SHAPE–Seq: High-Throughput RNA Structure Analysis

Stefanie A. Mortimer,[1] Cole Trapnell,[2,6] Sharon Aviran,[3,6] Lior Pachter,[1,5] and Julius B. Lucks[4]

[1]Department of Molecular and Cell Biology, University of California, Berkeley, California
[2]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts
[3]Center for Computational Biology and the California Institute for Quantitative Biomedical Research, University of California, Berkeley, California
[4]School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, New York
[5]Department of Mathematics, University of California, Berkeley, California
[6] These authors contributed equally to this work.

ABSTRACT

Knowledge of RNA structure is critical to understanding both the important functional roles of RNA in biology and the engineering of RNA to control biological systems. This article contains a protocol for selective 2′-hydroxyl acylation analyzed by primer extension and sequencing (SHAPE-Seq) that, through a combination of structure-dependent chemical probing and next-generation sequencing technologies, achieves structural characterization of hundreds of RNAs in a single experiment. This protocol is applicable in a variety of conditions, and represents an important tool for understanding RNA biology. The protocol includes methods for the design and synthesis of RNA mixtures for study, and the construction and analysis of structure-dependent sequencing libraries that reveal structural information of the RNAs in the mixtures. The methods are generally applicable to studying RNA structure and interactions in vitro in a variety of conditions, and allows for the rapid characterization of RNA structures in a high-throughput manner. *Curr. Protoc. Chem. Biol.* 4:275-297 © 2012 by John Wiley & Sons, Inc.

Keywords: RNA structure • next-generation sequencing • chemical probing • high throughput

INTRODUCTION

The emergence of high-throughput technologies that measure the sequence diversity and abundance of RNAs has revealed a remarkable array of RNA function that affects almost every aspect of the cell (Amaral et al., 2008; Mortazavi et al., 2008). However, comparatively little is known about how these sequences determine the structures of these RNAs, which is a key aspect in understanding their function. The structures of many RNAs have been individually studied in great detail using high-resolution techniques such as nuclear magnetic resonance spectroscopy and X-ray crystallography (Butcher and Pyle, 2011). However, these techniques can be laborious and limited in their applicability, since they require RNAs to be short, highly structured, and in high abundance and purity. In order to comprehensively understand the RNA sequence-structure-function relationship that is so important to biology, and use this relationship in engineering biological systems (Isaacs et al., 2006), structure characterization techniques are needed that can work on complex mixtures of RNAs in a high-throughput fashion.

Selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq) is a technique that combines advances in structure-dependent chemical probing of RNAs (Weeks and Mauger, 2011) with the throughput of next-generation sequencing (NGS)

Current Protocols in Chemical Biology 4: 275-297, December 2012
Published online December 2012 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/9780470559277.ch120019

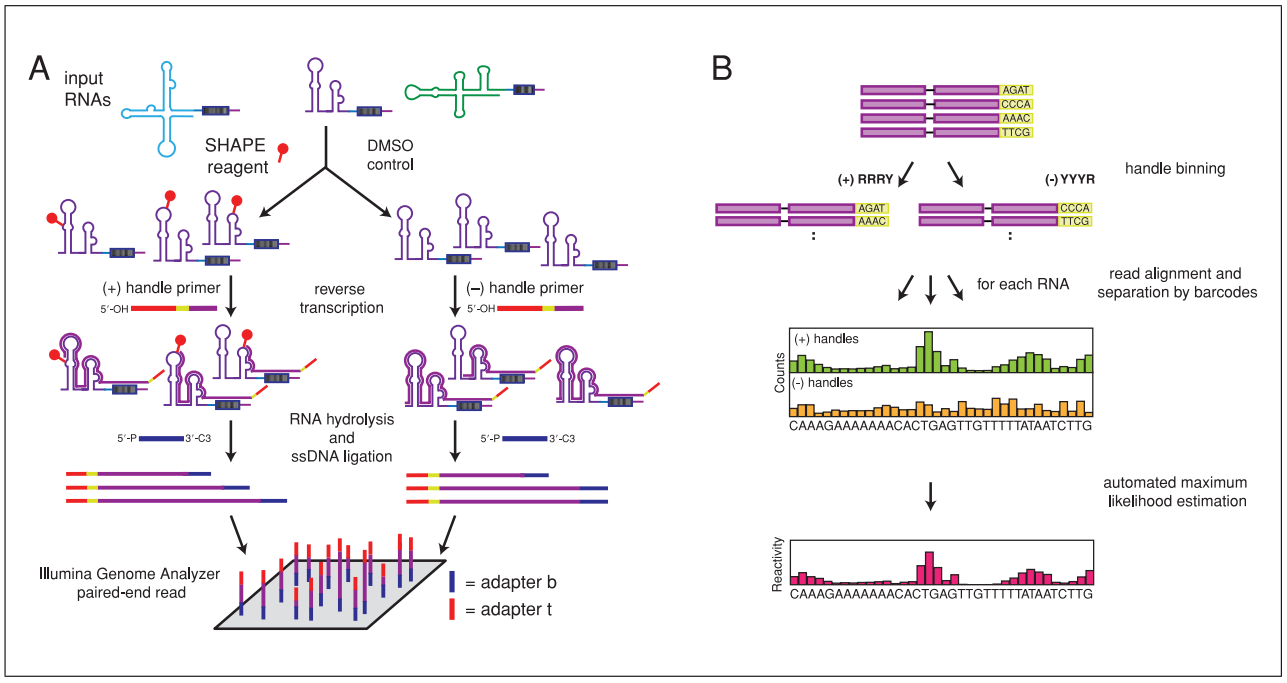

Figure 1 An overview of the SHAPE-Seq experiment. (**A**) The experimental protocol starting after in vitro RNA synthesis (Basic Protocol 1), and including the steps of Basic Protocol 2. (**B**) SHAPE-Seq data analysis as described in Basic Protocol 3. Figure adapted from Lucks et al. (2011).

to simultaneously gather nucleotide-resolution structural information for each RNA sequence in a mixture (Fig. 1). An in vitro SHAPE-Seq experiment begins by folding a mixture of RNAs in defined conditions and subjecting this mixture to a SHAPE reagent (+), or solvent without reagent as a control (–). The SHAPE reagent preferentially modifies the RNA molecules at the 2′-ribose hydroxyl groups of nucleotides that are conformationally flexible, or unconstrained by intra- or inter-molecular interactions. Reverse transcription (RT) of these RNAs is blocked by the modification, resulting in a pool of cDNAs whose length distribution reflects the distribution of modifications across each RNA.

In SHAPE-Seq, these cDNAs are then analyzed with NGS. NGS is a suite of sequencing platforms that can simultaneously sequence billions of short DNA fragments, and that has been used in wide range of applications such as genome sequencing and transcriptome profiling (Mortazavi et al., 2008). In order to apply NGS to sequencing of cDNAs, specific adapter DNA sequences must be attached to each end. These adapter sequences are used to immobilize DNA fragments at specific locations so that the sequencing process of many fragments can be simultaneously imaged. For an excellent review of NGS technology, see Metzker (2009). Readers interested in a review of applications of NGS for DNA and cDNA sequencing should consult Garber et al. (2011) and Shendure and Ji (2008). In SHAPE-Seq, one adapter is included as a tail of the RT primer, while the other is added via a single-stranded DNA ligation step. After only several rounds of PCR amplification, the double-stranded library is sequenced using Illumina's paired-end read capability.

Paired-end sequencing returns up to 150 bp of sequence reads from each end of the cDNAs. These paired-end reads are then mapped to the RNA sequences. One read of the pair (3′ cDNA end) is used to identify the location of the RT stop, while the other (5′ cDNA end) is used to map a barcode embedded in the RNA sequence (see Strategic Planning) that uniquely identifies the RNA in the pool from which the pair was generated (see Fig. 2). A second index barcode called a 'handle' in one of the

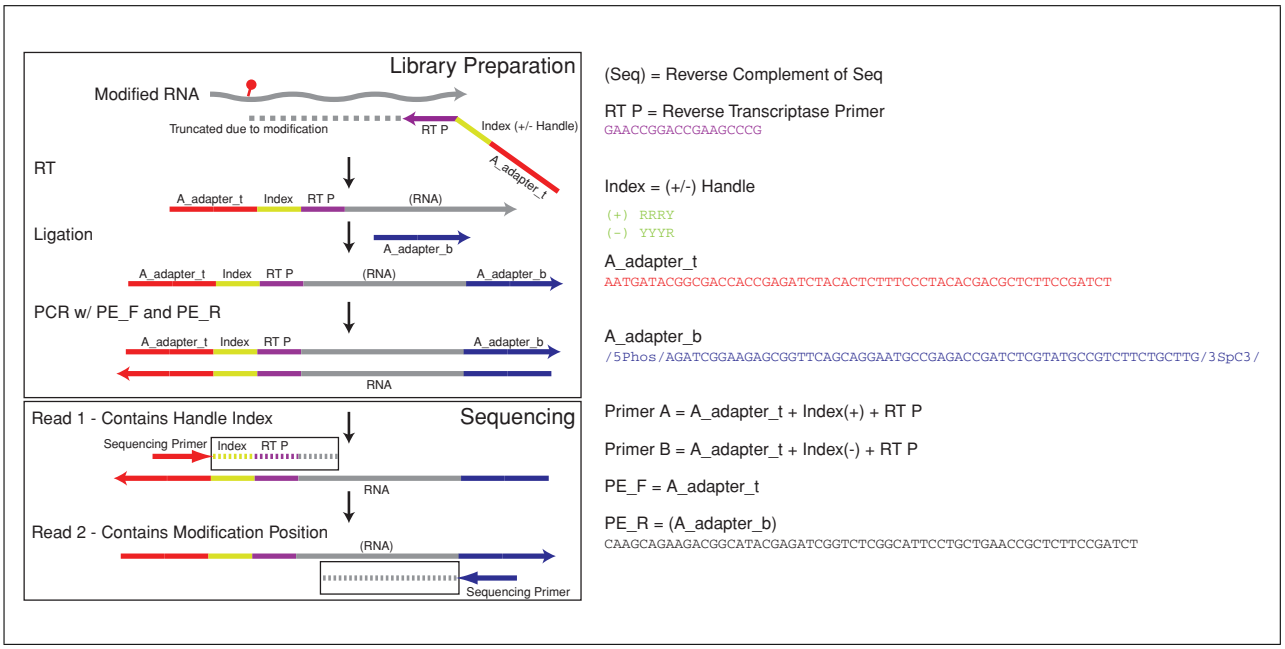

Figure 2 Preparation of a structure-specific cDNA library followed by paired-end sequencing. Arrows indicate 3′ ends. 5′ and 3′ modification codes (IDT) are included in the A_adapter_b sequence. See Basic Protocols 2 and 3 for details.

