

# Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE

Ryan A Flynn<sup>1,2</sup>, Qiangfeng Cliff Zhang<sup>1,2</sup>, Robert C Spitale<sup>1</sup>, Byron Lee<sup>1</sup>, Maxwell R Mumbach<sup>1</sup> & Howard Y Chang<sup>1</sup>

<sup>1</sup>Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, California, USA. <sup>2</sup>These authors contributed equally to this work. Correspondence should be addressed to H.Y.C. ([howchang@stanford.edu](mailto:howchang@stanford.edu)).

Published online 14 January 2016; doi:10.1038/nprot.2016.011

**icSHAPE** (*in vivo* click selective 2-hydroxyl acylation and profiling experiment) captures RNA secondary structure at a transcriptome-wide level by measuring nucleotide flexibility at base resolution. Living cells are treated with the icSHAPE chemical NAI-N<sub>3</sub> followed by selective chemical enrichment of NAI-N<sub>3</sub>-modified RNA, which provides an improved signal-to-noise ratio compared with similar methods leveraging deep sequencing. Purified RNA is then reverse-transcribed to produce cDNA, with SHAPE-modified bases leading to truncated cDNA. After deep sequencing of cDNA, computational analysis yields flexibility scores for every base across the starting RNA population. The entire experimental procedure can be completed in ~5 d, and the sequencing and bioinformatics data analysis take an additional 4–5 d with no extensive computational skills required. Comparing *in vivo* and *in vitro* icSHAPE measurements can reveal *in vivo* RNA-binding protein imprints or facilitate the dissection of RNA post-transcriptional modifications. icSHAPE reactivities can additionally be used to constrain and improve RNA secondary structure prediction models.

## INTRODUCTION

Properly regulated gene expression is critical to the function of normal cells, and alterations in the mechanisms that control gene expression have been linked to several diseases<sup>1</sup>. Although most efforts to understand gene regulation have focused on the role of DNA modification and structure, diverse biological processes are also controlled at the post-transcriptional level.

Mechanisms involving RNA can be found at the heart of many important cellular functions. mRNA is central in the conversion of genetic information in chromatin into protein expression. The ribosome uses RNA functional groups to control peptidyl-transferase activity<sup>2</sup>. tRNA maturation is controlled by the RNA enzyme RNaseP<sup>3</sup>. Small RNAs, including miRNAs and piwi-interacting RNAs, have been demonstrated to control the steady-state level of mRNA expression and chromatin remodeling<sup>4</sup>. Elucidating the biochemical basis of such functions is crucial for a complete understanding of how these processes work.

Arguably, all functions of RNA are controlled by RNA's ability to fold into complex secondary and tertiary structures. The majority of RNA structure probing experiments are performed on just a single RNA at a time. These investigations have revealed the intricate details of RNA folding, RNA-protein contacts and metabolite-RNA interactions<sup>5</sup>. In addition, structural studies have also shown that RNA structure has a role in regulating single-gene functions; however, it was only recently indicated that these mechanisms might extend to large groups of RNAs with the observation that RNA structure elements can control the translation of functionally related genes<sup>6</sup>. RNA secondary structure motifs are critical to the regulation of protein binding, subcellular localization and RNA decay. As such, extending our understanding of the structural content of RNA, transcriptome wide, will reveal unique facets of how diverse classes of RNA control the biology of the cell. Recent efforts have been focused on marrying conventional RNA structure probing techniques with genome-scale technologies.

### Conventional RNA structure probing techniques

RNA structure is frequently measured by chemical modification or RNase digestion. Whereas RNase S1 recognizes and cuts

single-stranded regions in RNA, RNase V1 will cleave at double-stranded residues<sup>7</sup>. Small-molecule chemical reagents can instead be used to increase the resolution of structure probing experiments. For example, dimethylsulfate (DMS) can alkylate the Watson-Crick face of adenosine and cytosine, as well as the N7 position of guanosine when not base-paired<sup>8</sup>. Kethoxal is used to modify the N1- and C2-exocyclic amines to reveal single-stranded residues in RNA<sup>9</sup>. Hydroxyl radical probing is widely used to measure RNA compaction and folding; the mechanism used cleaves RNA by hydrogen abstraction at the 5' position<sup>10</sup>. The recent development of Selective Hydroxyl Acylation analyzed by Primer Extension, or SHAPE, has greatly extended chemical probing experiments because of its ability to interrogate every position in the RNA through 2'-OH reactivity<sup>11</sup>. Overall, the field is well equipped to interrogate the structure of single RNAs.

The next step in the evolution of RNA structure probing is to transition from single-gene analysis to whole-transcriptome interrogation. A corollary can be made to our understanding of gene expression mechanisms and chromatin structure, which has greatly matured since the advent of deep sequencing<sup>12</sup>. Knowledge of the structural content of RNA across the transcriptome will reveal unique facets of how diverse classes of RNA control the biology of the cell.

### *In vitro* transcriptome-wide structure analysis

The first demonstration of the power of transcriptome-wide analysis of RNA structure was the development of Parallel Analysis of RNA Structure (PARS). PARS uses both single (nuclease S1)- and double-stranded (nuclease V1) nucleases to digest motifs in RNA and to generate a structure score<sup>13</sup>. The PARS scores from both experiments can be used to constrain the RNA structure predictions of thousands of transcripts in parallel. In a similar manner, Frag-seq uses nuclease P1 for the cleavage and identification of single-stranded residues<sup>14</sup>. A complementary method, using RNase One and RNase V1, has been used to describe the secondary structures of multiple metazoans<sup>15</sup>. The first chemical-based method to transform structural information

into sequencing-based results for understanding the structures of thousands of RNAs in parallel was SHAPE-seq. In SHAPE-seq<sup>16</sup>, a pool of *in vitro*-transcribed RNAs are subjected to SHAPE chemical probing and deep sequencing. This approach has revealed structural changes that occur because of primary sequence alterations<sup>16</sup>. Although these methods have been very useful for determining the structure of refolded, *in vitro* transcripts, they are not suitable for describing the structures of RNA inside living cells.

**In vivo transcriptome-wide structure analysis**

DMS was the first chemical that was successfully applied to transcriptome-wide structure probing in living cells, in similar approaches called DMS-seq<sup>17</sup>, Structure-seq<sup>18</sup> and Mod-seq<sup>19</sup>. These studies demonstrated the utility of chemical probes for understanding the structural elements in mRNAs and noncoding RNAs on a large scale. Rouskin *et al.*<sup>17</sup> demonstrated that mRNA structures seem to sample many different conformations; however, highly important RNA structures, such as those involved in localization, have similar profiles within and outside the cell. In addition, they demonstrated that ATP-dependent RNA remodeling enzymes (probably helicases) were responsible for the structure probing variation. Ding *et al.*<sup>18</sup> demonstrated with *in vivo* DMS profiling that mRNAs associated with stress responses tend to have more single-strandedness, longer maximal loop length and higher free energy per nucleotide. Such features may allow these RNAs to undergo conformational changes in response to environmental conditions. These key studies highlight important examples of information that *in vivo* RNA structural measurements garner about basic RNA biology. One important limitation of DMS is that it only interrogates two out of four bases in RNA, adenosine (A) and cytosine (C); hence, the structural information is incomplete. To address this limitation, we developed and previously described *in vivo* click-selective 2'-hydroxyl acylation and profiling experiment or icSHAPE<sup>20</sup>; here we provide a detailed protocol for icSHAPE. Distinguishing features of our method include the ability to interrogate every

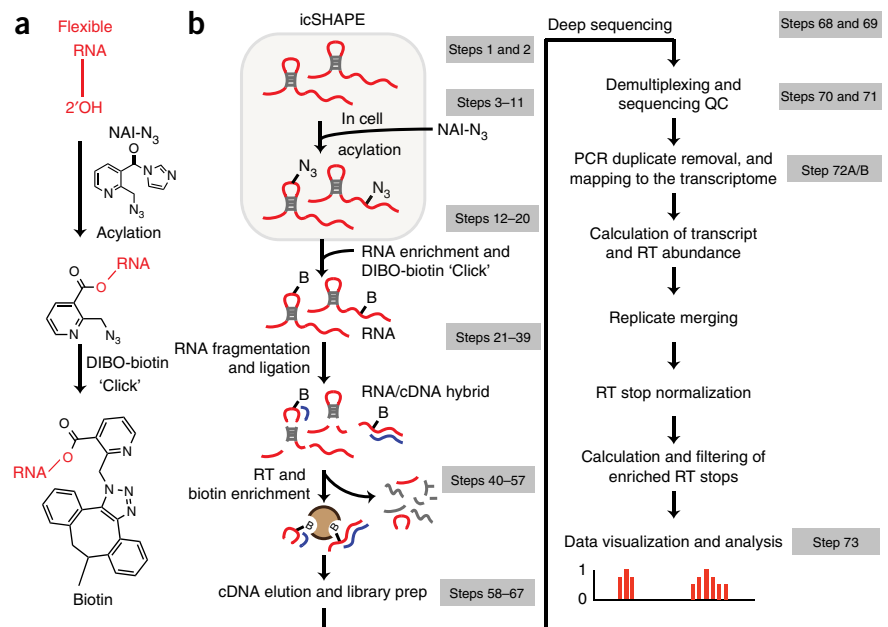
position in an RNA, independently of its primary sequence content, and to enrich for modified RNAs to markedly increase the signal-to-noise ratio in transcriptome-wide measurements.

**Development and overview of icSHAPE**

icSHAPE was developed to provide both the experimental and computational tools necessary to study RNA structure in a transcriptome-wide manner. Existing high-throughput, chemical-based structure determination methods suffer from high background because of sequencing unmodified RNA<sup>16–18</sup>. To address this issue, we took inspiration from methods such as RNA-binding protein immunoprecipitation (RIP)<sup>21</sup> and individual-nucleotide resolution cross-linking and immunoprecipitation (iCLIP)<sup>22</sup>, in which the signal for RNA binding can be highly enriched by selective purification. We transformed an existing *in vivo* SHAPE probe 2-methylnicotinic acid imidazole (NAI)<sup>23</sup> into a dual-function SHAPE probe (RNA structure probe and enrichment), by the addition of an azide to the nicotinic acid ring (at position 2), and we named this molecule NAI-N<sub>3</sub>. Incorporation of the azido-moiety was crucial, as it allowed the 'click' of biotin to the SHAPE reagent using copper-free 'click' chemistry, which proceeds under conditions that are not detrimental to the SHAPE-modified RNA (Fig. 1a). By the addition of a biotin moiety to each RNA that was successfully SHAPE modified, we are able to subsequently enrich specifically for RNA fragments that have NAI-N<sub>3</sub> covalently attached. We construct icSHAPE cDNA libraries on enriched material, thereby only sequencing structurally informative regions of RNA, thus increasing our signal to noise.

An overview of the icSHAPE procedure is provided in Figure 1b. In brief, the bifunctional NAI-N<sub>3</sub> SHAPE reagent is used to label all RNAs inside living cells (or *in vitro* refolded RNA) to generate covalent modifications on targeted RNAs at single-nucleotide resolution. RNA is isolated, biotinylated using click chemistry and randomly fragmented to a size of 50–100 nt. To generate cDNA sequencing libraries, the RNA fragments are subjected to 3'-end repair, adapter ligation, size selection, biotin enrichment, reverse transcription (RT)

**Figure 1 | Schematic overview of icSHAPE.** (a) Stepwise scheme for acylating flexible RNA nucleobases with the SHAPE reagent NAI-N<sub>3</sub> and subsequent 'click' reaction of DIBO-biotin specifically to icSHAPE-modified nucleotides<sup>20</sup>. (b) Experimental and computational steps in the icSHAPE procedure. Living cells are first treated with NAI-N<sub>3</sub> to record the *in vivo* flexibility of RNA nucleotides across the transcriptome. Modified (and mock-modified) RNA is isolated, and 'click' reactions add DIBO-biotin specifically to modified RNAs. RNA fragmentation and RT 'reads out' the NAI-N<sub>3</sub> modification, after which only NAI-N<sub>3</sub>-modified molecules are selectively isolated and sequenced. Deep-sequencing reads are separated for individual experiments, removed of PCR artifacts and mapped to the appropriate transcriptome build. RT stop sites and abundances are calculated and biological replicate experiments are merged to ensure reproducibility. RT stops are normalized for each transcript, and the resulting data are visualized in the UCSC genome browser or used for downstream analysis.



