In-cell RNA structure probing with SHAPE-MaP

Matthew J Smola & Kevin M Weeks

Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. Correspondence should be addressed to K.M.W. (weeks@unc.edu).

Published online 3 May 2018; doi:10.1038/nprot.2018.010

RNAs play key roles in many cellular processes. The underlying structure of RNA is an important determinant of how transcripts function, are processed, and interact with RNA-binding proteins and ligands. RNA structure analysis by selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) takes advantage of the reactivity of small electrophilic chemical probes that react with the 2′-hydroxyl group to assess RNA structure at nucleotide resolution. When coupled with mutational profiling (MaP), in which modified nucleotides are detected as internal miscodings during reverse transcription and then read out by massively parallel sequencing, SHAPE yields quantitative per-nucleotide measurements of RNA structure. Here, we provide an extension to our previous *in vitro* SHAPE-MaP protocol with detailed guidance for undertaking and analyzing SHAPE-MaP probing experiments in live cells. The MaP strategy works for both abundant-transcriptome experiments and for cellular RNAs of low to moderate abundance, which are not well examined by whole-transcriptome methods. In-cell SHAPE-MaP, performed in roughly 3 d, can be applied in cell types ranging from bacteria to cultured mammalian cells and is compatible with a variety of structure-probing reagents. We detail several strategies by which in-cell SHAPE-MaP can inform new biological hypotheses and emphasize downstream analyses that reveal sequence or structure motifs important for RNA interactions in cells.

INTRODUCTION

RNA is a critical regulator of cellular processes, operating through diverse mechanisms to modulate gene expression in all forms of life[1]. RNA can act in *cis* or in *trans* and may function alone or as part of ribonucleoprotein (RNP) complexes. RNA regulates alternative splicing[2], small RNA-mediated silencing[3], and metabolite sensing (through riboswitches)[4]; is catalytic (ribozymes)[5]; and has diverse actions in the form of long noncoding RNA (lncRNA)[6–8]. With recent advances in high-throughput biology, understanding of the capacity of RNA to influence cellular activities is expanding rapidly.

The ability of RNA molecules to form complex secondary and tertiary structures underlies many of its cellular functions[9–12]. These RNA structures are usually difficult to accurately predict from sequence information alone, especially for long transcripts. Adding an additional layer of complexity is the fact that most, if not all, RNAs interact with cellular partners either transiently or in stable RNP complexes[13]. Identifying the location and nature of these RNA–protein interactions *de novo* is difficult. Several approaches for studying RNA structure *in vitro* and in living cells have been described[14–22].

We previously shared an in-depth protocol for probing RNA structure *in vitro* using selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP)[23]. In this protocol extension, we highlight the utility of SHAPE-MaP as an approach for in-cell probing of both RNA structure and intermolecular interactions between RNA and other molecules in a native context.

Utility of in-cell RNA structure probing

In-cell SHAPE-MaP yields quantitative data describing local RNA flexibility at nucleotide resolution. In the native cellular environment, nucleotide reactivity to the chemical probe is influenced both by RNA structure and by interactions with proteins and other molecules. In-cell experiments can reveal complex sets of interactions and are particularly useful when comparing different experimental states. For example, our laboratory used in-cell SHAPE to analyze the conformations of the RNA in the bacterial 30S ribosome subunit in various stages of translation, revealing distinct assembly states[21] and a novel regulatory RNA conformational change[19]. In-cell SHAPE data can also be paired with SHAPE reactivities derived from cell-free probing, in which the RNA is gently extracted from cells and deproteinized before probing. By rigorously analyzing differences between in-cell and cell-free data, sites of RNA–protein interactions within RNP complexes can be identified with high confidence and with relatively high resolution[18]. In-cell SHAPE-MaP can be applied in a targeted, gene-specific way. Thus, SHAPE-MaP makes it possible to obtain highly quantitative per-nucleotide structure information about both abundant RNAs, such as the cytoplasmic 5S rRNA, signal-recognition particle RNA, and the U1 small nuclear RNA (snRNA)[18]; and rare nuclear transcripts such as the *Xist* lncRNA[24].

In a SHAPE-MaP experiment, RNA molecules are treated with a hydroxyl-selective electrophile that reacts with the 2′-hydroxyl position via a mechanism that primarily reports local nucleotide flexibility[25]. During the MaP readout stage, 2′-*O*-adducts are encoded as apparent sequence changes or deletions in the cDNAs produced during reverse transcription. MaP readout is compatible with every structure-probing reagent we have tested, including 1-methyl-7-nitroisatoic anyhydride (1M7), 1-methyl-6-nitroisatoic anhydride (1M6) and N-methyl-isatoic anhydride (NMIA) (refs. 17,24), 2-methylnicotinic acid imidazolide (NAI) (refs. 14,26), and dimethyl sulfate (DMS)[27,28]. Other sequencing-based RNA structure-probing experiments rely on faithfully preserving and identifying cDNA ends corresponding to adduct-induced reverse transcriptase termination or RNase cleavage sites[15,26,29–33]. The advantage of MaP over these approaches is that MaP eliminates the need to perform multiple ligation steps, which markedly simplifies readout of the chemical probing information and appears to substantially increase the accuracy[18] and sensitivity[34] of information recovery. The MaP step can be performed with random

primers to recover information from complex viral or cellular systems. In addition, MaP can be performed with gene- or region-specific primers[23]. Use of gene-specific primers makes it possible to target transcripts of low to moderate abundance or specific regions of large RNAs, which are otherwise difficult to examine comprehensively by methods that use random priming.

