can be isolated using a magnetic separation rack (illustrated with the horseshoe magnet shape).

sequencing approaches (Adiconis et al., 2013) and could be used in prior to probe annealing in capture enrichment methodologies to reduce off-target binding of probes. Commercial kits for rRNA depletion, such as the NEBNext rRNA depletion kit (New England Biolabs) are available, as well as more cost-effective protocols for frequent users of this methodology (Baldwin, Morris, & Mukherjee, 2021). However, it should be noted that due to differences in the rRNA sequences between species, studies of

RNA isolated from model organisms may require custom rRNA depletion methods (ex. *C. elegans* (Price, Wagner, Pastore, Hertz, & Tang, 2023)).

4. Step-by-step method details

The protocol described below (Fig. 2) provides a general procedure for capturing intron-containing transcripts and may require additional optimization. As a reminder, all steps of this procedure need to adhere to the RNA-safe conditions described in Section 3 above. Furthermore, as RNA is sensitive to pH and RNase contamination, always use gloves and, as much as possible, use equipment and solutions dedicated solely to RNA work.

4.1 Materials and equipment

- Thermocycler (ex. Applied Biosystems ProFlex PCR System)
- Thermomixer (ex. Benchmark Scientific, cat. no. H5000-HC)
- Magnetic separation rack (ex. DynaMag-2, Fisher, cat. no. 12321D)
- Picofuge (ex. Benchmark Scientific MyFuge 12, cat. no. 31-501)
- 1.5 mL siliconized microcentrifuge tubes (ex. Invitrogen RNase-free Microfuge Tubes, cat. no. AM12450)
- 0.5 mL PCR tubes (ex. Thermo Scientific Tubes, 0.5 mL, flat cap, cat. no. AB0350)
- Vortex mixer (ex. Fisher, cat. no. 02215365)
- Dynabeads MyOne Streptavidin C1 magnetic beads (Fisher, cat. no. 6500)

1. Streptavidin-coated Bead Preparation

1. Reagents and Buffers

- Solution A: 0.1 M NaOH, 0.05 M NaCl
- Solution B: 0.1 M NaCl
- 1X Binding and Wash buffer: 5 mM Tris HCl, 0.5 mM EDTA, 20 mM NaCl
- 2X Binding and Wash buffer: 10 mM Tris HCl, 1 mM EDTA, 40 mM NaCl

2. Procedure

2.1. Determine the number of samples to be subjected to enrichment. Vortex the Streptavidin-coated magnetic beads to resuspend. Aliquot the volume corresponding to the number of samples x 50 μL into a 1.5 mL siliconized microcentrifuge tube and keep on ice.

2.2. Add an equal volume of Solution A to the beads. Allow the magnetic beads to sit in solution for 1 min.

2.3. Place the tube containing the beads on the magnetic separation rack. Upon clearing of the solution, pipette off the supernatant.

2.4. Repeat steps 1.2.2. and 1.2.3. with the volume of Solution A used in 1.2.2., twice.

2.5. Add Solution B to the beads using the same volume used for Solution A in step 1.2.2. Allow the beads to sit in solution for 1 min.

2.6. Place the tube containing the beads on a magnetic tube rack. Upon clearing of the solution, pipette off the supernatant.

2.7. Add 1X Binding and Wash buffer to the beads using the same volume as used for Solution A in step 1.2.2. Vortex the sample and immediately place the tube with the beads on the magnetic separation rack for 1 min.

2.8. Turn the tube 180° on the magnetic separation rack and wait for 1 min before pipetting off the supernatant.

2.9. Repeat steps 1.2.7. and 1.2.8. five times.

2.10. Add 2X Binding and Wash buffer to the beads at twice the initial bead volume and store the tubes on ice. The Streptavidin-coated beads are now considered safe for working with RNA.

2. Biotinylated probe preparation

2.1. Reagents and Buffers

- RNA-safe water
- Custom commercial biotinylated DNA oligos (IDT, 100 nmol, standard desalting)

2.2. Procedure

2.2.1. Upon arrival, quickly spin the tubes containing the lyophilized oligo for 30 s in a picofuge.

2.2.2. Resuspend the pellet in RNA-safe water to yield a final concentration of 500 μM.

Note: exact amount of the oligo should be provided by the manufacturer.

2.2.3. Pool all oligos that will be used in the capture experiment. The exact concentration of this pool may vary depending on the number of oligos used in the experiments. The manufacturer recommends 500 picomoles of biotinylated single stranded oligos should be applied to 50 μL of

Streptavidin-coated magnetic beads. As the hybridization reaction will occur in a total volume of 100 μL and includes the Oligo hybridization buffer and the RNA sample, the oligo pool should be at a total concentration between 50-100 μM.