© 2016 Nature America, Inc. All rights reserved.



and PCR. The end result is selective amplification of fragments corresponding to modified bases in the NAI-N<sub>3</sub>-treated samples or background expression values from mock-treated samples. Sequencing reads generated from the NAI-N<sub>3</sub>- and mock (DMSO)-treated samples are mapped to a transcriptome annotation, followed by computational identification of RT stops corresponding to SHAPE modifications—and hence reactive nucleotides—across the entire transcriptome (Fig. 1b). Traditionally, SHAPE reactivity directly reports on the structural flexibility of an individual nucleotide; however, *in vivo* icSHAPE reactivities can be influenced by protein binding<sup>20</sup>, among other things.

We developed a highly modular and customizable computational pipeline to process raw icSHAPE sequencing data and to provide structural output files directly ready for interpretation. The pipeline was designed to take instructions from a centralized configuration file, and users can modify the parameters for each of the processing steps. To accommodate the numerous different strategies that laboratories use to analyze deep-sequencing data, each step taken by the icSHAPE pipeline can be achieved using alternative programs with similar functionality to those described here. The icSHAPE pipeline is capable of running in either ‘automatic’ or ‘manual’ modes, thus giving the user speed (automatic) or flexibility (manual) when implementing the analysis package. It should be noted that icSHAPE differs from regular RNA-seq experiments in that (i) icSHAPE profiles a transcript at base resolution, whereas RNA-seq experiments usually quantify transcriptomes and abundances of whole transcripts; (ii) the icSHAPE protocol generates relatively short cDNA inserts of ~20–50 nt; and (iii) in order to obtain base-resolution RNA structure profiles, the sequencing of icSHAPE should be very deep so as to generate sufficient coverage of individual bases. These features require analysis tools that can handle multiply mapping sequencing reads and that can efficiently process billions of reads with potentially limited computational resources and time.

#### Applications, advantages and limitations of icSHAPE

The use of NAI-N<sub>3</sub> provides an experimental framework for measuring the base flexibility of all four nucleotides in a transcriptome-wide experiment. This principally comes by leveraging SHAPE chemistry, whereas previous *in vivo* chemical methods have been restricted to interrogating only two of the four bases (As and Cs). Although integrating structural data from multiple probing strategies is ideal, icSHAPE provides a single experimental strategy to capture a global landscape of RNA structure, which is not biased for nucleotide content. An additional advantage of icSHAPE is the specific enrichment of NAI-N<sub>3</sub>-modified RNAs; however, it is formally possible that reverse transcriptase does not always fully extend to the nucleotides modified with NAI-N<sub>3</sub>, thereby possibly generating a false-positive signal. A recent study reported a method for mitigating these effects by digesting non-fully-extended RNA:cDNA hybrids with RNase One<sup>24</sup>. Although this feature is not a part of this protocol, future designs could easily implement this concept.

The principle that ‘read counting’ allows the user to accurately and quantitatively assess the structural state of a given nucleobase is common to all sequencing-based structure determination methods. Variable amounts of starting material and differential complexity between samples are examples of parameters that contribute to variation across experiments. Generating icSHAPE (or DMS-seq)

samples for sequencing requires PCR, and the above-mentioned experimental variables could contribute to differential amplification and/or overamplification of library samples. To combat this, we have incorporated a molecular barcode into the RT primer, which allows for accurate computational removal of PCR duplicates. The molecular barcode therefore allows reads with exactly the same sequence, but generated from unique RNA molecules, to be accurately counted as unique. Our barcode design allows for the precise resolution of >110,000 molecules with the exact same sequence.

Beyond comparing the reactivity profiles of specific transcripts, it is possible to use icSHAPE reactivity values to guide RNA secondary structure prediction. Software, such as RNAstructure<sup>25</sup>, can incorporate icSHAPE and other chemical probing data to constrain the predicted secondary structures, thus improving the accuracy of the prediction. Future work applying novel<sup>26</sup> and integrated computational methods (reviewed in refs. 27–29) will allow for a global view of RNA secondary structure as it exists *in vivo* and *in vitro*.

Finally, icSHAPE can be applied to a wide variety of biological systems and questions. Implementation of icSHAPE *in vitro* and *in vivo* recapitulated known global structural features of mammalian transcriptomes; however, we also uncovered the ability to read out RNA-protein interactions<sup>20</sup>. We demonstrated the ability to accurately predict protein-binding sites with a high true positive rate. This was accomplished for the proteins RBFOX and HuR, which are proteins that are important for regulating RNA splicing and stability, respectively<sup>20</sup>. In addition, we found that using only icSHAPE reactivities accurately predicted 85% of known m<sup>6</sup>A chemical modifications across the mouse embryonic stem cell (ESC) transcriptome<sup>20</sup>. An independent study further corroborated our finding, using focused biochemistry and traditional RNA structure footprinting<sup>30</sup>. Concurrently, another group developed a biotin-conjugated SHAPE probe that enriches for RNA structural signal, although whether this probe works in living cells was not reported<sup>24</sup>. Overall, these studies demonstrate the power and importance of chemical enrichment strategies, such as icSHAPE, for studying RNA structure *in vivo*.

#### Experimental design

**Controls and replicates.** As with all techniques, replication is crucial to ensuring accurate results. We find that performing icSHAPE with at least two biological replicates (tissue culture cells grown and NAI-N<sub>3</sub> modified on different days) is sufficient. Additional replication can help if the sequencing depth is low and the computational tools provided to analyze icSHAPE data can support an arbitrary number of both mock (DMSO) and modified (NAI-N<sub>3</sub>) replicates. The design of each icSHAPE experiment should include DMSO samples. The DMSO sample provides an ‘input’ sample, which allows the computational pipeline to determine whether a given region of RNA was present in the sample. This is critical, as low signal in the NAI-N<sub>3</sub> sample could be due to poor coverage (not enough of that RNA in the sample) or low icSHAPE reactivity (an RNA region with low reactivity toward NAI-N<sub>3</sub>). In addition to carefully selecting the system and modification parameters, we have found that performing icSHAPE on living cells and deproteinized RNA provides a valuable comparison. Base reactivity of deproteinized, *in vitro*-folded and NAI-N<sub>3</sub>-modified RNA can provide a transcriptome-wide view

## PROTOCOL

of intrinsic RNA structures that form. Base reactivity of *in vivo* NAI-N<sub>3</sub>-modified RNA will integrate RNA structure and protein-binding events, among other things. The PROCEDURE describes how to obtain both *in vivo*- and *in vitro*-modified RNA. The careful comparison of these independent experiments can yield insights into RNA-binding protein and post-transcriptional RNA modification functions<sup>20</sup>.

**RNA amount and source.** Transcriptome-wide icSHAPE maps require sufficiently complex pools of input RNA. We have found that 500 ng of input RNA (e.g., poly(A)<sup>+</sup> or ribosome-depleted) allowed the measurement of base flexibility for >13,000 transcripts<sup>20</sup>. Although the library construction method for icSHAPE is capable of producing high-quality libraries from at least 100× less starting material, we recommend using at least 500 ng when attempting to measure global RNA secondary structure. The PROCEDURE details poly(A)<sup>+</sup> enrichment coupled to the icSHAPE strategy; however, in principal, any starting RNA material can be used to generate icSHAPE libraries. Specific experimental designs may require purifications to capture non-poly(A)<sup>+</sup> RNAs (e.g., ribosome depletion strategies) or to focus on RNA classes that are specific to particular sizes (e.g., small-RNA fractionation).

icSHAPE was developed to be broadly applicable and easily implemented. We have tested the NAI-N<sub>3</sub> reagent in bacterial (Gram-negative *Escherichia coli*), yeast, insect and mammalian cells for SHAPE experiments<sup>23</sup>. All of these systems proved to be amenable for icSHAPE modification and RNA structure determination. We have successfully generated icSHAPE data from experiments in which the *in vivo* modification occurs at temperatures between 30 and 37 °C, to accommodate the specific culture temperatures. Thus, the procedure can be easily adapted to a wide variety of culture systems to allow transcriptome-wide RNA structure determination in most model systems.

**RNA fragmentation.** icSHAPE initially is performed on full-length transcripts, probably generating multiple NAI-N<sub>3</sub> modifications per RNA molecule. To identify these SHAPE modification sites via RT-PCR and sequencing, the RNA must be fragmented to smaller sizes. Most strategies have been optimized to achieve single-hit kinetics of chemical modification on short (~100 nt) RNAs, which results in only ~10% of the molecules being structurally informative<sup>11</sup>. Therefore, the procedure describes parameters for metal-based RNA fragmentation to generate small RNAs of ~100 nt. Because icSHAPE reads out truncated cDNA molecules, icSHAPE libraries will be, on average, shorter than the starting RNA length. The final insert size should be taken into consideration during the design of each experiment, as longer or shorter fragments may be advantageous in specific conditions.

**Sequencing depth.** The mass amount of starting material and expected complexity (number of unique transcripts) is important when estimating the depth at which a sample will be sufficiently sequenced. Starting with very little initial RNA mass will probably reduce the complexity of the RNA pool and require the user to perform additional PCR cycles, thereby causing more non-unique reads to exist in the final library sample. In contrast, if the icSHAPE experiment is designed to examine the structure of one RNA or a small number of RNAs, fewer sequencing reads are required to achieve sufficient coverage of the RNA for high-quality structure determination. Importantly, the ability to recover only the icSHAPE-modified RNA fragments from the NAI-N<sub>3</sub> sample results in less RNA mass in these samples. Therefore, the DMSO and NAI-N<sub>3</sub> samples will amplify at different times (NAI-N<sub>3</sub> at ~3 PCR cycles after the DMSO), and thus we often dedicate more sequencing depth to the NAI-N<sub>3</sub> samples so as to ultimately achieve the same unique read count as that of the DMSO sample.