Gene-specific SHAPE-MaP has enabled analysis of transcripts such as the 18-kb *Xist* lncRNA[24], which are too rare to be comprehensively detected in whole-transcriptome experiments. *Xist* adopts complex structures and interacts with many different proteins through diverse mechanisms, and the ability to compare the in-cell and cell-free structures of such transcripts proved highly informative. For example, we analyzed differences between cell-free and in-cell SHAPE reactivities of *Xist* and identified hundreds of potential RNA–protein interaction sites. By considering these sites in conjunction with SHAPE-directed secondary structure models, we identified conformational changes induced by the cellular environment, as well as sequence- and structure-selective RNA–protein interactions. A key finding of *Xist* probing in cells was that many RNA–protein interactions are governed by the underlying RNA structure, or the lack thereof.

Comparisons with cell-free probing

The experimental approach for in-cell SHAPE-MaP is similar to that used for *in vitro* experiments (**Fig. 1**). There are, however, unique considerations when planning an in-cell probing experiment: cells and culture media must be compatible with SHAPE probing conditions, and an effective strategy for enrichment of target RNAs should be considered. These aspects are further described in the Experimental design section below.

Although experimentally similar, cell-free and in-cell SHAPE probing methods are distinct in terms of data interpretation and analysis. SHAPE fundamentally measures local nucleotide flexibility. In a cell-free experiment, in which the RNA is probed in buffer in the absence of other biomolecules, the reactivity of a given nucleotide is strongly and specifically correlated with its local structure[35,36]. For this reason, SHAPE reactivities can be used to develop highly accurate structural models for many RNAs[37,38]. By contrast, the SHAPE reactivities of nucleotides probed in living cells are dependent not only on local structure, but also on interactions with other biomolecules. Thus, in-cell probing may yield reactivities that are reduced or enhanced relative to the cell-free state. For example, nucleotides in the apical loop of an RNA hairpin may exhibit high SHAPE reactivity in the cell-free state but, in cells, may interact with RNA-binding proteins in a way that decreases local SHAPE reactivity. If the in-cell data are used to guide secondary structure modeling, it is likely that the RNA structure model will be incorrect. In the case of large RNAs, local prediction errors can propagate through structure prediction calculations and cause widespread inaccuracies in the resulting model. Therefore, we caution against using SHAPE data derived from in-cell probing as the sole source of biochemical data for RNA structure modeling.

Although not ideal for secondary structure prediction, in-cell SHAPE experiments provide abundant useful data regarding the effects of the cellular environment on RNA structure. Because the cellular environment can have confounding influences on SHAPE reactivity that are not apparent in single experiments, in-cell data are usually best analyzed in a comparative manner (**Fig. 2**).

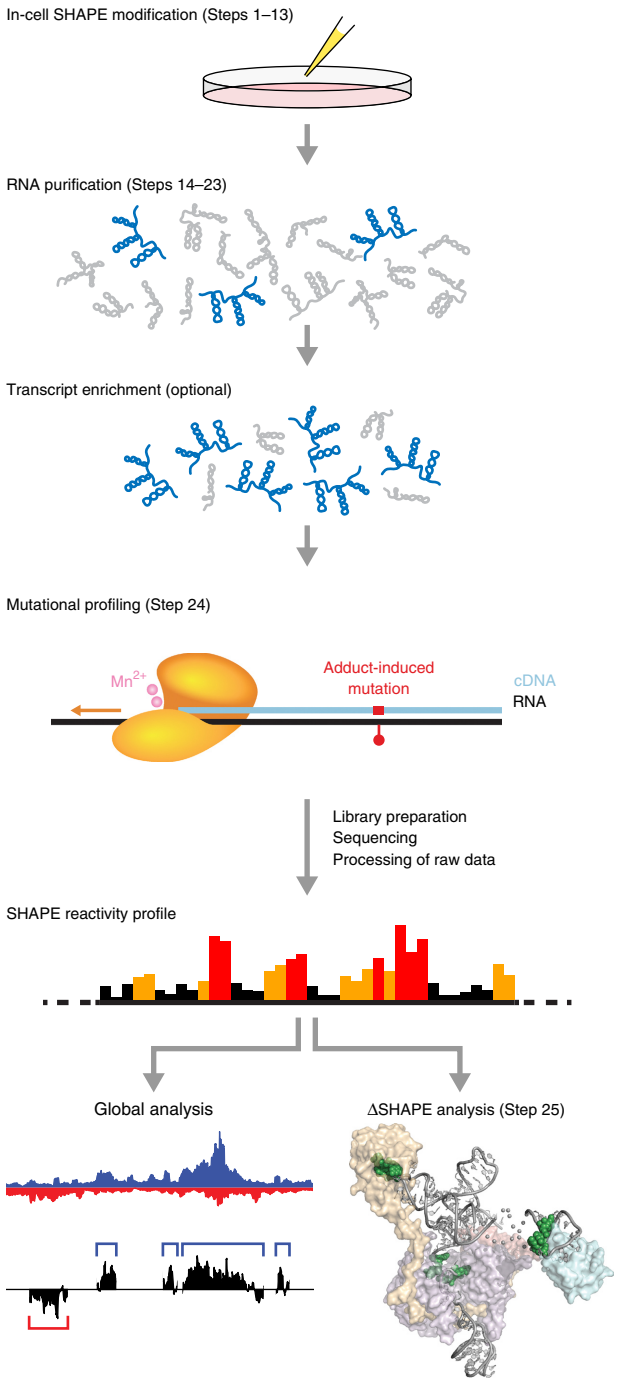

Figure 1 | Overview of in-cell SHAPE-MaP. (Top) Live cells are treated with a SHAPE reagent that reacts preferentially with flexible nucleotides. Then modified RNA is extracted from cells and purified, followed by optional enrichment of transcripts of interest by one (or more) methods. Then mutational profiling (MaP) conditions are used to reverse-transcribe the RNA into cDNA such that SHAPE adducts are encoded as noncomplementary sequences or deletions in the synthesized cDNA strand. Massively parallel sequencing libraries are then prepared from the cDNA and sequenced, followed by processing of raw data with ShapeMapper[17,48] to generate SHAPE reactivity profiles. Finally, in-cell reactivity profiles are analyzed to identify the effects of the cellular environment on RNA structure. Parts of the figure adapted with permission from ref. 23, Springer Nature; ref. 24, National Academy of Sciences; and ref. 18, American Chemical Society.