Note: from this point forward the biotinylated oligos will be referred to as probes.

2.2.4. The concentrated probe stocks can be stored at −80 °C for more stability. The pooled probe mixtures for the capture experiments can be aliquoted and stored at −20 °C.

3. Hybridization of RNA and Biotinylated Probes

3.1. Reagents and Buffers

- 5X Oligo hybridization buffer: 500 mM Tris-HCl (pH 7.4), 1 M NaCl
- Biotinylated probe mixture

3.2. Procedure

Note: in order to determine the enrichment of different RNA species upon completion of this protocol, an RNA sample matching the one added to the PCR tube in the first step described below should be set aside for later analysis. In later sections of the protocol, this RNA sample is referred to as the "input".

3.2.1. Add your RNA sample to a PCR tube.

Note: the amount of RNA sample to be used varies depending on the expression of the targets attempting to capture and probe binding efficiency. The concentration of RNA sample, as well as the depletion of rRNA (described above), needs to be optimized.

3.2.2. Add 500 picomoles of biotinylated probe mixture to the same tube.

Note: this amount is reflective of the manufacturers' suggestion to use twice the binding capacity for 50 μL of beads, which was the starting point of this protocol. If scaling up the capture experiment, the molecules of biotinylated probe used should also be scaled accordingly.

3.2.3. Add 5X Oligo hybridization buffer for a final concentration of 1X. Bring volume up to 100 μL with RNA-safe water.

3.2.4. Repeat steps 3.2.1.−3.2.3. for each sample.

3.2.5. Using a thermocycler, slowly cool the RNA/probe mixtures from 65 °C to 25 °C dropping in temperature 1 °C every 5 min. Immediately proceed to Step 4.2.1. upon reaching 25 °C.

4. Capture of hybridized RNA-probe species

4.1. Reagents and Buffers

- 1X Binding and Wash buffer: 5 mM Tris HCl, 0.5 mM EDTA, 20 mM NaCl

4.2. Procedure

4.2.1. Place microcentrifuge tubes containing washed Streptavidin-coated magnetic beads (from step 1.2.10.) into a thermomixer set at 25 °C.

4.2.2. Transfer the RNA/probe mixture (from step 3.2.5.) to a new 1.5 mL siliconized microcentrifuge tube.

4.2.3. Add 100 μL of the pre-warmed Streptavidin beads to the tubes containing the RNA/probe mixture.

4.2.4. Incubate samples in a thermomixer at 25 °C for a total of 15 min; shaking at 1200 rpm for 90 s with a 15 second rest between each shaking.

4.2.5. Place the tube containing the beads, probes, and RNA on the magnetic separation rack. Upon clearing of the solution, pipette off the supernatant.

4.2.6. Wash the beads with 300 μL of 1X Binding and Wash buffer by turning the tube 180° on the magnetic separation rack. Upon clearing of the solution, pipette off the supernatant.

4.2.7. Wash the beads with 150 μL of 1X Binding and Wash buffer by turning the tube 180° on the magnetic separation rack. Upon clearing of the solution, pipette off the supernatant.

Note: for efficient DNase treatment in the steps below, it is critical to remove all the 1X Binding and Wash buffer. If necessary, quick spin the microcentrifuge tube in a picofuge, replace on magnet, wait 1 min and then remove last of supernatant.

4.2.8. Immediately proceed to step 5.2.1. below.

5. RNA elution

5.1. Reagents and Buffers

- TURBO DNase (2 U/μL) (Invitrogen cat. no. AM2238)
  - Supplied with 10X TURBO DNase buffer
- RNasin Plus Ribonuclease Inhibitor (Promega cat. no. N2611)

5.2. Procedure

5.2.1. To each sample add the following:

2.1 μL 10X Turbo DNase buffer
1 μL RNasin
2 μL TURBO DNase
15.9 μL RNA water

5.2.2. Incubate the samples at 37 °C for 1 h in a thermomixer shaking at 1200 rpm for 1 min and 30 s every 15 s.

5.2.3. Place the tubes on the magnetic separation rack.

5.2.4. Upon clearing of the supernatant, pipette the supernatant into new 1.5 mL siliconized microcentrifuge tubes. These tubes now contain your captured RNA.

5.2.5. Store the RNA samples at −80 °C or proceed to library preparation for RNA sequencing.

Note: during optimization of this protocol, it may be useful to perform quantitative real-time PCR (qPCR) to assess whether enrichment of target RNAs is occurring with the specific probes and RNA samples of interest. As an example, capture enrichment was performed using a mixture containing ten probes that are complementary to the intronic regions of the C. elegans gene, mrpl-38 (Fig. 3). Enrichment of the pre-mrpl-38 transcript was improved about 2-fold by prior rRNA depletion of the RNA subjected to the capture protocol.