## MATERIALS

### REAGENTS

- Cells. icSHAPE has been implemented on bacterial, yeast, insect and mammalian cells<sup>23</sup> **! CAUTION** It is important to regularly check cell cultures to ensure that they are authentic and are not infected with *Mycoplasma*.
- PBS (Life Technologies, cat. no. 10010049)
- Trypsin-EDTA, 0.25% (wt/vol) (Life Technologies, cat. no. 25200-114)
- Trypan blue solution, 0.4% (wt/vol) (Life Technologies, cat. no. 15250-061)
- UltraPure DNase/RNase-free dH<sub>2</sub>O (Life Technologies, cat. no. 10977-023)  
**▲ CRITICAL** All experimental procedures must be performed under DNase- and RNase-free conditions. All reagents should be made or purchased so as to be free of DNases and RNases.
- DMSO (Sigma-Aldrich, cat. no. D2650-5×10ML)
- NAI-N<sub>3</sub>, 2 M (4-step organic synthesis is described<sup>20</sup>)
- Trizol LS reagent (Life Technologies, cat. no. 10296-028)
- Chloroform (Ricca Chemical, cat. no. RSOC0020-500C) **! CAUTION** Trizol and chloroform are hazardous, and caution should be taken when handling them; work should be performed in a chemical fume hood.
- Ethanol (200 proof; Gold Shield, cat. no. 412804)
- Buffer RWT (Qiagen, cat. no. 1067933)
- HEPES, 1 M (Life Technologies, cat. no. 15630-080)
- Magnesium chloride, 1 M (MgCl<sub>2</sub>; Ambion, cat. no. AM9530G)
- Sodium chloride, 5 M (NaCl; Ambion, cat. no. AM9759)
- Click-IT biotin DIBO alkyne (Life Technologies, cat. no. C-10412)
- RiboLock RNase inhibitor (40 U/μl; Thermo Scientific, cat. no. EO0384)
- RNA fragmentation reagents (Ambion, cat. no. AM8740)
- T4 polynucleotide kinase (PNK; New England BioLabs, cat. no. M0201L)

- Tris, pH 7.0, 1 M (Ambion, cat. no. AM9850G)
- FastAP thermosensitive alkaline phosphatase (1 U/μl; Thermo Scientific, cat. no. EF0651)
- T4 RNA ligase 1 (ssRNA ligase), high concentration (New England BioLabs, cat. no. M0437M)
- Preadenylylated and 3'-biotin blocked RNA linker (DMSO samples, IDT, PAGE purified): /5rApp/AGATCGGAAGAGCGGTTCAG/3Biotin/  
**▲ CRITICAL** 5' preadenylylated and 3' blocked RNA linkers are important for efficient 3' end RNA ligations.
- Preadenylylated and 3'-ddC blocked RNA linker (NAI-N<sub>3</sub> samples, IDT, PAGE purified): /5rApp/AGATCGGAAGAGCGGTTCAG/3ddC/
- UltraPure DTT (Invitrogen, cat. no. 15508-013)
- SequaGel UreaGel System, 2.2 liter kit (National Diagnostics, cat. no. EC-833-2.2LTR)
- Ultrapure TEMED (Invitrogen, cat. no. 15524-010)
- Ammonium persulfate (APS; Thomas Scientific, cat. no. 0149N94)
- Ultrapure TBE buffer, 10× (Life Technologies, cat. no. 15581-044)
- EDTA (Ambion, cat. no. AM9260G)
- UltraPure 10% (wt/vol) SDS solution (Life Technologies, cat. no. 15553-027)
- SuperScript III reverse transcriptase (Invitrogen, cat. no. 18080-044)
- Deoxynucleotide solution mix, 10 mM (dNTPs; New England BioLabs, cat. no. N0447L)
- RT-primer-1 (IDT, standard desalting purification, underlined portion is the experimental barcode<sup>31</sup>) /5phos/DDDNNAAACNNNNAG ATCGGAAGAGCGTTCGTTGGA/iSp18/GGATCC/iSp18/TACTGAACCGC; D = A/G/T and N = A/T/G/C are used to discriminate PCR duplicates,

'AACC' is the specific experimental barcode) ▲ **CRITICAL** The quality of the oligonucleotide synthesis should be verified by PAGE analysis; however, bulk PAGE purification before use in RT reactions is not necessary.

- Dynabeads MyOne streptavidin C1 (Life Technologies, cat. no. 65002)
- Tween 20 (Sigma-Aldrich, cat. no. P1379-500ML)
- UltraPure 1 M Tris-HCl buffer, pH 7.5 (Invitrogen, cat. no. 15567-027)
- Nonidet P40 (NP-40) (or substitute, IGEPAL CA-630; Roche, cat. no. 11332473001)
- *N*-Lauroylsarcosine sodium salt solution (20%, for molecular biology; Sigma-Aldrich, cat. no. L7414-10ML)
- Deoxycholic acid sodium salt (Fisher Scientific, cat. no. BP349-100)
- D-Biotin (Molecular Probes, cat. no. B20656)
- RNase cocktail enzyme mix (Ambion, cat. no. AM2286)
- RNaseH (Enzymatics, cat. no. Y9220L)
- Isopropanol (Molecular Biology Grade; Fisher Scientific, cat. no. BP2618-1)
- DNA ladder, 25 bp (Invitrogen, cat. no. 10597-011)
- Gel loading buffer II (Denaturing PAGE; Ambion, cat. no. AM8547)
- SYBR Gold nucleic acid gel stain, 10,000× (Life Technologies, cat. no. S-11494)
- BlueJuice gel loading buffer, 10× (Invitrogen, cat. no. 10816-015)
- 40% (wt/vol) acrylamide/Bis solution, 29:1 (Bio-Rad, cat. no. 161-0146)
- CircLigase II ssDNA ligase (Epicentre, cat. no. CL9025K)
- Phusion high-fidelity (HF) PCR master mix with HF buffer (New England BioLabs, cat. no. M0531L)
- SYBR Green I nucleic acid gel stain, 10,000× (Life Technologies, cat. no. S-7563)
- P5-Solexa PCR primer (IDT, PAGE purified):
- 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'
- P3-Solexa PCR primer (IDT, PAGE purified): 5'-CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCTCGTGAACCGCTCTCCGATCT-3'

#### EQUIPMENT

- Tissue culture plates (Denville, cat. no. T1115)
- Pipette tips (Denville, cat. nos. P1733, P1734, P1735, P1737)
- Micropipettors (Rainin, cat. nos. 17014382, 17014384, 17014388)
- Automatic pipettor (Drummond Pipet-aid, cat. no. 53498-001)
- Conical tubes, 15 and 50 ml (Fisher, cat. nos. 05-538-59A, 07-203-510)
- Filter, 500 ml, 0.22 μm (Millipore, cat. no. SCGVU05RE)
- Inverted light microscope (Leica DMi1)
- Hemocytometer (Reichert Bright-line)
- Cell culture hood (Thermo Forma, Class IIA/B3 Biosafety Cabinet)
- Tabletop centrifuge (Beckman Coulter Allegra 6R Centrifuge, cat. no. 366816)
- Benchtop centrifuge (Eppendorf Centrifuge 5424R)
- Cell scrapers (GeneMate, cat. no. T-2443-4)
- Razor blades (VWR, cat. no. 55411-050)
- Hybridization incubator (SciGene Model 400, cat. no. 1040-60-1)
- Vortexer (Scientific Industries Vortex Genie 2, cat. no. SI-0236)
- RNeasy mini and midi kits (Qiagen, cat. nos. 74106 and 75144)
- Silicized microcentrifuge tubes, 1.5 ml (MidSci, cat. no. AVSC1510)
- Freeze dryer/lyophilizer (Labconco, cat. no. 7522700)
- NanoDrop spectrophotometer (Thermo Scientific)
- Poly(A)Purist MAG kit (Ambion, cat. no. AM1922)
- Thermomixer (Eppendorf)
- Rotator (Thermo LabQuake, with clip bar, cat. no. 400110Q)
- PCR machine (Eppendorf Mastercycler Pro, cat. no. 6321 000.515)
- Amicon Ultra-0.5 Centrifugal Filter Unit (Millipore, cat. no. UFC5010BK)
- High-voltage electrophoresis power supply (Bio-Rad, cat. no. 164-5056)
- PAGE apparatus (CSB Scientific, cat. no. MGCV-102-U)
- Glass plates, 8 × 10 cm (R. Shadel, custom order)
- Combs, 0.75-mm-thick (R. Shadel, custom order)
- Gel spacers, 0.75-mm-thick (R. Shadel, custom order)
- Steel heating plate, 8 × 10 cm (R. Shadel, custom order)
- Binder clips, 2 inch (OfficeMax, cat. no. 09015772)
- Milli-Q filtered water (EMD Millipore Milli-Q, for water purification)
- Scalpels (Fisher, cat. no. 14-823-45)
- Blue light transilluminator (Clare Chemical Research Dark Reader, cat. no. DR22A)

- Tubes, 0.5 ml (BioExpress GeneMate, cat. no. C-3267-1)
- Tubes, 2 ml (BioExpress GeneMate, cat. no. C-3261-1)
- Corning Costar Spin-X centrifuge tube filters, 0.45 μm (Sigma-Aldrich, cat. no. CLS8162-96)
- PCR tubes, 0.2 ml (Eppendorf, cat. no. 951010022)
- DNA Clean & Concentrator-5 columns (Zymo Research, cat. no. D4013)
- RNA Clean & Concentrator-5 (Zymo Research, cat. no. R1016)
- Plastic wrap (E&K Scientific, cat. no. EK-1344)
- Gel imager (Bio-Rad Gel Doc XR+, cat. no. 1708195)
- qPCR machine (Stratagene Mx3005p, capable of mid reaction removal of individual samples)
- Optically Pure 0.2-ml PCR tubes (Agilent Life Sciences, cat. no. 401428)
- Magnetic stand, for 1.5-ml tubes (Invitrogen DynaMAG-2, cat. no. 12321D)
- Magnetic stand, for 0.2-ml tubes (ThermoFisher Magnetic Stand-96, cat. no. AM10027)

#### REAGENT SETUP

**3.3× SHAPE reaction buffer** Final component concentrations are 333 mM HEPES (pH 7.5), 20 mM MgCl<sub>2</sub> and 333 mM NaCl. Prepare the buffer using nuclease-free water. Store the buffer at 25 °C for up to 6 months.

**100 mM NAI-N<sub>3</sub>** Mix 25 μl of 2 M NAI-N<sub>3</sub> with 25 μl of DMSO in a final volume of 500 μl of PBS. Freshly prepare the solution. Before use, thaw the 2 M NAI-N<sub>3</sub> completely to 25 °C, and then mix it with DMSO and PBS.

**DTT solution, 100 mM and 1 M** To make a 1 M stock of DTT, dissolve 154.2 mg in a final volume of 1 ml of nuclease-free water. Dilute it 1:10 to make 100 mM stocks. Divide the solution into aliquots and store them at -20 °C for up to 6 months.

**5× PNK buffer** Final component concentrations are 350 mM Tris (pH 7.0), 90 mM MgCl<sub>2</sub> and 25 mM DTT. Prepare the buffer using nuclease-free water. Divide the solution into aliquots and store them at -20 °C for up to 3 months.

**APS, 10% (wt/vol)** Dissolve 5 g of APS in a final volume of 50 ml of nuclease-free water. Store it at 4 °C for up to 6 months.

**UreaGel denaturing PAGE solution, 6% (wt/vol)** Prepare 1 liter of UreaGel denaturing polyacrylamide gel stock solution by mixing 240 ml of UreaGel System concentrate, 660 ml of UreaGel System diluent and 100 ml of UreaGel System buffer. Store the solution at 4 °C for up to 6 months.

**Native PAGE solution, 1 liter, 6% (wt/vol)** Prepare gel stock solution by mixing 100 ml of 10× TBE solution and 150 ml of 40% (wt/vol) acrylamide/Bis solution in a final volume of 1 liter of nuclease-free water. Store the solution at 4 °C for up to 6 months.

**Bead binding buffer** The final component concentrations of SHAPE binding buffer are 100 mM Tris (pH 7.0), 1 M NaCl, 10 mM EDTA and 0.2% (vol/vol) Tween 20. Prepare the buffer in nuclease-free water. Store it at 4 °C for up to 6 months.

**Streptavidin bead solution** Prepare 20 μl of Dynabeads MyOne streptavidin C1 for each icSHAPE sample by washing the beads twice with 1 ml of bead binding buffer. After washing, resuspend the beads in 10 μl of bead binding buffer per icSHAPE sample, and store them on ice until needed. Freshly prepare the solution for each trial of icSHAPE.

**4 M SHAPE wash buffer** Final component concentrations of SHAPE wash buffer are 100 mM Tris (pH 7.0), 4 M NaCl, 10 mM EDTA and 0.2% (vol/vol) Tween 20. Prepare the buffer in nuclease-free water. Store it at 4 °C for up to 6 months.