reads is used to distinguish the modification (+) from the control (–) reaction. For each RNA in the pool, the distribution of RT stop locations for the (+) and (–) reactions are then fed into an automated maximum-likelihood estimation framework that calculates the SHAPE 'reactivity' for each nucleotide that is most consistent with the observed data. A nucleotide's SHAPE reactivity reflects its propensity to be modified by the SHAPE reagent. High reactivities (reflecting more observed modifications) are thus associated with nucleotides that are unconstrained and most likely single-stranded, while low reactivities (reflecting fewer observed modifications) are associated with nucleotides that are constrained by interactions, such as Watson-Crick base-pairing.

The end result of a SHAPE-Seq experiment is a nucleotide-resolution reactivity spectrum for each RNA sequence in the pool. With the techniques described in the following protocols, this may be done for hundreds of RNAs present in a single mixture. These reactivities reflect structural information about the RNAs, and can be combined with other biological insights to determine key structural features that may be responsible for a particular function. There have been numerous applications of SHAPE in the literature to examine areas such as the structure of small RNAs (Kladwang et al., 2011), the structure of ribosomal RNAs (Deigan et al., 2008), the dynamics of tertiary folding (Mortimer and Weeks, 2008), ligand binding of RNA aptamers (Steen et al., 2010), and even the structure of whole viral RNA genomes (Wilkinson et al., 2008). In some cases, it is also appropriate to use experimental reactivities as constraints in RNA folding algorithms to improve their accuracy (Deigan et al., 2008; Kladwang et al., 2011).

This article describes how to perform a typical SHAPE-Seq experiment and data analysis in several protocols. Basic Protocol 1 describes the preparation of an RNA library suitable for SHAPE-Seq. Basic Protocol 2 describes the folding, modification, conversion to cDNA, and preparation of a SHAPE-Seq library. It contains specific protocols for each major step and represents the experimental portion of the SHAPE-Seq protocol. Basic Protocol 3 describes how to generate SHAPE reactivities for each RNA in a mixture using a bioinformatic pipeline that first aligns the sequencing reads to the known RNA sequences and then uses statistics of the alignment with a mathematical model of the SHAPE-Seq experiment to calculate a reactivity spectrum for each RNA. While this

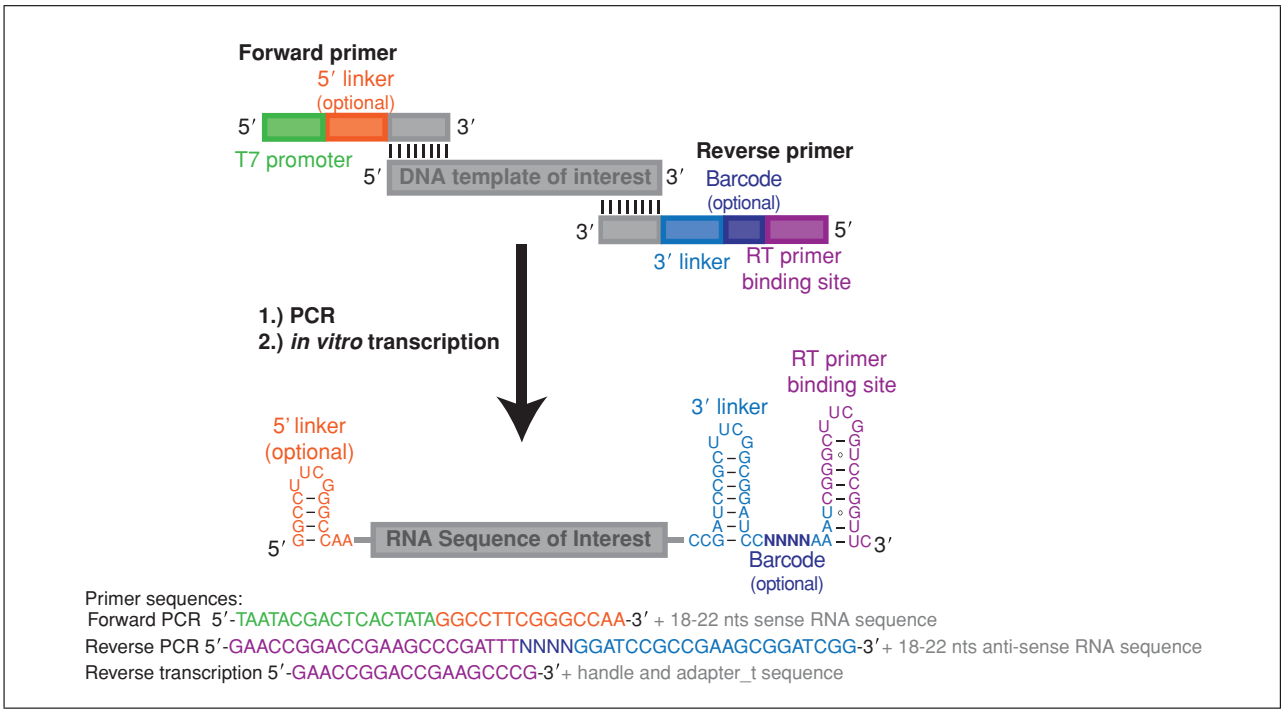

Figure 3 The structure cassette used in SHAPE-Seq library preparation. 1. PCR is used to add desired structure cassette sequences to the DNA template of the RNA sequence of interest; and 2. In vitro transcription of the PCR-generated DNA template generates the RNA with desired flanking sequences. If the 5′ structure cassette is not desired, the full T7 promoter sequence (TAATACGACTCACTATAGG) should be included in the forward PCR primer.

article is focused on RNA structure as characterized by structure-dependent modification by SHAPE chemistry, it should be generally applicable to the wide array of existing structure-dependent chemical probes (Weeks, 2010).

STRATEGIC PLANNING

RNA Library Construct Design

SHAPE-Seq requires modified RNAs to be converted into cDNAs using reverse transcription. Reverse transcriptase in turn requires a single-stranded DNA (ssDNA) primer, which binds to a specific location on the RNA template to initiate reverse transcription (RT). Designing an RT primer involves taking into consideration the binding energy between the primer and template, and any RNA structures that may compete with primer binding, both within the primer and within the RNA template. It is highly recommended that RT primers be verified to be able to bind and initiate extension on the RNAs being examined, using a technique such as polyacrylamide gel electrophoresis (PAGE), before proceeding with full SHAPE-Seq library construction.

To perform SHAPE-Seq on a mixture of different RNAs requires multiple RT primers that can bind to each RNA. One way to simplify and streamline SHAPE-Seq is to incorporate a single convenient RT primer binding site into each RNA so that the same RT primer may be used on the whole mixture. This protocol relies on the 3′ structure cassette initially developed for SHAPE analysis using capillary electrophoresis for this purpose (Merino et al., 2005; also see Fig. 3). The 3′ structure cassette is incorporated into the 3′ end of each RNA in the mixture by including the DNA sequence as part of the template used for in vitro transcription. The structure cassette itself is designed to fold into two GC-rich hairpin structures so as to minimize structural interference with the RNA of interest. The end of the 3′ structure cassette contains the binding site of an

RT primer that has been used extensively in SHAPE analysis, and is highly robust when used with the RT protocol outlined here.

We emphasize that use of the 3′ structure cassette is not required to perform SHAPE-Seq, but it is highly convenient since the same RT primer can be used for the whole mixture at once. The 3′ structure cassette also provides a convenient location to tag RNAs with individual barcodes. If the mixture of RNAs is diverse enough to allow sequencing reads to be uniquely aligned to each RNA, then barcodes are not required. If this is not the case, for example with a mixture of RNAs that only differ by a few nucleotides, a barcoded structure cassette may be used for each RNA, which facilitates read assignment and alignment as described below (Fig. 3).

It is always possible that addition of the 3′ structure cassette may interfere with the RNA fold, and care should be taken during data analysis to check for this. One strategy to both assess and circumvent this problem would be to create internal replicates by having the same RNA tagged with multiple barcodes so that interference can be explicitly monitored and data for non-interfering barcodes kept for further analysis. For RNAs that are particularly small (<100 nt), it may also be advantageous to include the 5′ structure cassette developed for SHAPE analysis to displace the full-length RT product and increase the quality of the reverse transcription for the RNA sequence of interest (Merino et al., 2005; Fig. 3).