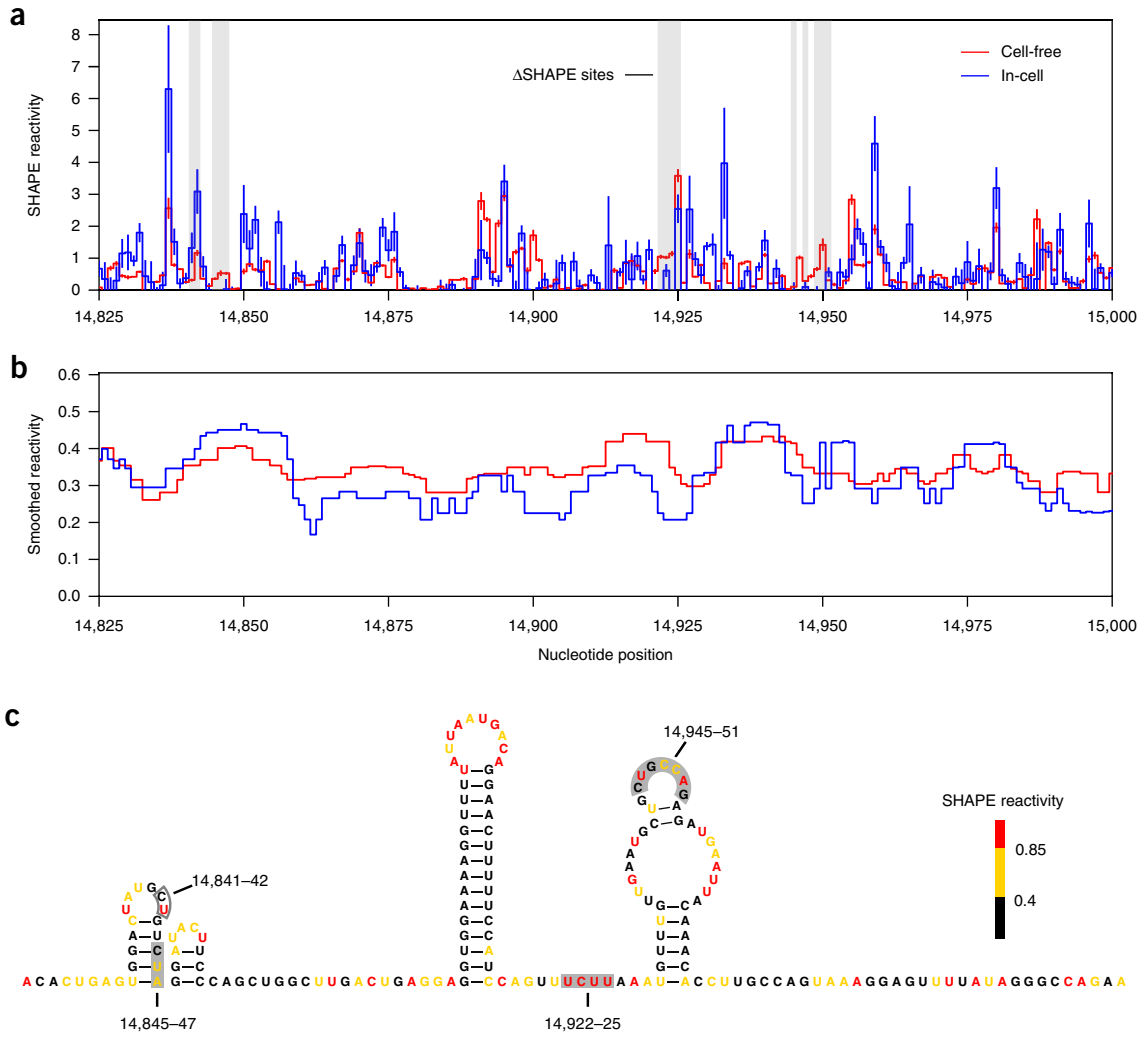

Figure 2 | Visualization and comparison of in-cell and cell-free SHAPE reactivities. (**a**) In-cell (blue) and cell-free (red) SHAPE reactivities for a portion of the *Xist* lncRNA are shown. Estimated standard errors for each nucleotide are shown as vertical bars. In-cell reactivity shows both protections and enhancements relative to the cell-free state. ΔSHAPE sites (locations of strong SHAPE reactivity change between samples) are shown in gray. This experiment was performed using ~1–6 million mouse embryonic stem cells per condition. (**b**) Windowed median reactivities, calculated over 50-nt sliding windows. (**c**) Superimposition of ΔSHAPE sites on the secondary structure model for a region of the *Xist* RNA. The region shown corresponds to a portion of the FUS interaction region[24]. Positive ΔSHAPE sites (corresponding to protection in-cell) are shaded gray, and one negative ΔSHAPE site is boxed.

High-value analyses include, for example, comparison of in-cell and cell-free states, comparison of cellular knockouts of individual RNA-binding proteins to a wild-type reference, comparison of wild-type cells versus a cell line deficient in a protein of interest, comparison of wild-type cells versus cells harboring an instructional mutation in the RNA of interest, and analysis of the consequences of adding a bioactive small molecule[18–21,24,39–41]. In-cell SHAPE probing can be used to study the effects of essentially any cellular perturbation that impacts RNA structure and dynamics relative to a reference state.

Limitations

Fundamentally, SHAPE-MaP measures local nucleotide flexibility and dynamics. Thus, any RNA conformational change or ligand-binding event that affects local nucleotide dynamics will be recorded as a change in SHAPE reactivity. The major limitation is that, in a cellular setting, it can be difficult to distinguish between intramolecular RNA structural constraints and intermolecular protein–RNA or other ligand–RNA interactions as the cause of specific reactivities or reactivity changes at a given nucleotide.