**Sodium deoxycholic acid solution (NaDOC), 10% (wt/vol)** Mix 5 g of deoxycholic acid sodium salt in 50 ml of nuclease-free water. Store it at 25 °C for up to 3 months.

**10× RNaseH buffer** Final component concentrations of RNaseH buffer are 0.5 M HEPES (pH 7.5), 0.75 M NaCl, 30 mM MgCl<sub>2</sub>, 1.25% (wt/vol) *N*-lauroylsarcosine sodium salt, 0.25% (wt/vol) NaDOC and 50 mM DTT. Prepare the buffer in nuclease-free water. If a precipitate forms, briefly warm it at 37 °C so that it dissolves into solution. Freshly prepare the solution.

**SYBR Green I, 25×** Make a 1:400 dilution from SYBR Green I nucleic acid gel stain (10,000×) in DMSO. Store it at -20 °C for up to 1 year.

#### EQUIPMENT SETUP

**Software configuration** To install and configure the necessary software, follow the instructions in **Box 1**.

## Box 1 | icSHAPE software installation and configuration ● TIMING ~4 h

The icSHAPE pipeline requires short read mapping software (e.g., Bowtie2) and transcriptome annotation files (for e.g., .gtf files from the Ensembl project).

1. Install Bowtie2 according to the manual from Bowtie2 website: <http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml>.
2. Create a Bowtie2 transcriptome index by downloading the mouse genome DNA sequence file (the primary\_assembly .fa file) from Ensembl website: <http://www.ensembl.org/info/data/ftp/index.html>. For example, currently the latest build of mouse genome can be downloaded from [ftp://ftp.ensembl.org/pub/release-81/fasta/mus\\_musculus/dna/Mus\\_musculus.GRCm38.dna\\_sm.primary\\_assembly.fa.gz](ftp://ftp.ensembl.org/pub/release-81/fasta/mus_musculus/dna/Mus_musculus.GRCm38.dna_sm.primary_assembly.fa.gz). For this protocol, we use the mouse transcriptome.
3. Download the mouse annotation file (.gtf file) from the same Ensembl website.
4. Retrieve the annotated transcript sequence by using the 'gffread' tool in the cufflinks program suite. If cufflinks is not installed, install it according to the manual from the cufflinks website: <http://cole-trapnell-lab.github.io/cufflinks/manual/>. Here 'genome.fa' is the input genome DNA sequence file in fasta format, 'genome.gtf' is the annotation file you downloaded from Ensembl in gtf format, 'genome.size' is a text file of two columns delimited by tab, with the first column representing the sequence name (the same as the one used in 'genome.fa' and 'genome.gtf') and the second column representing its length, and 'transcriptome.fa' is the output transcript sequence file. A typical command is as follows:

```
$ gffread -g genome.fa -s genome.size -W -M -F -G -A -O -E -w transcriptome.fa -d transcriptome.collapsed.info genome.gtf
```

▲ **CRITICAL STEP** In some cases, the annotation file is not complete, or experiment-specific RNA sequences are not included in the Ensembl build. After this step and before the next step, the 'transcriptome.fa' file can be edited to include any RNA sequence of interest as long as the file is kept in FASTA format.

5. Build 5h3 Bowtie2 transcriptome index by using the following command. Here transcriptome.fa is the input FASTA file of all transcripts, and transcriptome is the name of the index library.

```
$ bowtie2-build transcriptome.fa transcriptome
```

### icSHAPE pipeline installation

The icSHAPE pipeline is a Perl (5.10.1) program designed to be run in the UNIX/Linux environment, including Mac OS, and it has been tested on Red Hat Enterprise Linux Server release 6.3. Below, the prefixing '\$' character is the default command prompt. The user only needs to input the command after '\$' character for testing or running.

6. Download icSHAPE pipeline. This can be done using github to clone the icSHAPE pipeline from <https://github.com/qczhang/icSHAPE> and put it into any directory (referred to as installation\_directory in step 7 below) by using the following command.

```
$ git clone https://github.com/qczhang/icSHAPE
```

For more details on installing software or working with git and github, please see manuals or tutorials on the github website (<https://github.com/qczhang/icSHAPE/blob/master/README.md>). Alternatively, download a zipped directory of the icSHAPE pipeline from <https://github.com/qczhang/icSHAPE/archive/master.zip>, and put the package into any installation directory. Unzip it to install.

7. Define environment variables either by adding the following lines to your bash profiles

```
export ICSHAPE=installation_directory
export BOWTIELIB=bowtie_index_directory
export PATH=$PATH:$ICSHAPE/scripts
```

or by editing the configure file (in installation\_directory) so that the pipeline can find scripts and programs used in the pipeline without defining the above environment variables.

8. Open and edit the 'configure' file for additional configurations that specify how the pipeline can find the correct programs (e.g., java, bowtie and icSHAPE binaries and scripts) and scripts (e.g., de-multiplexing, PCR duplicates removal, RT-stop calculation, structural profile calculation, normalization and filtering). Parameters that control the behavior of these programs also can be found in the 'configure' file.

### Other tools for icSHAPE analysis

9. Download and install FastQC for quality control of sequencing data from: <http://www.bioinformatics.babraham.ac.uk/projects/download.html#fastqc>.
10. Download and install Trimmomatic for adapter trimming from: <http://www.usadellab.org/cms/?page=trimmomatic>.
11. Download and install ucsc\_tools if you are generating UCSC tracks for visualization from: [http://hgdownload.soe.ucsc.edu/admin/exe/linux.x86\\_64/](http://hgdownload.soe.ucsc.edu/admin/exe/linux.x86_64/).

## PROCEDURE

### Cell culture ● TIMING ~10 min

- 1| Grow adherent or suspension cells of choice under standard conditions to ~70–80% confluency. Count the number of cells per plate.

2| Collect adherent cells as described in Option A. Collect suspension cells as described in Option B.

▲ **CRITICAL STEP** It is important not to trypsinize the plates that will subsequently be used for icSHAPE experiments, as we have found that this results in partially degraded total RNA.

**(A) Collection of adherent cells**

- (i) Remove the culture medium and rinse the cells once with 1× PBS.
- (ii) Remove PBS and add 1.5 ml or 4 ml of 1× PBS to a 10-cm or 15-cm plate, respectively.
- (iii) Scrape the cells off the plate and prepare aliquots of  $20 \times 10^6$  cells per 15-ml conical tube.
- (iv) Centrifuge each 15-ml tube (1,000g for 2 min at 25 °C) and move directly to Step 3.

**(B) Collection of suspension cells**

- (i) Homogenize the cell suspension by gentle swirling of the culture.
- (ii) Prepare aliquots of  $20 \times 10^6$  cells per 15-ml conical tube.
- (iii) Centrifuge each 15-ml tube (1,000g for 2 min at 25 °C) and move directly to Step 3.

**icSHAPE modification and quality control ● TIMING ~2 h**

3| Aspirate PBS from the pelleted cells (from Step 2), resuspend the cell pellet in 500 µl of 100 mM NAI-N<sub>3</sub> solution or 500 µl of DMSO solution with a cut-tip p1000 and incubate at 37 °C with rotation for 5 min.

▲ **CRITICAL STEP** All NAI-N<sub>3</sub> experiments must be paired with a DMSO experiment. DMSO will not modify the RNA, and thus it provides a background signal for determining the coverage level of any given nucleotide in the sequencing experiment.

▲ **CRITICAL STEP** If you are intending to compare *in vivo* and *in vitro* icSHAPE reactivity profiles, which we have shown can reveal biological events such as protein imprints<sup>20</sup>, an additional DMSO sample should be included for subsequent *in vitro* modification at Step 17 (**Box 2**).

▲ **CRITICAL STEP** At least two independent biological replicates should be performed for each condition.

4| Stop the reaction by centrifuging the samples (2,500g for 1 min at 4 °C) and removing the supernatant.

5| Resuspend the cells in 500 µl of 1× PBS, and then transfer the cell solution to 6 ml of Trizol LS in a 15-ml tube.

6| Vortex the samples for ~15 s to homogenize, add 1.2 ml of chloroform (0.2× volumes), vortex again for ~15 s, and then let the samples incubate at 25 °C for 5 min.

7| Centrifuge the samples (2,500g for 15 min at 4 °C), move the aqueous phase to a new 15-ml tube, add 2× volumes of 96–100% ethanol, and vortex to mix.

8| Purify RNA using an RNeasy midi kit column according to the manufacturer’s protocol (starting at step 4 of the ‘RNA Cleanup’ section).

9| Elute RNA from the RNeasy midi column twice with 250 µl of water (final of 500 µl volume).

10| Quantify RNA yield, make 2-µg aliquots for quality control, and then freeze the RNA in 250-µg aliquots at –80 °C for storage.

**Box 2 | *In vitro* icSHAPE modification ● TIMING ~ 1 h**

1. Obtain 500 ng of poly(A)<sup>+</sup> RNA from DMSO-treated cells from Step 17 of the main PROCEDURE and freeze-dry them on a lyophilizer.
2. Resuspend the RNA sample in 5.2 µl of water and move the samples to a PCR tube.
3. Heat-denature the RNA samples at 95 °C for 2 min, and then transfer them to ice to cool.
4. Add 3.3 µl of 3.3× SHAPE reaction buffer, 0.5 µl of RiboLock, and then incubate the samples at 37 °C for 5 min.
5. Add 1 µl of 1 M NAI-N<sub>3</sub> to each sample, flick the tubes gently to mix, and incubate them at 37 °C for 10 min.
6. Stop the reactions by moving the samples to ice and by adding 35 µl of buffer RLT; mix the samples by inversion and then add 100 µl of 100% ethanol.
7. Purify RNA using an RNeasy mini column. Elute the RNA twice with 50 µl of water (100-µl final volume).
8. Proceed from Step 18 of the main PROCEDURE.

**? TROUBLESHOOTING**

## PROTOCOL

11| Assess RNA quality by agarose gel (1%, wt/vol) analysis. Clear 18S and 28S rRNA bands should be present.

### ? TROUBLESHOOTING

■ **PAUSE POINT** Purified RNA can be stored at  $-80^{\circ}\text{C}$  for >6 months.

### Isolation of poly(A)<sup>+</sup> RNA fraction ● **TIMING** ~5 h to overnight

12| For each biological replicate, thaw one 250- $\mu\text{g}$  aliquot of DMSO-treated RNA and *in vivo* NAI-N<sub>3</sub>-modified RNA (from Step 11) on ice. If you are intending to include *in vitro* NAI-N<sub>3</sub>-modified RNA in the analysis, thaw an additional DMSO-treated sample, which will be NAI-N<sub>3</sub>-modified at Step 17 after polyA<sup>+</sup> selection. For a standard experiment aimed at fully characterizing the RNA structure landscape, at least two biological replicates are needed, giving a total of six samples (2 $\times$  DMSO, 2 $\times$  *in vivo* NAI-N<sub>3</sub> and 2 $\times$  *in vitro* NAI-N<sub>3</sub>); however, additional biological replicates can be added for higher confidence.

13| Isolate poly(A)<sup>+</sup> RNA using the Poly(A)Purist MAG kit or an equivalent enrichment procedure. Bring each sample to a final elution volume of 400  $\mu\text{l}$ . When using the Poly(A)Purist MAG kit, this will be 400  $\mu\text{l}$  of RNA storage solution (part of the kit).

14| Wash the poly-dT beads once with 500  $\mu\text{l}$  of wash solution 1 and reuse the beads to perform a second poly(A)<sup>+</sup> enrichment, without purifying the first-step enriched poly(A)<sup>+</sup> RNA out of the RNA storage solution.