SHAPE Reagent Choice

There are several SHAPE reagents that have the same overall structure-dependent modification property, but differ greatly in the rate of hydrolysis with water and reactivity with RNA 2′-OH groups. In practice, two reagents are most commonly used. 1-Methyl-nitroisatoic anhydride (1M7; Mortimer and Weeks, 2007) is a versatile reagent with a hydrolysis half-life of 14 sec, meaning the modification reaction must be allowed to proceed for 70 sec at 37°C for the reaction to go to completion. Its advantages are that it can be used in a variety of buffer conditions (including small molecules or proteins that may bind to RNA aptamers), and that it is a user-friendly reagent with a reasonably long shelf life (∼2 weeks in DMSO and indefinitely when stored with desiccant at −20°C). Its primary disadvantage is that there are currently no commercially available sources; however, a synthesis method is included below in the Support Protocol.

Benzoyl cyanide (BzCN) has a much shorter hydrolysis half-life of 0.25 sec. Its advantages are that it is a useful reagent for probing RNA folding pathways on the second timescale (Mortimer and Weeks, 2008), and it is commercially available. Its primary disadvantage is that it can be harder to work with for the beginner due to its high sensitivity to trace amounts of water. It is recommended for beginners to first perform SHAPE-Seq experiments with 1M7, then move to BzCN if required once the technique has been mastered. Note that there may be slight differences in observed reactivities with each of these reagents (Mortimer and Weeks, 2008), making it necessary to specify the reagent used when reporting results. Another commercially available reagent for SHAPE is N-methylisatoic anhydride (NIMA; Merino et al., 2005). It is a slower reacting reagent and can be sensitive to Mg^{2+} concentration, but can also be used for this protocol.

Regardless of the reagent choice, it is typical to tune the concentration of SHAPE reagent used so that on average each RNA molecule is modified once, i.e., under ‘single-hit’ conditions. A lower modification rate will not yield sufficient signal, and a higher modification rate runs the risk of causing excessive signal decay that can confound data analysis. While the SHAPE-Seq data analysis algorithms (Aviran et al., 2011a,b; Basic Protocol 3) can account for such deviations from single-hit conditions, it is recommended

to at least aim for this condition when beginning SHAPE experiments on RNAs that have not previously been examined. This may take some trial and error; however, the concentrations suggested in Basic Protocol 2 may serve as a useful starting point.

BASIC PROTOCOL 1

PREPARATION OF A BARCODED RNA LIBRARY THROUGH PCR AND IN VITRO TRANSCRIPTION

This portion of the protocol focuses on creating RNA libraries via in vitro transcription using T7 RNA polymerase. A template for transcription that contains a T7 promoter and other flanking sequences necessary for downstream steps can be readily prepared via PCR. Once a template is created for each RNA, transcription with T7 RNA polymerase is used to generate RNA with appropriate sequences for bar coding (if necessary) and SHAPE-Seq library preparation.

Materials

Forward and reverse PCR primers at 100 μM (custom-synthesized by Integrated DNA Technologies, *https://www.idtdna.com/*)
PCR buffer (without Mg^{2+}; Invitrogen, cat. no. 18067-017)
1 M $MgCl_2$
10 mM dNTP mixture (dATP, dCTP, dGTP, dTTP; New England Biolabs, cat. no. N0447L)
Template DNA for PCR: dsDNA containing RNA coding sequence of interest (circular or linear)
Taq DNA polymerase (Invitrogen, cat. no. 10342-020)
4 M NaCl
100% ethanol
1 M Tris·Cl, pH 8.0
1 M DTT (Invitrogen, cat. no P2325)
Spermidine (Sigma, cat. no. 85558)
Triton X-100 (Sigma, cat. no. T8787)
NTPs: ATP, Sigma, cat no. A7699-1G), UTP (Sigma, cat. no. U6625-1G), CTP (Sigma, cat. no. C1506-1G), GTP (Sigma, cat. no. G8877-1G)
RNase Inhibitor (Promega, cat. no. N2615)
T7 RNA polymerase (Ambion, cat. no. 18033-019)
3 M sodium acetate, pH 5.2
Stop dye (see recipe)
Nuclease-free water
TE buffer (see recipe)

Additional reagents and equipment for PCR (Kramer and Coen, 2001), agarose gel electrophoresis (Voytas, 2001), and purification of RNA by denaturing polyacrylamide gel electrophoresis (Ellington and Pollard, 2001)

Perform PCR to generate a barcoded DNA template for transcription

1. Design PCR primers containing the 3′ structure cassette and desired barcodes (4 nt or longer, see Fig. 3). Have primers synthesized by a vendor such as Integrated DNA Technologies (IDT).

   It may be helpful to use more than one barcode per RNA for multiple replicates (i.e., create two barcoded versions of the same RNA), and sequences should be designed to minimize interaction with the target RNA sequence by looking for complementary sequences. Forward primers should contain the T7 promoter sequence (and 5′ structure cassette if desired; see Fig. 3).

2. Perform a large-scale (0.5 to 1 ml) PCR (Kramer and Coen, 2001) to create a DNA template for transcription containing the desired flanking sequences (T7 promoter and structure cassettes).

   We typically do 1-ml reactions, which are each enough for three transcription reactions (1.5 ml each).

   a. The assembled PCRs contain:
   1× PCR buffer
   2.5 mM $MgCl_2$
   200 μM each dNTP
   500 nM each forward and reverse primer
   1 μg template
   25 U *Taq* polymerase.

   b. Any typical PCR protocol should work, however we recommend performing the reaction with the following cycling conditions:

1 cycle:	3 min	94°C	(initial denaturation)
34 cycles:	45 sec	94°C	(denaturation)
	30 sec	55°C	(annealing)
	1 min	72°C	(extension)
1 cycle:	15 min	72°C	(final extension).

   The exact annealing temperature will depend upon primer design.

3. Recover the PCR product by ethanol precipitation.

   a. Add 1/10 vol 4 M NaCl and 3 vol of 100% ethanol (100%).
   b. Place at −80°C for 20 min, then microcentrifuge 20 min at maximum speed, 4°C.
   c. Discard supernatant with careful pipetting.

      The PCR product can be resuspended in 150 μl water or TE buffer for long-term storage.

   d. Before proceeding, confirm that the correctly sized product (RNA plus T7 promoter plus structure cassettes) has been obtained by using agarose gel electrophoresis (Voytas, 2001). Also, quantify the amount of DNA by measuring the A_{260}; typically, concentrations range from 1 to 2 μg/μl.

Perform in vitro transcription to prepare a barcoded RNA library

In vitro transcription with T7 RNA polymerase is used to generate at least 50 pmol of RNA. If a mixture of RNAs is being probed, then the 50 pmol should be split among them. We find that a 1.0- to 1.5-ml scale that uses 50 μl of PCR product from above is typically sufficient for several libraries.

4. In a 1.6-ml microcentrifuge tube, prepare transcription reactions containing:

   40 mM Tris·Cl, pH 8.0
   20 mM $MgCl_2$
   10 mM DTT
   2 mM spermidine
   0.01 % Triton X-100
   5 mM each NTP
   120 U RNase inhibitor
   0.1 mg/ml T7 RNA polymerase
   50 μl PCR product from step 3 above (∼50 to 100 μg).

   Keep ingredients on ice.

Recover, purify, and quantitate RNA

5. Incubate at 37°C for 6 hr to overnight; successful transcriptions will show a white precipitate as the reaction proceeds. Microcentrifuge 5 min maximum speed, room temperature. Retain the supernatant and discard the pellet.

6. Recover the RNA by ethanol precipitation as follows.

   a. Add 1/10 vol of 3 M sodium acetate, pH 5.2, and 3 volumes of 100% ethanol. Place at −80°C for 30 min, then microcentrifuge 30 min at maximum speed, 4°C. Carefully remove and discard supernatant by pipetting.

      It may be necessary to use several tubes per transcription for the ethanol precipitation.

   b. Add 50 µl stop dye to the pellet in each tube, and heat at 95°C, if necessary, to dissolve the RNA. Combine tubes from the same RNA before loading the purifying gel in the next step.

7. Purify the RNA by denaturing polyacrylamide gel electrophoresis (Ellington and Pollard, 2001).

   The appropriate percentage of acrylamide in the gel will depend on the size of the RNA, but 6% to 12% gels are commonly used for RNA ranging in length between 50 and 500 nucleotides. We use gels containing 7 M urea and a ratio of 29:1 acrylamide:bisacrylamide.

8. Visualize the RNA by UV shadowing and subsequently excise the RNA-containing pieces from the gel.

   UV shadowing is performed by shining a hand-held UV light over the gel while it is placed on top of a TLC plate that contains a fluorescent compound.

   In order to prevent RNA damage during this process, it is highly recommended to use a sacrifice lane (which contains a small amount of the same RNA that will not be recovered but be used to identify band locations) to identify band locations with UV shadowing to limit exposure to the RNAs that will be used in further experiments (Kladwang et al., 2012).

9. Crush the RNA-containing gel pieces and add 2 volumes of nuclease-free water. Allow RNA to passively elute overnight with gentle rocking at 4°C.

10. Filter off the gel pieces from the water and recover the RNA via ethanol precipitation according to step 3 above. Dissolve the resulting RNA pellet in a suitable amount of TE buffer (50 to 200 µl).

    This is the working stock of RNA for Basic Protocol 2. RNA can be stored for several months at −20°C or −80°C.

11. Quantify the RNA concentration using the following formula:

    $$[\text{RNA}, \mu\text{M}] = (A_{260\,\text{nm}}/8.6N) \times (\text{dilution factor}) \times (1000\ \mu\text{M}/1\ \text{mM})$$

    where N = number of nucleotides of the RNA.