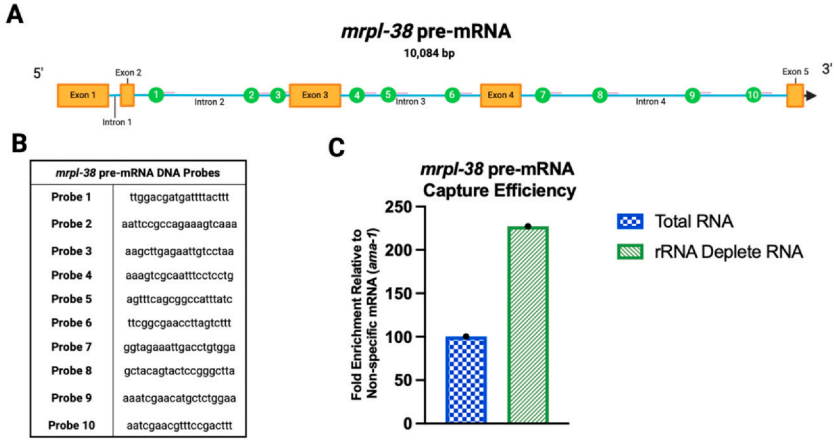

Fig. 3 Capture protocol enriches for pre-mRNA transcripts. (A) A schematic of the *C. elegans mrpl-38* transcript, where the lines represent intronic regions and the rectangles represent exonic regions. The location of the ten biotinylated DNA probes that were complementary to the intronic sequences are shown with a green circle and pink line. (B) Table of the sequences of the ten biotinylated DNA probes shown in panel A. (C) Total RNA was isolated from hatched, first stage larval wildtype animals and subjected either directly to the capture protocol (bar with blue hatches) or ribosomal RNA depletion and then the capture protocol (bar with green diagonals). The amount of precursor *mrpl-38* transcript was measured in the input RNA and the pulldown using qPCR. In addition, as a control for specificity, the presence of an abundant transcript corresponding to a subunit of RNA polymerase (*ama-1*) in each sample was also measured. The IP/input values for the probe target pre-*mrpl-38* was then normalized to the non-probe containing transcript, *ama-1*.

5. High-throughput sequencing and analysis

5.1 Library preparation

Examination of the sequence and editing state of the transcripts obtained via this capture enrichment method can be achieved by performing high-throughput RNA sequencing. To perform RNA sequencing using second generation sequencing methods, the captured RNA, as well as input RNA if assessing enrichment, should be converted to cDNA and sequencing libraries should be prepared. Commercial kits, such as the KAPA Stranded RNA-seq kit (cat. no. KK8400), can be employed for library preparation. For studies of RNA editing, it is critical to ensure that kits can both account for the RNA structure as well as retain strand information to allow for accurate identification of A-to-I editing sites. As ADARs target double-stranded regions of mRNA, library preparation that includes a fragmentation step prior to reverse transcription is important to obtain adequate coverage of the editing events. For this protocol in particular, the fragmentation time needs to be carefully optimized due to the low amount of captured RNA compared to a typical input/total RNA sample for library preparation. Another consideration for low input samples is the concentration of adapters that ligate to the fragments of interest; if too high concentration of adapter is used, adapter dimers will be formed in the ligation reaction and lead to a reduction in sequencing of the samples of interest.

5.2 Read processing and mapping

Computational analysis of high through-put sequencing results requires an annotated and indexed reference genome and appropriately processed reads to align to it. As sequencing utilizes adaptors that are ligated to the 3′ end of each RNA fragment, these adaptor sequences must be removed from each read before alignment to the reference genome. Once this is done, alignment software, such as STAR, can be used to align the acquired sequencing reads to the reference genome of choice (STAR –runThreadN 8 –outFilterMMultimap Nmax 1 –outFilterScoreMinOverLread.66 –outFilterMismatchNmax 10 –out FilterMismatchNoverLmax.3 –runMode alignReads –genomerDir genome/index –readFilesIn $FASTA –outFileNamePrefix $FILENAME –outSAM attributes All –outSAMtype BAM SortedByCoordinate) (Dobin et al., 2013). Note, some of these parameters are specific to the size of the reference genome and sequencing files to align (ex. runThreadN). However, others are important for either reducing false positive predictions for editing sites due to incorrect read alignment (ex. outFilterMMultimapNmax 1 which indicates the reads need to

uniquely map). In addition, other commands are important for ensuring accurate detection of editing sites (ex. outFilterMismatchNmax which sets the number of maximum mismatches between the sequencing read and the reference genome (remember that editing sites will appear as A-to-G mismatches in the alignment)). Users should understand the overall commands as well as specifics of the RNA-sequencing experiment analyzed (ex. read length) and adjust the alignment accordingly.