▲ **CRITICAL STEP** We perform two rounds of poly(A)<sup>+</sup> RNA enrichment. This ensures the lowest amount of rRNA contamination while still yielding high amounts of poly(A)<sup>+</sup> RNA for subsequent biochemical reactions.

15| After the second poly(A)<sup>+</sup> enrichment, freeze each RNA sample (~400  $\mu\text{l}$  each) for >10 min on dry ice and dry it on a lyophilizer (overnight for convenience).

16| Resuspend the enriched poly(A)<sup>+</sup> RNA in 100  $\mu\text{l}$  of water, add 350  $\mu\text{l}$  of buffer RLT, 900  $\mu\text{l}$  of 100% ethanol, and then clean up using an RNeasy mini column. Elute the RNA twice with 50  $\mu\text{l}$  of water (100- $\mu\text{l}$  final volume).

17| Quantify poly(A)<sup>+</sup> RNA and make 500-ng aliquots into individual 1.5-ml tubes. For each icSHAPE library reaction, 400–500 ng of RNA is needed for successful library generation.

▲ **CRITICAL STEP** At this point, two DMSO-treated samples from Step 17 can be used for *in vitro* NAI-N<sub>3</sub> modification, as described in **Box 2**.

■ **PAUSE POINT** Poly(A)<sup>+</sup> RNA can be stored at  $-80^{\circ}\text{C}$  for >6 months.

### Biotin click reaction of NAI-N<sub>3</sub> ● **TIMING** ~3 h

18| Thaw purified DMSO- and NAI-N<sub>3</sub>-modified poly(A)<sup>+</sup> RNA (from Step 17 and **Box 2**) on ice, and then add 2  $\mu\text{l}$  of 1.85 mM DIBO-biotin and 1  $\mu\text{l}$  of RiboLock.

19| Incubate the samples at 37  $^{\circ}\text{C}$  in a Thermomixer for 2 h at 1,000 r.p.m. Under these conditions, the reaction will proceed to completion with all of the SHAPE adducts finally obtaining a biotin handle<sup>20</sup>.

20| Stop the click reaction by purifying the RNA with a Zymo RNA Clean & Concentrator-5 column according to the manufacturer's protocol. Elute RNA twice in 10  $\mu\text{l}$  of water, and freeze-dry it on a lyophilizer.

■ **PAUSE POINT** Biotinylated RNA and DMSO samples can be stored at  $-80^{\circ}\text{C}$  for >6 months.

### RNA fragmentation and RNA end repair ● **TIMING** ~1.5 h

21| Resuspend the RNA samples from Step 20 in 9  $\mu\text{l}$  of water, preheat a PCR block to 95  $^{\circ}\text{C}$  and move RNA to a PCR tube.

22| Heat the RNA samples to 95  $^{\circ}\text{C}$  for 60 s, add 1  $\mu\text{l}$  of 10 $\times$  fragmentation reagent and incubate for 30–60 s.

▲ **CRITICAL STEP** RNA fragmentation time should be determined empirically by urea PAGE gel analysis. Once it is optimized, fragmenting conditions are highly reproducible. To facilitate accurate timing, only two samples should be processed in parallel for RNA fragmentation.

23| Add 1  $\mu\text{l}$  of 10 $\times$  stop solution to each sample and immediately place the samples on ice.

24| Purify fragmented RNA with a Zymo RNA Clean & Concentrator-5 column; freeze-dry as in Step 20.

■ **PAUSE POINT** Freeze-dried RNA samples can be stored at  $-80^{\circ}\text{C}$  for >6 months.



25| Resuspend RNA in 10  $\mu$ l of RNA end-repair mix (tabulated below) and incubate it at 37 °C for 1 h.

Component	Amount ( $\mu$ l)	Final concentration
T4 PNK buffer, 5 $\times$	2	1 $\times$
RiboLock, 40 U/ $\mu$ l	1	4 U/ $\mu$ l
FastAP 1 U/ $\mu$ l	1	0.1 U/ $\mu$ l
T4 PNK 10 U/ $\mu$ l	2	2 U/ $\mu$ l
RNase-free water	4	
Total	10 (per reaction)	

■ **PAUSE POINT** End-repaired RNA samples can be stored at -80 °C for ~1 month.

**RNA 3'-end ligation and size selection** ● **TIMING** ~5 h

26| Add 10  $\mu$ l of 3' RNA ligation mix (tabulated below) to each sample from Step 25, mix by flicking, briefly centrifuge, and incubate at 25 °C for 3 h.

Component	Amount ( $\mu$ l)	Final concentration
RNA ligase buffer, 10 $\times$	1	1 $\times$
DTT, 100 mM	1	5 mM
Preadenylylated and 3'-blocked RNA linker, 10 $\mu$ M	1	0.5 $\mu$ M
T4 RNA ligase 1, high concentration	1	0.66 U/ $\mu$ l
PEG8000, 50% (vol/vol)	6	15%
Completed end-repair reaction	10	
Total	20 (per reaction)	

▲ **CRITICAL STEP** DMSO-treated RNA samples require a 3'-biotin RNA linker, whereas NAI-N<sub>3</sub>-treated RNA samples require a 3'-ddC RNA linker. This ensures that every RNA molecule from the DMSO sample is recovered, whereas only those RNAs that were icSHAPE-modified are recovered from the NAI-N<sub>3</sub> sample.

27| Add 30  $\mu$ l of water to each sample (total volume of 50  $\mu$ l) and purify RNA with a Zymo RNA Clean & Concentrator-5 column; freeze-dry the sample as in Step 20.

28| Resuspend RNA in 7  $\mu$ l of gel loading buffer II, and flick the tube to mix the samples.

29| Prerun 6% (wt/vol) UreaGel denaturing PAGE, first at 25 W for 3 min and then at 50 W for 5 min. Use a metal (e.g., steel or aluminum) heating plate to evenly disperse the heat across the gel.

30| Load RNA samples onto the gel, and run them at 50 W for 4–10 min.

▲ **CRITICAL STEP** These pre-run and run conditions apply to home-poured mini-gels (~8  $\times$  10 cm); longer times will be required to sufficiently heat up larger gels. The duration of the gel run should be sufficient to separate the 3' RNA linker (~20 nt) from ligated RNA products (more than ~40 nt). However, excessive running times will require large gel slices to be cut at Step 32 and could lower the efficiency of RNA elution.

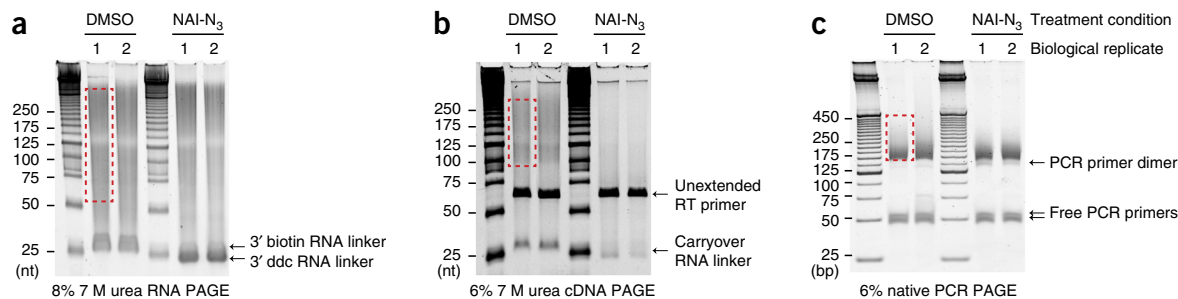
31| Stain the gel in SYBR Gold mix for 3 min at 25 °C and image for 0.3 s on a UV transilluminator for record keeping.

32| Move the gel to a blue light transilluminator, cut the samples above ~45 nt (**Fig. 2a**) and move the gel slices into 0.65-ml tubes nested in 2-ml tubes. Each 0.65-ml tube should be prepared with a needle (18–22 G) hole at the bottom of the tube.

33| Centrifuge the samples (16,000g for 5 min at 25 °C) and discard the 0.65-ml tube.

34| Add 250  $\mu$ l of water to each sample and heat it to 70 °C for 10 min, with shaking at 1,000 r.p.m. in a Thermomixer.

## PROTOCOL



**Figure 2** | Experimental data from representative icSHAPE experiments. (a) Urea PAGE analysis of fragmented and 3'-end-ligated RNA generated in Step 31. The free RNA linkers are labeled (black arrows) and the region of the gel cut for subsequent steps is shown within a dashed red box for one sample. (b) Urea PAGE analysis of cDNAs generated in Step 55. Carry-over RNA linkers and unextended RT primers are labeled (black arrows), and the red box (dashed red line) denotes the region cut for subsequent steps. (c) Native PAGE analysis of icSHAPE library DNA generated in Step 64. Un-incorporated PCR primers and PCR primer-dimer products are labeled (black arrows). The final library DNA, above the PCR primer dimer, is highlighted (dashed red box) and subsequently cut from the gel and analyzed on an Illumina deep-sequencing platform.

**35** | Pulse-centrifuge each sample, transfer the gel slurry into a Spin-X column with a cut-tip p1000 and centrifuge (6,500*g* for 1 min at 25 °C) the sample.

**36** | Resuspend the gel in the top chamber of the Spin-X column with 250  $\mu$ l of water, and repeat Steps 34 and 35 to obtain a final 500- $\mu$ l elution volume. Reuse the Spin-X columns for the second elution.

**37** | Transfer the 500- $\mu$ l eluate to a 0.5-ml 10K Amicon column and centrifuge the sample (13,500*g* for 5 min at 25 °C). This should concentrate the sample to  $\sim$ 100  $\mu$ l.

**38** | Elute the samples as per the manufacturer's instructions: flip the Amicon column over into a new 2-ml elution tube and centrifuge the samples (2,000*g* for 2 min at 25 °C).

**39** | Purify RNA with a Zymo RNA Clean & Concentrator-5 column; freeze-dry it as in Step 20.

■ **PAUSE POINT** Freeze-dried RNA samples can be stored at  $-80$  °C for >6 months.

### cDNA synthesis, isolation of NAI-N<sub>3</sub>-modified molecules, and cDNA purification ● **TIMING** $\sim$ 4.5 h

**40** | Resuspend RNA samples from Step 39 in 11  $\mu$ l of RNase-free water and 1  $\mu$ l of 1  $\mu$ M RT primer; flick the tubes to mix, and move the samples into PCR tubes.

**41** | Heat the samples to 70 °C for 5 min in a PCR block, cool the samples to 25 °C by stepping down 1 °C every 3 s (45 steps); hold at 25 °C.

**42** | Add 8  $\mu$ l of RT enzyme mix (tabulated below) and heat the samples to 25 °C for 3 min, 42 °C for 5 min, 52 °C for 30 min; hold at 4 °C.

Component	Amount ( $\mu$ l)	Final concentration
First strand buffer, 5 $\times$	4	1 $\times$
RiboLock, 40 U/ $\mu$ l	0.75	1.5 U/ $\mu$ l
DTT, 100 mM	1	5 mM
dNTPs, 10 mM (10 mM each)	1	0.5 mM
SuperScript III	1.25	12.5 U/ $\mu$ l
Total	20 (per reaction)	

▲ **CRITICAL STEP** Do not heat samples above 37 °C from this point until cDNA elution at Step 50. Premature heating could cause unwanted elution of samples.

■ **PAUSE POINT** Extended cDNA samples can be stored at  $-80$  °C for  $\sim$ 1 month.