BASIC PROTOCOL 2

PREPARATION OF A STRUCTURE-SPECIFIC cDNA LIBRARY FOR PAIRED-END SEQUENCING

This protocol describes how to modify the RNA in a structure-dependent fashion, and subsequently transform the modified RNA into a cDNA library suitable for deep sequencing (Fig. 2). After the RNA is modified by a SHAPE reagent, it undergoes first-strand synthesis (reverse transcription) to generate cDNAs of various lengths corresponding to the sites of modification. This first-strand synthesis step incorporates handle sequences necessary to identify fragments as coming from either the (+) or (–) pools, and also one of two adapters necessary for the Illumina paired end sequencing, A_adapter_t (Fig. 2).

Following this is an ssDNA ligation to incorporate the second adapter required for sequencing, A_adapter_b. We recommend performing all steps below using a thermal cycler with heated lid to facilitate experimental ease and to prevent sample evaporation.

Materials

RNA (Basic Protocol 1)
Nuclease-free water
3.3× folding buffer (see recipe)
SHAPE reagent (one of the following):
  65 mM 1-methyl-7-nitroisatoic anhydride (1M7; see recipe)
  65 mM NMIA (see recipe)
  400 mM benzoyl cyanide (BzCN; see recipe)
3 M sodium acetate, pH 5.2
20 mg/ml glycogen (Invitrogen, cat. no 10814-010)
100% ethanol
3 μM Primer A and B (custom-synthesized by Integrated DNA Technologies, *https://www.idtdna.com/*; see Fig. 2)
Enzyme master mix (see recipe)
SuperScript III reverse transcriptase (Invitrogen, cat. no. 18080-044)
4 M NaOH
10× CircLigase ligation buffer (see recipe)
50 mM $MnCl_2$
1 mM ATP
A_Adapter_b with 5′ phosphate and 3′ C3 (propyl) modification at 100 μM (custom-synthesized by Integrated DNA Technologies, *https://www.idtdna.com/*, PAGE-purified; see Fig. 2)
CircLigase ssDNA ligase (Epicentre Biotechnologies, cat. no. CL4111K)
Agencourt XP purification beads (Beckman Coulter, cat. no. A63880)
70% ethanol
TE buffer (see recipe)
5× Phusion high-fidelity buffer (NEB, cat. no M0530S)
10 mM dNTP mixture (dATP, dCTP, dGTP, dTTP; New England Biolabs, cat. no. N0447L)
100 μM Primers PE_F and PE_R at 100 μM (IDT) (Fig. 2)
Phusion high fidelity DNA polymerase (NEB, cat. no M0530S)

Thermal cycler with heated lid
Agencourt XP magnetic stand (Beckman Coulter, cat. no. A29182)
Agilent Bioanalyzer high-sensitivity DNA kit (Agilent, cat. no 5067-4626)

RNA folding and modification

1. Combine 50 pmol of RNA from Basic Protocol 1 with nuclease-free water for a total volume of 60 μl.
2. Heat at 95°C in a thermal cycler for 2 min to denature the RNA.
3. Cool on ice for 1 min.
4. Add 30 μl 3.3× folding buffer and incubate at 37°C for 20 min.
5. During step 4, aliquot 5 μl of 65 mM NMIA, 65 mM 1M7, or 400 mM BzCN dissolved in anhydrous DMSO in a tube labeled "+", and 5 μl anhydrous DMSO in a tube labeled "–".

6. After the completion of step 4, add 45 μl of the folded RNA solution into each tube from step 5. Place at 37°C for an additional 70 sec (for 1M7), 45 min (for NMIA), or no additional time (for BzCN).

   If 1M7 is used, then at the end of this time, the (+) solution should have changed color from a bright yellow to a dull orange, indicating hydrolysis of the SHAPE reagent. There are no accompanying color changes associated with NMIA or BzCN.

7. Add 50 μl of nuclease-free water to each tube. Ethanol-precipitate the modified RNA by adding 5 μl sodium acetate (3 M), 1 μl glycogen (20 mg/ml), and 400 μl ice-cold ethanol (100%). Leave at −80°C for 30 min, then microcentrifuge 30 min at maximum speed, 4°C (it may be necessary to transfer the solutions to larger tubes for ethanol precipitation). Be careful when removing the supernatant; the pellet will be very small and easily dislodged. Resuspend each pellet in 10 μl of nuclease-free water, and place in tubes appropriate for further steps in the thermal cycler as outlined below. Place tubes on ice while preparing for the subsequent cDNA synthesis.

   The resuspended RNA pellet can be stored at −20°C overnight prior to cDNA synthesis if necessary.

Addition of adapter and handle sequences via first strand cDNA synthesis

8. Add 3 μl of 3 μM primer A to the "+" tube and 3 μl of primer B to the "−"tube on ice (i.e., only one primer is added per tube).

9. Place tubes at 95°C for 2 min to denature primers and RNA, then heat at 65°C for 5 min to anneal the primers.

10. Place tubes on ice and add 6 μl of enzyme master mix followed by 1 μl of SuperScript III reverse transcriptase to each tube. In the thermal cycler, place tubes at 45°C for 1 min, then incubate at 52°C for 25 min, 65°C for 5 min, and hold at 4°C.

11. Add 1 μl 4 M NaOH to each tube and incubate at 95°C for 5 min to hydrolyze the RNA.

12. Ethanol precipitate the cDNA in each tube separately by adding 63 μl of 100% ethanol, keeping at −80°C for 20 min, and then microcentrifuging 20 min at maximum speed, 4°C. Remove the supernatant and resuspend the pellet in each tube in 73 μl of nuclease-free water.

    The cDNA can be stored at −20°C for several days if necessary.

Attach second adapter via ssDNA ligation

13. Add the following ligation reagents to the 73 μl of cDNA:

    10 μl 10× CircLigase ligation buffer
    5 μl 50 mM $MnCl_2$
    5 μl 1 mM ATP
    5 μl 100 μM A_adapter_b
    2 μl 100 U/μl CircLigase ssDNA ligase.

14. Incubate at 68°C for 6 hr, followed by 80°C for 10 min to inactivate ligase. Place at 4°C.

15. Ethanol precipitate tubes separately by adding 10 μl of 3 M sodium acetate and 300 μl 100% ethanol, incubate at −80°C for 30 min and centrifuge at maximum speed at 4°C for 30 min, then remove supernatant from the pellets. Resuspend each pellet in 10 μl water, and then combine solutions for a total of 20 μl.

16. Remove excess A_adapter_b using Agencourt AmpureXP bead purification as follows.

    a. Add 36 μl of beads to the solution and gently pipet ten times to mix.

       IMPORTANT NOTE: *Gently invert the bead solution before this step to ensure proper bead suspension.*

    b. Incubate the mixture at room temperature for 5 min.

    c. Carefully place the tube on a magnetic stand and let beads separate for 2 min.

    d. While the tube is still on the stand, remove supernatant and gently wash the beads twice using 200 μl of freshly made 70% ethanol. Incubate for 30 sec before aspirating and discarding the ethanol. Let dry for 3 min.

    e. Remove the tube from the magnetic stand, add 40 μl of TE buffer, place tube back on magnetic stand, and let beads separate. Remove supernatant which contains the library.

       The library can be stored at −20°C for several days to months if necessary.

Library enrichment, quality control, and paired-end sequencing

17. Assemble a 100-μl PCR reaction that contains:

    20 μl of 5× Phusion buffer
    71 μl water
    2 μl 10 mM dNTP mix (10 mM each dNTP)
    2 μl 100 μM PE_F primer
    2 μl 100 μM PE_R primer
    2 μl SHAPE-Seq library from step 16
    1 μl Phusion DNA polymerase.

    Split into two tubes labeled 9 and 12.

    The tube labeled 9 will be subjected to 9 cycles of PCR in step 18, while the tube labeled 12 will be subjected to 12 cycles of PCR.

18. Perform PCR with the following thermal cycling program:

1 cycle:	30 sec	98°C	(initial denaturation)
9 or 12 cycles:	10 sec	98°C	(denaturation)
	30 sec	65°C	(annealing)
	30 sec	72°C	(extension)
1 cycle:	15 min	72°C	(final extension).

19. Purify PCR reactions using Agencourt AmpureXP beads according to step 16, except use 90 μl beads.

20. Check the quality of the products: run 1 μl of each reaction on an Agilent Bioanalyzer using a high-sensitivity DNA chip.

    A good library is expected to yield distinct peaks that show clear enrichment between the 9- and 12-cycle amplifications (Fig. 4). Some sequencing facilities will require concentration quantification with qPCR before sequencing. In our experience, library concentrations in the range of 10 to 20 nM have been adequate.

    Standard Illumina paired-end sequencing can now be performed on the 9-cycle amplification product using 36 to 110 bp reads per end. Some sequencing centers may require their own quality control procedures prior to sequencing.