SHAPE also requires that the 2′-hydroxyl probing reagent be able to enter target cells and react broadly with its constituent RNAs. We have consistently found that 1M7, NMIA, and closely related reagents enter most cells and yield reproducible reactivities above background. In some cases, researchers might wish to explore other reagents. However, work from our group[19,21,24] and from independent groups[39,40,42] does not support the claim[14,43] that 1M7 does not react with RNAs in bacterial cells and cultured mammalian cells. We note that NAI is a useful SHAPE reagent as long as reactivity is specifically quenched (e.g., using 2-mercaptoethanol) after a reasonable in-cell time frame of 2–10 min. Some cell types are less permeable to SHAPE reagents, and so cell permeability should be specifically tested early in any in-cell SHAPE probing experiment.

SHAPE-MaP is generally far less sensitive to the vagaries of sequencing library construction than methods that require adapter ligation to read out structural information. Nevertheless, MaP relies on the ability, first, to construct sequencing libraries and, second, to align these with a reference sequence in order to detect sites of chemical adduct formation. For difficult-to-amplify sequences, such as tandem repeats or GC-rich regions, extensive optimization may be required both in library construction and computational sequence alignment. For regions with highly repetitive sequences, which are common in some lncRNAs, data readout can be severely limited by the inability to uniquely align sequencing reads, although this limitation affects all chemical probing approaches read out by massively parallel sequencing.

Applications

The ability to interrogate the structural state of RNAs in living cells has numerous applications. Broadly, in-cell SHAPE-MaP allows researchers to study the effects of biological phenomena on RNA structure. Examples include assessment of the impacts of knocking down or overexpressing an RNA-binding protein, monitoring of viral transcript structure over the course of an infection, or identification of structural changes associated with small-molecule ligands. Regions with dense segments of differences between in-cell and cell-free states are good candidates for likely centers of functional elements in large RNAs. SHAPE-MaP allows such studies to be conducted on a broad, transcriptome-wide scale or on a focused set of transcripts through the use of appropriate region- or gene-specific primers. Comparative experiments between distinct conditions are highly instructive and can be especially diagnostic of RNA–protein interactions and RNA assembly states.

Experimental design

Cell choice and culture conditions. Cells used for SHAPE-MaP experiments must tolerate brief exposure to SHAPE reagents in DMSO without exhibiting a marked response that substantially changes the expected cellular state. We and others have found that most bacteria, yeasts, and cultured mammalian cell lines are tolerant of brief probing[14,18,19,24,26,39,44]. Our lab has obtained good results with the following cell types: *Escherichia coli* (strain K12 and others), HeLa, Jurkat, B lymphoblasts, mouse embryonic stem cells, and mouse myoblast cells; we have found HEK293 cells to be challenging for in-cell probing. SHAPE chemistry is dependent on pH[35] and performs best within a pH range of 7.4–8.3. Most mammalian cell lines have an intracellular pH compatible with SHAPE probing; however, certain cell types are often grown at lower pH (e.g., Sf9 cells are typically cultured at pH 6.2). In some cases, it may be necessary to replace the growth media with pH-adjusted media immediately before SHAPE probing. When planning a SHAPE-MaP experiment with a new cell line, we recommend assessing both the culture conditions and the cellular response to treatment with SHAPE reagent before undertaking a complete experiment.

SHAPE probing (Steps 1–13). SHAPE experiments are inherently ensemble measurements of RNA structure. It is critical that the SHAPE reagent be rapidly and uniformly mixed with the sample to ensure homogeneous probing. It is also advantageous to probe cells over a relatively short time period. This protocol will focus on experiments performed using 1M7, but it has been tested and works well for all commonly used SHAPE reagents and with DMS. Cells that grow in monodisperse suspension or as an adherent monolayer are well-suited for in-cell SHAPE-MaP, as the entire culture can be quickly and evenly exposed to the SHAPE reagent. When probing either liquid bacterial cultures or eukaryotic cells grown in suspension cultures, thorough mixing can be achieved by adding suspended cells in a large volume directly to a smaller volume of SHAPE reagent, ensuring rapid mixing. Adherent cells can be uniformly probed by thoroughly swirling the culture vessel immediately after addition of the SHAPE reagent. Cells that grow in large clusters or clumps (e.g., embryoid bodies or bacteria in certain stationary stages of growth) are likely to exhibit an apparent lower level of modification, as diffusion of the reagent through the cell mass competes with the rate of inactivating hydrolysis. In such cases, cell clusters can be trypsinized before SHAPE probing, and it is recommended that controls be implemented to examine whether such treatment alters the expected internal cellular state.

RNA isolation and enrichment (Steps 14–23). Acylation of RNA by SHAPE reagents results in covalent 2′-*O*-adducts that are stable in most biochemical buffers. Thus, after SHAPE probing is complete, RNA can be isolated from cells using a variety of methods. Typically, high-purity isolation is accomplished by extraction with phenol:chloroform or similar commercial reagents. The MaP approach is highly sensitive to low levels of contaminating genomic DNA, and it is critical that the purified RNA be free of DNA. DNase digestion should be performed after RNA extraction.

rRNA can account for >90% of cellular RNA; RNAs of interest thus often represent only a very small percentage of the total RNA. Therefore, it is usually critical that the RNA of interest be isolated or enriched from the cellular RNA pool. There are many strategies to accomplish this, and we outline a few here. First, a common approach is to selectively remove or degrade rRNA from the sample. Several commercial kits are available for this purpose. An alternative is to enrich for poly-adenylated transcripts using biotinylated oligo-dT probes. Use of both of these approaches may be essential for transcriptome-scale probing experiments. Another strategy is to selectively isolate the RNA of interest with biotinylated antisense oligonucleotides. The efficiency of this method depends on the size, abundance, and sequence of the transcript of interest, and enrichment probes must be designed carefully to ensure sufficient specificity. Methods for designing probes and carrying out affinity RNA purification have been described[45,46] and will not be covered in detail here.