- 43|** Add 10  $\mu\text{l}$  of streptavidin bead solution, mix by pipetting up and down until the sample is homogeneous, and then incubate the sample at 25 °C for 45 min with rotation.
- 44|** Apply the samples to DynaMag-96 side magnet for 1 min, and remove 25  $\mu\text{l}$  of supernatant.
- 45|** Add 100  $\mu\text{l}$  of 4 M SHAPE wash buffer to each sample and transfer the samples to 1.5-ml siliconized tubes. Add an additional 400  $\mu\text{l}$  (500  $\mu\text{l}$  final) of 4 M SHAPE wash buffer and invert the tubes four times to mix.
- 46|** Apply the samples to a DynaMag2 magnet for 1 min, and then remove the supernatant.
- 47|** Add 500  $\mu\text{l}$  of 4 M SHAPE wash buffer to each sample and repeat Steps 46 and 47 four more times (a total of 5 washes).
- 48|** Add 500  $\mu\text{l}$  of 1 $\times$  PBS to each sample and repeat once more (a total of two washes).
- 49|** Resuspend the beads in 50  $\mu\text{l}$  of RNaseA/T1/H elution mix (tabulated below) and incubate them at 37 °C for 30 min at 1,000 r.p.m. in a Thermomixer.

Component	Amount ( $\mu\text{l}$ )	Final concentration
RNaseH buffer, 10 $\times$	5	1 $\times$
D-biotin, 50 mM	12.5	12.5 mM
P3-Solexa-PCR primer, 1 $\mu\text{M}$	1	20 mM
RNase cocktail enzyme mix	1	
RnaseH	1	0.1 U/ $\mu\text{l}$
RNase-free water	30.5	
Total	50 (per reaction)	

- 50|** Add 1  $\mu\text{l}$  of 100% DMSO to each sample and heat to 95 °C for 4 min. Pulse-centrifuge the samples, apply them to DynaMag2 magnet for 1 min, and transfer the eluted cDNA to new 1.5-ml siliconized tubes.
- 51|** Purify cDNA with a Zymo DNA Clean & Concentrator-5 column. Add a 7 $\times$  volume (350  $\mu\text{l}$ ) of DNA binding buffer, mix by vortexing, and then add 350  $\mu\text{l}$  (1 volume) of 96–100% ethanol and mix by vortexing. Wash the columns according to the manufacturer's protocol. Elute the cDNA twice with 10  $\mu\text{l}$  of water (final of 20  $\mu\text{l}$ ) and freeze-dry as in Step 20.
- ▲ CRITICAL STEP** Use of the Zymo DNA Clean & Concentrator-5 columns under the conditions detailed in Step 51 will recover single-stranded DNA (cDNA) as small as 20 nt. Standard binding conditions recover DNA of 50 bp or larger, which would result in a loss of short single-stranded cDNA molecules.
- PAUSE POINT** Biotin-enriched cDNA samples can be stored at –80 °C for >6 months.
- 52|** Resuspend cDNA in 7  $\mu\text{l}$  of Gel loading buffer II, and flick to mix the samples.
- 53|** Prerun 6% (wt/vol) denaturing PAGE, first at 25 W for 3 min, and then at 50 W for 5 min. During the pre-run, boil samples for 2 min at 95 °C, and then place them on ice for 1 min.
- 54|** Load cDNA samples onto the gel and run them at 50 W for 4–10 min.
- 55|** Stain, image and cut the gel as above for the RNA size selection (Steps 31 and 32). See **Figure 2b** for representative results.
- ▲ CRITICAL STEP** It is important to separate the cDNA products from any excess unextended RT primer, which will run at ~50 nt. Isolation of cDNA >70 nt (~20 nt of cDNA and ~50 nt of RT primer) will provide sufficient insert sequence length for high-confidence mapping to mammalian genomes.
- 56|** Crush the gel slices, elute cDNA, filter over a Spin-X and concentrate with 10K Amicon columns, as in Steps 37 and 38.

## PROTOCOL

57| Purify the concentrated cDNA samples with a Zymo DNA Clean & Concentrator-5 column, as in Step 51. Elute cDNA twice in 8.5  $\mu$ l of water (final of  $\sim$ 16  $\mu$ l). Transfer the cDNA to PCR tubes.

■ **PAUSE POINT** PAGE-purified cDNA samples can be stored at  $-80$  °C for >6 months.

### cDNA circularization, library amplification and size selection ● **TIMING** $\sim$ 5 h

58| Add 4  $\mu$ l of circularization reaction mix (tabulated below) to each sample from Step 57, and incubate them at 60 °C for 2 h.

Component	Amount ( $\mu$ l)	Final concentration
CirLigase II buffer, 10 $\times$	2	1 $\times$
MnCl <sub>2</sub> , 50 mM	1	2.5 mM
CirLigase II	1	5 U/ $\mu$ l
Total	20 (per reaction)	

59| Purify cDNA with a Zymo DNA Clean & Concentrator-5 column as in Step 51 and elute twice in 10  $\mu$ l of water (final of  $\sim$ 20  $\mu$ l). Transfer cDNA to optical PCR tubes.

■ **PAUSE POINT** Purified circularized cDNA samples can be stored at  $-80$  °C for >6 months.

60| Add 21.4  $\mu$ l of qPCR reaction mix (tabulated below) to each circularized cDNA sample.

Component	Amount ( $\mu$ l)	Final concentration
Phusion HF PCR master mix, 2 $\times$	20	1 $\times$
SYBR Green I, 25 $\times$	0.4	0.25 $\times$
Solexa PCR primer, 10 $\mu$ M (10 $\mu$ M forward and 10 $\mu$ M reverse)	1.0	0.5 $\mu$ M
Total	21.4 (per reaction)	

61| Perform quantitative PCR (qPCR) on the samples from Step 60. Track the amplification of each library in real time on the qPCR machine and stop each sample as is necessary.

Cycle number	Denature	Anneal	Extend
1	98 °C, 45 s	—	—
2– <i>n</i>	98 °C, 15 s	65 °C, 20 s	72 °C, 60 s

▲ **CRITICAL STEP** To avoid overamplification, real-time qPCR is used to monitor the progress of each library individually. Samples become sufficiently amplified after reaching a fluorescence unit value of  $\sim$ 8,000–10,000 on the Agilent Mx3000P real-time PCR system. Amplification values for machines need to be determined empirically; however, for DMSO- and NAI-N<sub>3</sub>-treated samples, *n* is expected to be cycles 5–8 and 8–11, respectively.

### ? **TROUBLESHOOTING**

62| Purify PCR DNA with a Zymo DNA Clean & Concentrator-5 column by adding 5 $\times$  DNA binding buffer, according to the manufacturer's protocol, eluting twice in 10  $\mu$ l (final of  $\sim$ 18  $\mu$ l) and freeze-drying the samples.

63| Load the samples onto a 6% (wt/vol) native PAGE gel and run them at 160 V for  $\sim$ 25 min.

64| Stain the gel in SYBR Gold mix for 3 min at 25 °C and image for 0.3 s on a UV transilluminator for record keeping.

65| Move the gel to a blue light transilluminator and cut the samples above  $\sim$ 150 bp. See **Figure 2c** for representative results.

▲ **CRITICAL STEP** Library PCR may generate other products, such as primer-dimer species. These will run at  $\sim$ 120–130 bp and should be avoided; contamination of library material with primer-dimer will lower the clustering density of the final icSHAPE library on deep-sequencing platforms.

### ? **TROUBLESHOOTING**

**66** | Crush the gel slices, elute the DNA twice and Amicon-concentrate the PCR DNA as in Steps 37 and 38.

**67** | Clean up the PCR samples with a Zymo DNA Clean & Concentrator-5 column as in Step 51, and elute twice in 13 µl of water (final of ~24 µl).

**68** | Quantify 1 µl of each library using a high-sensitivity Bioanalyzer chip, pool the desired samples for multiplexing if necessary, and sequence icSHAPE libraries on an Illumina-based machine using standard conditions. icSHAPE libraries are paired-end compatible; however, 75-bp single-end sequencing is sufficient to identify RT stops and to confidently map to mammalian transcriptomes. As the random and experimental barcodes are ‘in-line’, the standard Illumina 6- or 8-nt barcode read should be omitted.

**icSHAPE data analysis** ● **TIMING ~1–2 d**

**69** | Obtain the raw sequencing reads from Step 68 and unzip if necessary. Sequencing reads usually will be in the FASTQ file format. Sometimes they are zipped to save space (as in our example .fastq file, provided at <https://github.com/qczhang/icSHAPE/blob/master/data/test/hiseq.nanog.fastq>). If this is the case, they need to be unzipped by using the ‘unzip’ command (or ‘gunzip’ if the file is ‘gzipped’, which is usually implied by the suffix ‘.gz’):

```
$ unzip example.fastq.zip
```

**70** | If the sequencing data contain multiple libraries, they should be demultiplexed according to their barcodes. To split the sequencing data, use the following command:

```
$ splitFastq.pl -U your_fastq_file1:your_fastq_file2:...:your_fastq_fileN -l
GGTT:LIB_DMSO1::TTGT:LIB_DMSO2::ACCT:LIB_NAI1.invivo::CAAT:LIB_NAI2.invivo::TGGC:LIB_N
AI1.invitro::GGTC:LIB_NAI2.invitro::others:unmatched -b 5:4 -d library_split -s
hiseq_barcode.stat
```

**▲ CRITICAL STEP** In this example, single-end sequencing reads are contained in *N* files, and they actually belong to two DMSO-treated samples and four NAI-N<sub>3</sub>-modified samples (six total samples). The two DMSO libraries contain the barcodes GGTT and TTGT, and the four NAI-N<sub>3</sub> libraries contain the barcodes ACCT, CAAT, TGGC and GGTC at positions 5–8 (-b 5:4). The split files will be stored in the directory library\_split, and the statistics on the number of reads of different barcodes are stored in the file hiseq\_barcode.stat.

**71** | Check the quality of the reads from the FASTQ files using a sequencing quality control program such as FastQC. This tool can be used to produce summary graphs and tables that inform the user of potential problems in the data. For example, the software can readily identify low-quality sequence, and it usually suggests that the user trim the read to exclude this region. Use the following command to run FastQC:

```
$ fastqc -o out_directory libX.fastq
```

**72** | Run the icSHAPE pipeline either automatically, as described in Step 72A, or manually, as described in Step 72B.

**(A) Running the icSHAPE pipeline ‘automatically’**

(i) Run the pipeline by typing the following example command:

```
$ icSHAPE_pipeline.pl -i LIB_DMSO1.fastq:LIB_DMSO2.fastq -t
LIB_NAI1.invivo.fastq:LIB_NAI2.invivo.fastq -o test -c configure
```

Here LIB\_DMSO1.fastq and LIB\_DMSO2.fastq are two replicates of the DMSO (input) background libraries, and LIB\_NAI1.invivo.fastq and LIB\_NAI2.invivo.fastq are two replicates of the *in vivo* NAI-N<sub>3</sub>-modified libraries. The configure file is a tab-delimited text file, and it contains information for the input data, and on where to look for specific programs and their associated parameters. For example, the key ‘FASTQCODING’ defines the coding of input FASTQ reads, and the key ‘COLLAPSEBIN’ specifies which program to use to remove PCR duplicates. Although many important keys will be discussed in the following sections, please see the associated ‘readme’ file (<https://github.com/qczhang/icSHAPE/blob/master/README.md>) for a complete description of the icSHAPE pipeline.

The pipeline will create a directory called ‘test’ and put all intermediate files and final output into the directory. Please see **Figure 1b** for the workflow of the pipeline and **Table 1** for the intermediate files generated in each step.