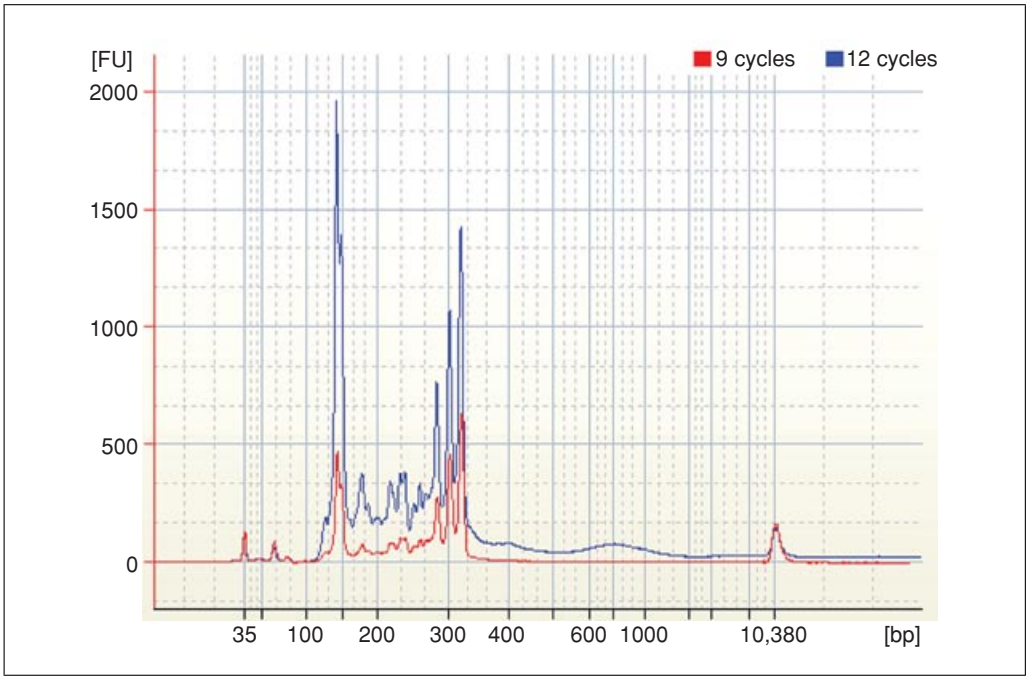

Figure 4 SHAPE-Seq library quality control using an Agilent Bioanalyzer. A library of the RNase P specificity domain (Lucks et al., 2011) was analyzed following step 20 of Basic Protocol 2. Clearly defined peaks that are enriched from 9 (red) to 12 (blue) cycles of PCR amplification are an indication of a high-quality library. Furthermore, a large peak at approximately the length of fully extended product (here 326 bp) is expected. Other well-defined peaks will be dependent on the particular RNAs under study.

BASIC PROTOCOL 3

READ ALIGNMENT AND CALCULATION OF SHAPE REACTIVITIES

This protocol describes how to process the reads returned from a paired-end Illumina sequencing run into nucleotide reactivities for each RNA. Paired-end sequencing returns two text files in FASTQ format typically called `s_X_sequence_1.fq` and `s_X_sequence_2.fq`. The FASTQ format is an augmented form of the ubiquitous FASTA sequence format, and consists of four lines of information for each sequencing read, as illustrated below:

```
@HWI-EAS39X_10292_FC62DRM:6:1:1236:1010#0/1

NGGTGAACCGGACCGAAGCCCGATTTGGATCAGCCGAAGCGGATCGGTCG

+HWI-EAS39X_10292_FC62DRM:6:1:1236:1010#0/1

BIQIMMTTTT_____BBBBBBBBBBBBBBBBBBBBBBBB
```

The first line, prefixed with an @ character, contains the name of the read. The second line contains the sequence. The third line, prefixed by a +, holds a secondary name for the read, and may contain no information beyond the + character. The last line contains per-base quality scores on the Phred quality scale, encoded in ASCII. The FASTQ files may be opened with any text editor for manual inspection, but note that these files are typically large (10 GB or more), so manual editing is not practical. Manipulating the files requires dedicated programs like the FASTX toolkit (see below) or custom scripts.

The two read files contain an identical number of reads, and they are ordered by the cDNA fragment from which they were sequenced. That is, the first reads in each file form a "mate pair," as do the second reads in each file, and so on. It is critical for downstream analysis that this matched order be preserved during any manipulation of these files. That

is, if any reads are removed from one of the files, the mated reads must be removed in the other file.

As described below, read analysis is performed in three stages using the Spats data analysis pipeline. The first stage splits the provided read files into the (+) and (−) channels by examining the index handle sequence in each handle read (see Fig. 2). After this stage, the two input FASTQ files are split into four temporary FASTQ files that are deleted after the analysis is complete. The second stage uses the Bowtie read aligner to map each read to the RNA sequence from which it was generated, using barcodes if included (Langmead et al., 2009). The resulting read mapping tells Spats where the cDNAs in each channel begin and end. In the third step, Spats uses the beginning and end sites to infer nucleotide reactivities as described in Aviran et al. (2011a).

The Spats pipeline runs on Intel-based Linux and Mac OS X PCs, and depends on several software components listed in the Materials section below. If practice with the pipeline is desired, users are encouraged to download the sample SHAPE-Seq dataset from the Gene Expression Omnibus (GEO), dataset accession GSM783863 *(http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM783863)*. This dataset is stored in `.sra` format, which can be converted to fastq format for use with the protocol below, using the SRA toolkit as described at *http://www.ncbi.nlm.nih.gov/books/NBK47540/*.

Materials

Python: Python is a common programming language found on many platforms by default. Python installation instructions can be found at *http://python.org* for a variety of systems.

Bowtie: The bowtie short read aligner and installation instructions can be downloaded from *http://bowtie-bio.sourceforge.net/index.shtml*.

The Boost Libraries: Boost is a collection of open-source programming utilities that provide advanced numerical capabilities relied upon by the Spats SHAPE-Seq pipeline. Boost can be downloaded by following instructions at *http://www.boost.org*.

Spats package: The Spats package can be downloaded from *http://spats.sourceforge.net/*. Complete installation instructions for Spats can be found in the documentation on the Web site.

1. Prepare a target RNA sequence file for alignment. This will be used by Spats to uniquely align reads to the RNAs in the library. If barcodes are used, include an entry for each barcoded RNA (including structure cassette) in this file. A sample FASTA file containing a target RNA sequence for each RNA in the mixture in the experiment is included as an online supplementary file (go to *http://www.currentprotocols.com/protocol/ch120019*; filename is `target_pool.fa`), and will work with the sample data set indicated above. Each RNA sequence should have a single record in the file, and each record should be in FASTA format:

```
>target_GTAC
GGTCGTGCCTAGCGAAGTCATAAGCTAGGGCAGTCTTTAGAGGCTGACGGCAG
GAAAAAAGCCTACGTCTTCGGATATGGCTGAGTATCCTTGAAAGTGCCACA
GTGACGAAGTCTCACTAGAAATGGTGAGAGTGGAACGCGGTAAACCCCTCG
ACCGATCCGCTTCGGCGGATCCGTACAAATCGGGCTTCGGTCCGGTTC
```

The first line in each record should be prefixed by `>`, followed immediately by the name of the target (with no space between the `>` and the name). In this case, this specifies the target as containing the 'GTAC' barcode, but any other text string will suffice. The sequence of the target RNA should begin on the next line, and can

span multiple subsequent lines. The FASTA specification recommends that each line contain no more than 60 characters of sequence, but this is not a strict requirement for Spats. RNA sequences should, however, be provided in the DNA alphabet, so uracil bases should be written as thymidine (t/T). A single blank line should separate multiple entries in the target sequence file.

2. Run the Spats pipeline. At the UNIX shell prompt, type in the following command:

   ```
   spats -p <num_threads> -o <your_output_dir> tar-
   get_pool.fa RRRY YYYR s_X_sequence_1.fq s_X_sequence_2.fq
   ```

   where:

   `<num_threads>` is the number of threads of the processor to use. Increasing num_threads will accelerate the analysis run. We recommend using one thread per available CPU core.

   `<your_output_dir>` is the location where the output of the analysis will be stored.

   `target_pool.fa` is the name of the target FASTA file prepared in step 1. Change this name accordingly.

   `RRRY YYYR` specifies the format for the handles for the (+) and (-) channel [R=Purine (A,G), Y = Pyrimidine (C,T)]. If the protocol above is followed, this is left unchanged. If the handle barcodes for the (+) and (-) channel have been altered, see the Spats manual for details on how to specify these sequences. However, it should be noted that these handle sequences were chosen to introduce randomness into the first few bases sequenced which can aid in high quality sequencing.

   The last two strings are the names of the two paired-end reads FASTQ files (`.fq`), and may vary from run to run.

3. Analyze the reactivity data. In the specified output directory (by default, called `spats_out`) is a file called `reactivities.out`. This text file may be opened using any standard text editor, or input into programs such as Microsoft Excel. The file is of the format shown in Figure 5, where the columns are as follows:

   `sequence`: name of target RNA, taken from the `target_pool.fa` file
   `five_prime_offset`: position in target RNA, 0 = full length
   `nucleotide`: base at position in target, * = full length
   `treated_mods`: number of (+) fragments ending at position in target
   `untreated_mods`: number of (-) fragments ending at position in target
   `beta`: beta value at position in target
   `theta`: theta value at position in target
   `c`: estimated SHAPE modifications per target

sequence	five_prime_offset	nucleotide	treated_mods	untreated_mods	beta	theta	c
RNaseP_WT	0	*	77211	55078	-	-	0.840988
RNaseP_WT	1	G	10779	5532	0.0422978	0.041584	0.840988
RNaseP_WT	2	G	5505	3141	0.012441	0.0120808	0.840988
RNaseP_WT	3	T	289	249	0	0	0.840988
RNaseP_WT	4	C	381	306	0	0	0.840988
…							

Figure 5 Text file `reactivities.out`.