The alignment process will result in an output of compressed sequence alignment files, or BAM files. These BAM files can be viewed in a genome browser to see where the reads accumulate with respect to the reference genome; thus, providing insight into the RNA species captured in the experiment. The Bioconda subreads software, featureCounts, can be used to summarize information about the aligned reads in terms of read count per gene into comma separated value format for easier digestion and quantification of the results (Liao, Smyth, & Shi, 2013).

5.3 RNA editing site identification

After read mapping, RNA editing sites can be computationally identified using the aligned reads. Alternatively, if editing site locations are already known from prior studies and *de novo* identification is not of interest, it is possible to immediately proceed to the following section on quantification of RNA editing levels. The most basic form of A-to-I editing site detection is to simply identify single nucleotide variants (SNVs) of guanosines in RNA-seq data at positions where an adenosine is present in the reference genome. However, there are also many reasons other than RNA editing that could give rise to single nucleotide changes, such as PCR errors, sequencing errors and single nucleotide polymorphisms (SNPs). Therefore, several considerations need to be employed for accurate identification of A-to-I editing sites (Karagianni et al., 2024; Pinto & Levanon, 2019). To address these concerns, a number of different programs (ex. REDItools (Flati et al., 2020) and JACUSA (Piechotta, Naarmann-de Vries, Wang, Altmuller, & Dieterich, 2022)) have been developed. Our preferred method is to use the SAILOR software (Deffit et al., 2017) which has recently been updated to use a Snakemake implementation (Kofman, Yee, Medina-Munoz, & Yeo, 2023). The output of the SAILOR pipelines provides a confidence score in the prediction of each nucleotide as a true A-to-I editing event. The confidence score uses a Bayesian model that considers the read depth and number of variant reads at each nucleotide. In the absence of individual biological replicates of sufficient sequencing depth for A-to-I editing detection, it is

advisable to only proceed with sites of very high confidence (> 0.95–0.99). However, the presence of editing sites in multiple biological replicates may allow for a lower confidence score to be employed. In addition, it is recommended to sequence RNA isolated from a control animal or cell line lacking ADARs to remove false-positive sites. However, it should be noted that the SAILOR pipelines employ a number of filters to avoid false identification of editing sites including removing known SNPs (provided by the user), and predictions that reside in low-quality reads (ex. have > 1 non A-to-G mismatch), reads that have a splice junction overhang < 10 nucleotides, and reads that have mismatches only near the ends.

5.4 Quantification of RNA editing levels

To accurately quantify RNA editing levels from high-throughput sequencing data, it is important to limit the examination of sequencing reads that arise due to PCR duplication in the library preparation. Therefore, any attempt to quantify RNA editing should begin with removal of duplicated reads using samtools rmdup (Li et al., 2009) (samtools rmdup inputfilename.bam outputfilename_nodup.bam). After removing PCR duplicates, one of the ways to determine the percent editing at a given position is to call variants in the sequencing data using samtools mpileup (Li et al., 2009) (samtools mpileup -u -I -t DP4 -v -f genomeassembly.fasta inputfilename_nodup.bam | awk '$5 != "<*>"' | tail -n +30 > outputfilename.rmdup.mpileup.vcf).

It is important to note that the SAILOR pipeline also outputs the percent editing observed at each predicted A-to-I editing site. However, if an RNA sample has reduced or no editing at a given site (such as that isolated from a mutant animal/cell line lacking a potential editing regulator), it would not be possible to obtain the percentage editing at that site using SAILOR. In addition, as the SAILOR pipeline limits editing detection to reads filtered by a number of criteria (described above), it is possible that the percent editing output from these pipelines differs from variant calling. Therefore, the choice of method to determine percent editing should take all these into consideration and can vary depending on experimental design.

6. Advantages and limitations of the protocol

This method is designed to provide additional read coverage of low abundance transcripts to allow for more accurate quantification of editing levels.

Applying this method to RNA isolated from different genetic backgrounds and disease conditions should allow for identification of novel regulators of RNA editing, particularly in transient RNA species, such as intron-containing RNAs. However, there are also several limitations to this approach. To use this protocol, one must have prior knowledge of the edited transcripts to examine. It is possible that the protocol can reveal novel editing sites within these transcripts, but identification of novel edited genes is not a potential outcome. There are several limitations concerning the probes, including cost, length and specificity (discussed in the probe design section above). However, these issues are being actively tackled as capture enrichment methods are being used to study many different aspects of RNA biology, and researchers are increasingly becoming more creative in ways to develop cheaper probes, such as the strand displacement amplification method recently provided in the TEQUILA-seq approach (Wang et al., 2023).