An alternative approach to physical transcript enrichment—unique to the MaP strategy—is the use of gene- or region-specific primers during the mutational profiling reverse transcription process. Because MaP experiments do not depend on preservation of cDNA ends for signal detection, an RNA of interest can be enriched by amplification from total RNA using target-specific reverse transcription primers followed by specific PCR amplification during library preparation. Thus, SHAPE-MaP experiments targeting low-abundance RNAs or specific subregions of large transcripts can be performed without optimized enrichment protocols. Similarly, if 3′-UTRs are of particular interest, oligo-dT reverse transcription primers may be used without prior enrichment. Using these enrichment-by-amplification approaches, sequencing resources

can be focused specifically on the transcripts of interest. Detailed advice pertaining to targeted amplification can be found in the original protocol (see the amplicon workflow)[23].

Library construction, sequencing, and data processing (Step 24). After completion of RNA enrichment steps, in-cell SHAPE-MaP experiments proceed through MaP reverse transcription, library construction, sequencing, and raw-data-processing steps in exactly the same way as cell-free SHAPE-MaP experiments. For in-cell experiments with a very small amount of starting material, it may be advantageous to reduce the number of PCR cycles during library preparation to minimize PCR duplication events and other amplification artifacts. Alternatively, a molecular barcoding approach may be used to detect and filter duplicate sequences[47]. Further experimental considerations and a step-by-step guide regarding this portion of the experiment are described in detail in the original protocol[23]. In our experience, if the total number of PCR cycles is kept below roughly 15, PCR artifacts do not appear to be a serious problem. Importantly, gene- and target-specific SHAPE-MaP is able to probe RNA structure at orders-of-magnitude lower cellular concentrations than random priming or whole-transcriptome methods.

The sites of chemical modification-induced mutations or sequence changes are identified using the ShapeMapper software, which has recently been updated[48]. ShapeMapper produces a number of useful quality-control metrics in addition to SHAPE reactivity profiles. For cell-free experiments, the distribution of mutation rates in the SHAPE-modified sample should be distinctly higher than that in the DMSO control. In-cell experiments generally have a lower signal-to-noise ratio than cell-free experiments, and it is important to ensure that the plus-reagent experiment has a higher modification rate than the no-reagent control. Because of the high sensitivity of the MaP readout strategy[17,34], small reactivity differences above background are more easily detected than with reverse transcription truncation-and-ligation approaches, and high-quality analysis of in-cell structure probing is often readily achieved (**Fig. 2**). For experiments comparing in-cell and cell-free samples, we recommend including a well-characterized control RNA, such as the U1 snRNA, RNase MRP, or 5S RNA, in each condition. These RNAs are small, abundant, and form stable RNP complexes that can be compared with published in-cell profiles[18]. Alternatively, a short amplicon targeting rRNA can be used to compare in-cell reactivity with published crystal structures to ensure the in-cell experiment is performing well.

Data analysis and interpretation (Step 25). The utility of in-cell SHAPE-MaP data is greatest when used for comparative analysis between experimental conditions, and the majority of work to date has focused on identifying effects of the cellular environment relative to a cell-free-extracted RNA state. However, in-cell SHAPE-MaP can be applied to any two (or more) experimental states, including the following: inhibition of transcription or translation, addition of an RNA-binding small molecule, knockdown or deletion of an RNA-binding protein, and mutation of an RNA. We have created analysis strategies that identify both large- and small-scale changes in SHAPE reactivity between experiments. The results of these analyses can be interpreted both individually and collectively.

In-cell experiments can have diverse goals. When the structure and overall architecture of an RNA are poorly understood, the goal is often to identify novel areas of interest along a transcript by examining where cellular influences alter the structural profile of the RNA of interest. In many cases, especially when evaluating large RNAs, it can be useful to hone in on large-scale regions or domains of a transcript that are most different between conditions. One approach to identifying these domains is to implement a global difference analysis (**Fig. 3**). First, the absolute change in SHAPE reactivity is calculated at each nucleotide along the transcript of interest. These values are then smoothed by calculating the 50-nt sliding median (**Fig. 3a**). Regions that undergo the greatest change in SHAPE reactivity are thus identified by high-difference values. The extent of these regions is determined in part by filtering; as a starting point, we suggest seeking out regions in which the smoothed difference values are greater than the global median for at least 100 consecutive nucleotides (**Fig. 3a**, purple shading). These cutoffs can be adjusted empirically.

The large-scale impacts of the cellular environment can be further evaluated on the basis of positive or negative reactivity changes (**Fig. 3b**,**c**). For example, if in-cell reactivities are subtracted from cell-free reactivities, a region exhibiting an overall positive difference (corresponding to reduced reactivity in cells) may interact with proteins or other ligands, or may experience a local conformational change, in a way that reduces per-nucleotide reactivity with the SHAPE reagent. Similarly, we have observed that overall negative differences (corresponding to enhanced reactivity in cells) tend to indicate RNA structural rearrangements or other cellular effects. These trends may obviously vary depending on the RNA or in-cell state of interest. Regions of large-scale SHAPE reactivity change can be further analyzed computationally and experimentally. It may be possible to identify sequence motifs present in distinct regions or, if a high-quality structural model is available, analysis of the underlying features may reveal critical structural motifs. In particular, it is often useful to focus on RNA regions with high levels of well-determined structure, as determined empirically by low SHAPE/low Shannon entropy regions (as discussed in prior work[17,24] and in the original protocol[23]).