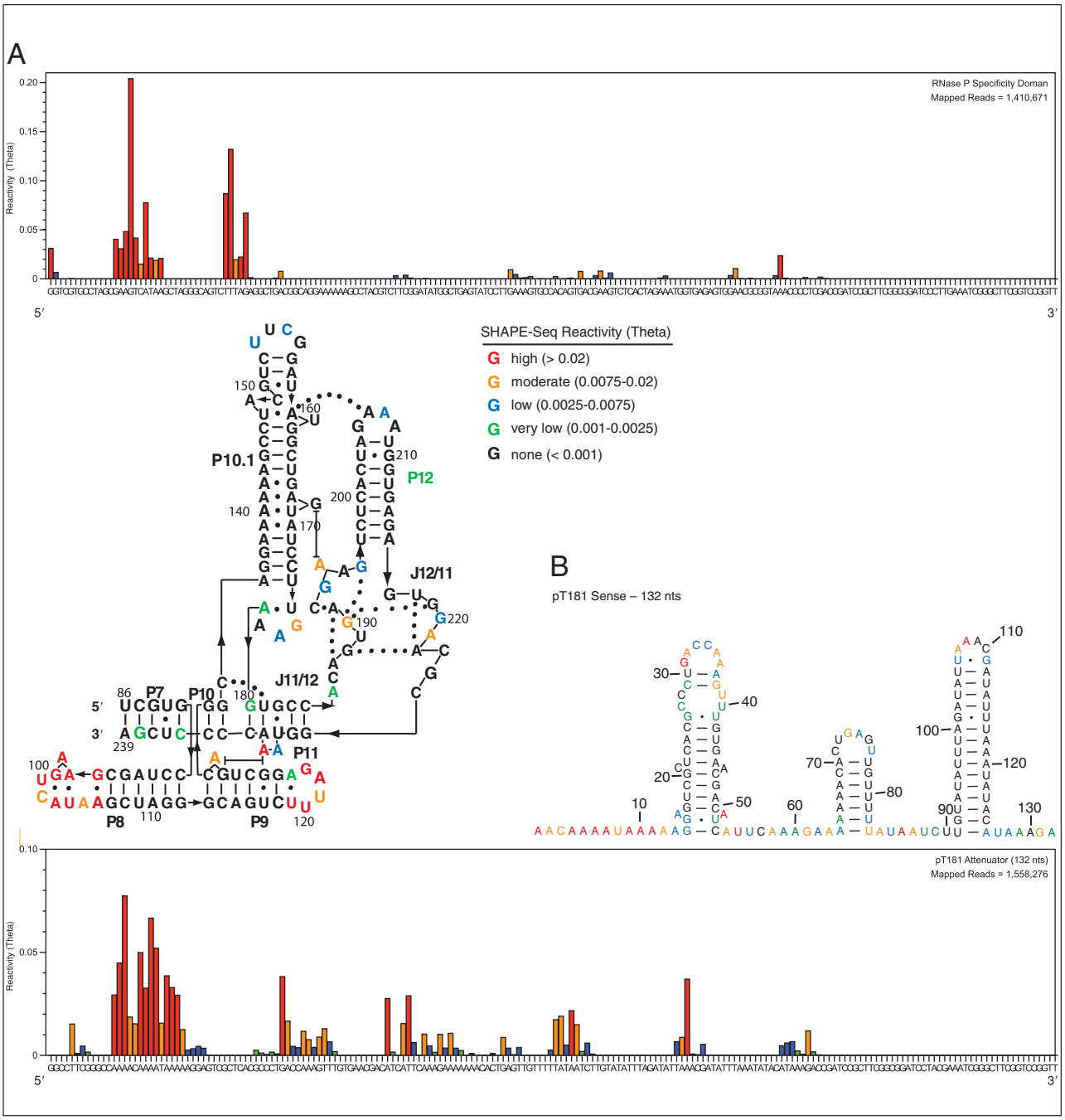

Figure 6 Representative SHAPE-Seq reactivity data. (**A**) Reactivities for the *B. subtilis* RNase P specificity domain showing two central clusters of reactivities corresponding to hairpin loops. Color-coded reactivities are shown alongside the known secondary structure of RNase P (Krasilnikov et al., 2003), which is also color coded by experimental reactivities. (**B**) Reactivities for 132 nt of the pT181 transcriptional attenuator showing more regions of flexible, unpaired nucleotides. Color-coded reactivities are shown alongside a SHAPE-constrained computational structure prediction (Low and Weeks, 2010), which is also color coded by experimental reactivities. Data is replotted from (Aviran et al., 2011b; Lucks et al., 2011).

Example outputs for the single WT RNaseP sequence as well as one for a mixed library can be found as supplementary files at *http://www.currentprotocols.com/protocol/ch120019* (filenames are `Sample_output_mixed_library.out` *and* `Sample_output_RNaseP_WT.out`).

The fragment counts in these files are mathematically transformed into nucleotide *reactivities*, denoted *thetas*, which represent the *relative* propensity of each nucleotide

to be modified by the SHAPE reagent. The applied mathematical transformation extracts estimates of these propensities that maximize the likelihood of observing the obtained sequencing data, under a novel statistical model of the SHAPE-Seq experiment that we have previously introduced (Aviran et al., 2011b). Note that in the accompanying Spats code, the theta estimates are computed directly via a simplified closed-form maximum-likelihood solution (Aviran et al., 2011a). These publications provide complete details and interpretation of our statistical analysis methodology and provide for a broader statistical framework as well, which also applies to analysis of traditional capillary-based SHAPE data.

To obtain a plot of reactivities, plot theta values versus the nucleotide identity or position as in Figure 6. These values can be compared to SHAPE reactivities in other work by renormalizing according to the 2%/8% rule (Lucks et al., 2011): exclude the top 2% of thetas, calculate a normalization factor by averaging the next 8%, and divide all thetas by this normalization factor [see Low and Weeks (2010) for details and for an alternative normalization approach].

It should be noted that the data analysis pipeline as described here does not perform adapter sequence trimming from sequencing reads prior to alignment. The result is that sequences that contain adapter sequence (for example, from a cDNA fragment that is shorter than the sequencing read length used in the experiment) will not be mapped to the RNA sequences in the spats target FASTA file. This will result in an absence of reactivity information in the very 3′ ends of the molecules. For detailed instructions on including adapter trimming when running spats, please consult the Spats Web site. We recommend sequencing libraries with 35-bp reads on each end.

SUPPORT PROTOCOL

SYNTHESIS OF 1-METHYL-7-NITROISATOIC ANHYDRIDE (1M7)

The most commonly used SHAPE reagent is currently the 2nd-generation reagent 1M7 (Mortimer and Weeks, 2007; Fig. 7). Unfortunately, it is not yet commercially available; however, it is readily synthesized with basic equipment. One does need access to a fume hood equipped with a nitrogen line, but otherwise it requires materials that are readily available. It is imperative that the reaction remain free of water, and it should therefore be performed under an inert atmosphere of dry nitrogen. All glassware should be oven or flame dried beforehand, including a magnetic stir bar. Another aspect that makes the synthesis straightforward is that every step is done at room temperature. The protocol below describes the synthesis performed on a 5 g scale, but it can be scaled up or down as necessary to reasonable limits (for example, 2 or 3 fold differences in mass).

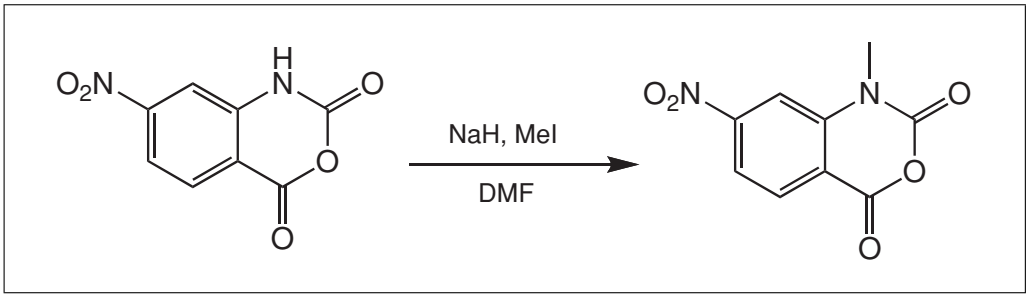

Figure 7 Reaction scheme for the synthesis of 1M7 from 4-nitro-isatoic anhydride.

Materials

4-Nitro-isatoic anhydride (AstaTech, cat. no. 69441; *http://www.astatechinc.com/*)
Sodium hydride (60% dispersion in mineral oil; Sigma-Aldrich, cat. no. 452912-100G)
Iodomethane (Sigma, cat. no. 289566)
N,*N*-dimethylformamide (DMF) (anhydrous, Sigma, cat. no. 227056)
1 N hydrochloric acid solution (1 N Certified; Fisher, cat. no. 7647-01-0), cold
Ether

Flame or oven-dried round bottom flasks (2) fitted with magnetic stir bar
Magnetic stir plate
Watch glass

Additional reagents and equipment for mass spectrometry and NMR

1. Deprotonate starting material by adding 5.0 g (24 mmol) of 4-nitroisatoic anhydride dissolved in 60 ml DMF to a suspension of 1.0 g (24 mmol) of sodium hydride (60% in mineral oil) in 20 ml DMF, in a round-bottom flask under a dry nitrogen atmosphere.
2. Stir at room temperature for 5 to 10 min.

   The initially cloudy, reddish orange solution should slowly become a clear solution of the same color.
3. Methylate the deprotonated nitrogen by the dropwise addition of 1.5 ml (24 mmol) of iodomethane into the reaction mixture, using a needle inserted through the rubber septum.
4. Allow the reaction to stir at room temperature for 4 hr.
5. To isolate 1M7, precipitate the reaction product by pouring the reaction mixture into 100 ml of cold 1 N HCl on ice.
6. Immediately filter the resulting bright orange precipitate, using filter paper, and wash sequentially with cold water and ether.
7. Air dry the filtered precipitant on a dry watch glass (for few hours to overnight). Weigh the dried product to obtain the yield of 1M7, which should be about a 79% yield (just over 4 g).
8. Characterize the product by mass spectrometry (molecular weight 222 g/mol) and proton NMR. ^{1}H NMR ($CO(CD_3)_2$, 400 MHz) δ 3.69 (s, 3 H, -NC*H*$_3$-), 8.12 (dd, $J =$ 8.8 Hz, 2 Hz, 1 H, Ar*H*), 8.2 (d, $J = 2$ Hz, 1H, Ar*H*), 8.34 (d, $J = 8.4$ Hz, 1 H, Ar*H*).