7. Conclusions and future perspectives

We hope that use of this capture methodology combined with genetic mutants can lead to identification of novel regulators of editing. Editing levels are altered in many diseases and these changes do not correlate directly with the expression level of the ADAR enzymes; thus, identification of editing regulators can provide potential therapeutic targets. A few studies have identified disease-relevant RNA binding proteins (RBPs) that regulate global intronic editing as well as individual intronic editing events that contribute to disease. The most well-studied of these is editing of the branch point adenosine in the tyrosine phosphatase, PTNP6, which is thought to play a role in acute myeloid leukemia (Beghini et al., 2000). In contrast to this one specific intronic editing event playing a role in disease, a recent study revealed that general ADAR-mediated editing of intronic RNA synergizes with the activity of the RBP hnRNPC to prevent autoinflammatory activation of dsRNA sensors in human monocytes (Herzner et al., 2021). Similarly, TDP-43 has been identified as another important global regulator of intronic editing in human cells and model organisms (Saldi et al., 2014). As TDP-43 and A-to-I RNA editing are implicated in a number of neuropathological diseases, including amyotrophic lateral sclerosis (ALS), further dissecting the transcripts that arise upon TDP-43 loss, particularly the editing sites within and how these editing sites impact gene expression, is important. This protocol could be used to gain insight into these open questions.

In addition to identifying novel regulators of editing, this protocol could further our understanding of the transcript diversity generated by RNA editing. The current protocol couples enrichment capture to second generation RNA sequencing methodology; however, an important future direction will be to obtain transcript level information via use of third generation sequencing technologies (Goodwin, McPherson, & McCombie, 2016). These technologies offer the ability to sequence longer regions of RNA, including entire transcripts, and in a single molecule fashion (Zhang et al., 2024). In addition, third generation sequencers, such as the Oxford Nanopore system, offer the ability for direct RNA sequencing (i.e. without the need to convert to cDNA). Combining long-read sequencing and editing detection will be critical for fully understanding the impact of editing and potential regulators on diversity of the transcriptome.

Acknowledgments

Work in the Hundley lab is supported by the National Institutes of Health (NIH) (R01GM130759). E.L. is supported by the Graduate Training Program in Quantitative and Chemical Biology (NIH Award Number T32GM131994) and Indiana University. B.Y. is supported by the Indiana University College of Arts and Sciences Dissertation Research Fellowship.

References

Adiconis, X., Borges-Rivera, D., Satija, R., DeLuca, D. S., Busby, M. A., Berlin, A. M., ... Levin, J. Z. (2013). Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nature Methods, 10*(7), 623–629. https://doi.org/10.1038/nmeth.2483.

Athanasiadis, A., Rich, A., & Maas, S. (2004). Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biology, 2*(12), e391. https://doi.org/10.1371/journal.pbio.0020391.

Baldwin, A., Morris, A. R., & Mukherjee, N. (2021). An easy, cost-effective, and scalable method to deplete human ribosomal RNA for RNA-seq. *Current Protocols, 1*(6), e176. https://doi.org/10.1002/cpz1.176.

Bazak, L., Haviv, A., Barak, M., Jacob-Hirsch, J., Deng, P., Zhang, R., ... Levanon, E. Y. (2014). A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Research, 24*(3), 365–376. https://doi.org/10.1101/gr.164749.113.

Beghini, A., Ripamonti, C. B., Peterlongo, P., Roversi, G., Cairoli, R., Morra, E., & Larizza, L. (2000). RNA hyperediting and alternative splicing of hematopoietic cell phosphatase (PTPN6) gene in acute myeloid leukemia. *Human Molecular Genetics, 9*(15), 2297–2304. https://www.ncbi.nlm.nih.gov/pubmed/11001933.

Blow, M., Futreal, P. A., Wooster, R., & Stratton, M. R. (2004). A survey of RNA editing in human brain. *Genome Research, 14*(12), 2379–2387. https://doi.org/10.1101/gr.2951204.

Chivers, C. E., Koner, A. L., Lowe, E. D., & Howarth, M. (2011). How the biotin-streptavidin interaction was made even stronger: Investigation via crystallography and a chimaeric tetramer. *The Biochemical Journal, 435*(1), 55–63. https://doi.org/10.1042/BJ20101593.

Chu, C., Quinn, J., & Chang, H. Y. (2012). Chromatin isolation by RNA purification (ChIRP). *Journal of Visualized Experiments (61)*. https://doi.org/10.3791/3912.