**TABLE 1** | icSHAPE pipeline file input and output names.

Step	Input	Output
Read collapse	libX.fastq	libX.rmdup.fastq
Read trimming	libX.rmdup.fastq	libX.trimmed.fastq
Read mapping	libX.trimmed.fastq	libX.sam (*)
Transcript abundance estimation	libX.sam	libX.rpkm (*)
RT-stop calculation	libX.sam, libX.rpkm	libX.rt (*)
Replicate combining	libX.rt, libY.rt, libZ.rt ...	background.rt or target.rt
RT-stop normalization	background.rt Or Target.rt	background.normalized.rt Or target.normalized.rt
Enrichment reactivity score calculation	background.normalized.rt and target.normalized.rt	icSHAPE.tmp.out (*)
Valid score filtering	icSHAPE.tmp.out	icSHAPE.out (*)

We suggest keeping some important intermediate files (denoted with an asterisk (\*)) that will allow rapid rerunning of the icSHAPE pipeline under different parameter settings.

The final output, 'icSHAPE.out', contains the filtered valid structural profile for each transcript. All other files are intermediate results, and in principle they could be deleted to save space after the job is accomplished. We suggest keeping some important intermediate files (denoted with an asterisk (\*) in **Table 1**) that will allow rapid rerunning of the icSHAPE pipeline under different parameter settings. In this case, you do not need to run the whole pipeline from the beginning to the end and can choose specific steps to run.

**▲ CRITICAL STEP** Normally, if the pipeline is correctly configured, the 'automatic' running mode will run smoothly and generate the final results. However, incorrect configuration or error input data could result in failure of the pipeline. Each step in the pipeline is important for completion of the analysis and generation of structure profiles. The 'manual' implementation (Step 72B) can be used to perform each step individually if errors occur during the 'automatic' running mode.

**(B) Running the icSHAPE pipeline 'manually'**

(i) Remove PCR duplicates for each library with the following command:

```
$ readCollapse.pl -U libX.fastq -o libX.rmdup.fastq -f libX.fa
```

Here the input file 'libX.fastq' is a single-end sequencing data set and the output file 'libX.rmdup.fastq' contains collapsed reads after PCR duplicate removal. The frequency of each sequence in the input file is recorded in the 'libX.fa' file. Please check 'readCollapse.pl' for more options.

(ii) Trim adapter and barcode sequences from reads with the following command:

```
$ trimming.pl -U libX.rmdup.fastq -o libX.trimmed.fastq -l 13 -t 0 -c phred33 -a adapter.fa -m 25
```

Here the input file 'libX.rmdup.fastq' is the output-collapsed sequence file from the Step 72B(i), and the output file 'libX.trimmed.fastq' contains reads after trimming. This command will trim the leading 13-nt sequence (barcodes used in the current icSHAPE protocol) and the adapter sequence in the file 'adapter.fa'. If the remaining read is <25 nt, it will be discarded. The default adapter trimmer used in icSHAPE is 'Trimomatic', but it can be easily replaced by other tools such as 'cutadapt' (<https://cutadapt.readthedocs.org/en/stable/>).

(iii) Map the reads to the transcriptome with the following command:

```
$ bowtie2 -U libX.trimmed.fastq -S libX.sam -x mouse.transcriptome --non-deterministic --time
```

Here the input file 'libX.trimmed.fastq' is the output-trimmed sequence file from Step 72B(ii), 'libX.sam' is the output alignment file and 'mouse.transcriptome' should point to the location of a Bowtie2 index.

**? TROUBLESHOOTING**

(iv) Estimate transcript abundance with the following command:

```
$ estimateRPKM.pl -i libX.sam -o libX.rpkm
```

Here 'libX.sam' is the alignment file from the Step 72B(iii), and 'libX.rpkm' is the output transcript abundance file in the RPKM metric<sup>32</sup>.



- (v) Calculate the RT stop in each transcript with the following command:

```
$ calcRT.pl -i libX.sam -o libX.rt -r libX.rpkm -c 1
```

Here 'libX.sam' and 'libX.rpkm' are the alignment file and the transcript abundance file from the Step 72B(iii) and 72B(iv), respectively, and 'libX.rt' is the output RT-stop file. By setting '-c 1', transcripts of RPKM <1 will be skipped in the calculation.

- (vi) Merge background (DMSO) and target (NAI-N<sub>3</sub>) libraries into two separate files with the following command:

```
$ combineRTreplicates.pl -i libX.rt:libY.rt:libZ.rt -o background.rt
```

Here 'libX.rt', 'libY.rt' and 'libZ.rt' are the input RT-stop files of the background libraries, and 'background.rt' is the output combined RT-stop file from Step 72B(v). Repeat this also to combine and obtain 'target.rt', the target-modified RT-stop file.

**▲ CRITICAL STEP** It is important to make sure that the replicates agree with each other well before combining them. Use the script 'correlation.pl' to calculate the correlation of RT-stop profiles of different replicates:

```
$ correlationRT.pl -1 libX.rt -2 libY.rt 1> libX.Y.corr.txt
```

For each common transcript in libX and libY, the script will calculate their correlation of RT-stop profiles. Investigate and/or plot the distribution with statistical programs such as R (<http://www.r-project.org/>) to check the overall correlation between the two replicates. The script has other parameters that allow calculation of correlation of specific regions of a transcript. Explore them by checking the help from the command line.

### ? TROUBLESHOOTING

- (vii) Normalize the combined background and target RT-stop files with the following command:

```
$ normalizeRTfile.pl -i background.rt -o background.normalized.rt -m
mean:vigintile2 -d 32 -l 32
```

Here 'background.rt' is the input-combined background RT-stop file from the Step 72B(vi), and 'background.normalized.rt' is the output-normalized background RT-stop file. The other parameters normalize the mean of the RT stops of the 90–95% most reactive bases to 1 and scale all other RT stops proportionally. The leading 32 nt ('-d 32') and trailing 32 nt ('-l 32') will be excluded in the calculation of normalizing benchmarks.

- (viii) Calculate enrichment reactivity scores for all transcripts with the following command:

```
$ calcEnrich.pl -f target.normalized.rt -b background.normalized.rt -o
icSHAPE.tmp.out -w factor5:scaling1
```

Here 'target.normalized.rt' and 'background.normalized.rt' are the input-normalized target and background RT-stop files from the Step 72B(vii), and 'icSHAPE.tmp.out' is the output file that contains enrichment scores. Setting '-w factor5:scaling1', the enrichment reactivity score is scaled into the range of [0, 1], after removing the outliers by 90% Winsorization (the top 5th percentile is set to 1 and the bottom 5th percentile is set to 0).

- (ix) Filter valid enrichment reactivity scores for regions passing a certain threshold with the following command:

```
$ filterEnrich.pl -i icSHAPE.tmp.out -o icSHAPE.out -T 2 -t 200 -s 5 -e 30
```

Here 'icSHAPE.tmp.out' is the input raw icSHAPE enrichment score file from Step 72B(viii), and 'icSHAPE.out' is the output file that contains valid icSHAPE enrichment scores. The other parameters are defined to select the valid scores for those transcripts that have an average hit coverage higher than 2 ('-T 2') in the target library, and within regions that have a background base density higher than 200 ('-t 200'), and to skip the leading 5 ('-s 5') and the ending 32 nt ('-e 30').

### ? TROUBLESHOOTING

- 73|** Generate bigwig files for visualization with the University of California, Santa Cruz (UCSC) genome browser by using the following commands:

```
$ enrich2Bedgraph.pl icSHAPE.out > icSHAPE.bedgraph
$ sort -k1,1 -k2,3n icSHAPE.bedgraph -o icSHAPE.sorted.bedgraph
$ uniqueTrack.pl icSHAPE.sorted.bedgraph icSHAPE.sorted.uniq.bedgraph
$ cut -f1-4 icSHAPE.sorted.uniq.bedgraph | grep -v NULL > icSHAPE.sim.bedgraph
$ bedGraphToBigWig icSHAPE.sim.bedgraph genome.chr.size icSHAPE.sim.bw
```

## PROTOCOL

Here we first convert the reactivity score for each base in the transcript coordinates into genome coordinates (file icSHAPE.out from the Step 72B(ix)) and then merge genome positions that are covered by multiple transcripts. Finally, convert the .bedgraph file to a .bw file by using the UCSC utility 'bedGraphToBigWig'. The file 'genome.chr.size' contains the lengths of all chromosomes of the investigated genome.

### ? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 2**.

**TABLE 2** | Troubleshooting table.

Step	Problem	Possible reason	Solution
Step 11, <b>Box 2</b>	Low recovery of NAI-N <sub>3</sub> -modified poly(A) <sup>+</sup> RNA	The duration of NAI-N <sub>3</sub> modification is probably too long for the specific cell type. Overmodification with NAI-N <sub>3</sub> can lead to excess acylation of the poly(A) tail, which can reduce the efficiency of the poly(A) <sup>+</sup> RNA purification	Reduce the duration of NAI-N <sub>3</sub> modification to cells or <i>in vitro</i> RNA
61	Poor qPCR amplification of NAI-N <sub>3</sub> -modified samples	The initial icSHAPE modification with NAI-N <sub>3</sub> was not sufficiently long	After RNA fragmentation, ~10% of the NAI-N <sub>3</sub> sample should be recovered after biotin capture, relative to the DMSO samples, which should recover ~100% of the fragmented RNA. The NAI-N <sub>3</sub> samples should require three or four additional cycles of PCR compared with the DMSO samples. If there is a larger difference in PCR cycles between the DMSO and NAI-N <sub>3</sub> samples, increase the time of NAI-N <sub>3</sub> modification
	Sequencing library has low complexity	The final number of qPCR cycles is >12, especially for NAI-N <sub>3</sub> samples	Prepare multiple samples in the same tubes. Proceed with Steps 3–42 (cDNA synthesis). After RT, samples can be pooled for subsequent purification, as the individual samples are now barcoded
65	NAI-N <sub>3</sub> final library inserts are too short	NAI-N <sub>3</sub> causes truncated cDNA fragments to be shorter than in the DMSO sample	The RNA fragmentation (Steps 21–24) can be optimized to yield larger initial fragments, producing on average longer cDNAs and inserts
72B(iii)	Bowtie mapping rate is low	Trimming of adapters was not successful; mapping to wrong references	Make sure that the correct adapter sequences are used in running 'trimming.pl', and check the running log file to see whether the trimmer has been run correctly. Download or build the correct transcriptome library for Bowtie mapping
72(B)vi	Correlation of different replicates is low	Inconsistent preparation of RNA sequencing libraries between replicates	Check the differences in performing replicate experiments, and re-do the experiments, taking care to follow the protocol in a consistent manner
72(B)ix	Pipeline outputs no or few RNA structure profiles	Sequencing depth is low	Prepare enough sequencing sample and re-do sequencing to get enough reads that cover most, if not all, sequences in the library

### ● TIMING

Steps 1 and 2, cell culture: ~10 min

Steps 3–11, icSHAPE modification and quality control: ~2 h

Steps 12–17, isolation of poly(A)<sup>+</sup> RNA fraction: ~5 h (to overnight)

Steps 18–20, biotin click reaction of NAI-N<sub>3</sub>: ~3 h

Steps 21–25, RNA fragmentation and RNA end repair: ~1.5 h

Steps 26–39, RNA 3'-end ligation and size selection: ~5 h

Steps 40–57, cDNA synthesis, isolation of NAI-N<sub>3</sub>-modified molecules and cDNA purification: ~4.5 h

Steps 58–68, cDNA circularization, library amplification and size selection: ~5 h

Steps 69–73, icSHAPE data analysis: ~1–2 d

**Box 1**, icSHAPE software installation and configuration: ~4 h

**Box 2**, *in vitro* icSHAPE modification: ~1 h



**ANTICIPATED RESULTS**

The icSHAPE protocol generates many experimental data points, all of which are diagnostic of prior steps. RNA fragmentation and ligation can be assessed by urea PAGE (Step 31); an even smear of RNA signal is expected. The RNA linker will be easily identified as a band just larger than 25 nt, with the preadenylylated and 3'-biotin blocked RNA linker (DMSO samples) running more slowly than the preadenylylated and 3'-ddc blocked RNA linker (Fig. 2a). At Step 55, the cDNA Urea PAGE can visualize the RT and enrichment of NAI-N<sub>3</sub>-modified fragments (Fig. 2b). The cDNAs generated from the DMSO sample should closely match the distribution of the RNA fragments that were initially size-selected at Step 31; however, they will be characteristically longer owing to incorporation of the long RT primer that is capable of later intramolecular circularization. The cDNAs generated from the NAI-N<sub>3</sub>-modified sample will be both shorter and less abundant because of the NAI-N<sub>3</sub> causing structure-dependent stops along the RNA molecule and the 'single-hit' kinetics the cells were modified under, respectively. The final library native PAGE at Step 64 will confirm the qPCR amplification of specific icSHAPE products. It is important to remove any PCR primer-dimer products (Fig. 2c), as these will efficiently cluster on the deep-sequencing platforms and absorb valuable sequencing reads.