REAGENTS AND SOLUTIONS

Use deionized, distilled water in all recipes and protocol steps.

Benzoyl cyanide (BzCN), 400 mM

17.8 µl dimethylsulfoxide (DMSO; Sigma, cat. no. D8418)
1 mg benzoyl cyanide (BzCN; Sigma, cat. no. 115959)
Prepare fresh each day

CircLigase ligation buffer, 10×

500 mM MOPS, pH 7.5
10 mM dithiothreitol (DTT)
100 mM KCl
50 mM $MgCl_2$
Store up to several months at −20°C

Enzyme master mix

4 parts 5× SSIII first strand buffer (see recipe)
1 part 10 mM dNTP mixture (dATP, dCTP, dGTP, dTTP; New England Biolabs, cat. no. N0447L)
1 part 0.1 M DTT
Store up to several months at −20°C

Folding buffer, 3.3×

33.3 mM $MgCl_2$
333 mM NaCl
333 mM HEPES pH 8.0
Store up to several months at room temperature

1M7, 65 mM

69.3 μl DMSO (anhydrous; Sigma, cat. no. D8418)
1 mg 1M7 (see Support Protocol)
Store up to 2 weeks at room temperature

NMIA, 65 mM

86.9 μl DMSO (anhydrous; Sigma, cat. no. D8418)
1 mg NMIA (Invitrogen, cat. no. M25)
Store up to 2 weeks at room temperature

SSIII first strand buffer, 5×

250 mM Tris·Cl, pH 8.3
375 mM KCl
15 mM $MgCl_2$
Store up to 6 months at −20°C

Stop dye

85% (v/v) deionized formamide
0.5× TBE buffer (see recipe)
50 mM disodium EDTA, pH 8.0
0.01% (w/v) xylene cyanol
0.01% (w/v) bromphenol blue
Store up to several months at 4°C

TBE buffer

89 mM Tris·Cl, pH 8.0
89 mM boric acid
2 mM disodium EDTA, pH 8.0
Store up to several months at room temperature

TE buffer

10 mM Tris·Cl, pH 8.0
1 mM disodium EDTA, pH 8.0
Store up to several months at room temperature

COMMENTARY

Background Information

SHAPE chemistry is the latest advance in a long history of work using chemical probes to characterize RNA structures and RNA interactions with small molecules and proteins (Moazed et al., 1986; Weeks, 2010). The great advantage of SHAPE is that the structure-dependent modification occurs

at the backbone of the RNA and not the base, so that the same reagent can be used to probe every nucleotide simultaneously in each RNA (Merino et al., 2005). Furthermore, SHAPE reagents are extremely sensitive to local nucleotide flexibility and can reveal subtle changes in RNA structure such as the emergence of tertiary interactions upon addition of Mg^{2+} (Mortimer and Weeks, 2007), or the binding of a small molecule ligand to an aptamer (Steen et al., 2010). Additionally, SHAPE chemistry can reveal RNA-protein interactions, and even structural changes that result upon protein binding (Duncan and Weeks, 2008).

Previous versions of SHAPE used capillary electrophoresis (CE) as a read out of fluorescently labeled cDNA fragments (Vasa et al., 2008; Mortimer and Weeks, 2009). SHAPE-CE can thus only be performed on one RNA at a time, with a significant effort in data analysis required to convert the analog CE signal to quantitative reactivities (Aviran et al., 2011b). There are several major advantages to using direct sequencing of cDNAs with NGS in SHAPE-Seq. First, direct sequencing allows reads to be aligned to individual RNA sequences, which allows the multiplexing of the nucleotide-resolution SHAPE chemistry to be able to simultaneously characterize a complex mixture of RNAs in a single experiment. Second, this direct sequencing allows barcoding of individual RNAs, making it possible to detect structural changes that result from single nucleotide changes in a single experiment (Lucks et al., 2011). Third, there is an inherent sensitivity boost when using NGS, requiring an order of magnitude less RNA to be used compared to conventional CE-based detection methods. Fourth, SHAPE-Seq does not suffer from information loss near the ends of the RNAs due to CE detector overloading, and can thus give higher accuracy across the whole molecule (Aviran et al., 2011a,b). Furthermore, many of these advantages will immediately benefit from the continuous dramatic improvements in throughput and cost of NGS techniques.

Despite these advantages, we highly recommend performing SHAPE-CE on new RNAs under investigation to develop an intuition about how these RNAs may behave in more complex SHAPE-Seq libraries. There are several well-established protocols for performing SHAPE-CE for this purpose (Wilkinson et al., 2006; Mortimer and Weeks, 2009). It should also be noted that there are differences between reactivities generated with SHAPE-Seq and SHAPE-CE, which could be attributed to the extra steps involved in SHAPE-Seq (ligation, PCR, amplification during sequencing, etc.) that could cause additional biases. To date, our analysis shows that if these biases exist, they are minor (Lucks et al., 2011). The interested reader is encouraged to read the original work and commentary to learn more about these differences (Aviran et al., 2011b; Lucks et al., 2011; Weeks, 2011).

Compared to other RNA structure determination methods, SHAPE-Seq cannot directly provide the three-dimensional orientation of RNA atoms at the angstrom resolution afforded by X-ray crystallography and nuclear magnetic resonance. However, its dramatic increase in throughput, with hundreds of RNAs probed in a single experiment lasting about 1 day (before sequencing), makes it advantageous for many studies, and opens the door to using structure probing as an actual screening technique based on RNA structure and RNA interactions.

Two techniques related to SHAPE-Seq, parallel analysis of RNA structures (PARS; Kertesz et al., 2010) and fragmentation sequencing (Frag-Seq; Underwood et al., 2010), were recently developed. These techniques use NGS to analyze structure-dependent cleavage patterns resulting from different RNase enzymes. In their current implementation, they have the advantage of being more high-throughput than SHAPE-Seq, with structural information obtained for thousands of RNAs at once. The primary advantage of SHAPE-Seq is the flexibility and accuracy of the small molecule chemistry over nuclease probing, which allows it to be applied to a wide variety of conditions and experiments outlined above. A current limitation of all of these techniques is that the folding of the RNAs is done in vitro. However, we are hopeful that further advances will allow folding in conditions that better approximate the cellular environment.

In all but the simplest investigations, it is highly recommended to use a variety of conditions to tease out the specific structural changes and interactions of interest, for example by using titrations of component concentrations or time courses to look at the evolution of folds. It is also extremely advantageous to combine SHAPE reactivities with other functional evidence, complementary experimental data, or biological insight to interpret the reactivities. It is also possible to use experimental reactivities as constraints in RNA-folding

algorithms (Deigan et al., 2008; Kladwang et al., 2011).

Critical Parameters

It is imperative that the SHAPE reagents and anhydrous DMSO are kept dry; storing them in a desiccator at room temperature can accomplish this. For long-term storage of SHAPE reagents, we recommend using a desiccator at −20°C. The SHAPE reagent solutions should be prepared fresh when possible; however, solutions of 1M7 and NMIA may be stored in a desiccator at room temperature for up to 2 weeks. Solutions of BzCN should be prepared fresh before every experiment. A solution of 1M7 in DMSO undergoes a colorimetric reaction upon hydrolysis. Unhydrolyzed 1M7 is a pale yellow color but turns darker orange-yellow upon hydrolysis, which is a good indicator of whether the stock should be used for experiments. If using benzoyl cyanide, a fresh solution should be prepared before every experiment as it is extremely sensitive to hydrolysis.

As mentioned in the annotated steps during the recovery of modified RNA, this pellet is small and can be easily dislodged, so care must be taken to not lose this pellet but to remove as much liquid as possible. Maintaining this pellet is critical to downstream steps and the protocol will not work if all or part is lost.

It is extremely critical that the library be of good quality before sequencing. Removing the excess A_adapter_b post ligation is required to have high-quality sequencing runs, as this will compete for flow cell binding. The beads used in this protocol appear to do this sufficiently, but it is a step that cannot be skipped and if overlooked could be detrimental to obtaining good-quality results. Before sending libraries to be sequenced, we use the Agilent Bioanalyzer trace comparing the 9- and 12-cycle PCRs to assess overall library quality (Fig. 4). It is imperative that there be distinct peaks in the size range you expect, which will vary depending on the length of your RNA. The smallest size possible is the length of the two adapters ligated together (120 bp), while the largest size possible is this plus the length of your RNA. You should see increases in peak intensity in the size range expected when comparing the 9- and 12-cycle PCRs on the bioanalyzer. If you do not, it would be wise to not proceed to sequencing, as this step is costly and requires the longest time to complete. Reasons for poor library quality are addressed in the troubleshooting section below.

Troubleshooting

Table 1 discusses several problems commonly encountered with the SHAPE-Seq methodology along with possible causes and recommendations to avoid or overcome these problems.

Anticipated Results

SHAPE has been performed under a variety of conditions in experiments designed to answer a wide array of questions, including the nature of the structure of both large and small RNAs (Deigan et al., 2008; Kladwang et al., 2011), how RNAs bind small molecules (Steen et al., 2010), the nature of RNA-protein interactions (Duncan and Weeks, 2008), and how tertiary interactions evolve over time (Mortimer and Weeks, 2007). Therefore, SHAPE-Seq is anticipated to generate the same type of results for these questions, except in much higher throughput.