Cuddleston, W. H., Fan, X., Sloofman, L., Liang, L., Mossotto, E., Moore, K., ... Breen, M. S. (2022). Spatiotemporal and genetic regulation of A-to-I editing throughout human brain development. *Cell Reports, 41*(5), 111585. https://doi.org/10.1016/j.celrep.2022.111585.

Deffit, S. N., & Hundley, H. A. (2016). To edit or not to edit: Regulation of ADAR editing specificity and efficiency. *Wiley Interdisciplinary Reviews-RNA, 7*(1), 113–127. https://doi.org/10.1002/wrna.1319.

Deffit, S. N., Yee, B. A., Manning, A. C., Rajendren, S., Vadlamani, P., Wheeler, E. C., ... Hundley, H. A. (2017). The C. elegans neural editome reveals an ADAR target mRNA required for proper chemotaxis. *Elife, 6*. https://doi.org/10.7554/eLife.28625.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England), 29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Eisenberg, E., & Levanon, E. Y. (2018). A-to-I RNA editing – Immune protector and transcriptome diversifier. *Nature Reviews. Genetics, 19*(8), 473–490. https://doi.org/10.1038/s41576-018-0006-1.

Erdmann, E. A., Mahapatra, A., Mukherjee, P., Yang, B., & Hundley, H. A. (2021). To protect and modify double-stranded RNA – The critical roles of ADARs in development, immunity and oncogenesis. *Critical Reviews in Biochemistry and Molecular Biology, 56*(1), 54–87. https://doi.org/10.1080/10409238.2020.1856768.

Farrell, R. E. (2005). Nucleic acid probe technology. In In. R. E. Farrell (Ed.). *RNA methodologies* (pp. 285–316) Academic Press. https://doi.org/10.1016/B978-012249696-7/50013-5.

Flati, T., Gioiosa, S., Spallanzani, N., Tagliaferri, I., Diroma, M. A., Pesole, G., ... Castrignano, T. (2020). HPC-REDItools: A novel HPC-aware tool for improved large scale RNA-editing analysis. *BMC Bioinformatics, 21*(10), 353. https://doi.org/10.1186/s12859-020-03562-x.

Freund, E. C., Sapiro, A. L., Li, Q., Linder, S., Moresco, J. J., Yates, J. R., 3rd, & Li, J. B. (2020). Unbiased identification of trans regulators of ADAR and A-to-I RNA editing. *Cell Reports, 31*(7), 107656. https://doi.org/10.1016/j.celrep.2020.107656.

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews. Genetics, 17*(6), 333–351. https://doi.org/10.1038/nrg.2016.49.

Grabski, D. F., Broseus, L., Kumari, B., Rekosh, D., Hammarskjold, M. L., & Ritchie, W. (2021). Intron retention and its impact on gene expression and protein diversity: A review and a practical guide. *Wiley Interdisciplinary Reviews-RNA, 12*(1), e1631. https://doi.org/10.1002/wrna.1631.

Han, L., Diao, L., Yu, S., Xu, X., Li, J., Zhang, R., ... Liang, H. (2015). The genomic landscape and clinical relevance of A-to-I RNA editing in human cancers. *Cancer Cell, 28*(4), 515–528. https://doi.org/10.1016/j.ccell.2015.08.013.

Herzner, A. M., Khan, Z., Van Nostrand, E. L., Chan, S., Cuellar, T., Chen, R., ... Albert, M. L. (2021). ADAR and hnRNPC deficiency synergize in activating endogenous dsRNA-induced type I IFN responses. *The Journal of Experimental Medicine, 218*(9), https://doi.org/10.1084/jem.20201833.

Hsiao, Y. E., Bahn, J. H., Yang, Y., Lin, X., Tran, S., Yang, E. W., ... Xiao, X. (2018). RNA editing in nascent RNA affects pre-mRNA splicing. *Genome Research, 28*(6), 812–823. https://doi.org/10.1101/gr.231209.117.

Hwang, T., Park, C. K., Leung, A. K., Gao, Y., Hyde, T. M., Kleinman, J. E., ... Weinberger, D. R. (2016). Dynamic regulation of RNA editing in human brain development and disease. *Nature Neuroscience, 19*(8), 1093–1099. https://doi.org/10.1038/nn.4337.

Karagianni, K., Bibi, A., Made, A., Acharya, S., Parkkonen, M., Barbalata, T., ... CA, E. U.-C. C. A. (2024). Recommendations for detection, validation, and evaluation of RNA editing events in cardiovascular and neurological/neurodegenerative diseases. *Molecular Therapy Nucleic Acids, 35*(1), 102085. https://doi.org/10.1016/j.omtn.2023.102085.

Kim, D. D., Kim, T. T., Walsh, T., Kobayashi, Y., Matise, T. C., Buyske, S., & Gabriel, A. (2004). Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Research, 14*(9), 1719–1725. https://doi.org/10.1101/gr.2855504.