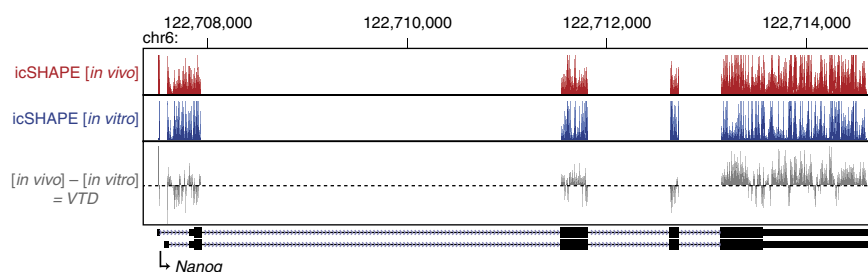
After deep sequencing, each icSHAPE read provides information about the reactive base (i.e., where the reverse transcriptase stops). Sequencing reads from each library are collapsed to remove PCR duplicates. In our protocol, we use a complex barcoding system that contains a fixed region to represent different libraries and two random regions (9 nt in total) to distinguish PCR duplicates from fragments with identical sequences. This is important to retain the quantitative nature of the experiment across a large number of PCR cycles between DMSO and NAI-N<sub>3</sub> experiments. As a consequence, usually there are many identical fragments that correspond to the same modification (i.e., reactive) site. These identical but meaningful fragments cannot be distinguished from duplicates that arise from subsequent PCR amplification steps without additional information. The random part of the barcodes is designed to solve this problem, because the chance for two random fragments to receive the same barcode is low (1/4<sup>9</sup>). The icSHAPE pipeline thus identifies duplicated PCR reads as those that are fully identical, including both the insert and the barcode regions.

Within the unique set of reads, RT stops generated from NAI-N<sub>3</sub>-induced truncations are identified as the first base in the sequencing read after the 13-nt barcode region. As reverse transcriptase characteristically stops one base downstream of the modified nucleotide, we computationally shift the RT stop to the (-) 1 position to correctly identify the flexible base. This position is the fragmentation position in the DMSO libraries. For each base, the RT stop count is then summed across the whole sequencing library, yielding an RT-stop coverage value. RT-stop coverage values of all bases across each transcript are then normalized with a 5–5% procedure, in which the means of the RT stops of top 90–95% of bases are normalized to 1, and all RT stops are scaled proportionally. It should be noted that because the experimental protocol usually introduces some degree of bias (difficulties in efficiently sequencing the 5' ends of RNAs and a 'blind 3'-end tail' because of low mapping of short reads), the current icSHAPE pipeline excludes the 32 bases at the 5' and 3' ends of the transcript in normalization. Reactivity scores for each base in a transcript are then calculated as the difference between background RT stops (DMSO samples) from icSHAPE RT stops (NAI-N<sub>3</sub> samples), and then adjusted by the background base density:

$$R = (\text{RT\_stop}_{\text{NAI}} - \alpha \text{RT\_stop}_{\text{DMSO}}) / \text{background\_base\_density}_{\text{DMSO}}$$

The reactivity score is then scaled to the range of [0, 1], after removing the outliers by 90% Winsorization (the top fifth percentile is set to 1 and the bottom fifth percentile is set to 0). This is performed for each transcript independently. The parameter  $\alpha$  is designed to remove the effect of natural fall-over of reverse transcriptase from a transcript in mock-treated libraries when calculating RT stops—e.g., RT stops caused by endogenous modifications in DMSO libraries. As the treated and mock-treated libraries are sequenced differently, this parameter does not equal to 1, and it was trained on mouse rRNA structures and set to 0.25 to maximize the correlation of reactivity score  $R$ , determined by deep sequencing and reactivity score measured in low-throughput gel-shift experiments<sup>23</sup>. Finally, the structural profile of each transcript is defined as the vector of reactivity score of every base from the beginning to the end, and the valid structural profile of a transcript is filtered by only selecting those regions that pass certain RT stop coverage threshold and background base density threshold.

**Figure 3** | icSHAPE reactivities of *Nanog* (*Nanog* homeobox) mRNA. A 7-kb region centered on the *Nanog* locus is shown with two isoforms of the *Nanog* mRNA in the mm10 build of the mouse genome. *In vivo* (red) and *in vitro* icSHAPE experimental data are represented as tracks with values ranging from 0 to 1. The *vivo-vitro* difference (VTD) score for each base is shown in gray.



Processed RT stops are visualized in the UCSC genome browser (**Fig. 3**). In the browser, each nucleotide that passes the filter cutoff and enrichment score threshold will have a value from 0 to 1, representing the icSHAPE flexibility of that base from either the *in vivo* (red) or *in vitro* (blue) experiment (**Fig. 3**, top and middle). To understand what regions of each RNA are remodeled inside cells, we devised a *vivo-vitro* difference metric that calculates the base-reactivity difference between the *in vitro*- and *in vivo*-modified samples<sup>20</sup>. This metric is represented as a score from -1 to 1 and termed the 'VTD' (**Fig. 3**, bottom). *In vivo* protein binding and RNA post-transcriptional modifications have been shown to change the VTD. It is possible that *in vivo* interactions such as RNA-RNA pairing could result in VTD changes. Thus, this calculation facilitates the discovery of important RNA regulatory domains.

**ACKNOWLEDGMENTS** We thank E. Kool and Kool laboratory members for synthesis of NAI-N<sub>3</sub>. This study was supported by National Institutes of Health grant nos. NIH R01-HG004361 and NIH P50-HG007735, and by the Howard Hughes Medical Institute (H.Y.C.).

**AUTHOR CONTRIBUTIONS** R.A.F., Q.C.Z., R.C.S. and H.Y.C. designed the experimental and computational strategy. R.A.F., R.C.S., B.L. and M.R.M. optimized experimental conditions. Q.C.Z. optimized computational parameters. R.A.F., Q.C.Z. and H.Y.C. wrote the manuscript with input from all authors.

**COMPETING FINANCIAL INTERESTS**

The authors declare competing financial interests: details are available in the [online version of the paper](#).

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Cooper, T.A., Wan, L. & Dreyfuss, G. RNA and disease. *Cell* **136**, 777–793 (2009).
2. Cech, T.R. Structural biology. The ribosome is a ribozyme. *Science* **289**, 878–879 (2000).
3. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).
4. Rana, T.M. Illuminating the silence: understanding the structure and function of small RNAs. *Nat. Rev. Mol. Cell Biol.* **8**, 23–36 (2007).
5. Weeks, K.M. Advances in RNA structure analysis by chemical probing. *Curr. Opin. Struct. Biol.* **20**, 295–304 (2010).
6. Goodman, D.B., Church, G.M. & Kosuri, S. Causes and effects of N-terminal codon bias in bacterial genes. *Science*. **342**, 475–479 (2013).
7. Wan, Y., Qu, K., Ouyang, Z. & Chang, H.Y. Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing. *Nat. Protoc.* **8**, 849–869 (2013).
8. Tijerina, P., Mohr, S. & Russell, R. DMS footprinting of structured RNAs and RNA-protein complexes. *Nat. Protoc.* **2**, 2608–2623 (2007).
9. Ziehler, W.A. & Engelke, D.R. Probing RNA structure with chemical reagents and enzymes. *Curr. Protoc. Nucleic Acid Chem.* Chapter 6, Unit 6.1 (2001).
10. Ingle, S., Azad, R.N., Jain, S.S. & Tullius, T.D. Chemical probing of RNA with the hydroxyl radical at single-atom resolution. *Nucleic Acids Res.* **42**, 12758–12767 (2014).
11. Merino, E.J., Wilkinson, K.A., Coughlan, J.L. & Weeks, K.M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **127**, 4223–4231 (2005).
12. Risca, V.I. & Greenleaf, W.J. Beyond the linear genome: paired-end sequencing as a biophysical tool. *Trends in Cell Biol.* **25**, 716 (2015).
13. Kertesz, M. *et al.* Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467**, 103–107 (2010).

14. Underwood, J.G. *et al.* FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* **7**, 995–1001 (2010).
15. Li, F. *et al.* Global analysis of RNA secondary structure in two metazoans. *Cell Rep.* **1**, 69–82 (2012).
16. Lucks, J.B. *et al.* Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-seq). *Proc. Natl. Acad. Sci. USA* **108**, 11063–11068 (2011).
17. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J.S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures *in vivo*. *Nature* **505**, 701–705 (2014).
18. Ding, Y. *et al.* *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**, 696–700 (2014).
19. Talkish, J., May, G., Lin, Y., Woolford, J.L. & McManus, C.J. Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *RNA* **20**, 713–720 (2014).
20. Spitale, R.C. *et al.* Structural imprints *in vivo* decode RNA regulatory mechanisms. *Nature* **519**, 486–490 (2015).
21. Kelley, D.R., Hendrickson, D.G. & Tenen, D. Transposable elements modulate human RNA abundance and splicing via specific RNA-protein interactions. *Genome* **15**, 537 (2014).
22. König, J. *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* **17**, 909–915 (2010).
23. Spitale, R.C. *et al.* RNA SHAPE analysis in living cells. *Nat. Chem. Biol.* **9**, 18–20 (2012).
24. Poulsen, L.D., Kielbinski, L.J., Salama, S.R., Krogh, A. & Vinther, J. SHAPE selection (SHAPES) enrich for RNA structure signal in SHAPE sequencing-based probing data. *RNA* **21**, 1042–1052 (2015).
25. Reuter, J.S. & Mathews, D.H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 129 (2010).
26. Ouyang, Z., Snyder, M.P. & Chang, H.Y. SeqFold: genome-scale reconstruction of RNA secondary structure integrating high-throughput sequencing data. *Genome Res.* **23**, 377–387 (2013).
27. Eddy, S.R. Computational analysis of conserved RNA secondary structure in transcriptomes and genomes. *Annu. Rev. Biophys.* **43**, 433–456 (2014).
28. Sloma, M.F. & Mathews, D.H. Chapter four-improving RNA secondary structure prediction with structure mapping data. *Methods Enzymol.* **553**, 91–114 (2015).
29. Bai, Y., Dai, X., Harrison, A., Johnston, C. & Chen, M. Toward a next-generation atlas of RNA secondary structure. *Brief. Bioinform.* <http://dx.doi.org/10.1093/bib/bbv026> (2015).
30. Liu, N. *et al.* N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518**, 560–564 (2015).
31. Flynn, R.A. *et al.* Dissecting noncoding and pathogen RNA-protein interactomes. *RNA* **21**, 135–43 (2015).
32. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* **5**, 621–628 (2008).