Specific to SHAPE-Seq results is the overall sequencing quality, which is reflected in the percent of reads that can be successfully aligned to the RNA sequences. Previous results indicate that 80% of reads should be aligned in high-quality libraries. Another SHAPE-Seq-specific result is the number of reads needed to accurately infer a reactivity spectrum. This is likely to vary, and a systematic study is needed to definitively determine this. An extensive analysis of this question for our previous results can be found in Aviran et al. (2011b). In general, we found for the highly compact fold of the RNase P specificity domain, as little as 10,000 reads were needed (Aviran et al., 2011b; Lucks et al., 2011). However, it is advisable to aim for 100,000 reads and to account for the greater number of reads needed for more flexible RNAs that show greater overall reactivity. In addition, the modification rate parameter, as determined from the maximum likelihood calculation, was found to be different for each RNA in a mixture, which likely reflects the differences in overall structure, and thus overall modification, of the different RNAs.

Time Considerations

Creating the RNA library takes 2 to 3 days depending on whether the PCR is performed the same day as the in vitro transcription, but at minimum an overnight step is required (passive elution), and therefore at least 2 days should be allocated for library creation. The RNA is stable for months at −20°C and there is no requirement to perform the rest of the

Table 1 Troubleshooting Guide for SHAPE-Seq Analysis

Problem	Possible cause	Solution
No or very small peaks of expected size range in Bioanalyzer trace, or no amplification of peaks in 9- and 12-cycle PCR	Poor or incomplete primer extension. Superscript III is sensitive to freezing and $MgCl_2$ concentration.	Repeat with fresh enzyme and ensure buffer is prepared correctly
	Insufficient modification of RNA. SHAPE reagent stock may be hydrolyzed.	Repeat using freshly prepared SHAPE reagent. Alternatively, use higher concentration of SHAPE reagent.
	No A_adapter_b sequence present caused by poor ligation efficiency	Repeat with fresh ligase
	Excess A_adapter_b present	Perform several rounds of bead purification
Extremely high counts in (-) compared to the (+)	RNase contamination	Check for degradation of input RNA library on a gel. Remake library if necessary. Check stock solutions for contamination.
	Structure-induced pausing of the reverse transcriptase enzyme	Increase the time of extension at both 52°C and 65°C
No full length counts in (+) or intense counts that disappear rapidly with length	Excessive modification of RNA	Perform modification using 2-fold lower concentration of SHAPE reagent
No full length counts in (+) or (−) channels	Insufficient RT extension	Increase the time of extension at both 52°C and 65°C
The automated maximum likelihood estimation is not working due to extreme signal attenuation in (+) or (−) channel	Excessive modification of RNA	Perform modification using 2-fold lower concentration of SHAPE reagent
	Insufficient modification of RNA. SHAPE reagent stock may be hydrolyzed.	Repeat using freshly prepare SHAPE reagent. Alternatively, use higher concentration of SHAPE reagent.
	Polymerase drop-off hot spots	Increase the time of extension at both 52°C and 65°C

protocol immediately following this step. The modification, first-strand synthesis, and ligation require one very long day, or alternatively the ligation can be performed overnight in a thermal cycler and held at 4°C afterwards. After each of these steps, the protocol can be safely stopped and RNA or DNA stored at −20°C. Removal of excess adapter, PCR, and bioanalyzer can be performed in a day. The sequencing in paired-end mode requires from 1 to 2 weeks, depending on the length of sequence per end requested. Analysis takes several hours once all software is installed properly, although most of this time is allocated for automatic processing of read mapping.

Acknowledgements

The authors acknowledge Adam P. Arkin, Jennifer A. Doudna, Shujun Luo, and Gary P. Schroth for critical advice and support during the development and writing of this protocol. We also thank David Loughrey for a careful reading of the manuscript. S.A.M. is a fellow of the Leukemia and Lymphoma Society. C.T. is supported by a Damon Runyon Fellowship. S. A. is supported by a Celera Innovation Fellowship and by Tata Consultancy Services (TCS), through grants to the Center for Computational Biology at the University of California, Berkeley.

Literature Cited

Amaral, P.P., Dinger, M.E., Mercer, T.R., and Mattick, J.S. 2008. The eukaryotic genome as an RNA machine. *Science* 319:1787-1789.

Aviran, S., Lucks, J.B., and Pachter, L. 2011a. RNA structure characterization from chemical mapping experiments. Proc. 49th Allerton Conf. on Communication, Control and Computing 1743-1750. IEEE Press, New York.

Aviran, S., Trapnell, C., Lucks, J.B., Mortimer, S.A., Luo, S., Schroth, G.P., Doudna, J.A., Arkin, A.A., and Pachter, L. 2011b. Modeling and automation of sequencing-based characterization of RNA structure. *Proc. Natl. Acad. Sci. U.S.A.* 108:11069-11074.

Butcher, S.E. and Pyle, A.M. 2011. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.* 44:1302-1311.

Deigan, K. E., Li, T. W., Mathews, D. H., and Weeks, K. M. 2008. Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A.* 106:97-102.

Duncan, C.D.S. and Weeks, K.M. 2008. SHAPE analysis of long-range interactions reveals extensive and thermodynamically preferred misfolding in a fragile group I intron RNA. *Biochemistry* 47:8504-8513.

Ellington, A. and Pollard, J.D. 2001. Purification of oligonucleotides using denaturing polyacrylamide gel electrophoresis. *Curr. Protoc. Mol. Biol.* 42:2.12.1-2.12.7.

Garber, M., Grabherr, M.G., Guttman, M., and Trapnell, C. 2011. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat. Methods* 8:469-477.

Isaacs, F.J., Dwyer, D.J., and Collins, J.J. 2006. RNA synthetic biology. *Nat. Biotechnol.* 24:545-554.

Kertesz, M., Wan, Y., Mazor, E., Rinn, J.L., Nutter, R.C., Chang, H.Y., and Segal, E. 2010. Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467:103-107.

Kladwang, W., Vanlang, C.C., Cordero, P., and Das, R. 2011. Understanding the errors of SHAPE-directed RNA structure modeling. *Biochemistry* 50:8049-8056.

Kladwang, W., Hum, J., and Das, R. 2012. Ultraviolet shadowing of RNA causes substantial non-Poissonian chemical damage in seconds. *Sci. Rep.* 2:517.

Kramer, M.F. and Coen, D.M. 2001. Enzymatic amplification of DNA by PCR: Standard procedures and optimization. *Curr. Protoc. Mol. Biol.* 56:15.1.1-15.1.14.

Krasilnikov, A.S., Yang, X., Pan, T., and Mondragón, A. 2003. Crystal structure of the specificity domain of ribonuclease P. *Nature* 421:760-764.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25.

Low, J.T. and Weeks, K.M. 2010. SHAPE-directed RNA secondary structure prediction. *Methods* 52:150-158.

Lucks, J.B., Mortimer, S.A., Trapnell, C., Luo, S., Aviran, S., Schroth, G.P., Pachter, L., Doudna, J.A., and Arkin, A.P. 2011. Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. U.S.A.* 108:11063-11068.

Merino, E.J., Wilkinson, K.A., Coughlan, J.L., and Weeks, K.M. 2005. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* 127:4223-4231.

Metzker, M.L. 2009. Sequencing technologies—the next generation. *Nat. Rev. Genetics* 11:31-46.

Moazed, D., Stern, S., and Noller, H.F. 1986. Rapid chemical probing of conformation in 16 S ribosomal RNA and 30 S ribosomal subunits using primer extension. *J. Mol. Biol.* 187:399-416.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5:621-628.

Mortimer, S.A. and Weeks, K.M. 2007. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* 129:4144-4145.

Mortimer, S.A. and Weeks, K.M. 2008. Time-resolved RNA SHAPE chemistry. *J. Am. Chem. Soc.* 130:16178-16180.

Mortimer, S.A. and Weeks, K.M. 2009. Time-resolved RNA SHAPE chemistry: Quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution. *Nat. Protoc.* 4:1413-1421.

Shendure, J. and Ji, H. 2008. Next-generation DNA sequencing. *Nat. Biotechnol.* 26:1135-1145.

Steen, K.-A., Arun, M., and Weeks, K.M. 2010. Selective 2'-hydroxyl acylation analyzed by protection from exoribonuclease. *J. Am. Chem. Soc.* 132:9940-9943.

Underwood, J.G., Uzilov, A.V., Katzman, S., Onodera, C.S., Mainzer, J.E., Mathews, D.H., Lowe, T.M., Salama, S.R., and Haussler, D. 2010. FragSeq: Transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* 7:995-1001.

Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M., and Giddings, M.C. 2008. ShapeFinder: A software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA* 14:1979-1990.

Voytas, D. 2001. Agarose gel electrophoresis. *Curr. Protoc. Mol. Biol.* 51:2.5A.1-2.5A.9.

Weeks, K.M. 2010. Advances in RNA structure analysis by chemical probing. *Curr. Opin. Struct. Biol.* 20:295-304.

Weeks, K.M. 2011. RNA structure probing dash seq. *Proc. Natl. Acad. Sci. U.S.A.* 108:10933-10934.

Weeks, K.M. and Mauger, D.M. 2011. Exploring RNA structural codes with SHAPE chemistry. *Acc. Chem. Res.* 44:1280-1291.

Wilkinson, K.A., Merino, E.J., and Weeks, K.M. 2006. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): Quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* 1:1610-1616.

Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., Rein, A., Mathews, D.H., Giddings, M.C., and Weeks, K.M. 2008. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.* 6:e96.

Internet Resources

http://spats.sourceforge.net/

Download site for Spats, a program for the analysis of sequencing information to determine SHAPE reactivities in a SHAPE-Seq experiment.