Kofman, E., Yee, B., Medina-Munoz, H. C., & Yeo, G. W. (2023). FLARE: A fast and flexible workflow for identifying RNA editing foci. *BMC Bioinformatics, 24*(1), 370. https://doi.org/10.1186/s12859-023-05452-4.

Levanon, E. Y., Eisenberg, E., Yelin, R., Nemzer, S., Hallegger, M., Shemesh, R., ... Jantsch, M. F. (2004). Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nature Biotechnology, 22*(8), 1001–1005. https://doi.org/10.1038/nbt996.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Genome Project Data Processing, S. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics (Oxford, England), 25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Li, J. B., Levanon, E. Y., Yoon, J. K., Aach, J., Xie, B., Leproust, E., ... Church, G. M. (2009). Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science (New York, N. Y.), 324*(5931), 1210–1213. https://doi.org/10.1126/science.1170995.

Liao, Y., Smyth, G. K., & Shi, W. (2013). The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research, 41*(10), e108. https://doi.org/10.1093/nar/gkt214.

Licht, K., Kapoor, U., Amman, F., Picardi, E., Martin, D., Bajad, P., & Jantsch, M. F. (2019). A high resolution A-to-I editing map in the mouse identifies editing events controlled by pre-mRNA splicing. *Genome Research, 29*(9), 1453–1463. https://doi.org/10.1101/gr.242636.118.

Ma, Y., Dammer, E. B., Felsky, D., Duong, D. M., Klein, H. U., White, C. C., ... De Jager, P. L. (2021). Atlas of RNA editing events affecting protein expression in aged and Alzheimer's disease human brain tissue. *Nature Communications, 12*(1), 7035. https://doi.org/10.1038/s41467-021-27204-9.

Morlan, J. D., Qu, K., & Sinicropi, D. V. (2012). Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. *PLoS One,* 7(8), e42882. https://doi.org/10.1371/journal.pone.0042882.

Paz, N., Levanon, E. Y., Amariglio, N., Heimberger, A. B., Ram, Z., Constantini, S., ... Rechavi, G. (2007). Altered adenosine-to-inosine RNA editing in human cancer. *Genome Research, 17*(11), 1586–1595. https://doi.org/10.1101/gr.6493107.

Paz-Yaacov, N., Bazak, L., Buchumenski, I., Porath, H. T., Danan-Gotthold, M., Knisbacher, B. A., ... Levanon, E. Y. (2015). Elevated RNA editing activity is a major contributor to transcriptomic diversity in tumors. *Cell Reports, 13*(2), 267–276. https://doi.org/10.1016/j.celrep.2015.08.080.

Piechotta, M., Naarmann-de Vries, I. S., Wang, Q., Altmuller, J., & Dieterich, C. (2022). RNA modification mapping with JACUSA2. *Genome Biology, 23*(1), 115. https://doi.org/10.1186/s13059-022-02676-0.

Pinto, Y., & Levanon, E. Y. (2019). Computational approaches for detection and quantification of A-to-I RNA-editing. *Methods (San Diego, Calif.), 156*, 25–31. https://doi.org/10.1016/j.ymeth.2018.11.011.

Price, I. F., Wagner, J. A., Pastore, B., Hertz, H. L., & Tang, W. (2023). C. elegans germ granules sculpt both germline and somatic RNAome. *Nature Communications, 14*(1), 5965. https://doi.org/10.1038/s41467-023-41556-4.

Pritchard, J. (2023). https://www.idtdna.com/pages/education/decoded/article/which-biotin-modification-to-use-.

Quinones-Valdez, G., Tran, S. S., Jun, H. I., Bahn, J. H., Yang, E. W., Zhan, L., ... Xiao, X. (2019). Regulation of RNA editing by RNA-binding proteins in human cells. *Communications Biology, 2*, 19. https://doi.org/10.1038/s42003-018-0271-8.

Rajendren, S., Dhakal, A., Vadlamani, P., Townsend, J., Deffit, S. N., & Hundley, H. A. (2021). Profiling neural editomes reveals a molecular mechanism to regulate RNA editing during development. *Genome Research, 31*(1), 27–39. https://doi.org/10.1101/gr.267575.120.

Rio, D. C. A., Jr, M., Hannon, G. J., & Nilsen, T. W. (2010). RNA – A laboratory manual. *Cold Spring Harbor Protocols, 2010*.

Rodriguez, J., Menet, J. S., & Rosbash, M. (2012). Nascent-seq indicates widespread cotranscriptional RNA editing in Drosophila. *Molecular Cell, 47*(1), 27–37. https://doi.org/10.1016/j.molcel.2012.05.002.