In addition to large-scale analyses of RNA regions, the single-nucleotide resolution of SHAPE probing can be leveraged to identify local changes in SHAPE reactivity with high confidence. The ShapeMapper data-processing software automatically computes an estimated standard error for the reactivity value at each nucleotide[17,48]. We developed an analysis strategy, termed ΔSHAPE, which uses these error estimates to identify RNA–protein interactions with high confidence and relatively high resolution (**Fig. 3d**)[18]. This approach is based on the idea that functionally important changes in RNA structure or interactions (e.g., conformational changes or protein binding) will impact SHAPE reactivity, but, due to the large number of comparisons and inherent measurement errors, not all differences are meaningful. The ΔSHAPE framework considers both the raw change in SHAPE reactivity at a given nucleotide and also the magnitudes of the estimated errors in both measurements and the distribution of reactivity changes across the transcript of interest (**Fig. 3d**). By identifying significant differences with the largest magnitudes, and with the requirement that differences be clustered, ΔSHAPE analysis produces a list of sites ~3–7 nt long that are likely to reflect interaction sites[18]. These sites can inform further analysis

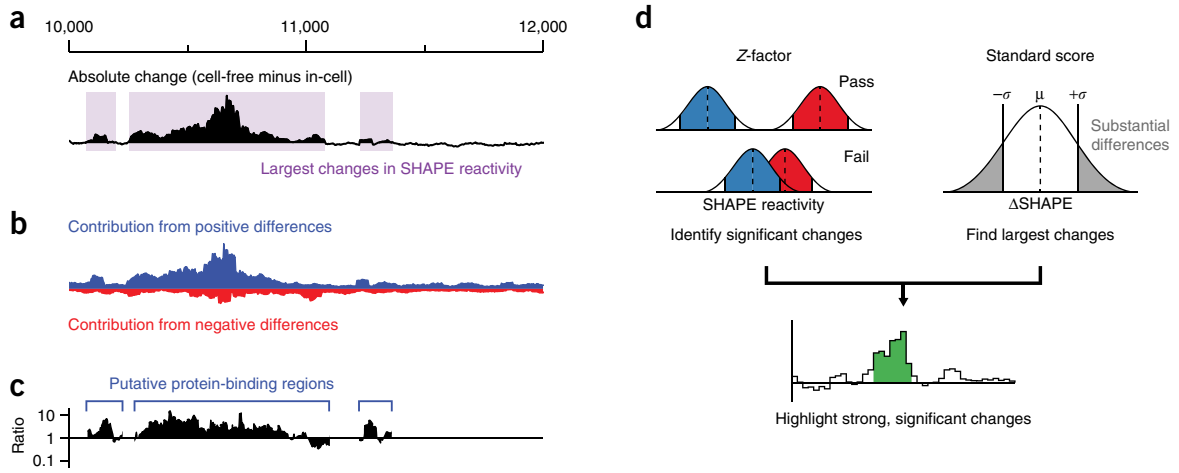

Figure 3 | In-cell SHAPE reactivity analyses. (**a**) Global difference analysis for a portion of the mouse *Xist* lncRNA. The absolute change in SHAPE reactivity between cell-free and in-cell states is shown, smoothed over 50-nt windows. Areas where the smoothed absolute change exceeds the global median for at least 100 nt are highlighted with purple shading. (**b**) Contributions of positive and negative differences (blue and red, respectively) to the absolute change highlight that much of the change in reactivity in this region of *Xist* reflects a general decrease in SHAPE reactivity in cells. (**c**) Ratio of positive to negative contributions calculated for regions highlighted in **a**, allowing for efficient categorization of regions as exhibiting mostly positive changes, mostly negative changes, or a combination of both. (**d**) The statistical filters for ΔSHAPE analysis. A *Z*-factor test (left) is implemented to identify the differences for which the estimated errors for each measurement (red and blue shading) do not overlap. A standard score (right) is calculated to identify nucleotides that undergo the largest absolute changes in SHAPE reactivity. Heuristically, we require that at least 3 nt within a 5-nt window pass both the *Z*-factor and standard score filters to be highlighted as a local region of significant SHAPE reactivity change (bottom). **a**–**c** adapted with permission from ref. 24, National Academy of Sciences; **d** adapted with permission from ref. 18, American Chemical Society.

in numerous useful ways. For example, ΔSHAPE was used to robustly identify protein-binding sites in several RNP complexes[18]. In addition, ΔSHAPE complements RNA immunoprecipitation (RIP) or cross-linking and immunoprecipitation (CLIP) strategies, and can be used to stringently filter RIP or CLIP data sets for high-confidence binding sites[24].

Box 1 | Creation of a cell-free SHAPE reference sample ● TIMING 6 h

We describe an example procedure for gently extracting natively folded nuclear RNA for the creation of a cell-free SHAPE reference sample. These conditions are generally applicable for ~10^6 adherent mammalian cells. The procedure may need to be adjusted depending on the subcellular localization of the RNA of interest and the type of cells being probed.

1. In a 15-ml conical tube, pellet 10^6 cells in ice-cold PBS at 500*g* for 5 min at 4 °C.
2. Resuspend the cells in 2.5 ml of lysis buffer and rotate at 4 °C for 5 min.
3. Pellet the cell nuclei at 2,250*g* for 2 min at 4 °C.
4. Resuspend the nuclei in 2.5 ml of proteinase K buffer and rotate at 20 °C for 45 min.
5. Add 2.5 ml of pre-equilibrated phenol/chloroform/isoamyl alcohol and mix until an emulsion forms (~20–30 s).

▲ **CRITICAL** TRIzol and similar reagents should be avoided during native purification of RNA, as they contain strong chemical denaturants that will disrupt RNA structures.