Saldi, T., Riemondy, K., Erickson, B., & Bentley, D. L. (2021). Alternative RNA structures formed during transcription depend on elongation rate and modify RNA processing. e1785 *Molecular Cell, 81*(8), 1789–1801. https://doi.org/10.1016/j.molcel.2021.01.040.

Saldi, T. K., Ash, P. E., Wilson, G., Gonzales, P., Garrido-Lecca, A., Roberts, C. M., ... Link, C. D. (2014). TDP-1, the Caenorhabditis elegans ortholog of TDP-43, limits the accumulation of double-stranded RNA. *The EMBO Journal, 33*(24), 2947–2966. https://doi.org/10.15252/embj.201488740.

Sapiro, A. L., Freund, E. C., Restrepo, L., Qiao, H. H., Bhate, A., Li, Q., ... Li, J. B. (2020). Zinc finger RNA-binding protein Zn72D regulates ADAR-mediated RNA editing in neurons. *Cell Reports, 31*(7), 107654. https://doi.org/10.1016/j.celrep.2020.107654.

Sheynkman, G. M., Tuttle, K. S., Laval, F., Tseng, E., Underwood, J. G., Yu, L., ... Vidal, M. (2020). ORF Capture-Seq as a versatile method for targeted identification of full-length isoforms. *Nature Communications, 11*(1), 2326. https://doi.org/10.1038/s41467-020-16174-z.

Sternberg, P. W., Van Auken, K., Wang, Q., Wright, A., Yook, K., Zarowiecki, M., ... Stein, L. (2024). WormBase 2024: Status and transitioning to Alliance infrastructure. *Genetics, 227*(1), https://doi.org/10.1093/genetics/iyae050.

Szabo, B., Mandl, T. C., Woldrich, B., Diensthuber, G., Martin, D., Jantsch, M. F., & Licht, K. (2024). RNA Pol II-dependent transcription efficiency fine-tunes A-to-I editing levels. *Genome Research, 34*(2), 231–242. https://doi.org/10.1101/gr.277686.123.

Tan, M. H., Li, Q., Shanmugam, R., Piskol, R., Kohler, J., Young, A. N., ... Li, J. B. (2017). Dynamic landscape and regulation of RNA editing in mammals. *Nature, 550*(7675), 249–254. https://doi.org/10.1038/nature24041.

Theil, K., Imami, K., & Rajewsky, N. (2019). Identification of proteins and miRNAs that specifically bind an mRNA in vivo. *Nature Communications, 10*(1), 4205. https://doi.org/10.1038/s41467-019-12050-7.

UniProt, C. (2023). UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research, 51*(D1), D523–D531. https://doi.org/10.1093/nar/gkac1052.

Verwilt, J., Mestdagh, P., & Vandesompele, J. (2023). Artifacts and biases of the reverse transcription reaction in RNA sequencing. *RNA (New York, N. Y.), 29*(7), 889–897. https://doi.org/10.1261/rna.079623.123.

Wang, F., Xu, Y., Wang, R., Zhang, B., Smith, N., Notaro, A., ... Lin, L. (2023). TEQUILA-seq: A versatile and low-cost method for targeted long-read RNA sequencing. *Nature Communications, 14*(1), 4760. https://doi.org/10.1038/s41467-023-40083-6.

Washburn, M. C., Kakaradov, B., Sundararaman, B., Wheeler, E., Hoon, S., Yeo, G. W., & Hundley, H. A. (2014). The dsRBP and inactive editor ADR-1 utilizes dsRNA binding to regulate A-to-I RNA editing across the C. elegans transcriptome. *Cell Reports, 6*(4), 599–607. https://doi.org/10.1016/j.celrep.2014.01.011.

Zhang, P., Zhu, Y., Guo, Q., Li, J., Zhan, X., Yu, H., ... Li, Q. (2023). On the origin and evolution of RNA editing in metazoans. *Cell Reports, 42*(2), 112112. https://doi.org/10.1016/j.celrep.2023.112112.

Zhang, R., Li, X., Ramaswami, G., Smith, K. S., Turecki, G., Montgomery, S. B., & Li, J. B. (2014). Quantifying RNA allelic ratios by microfluidic multiplex PCR and sequencing. *Nature Methods, 11*(1), 51–54. https://doi.org/10.1038/nmeth.2736.

Zhang, T., Li, H., Jiang, M., Hou, H., Gao, Y., Li, Y., ... Liu, Y. X. (2024). Nanopore sequencing: Flourishing in its teenage years. *Journal of Genetics and Genomics = Yi Chuan xue bao*. https://doi.org/10.1016/j.jgg.2024.09.007.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research, 31*(13), 3406–3415. https://doi.org/10.1093/nar/gkg595.