6. Centrifuge at 12,000*g* for 15 min at 4 °C, carefully collect the aqueous phase, and place in a fresh 15-ml conical tube.
7. Repeat steps 5 and 6 for a total of two extractions with phenol/chloroform/isoamyl alcohol.
8. Add 2.5 ml of chloroform to the aqueous phase and mix for 20–30 s until an emulsion forms.
9. Centrifuge as in step 6, carefully collect the aqueous phase, and place it in a fresh 15-ml conical tube.
10. Repeat steps 8 and 9 for a total of two extractions with chloroform.
11. Buffer-exchange the RNA into 1.1× folding buffer using a PD-10 column according to the manufacturer's instructions.
12. Incubate the buffer-exchanged RNA at 37 °C for 20–30 min to allow RNA structures to equilibrate.
13. Prepare two fresh 15-ml conical tubes. Label one as '+' and the other as '–'. Calculate 1/18 of the RNA volume and add this volume of 100 mM 1M7 to the '+' tube and add the same volume of neat DMSO to the '–' tube.
14. To each tube from step 13, add half of the equilibrated RNA from step 12. Quickly cap the tubes, mix thoroughly by inversion, and incubate at 37 °C for 5 min.
15. Add an equivalent volume of 100% isopropanol to each tube. Mix well and incubate at –20 °C for at least 30 min.
16. Centrifuge at 12,000*g* for 15 min at 4 °C. Remove the supernatant. A small RNA pellet should be visible on the side of the tube.
17. Wash the pellet with 75% ethanol, and then centrifuge at 12,000*g* for 5 min at 4 °C.

■ **PAUSE POINT** The RNA pellet can be stored in ethanol for up to a year at –20 °C.

18. Dry the pellets in air or under vacuum, and then resuspend the RNA in 88 µl of RNase-free water.

As ΔSHAPE requires a change in reactivity between experimental conditions, cellular interactions at sites that are strongly paired will likely be difficult to detect. In addition, the ΔSHAPE statistical thresholds were derived from studying a small set of well-understood RNP complexes[18] and may need to be adjusted to be less stringent for less-stable RNA–protein complexes. RNA–protein interactions may cause structural changes in a region of the RNA not directly involved in protein binding. Finally, other impacts of the cellular environment might induce punctate changes in SHAPE reactivity that are scored highly by ΔSHAPE but reflect events other than stable protein binding, including RNA conformational changes. Despite this complexity, we have found the ΔSHAPE approach to be a powerful tool for detecting and characterizing recognition sites for diverse RNA-binding proteins.

The ΔSHAPE analysis pipeline is available as a Python program that provides both graphical and plain-text output. It is preconfigured to use the default ΔSHAPE statistical cutoffs but allows the advanced user to adjust most parameters with command-line flags (**Box 1**).

MATERIALS

REAGENTS

▲ **CRITICAL** A complete list of reagents is given below for the in-cell and cell-free SHAPE-MaP probing and for the purification steps outlined in this protocol extension. For additional reagents required for library preparation and sequencing, please refer to the original protocol[23].

- Total cellular RNA at a concentration between 250 and 500 ng/μl in sterile water ▲ **CRITICAL** Total cellular RNA should be purified from the same cell line used for SHAPE probing and should be DNase-treated to prevent genomic DNA contamination and carryover.
- Cells (e.g., HeLa cells (ATCC, cat. no. CCL-2)) ! **CAUTION** The cell lines used in your research should be regularly checked to ensure that they are authentic and are not infected with mycoplasma.
- Cell culture medium ▲ **CRITICAL** The culture medium may vary by cell type. Ensure that the medium being used is optimal for SHAPE chemistry (pH 7.4–8.3).
- SHAPE reagent of choice. ▲ **CRITICAL** This protocol extension will focus on examples using 1M7 (synthesis is described in refs. 49–51), but other reagents, including NMIA, 1M6, and 5NIA (5-nitroisatoic anhydride), can also be used. NAI[14] can also be used, but NAI requires a specific quenching step (such as with 2-mercaptoethanol) to achieve accurate RNA structure probing in the optimal 2- to 10-min reaction time frame. ▲ **CRITICAL** SHAPE reagents should be stored in a desiccator at 4 °C. When properly stored, these reagents are stable for at least a year. Probing by DMS can also be analyzed with this protocol.
- HEPES (Fisher BioReagents, cat. no. BP310–500)
- Tris (Fisher BioReagents, cat. no. BP152)
- Sodium chloride (NaCl; Fisher BioReagents, cat. no. BP358)
- Magnesium chloride ($MgCl_2$, 1 M; Thermo Fisher Scientific, cat. no. AM9670G)
- Calcium chloride ($CaCl_2$, 2 M; Fisher BioReagents, cat. no. BP9742)
- Sucrose (Fisher BioReagents, cat. no. BP220)
- Triton X-100 (Sigma-Aldrich, cat. no. T8787)
- RNase inhibitor (Promega, cat. no. N2115)
- DNase I (Roche, cat. no. 04536282001)
- Proteinase K (20 mg/ml; Thermo Fisher Scientific, cat. no. AM2546)
- EDTA (pH 8.0, 0.5 M; Thermo Fisher Scientific, cat. no. AM9260G)
- PBS (pH 7.4; Thermo Fisher Scientific, cat. no. 10010023)
- DMSO (anhydrous; Sigma-Aldrich, cat. no. 276855) ! **CAUTION** DMSO readily passes through skin and latex gloves, and it can facilitate bodily absorption of dissolved substances. Avoid direct contact.
- Phenol/chloroform/isoamyl alcohol (25:24:1; Thermo Fisher Scientific, cat. no. AM9730)
- TRIzol reagent (Thermo Fisher Scientific, cat. no. 15596018)
- Chloroform (Fisher BioReagents, cat. no. BP1145-1)
- Formamide (highly deionized; Thermo Fisher Scientific, cat. no. 4311320)
- Absolute ethanol (Fisher BioReagents, cat. no. BP2818-500)
- Turbo DNase reaction buffer (10×; Thermo Fisher Scientific, cat. no. AM2238)
- Turbo DNase (2 U/μl; Thermo Fisher Scientific, cat. no. AM2238)
- RNeasy Mini Kit (Qiagen, cat. no. 74104)
- PD-10 desalting columns (GE Healthcare, cat. no. 17085101)

EQUIPMENT

- 3.5-cm Culture dishes (Nunc cell culture dishes; Thermo Fisher Scientific, cat. no. 150318)
- Cell culture incubator (air-jacketed incubator; VWR, cat. no. 10810-902)
- 15-ml Conical tubes (Corning; Sigma-Aldrich, cat. no. CLS430791)
- 1.7-ml Microcentrifuge tubes (VWR, cat. no. 87003-294)
- 0.65-ml Reaction tubes (microcentrifuge tubes; VWR, cat. no. 87003-290)
- Refrigerated microcentrifuge (Eppendorf 2427R; Eppendorf, cat. no. 022620702)
- UV spectrophotometer (NanoDrop 2000; Thermo Fisher Scientific, cat. no. ND-2000)
- Tube rotator (VWR, cat. no. 10136-084)

Computational requirements

- ShapeMapper and SuperFold, both available from http://www.chem.unc.edu/rna/software.html. Extensive documentation, installation instructions, and instructions for accessing sample data are described in the original protocol[23]. An updated and substantially more user-friendly version of ShapeMapper is now available[48]. These programs run from a command-line terminal (32- or 64-bit computer running Linux or OS X (v10.6 or greater); 4-GB RAM).

Optional computational resources (optional, used for running deltaSHAPE.py)

- deltaSHAPE.py, available from http://www.chem.unc.edu/rna/software.html. This small program will easily run on the same computer hardware used to run ShapeMapper.
- Python v2.7, available at https://www.python.org/download/releases/2.7
- Python module NumPy, v1.4 or greater, available at http://www.numpy.org
- Python module SciPy, available at http://www.scipy.org (deltaSHAPE.py has been validated to work with v0.14 or greater)
- Python module Matplotlib, available at http://www.matplotlib.org (deltaSHAPE.py has been validated to work with v1.5)

REAGENT SETUP

10× Denaturing control buffer 10× Denaturing control (DC) buffer consists of 500 mM HEPES (pH 8.0) and 40 mM EDTA. ▲ **CRITICAL** This buffer must be kept free of contamination by divalent ions such as Mg^{2+}, which will rapidly degrade RNA upon heating. This solution is stable at room temperature (20–25 °C) for at least 6 months.

Lysis buffer Lysis buffer consists of 40 mM Tris (pH 7.9), 25 mM NaCl, 6 mM $MgCl_2$, 1 mM $CaCl_2$, 256 mM sucrose, 0.5% (vol/vol) Triton X-100, 1,000 U/ml RNase inhibitor, and 450 U/ml DNase I. This solution should be stored at 4 °C and is stable for up to 6 months. For best results, add RNase inhibitor and DNase I immediately before use.

Protease K buffer Protease K buffer consists of 40 mM Tris (pH 7.9), 200 mM NaCl, 1.5% (wt/vol) SDS, and 500 μg/ml proteinase K. This solution should be stored at 4 °C and is stable for up to 6 months. For best results, proteinase K should be added immediately before use.

1.1× Folding buffer 1.1× Folding buffer consists of 110 mM HEPES (pH 8.0), 110 mM NaCl, and 4.4 mM $MgCl_2$. This buffer can be stored at room temperature for at least 6 months. ▲ **CRITICAL** The composition of this buffer—particularly the magnesium ion concentration—can be tailored to suit the specific needs of the system under study, but care should be taken to ensure that native RNA structures are preserved during cell-free extraction.

Pre-equilibrated phenol/chloroform/isoamyl alcohol Combine equal volumes of 1.1× folding buffer and phenol/chloroform/isoamyl alcohol in a glass bottle. Mix vigorously and allow the aqueous and organic phases to separate. Carefully aspirate the aqueous phase and store the equilibrated solvent mixture at 4 °C for up to 6 months. **! CAUTION** Phenol can cause severe burns. Always work with phenol in a fume hood and wear safety glasses, gloves, and a lab coat. Avoid direct contact.

PROCEDURE

In-cell RNA modification ● TIMING 30 min

▲ **CRITICAL** We describe in-cell modification conditions that are generally applicable for adherent mammalian cells in culture. The protocol may need to be adjusted depending on the type of cells being probed.

1| Culture the cells in two 3.5-cm culture dishes (or two wells of a six-well plate) to ~80% confluency. Label each dish/well as '+' or '–'.

2| Aspirate the medium from each dish and gently wash the cells with 1 ml of sterile PBS per dish.

3| Aspirate the PBS from each dish and add 900 µl of fresh prewarmed medium.

4| Add 100 µl of 100 mM 1M7 to the '+' dish while gently and promptly swirling to ensure rapid and thorough mixing. ▲ **CRITICAL STEP** It is important that the cell culture be rapidly and uniformly exposed to the SHAPE reagent. Continue to swirl or rock the dish back and forth until the culture medium appears homogeneous. If performing multiple probing experiments in parallel, thoroughly mix the SHAPE reagent in one dish before moving on to the next reaction.

5| Working quickly, add 100 µl of neat DMSO to the '–' dish while gently swirling to ensure thorough mixing.

6| Place the '+' and '–' dishes in an incubator at 37 °C for at least five 1M7 hydrolysis half-lives (~75 s at 37 °C).

7| For each dish, aspirate the medium and add 1 ml of TRIzol reagent. Pipette up and down to thoroughly lyse the cells. **! CAUTION** TRIzol contains phenol, which can cause severe burns. Always work with TRIzol in a fume hood and wear safety glasses, gloves, and a lab coat. Avoid direct contact.

8| Pipette the lysates into separate clean 1.5-ml microcentrifuge tubes and set aside at room temperature while performing the DC reaction (Steps 9–13).

9| Place 3 µl of previously prepared total cellular RNA, 5 µl of 100% formamide, and 1 µl of 10× DC buffer into a 0.65-ml reaction tube. Mix well by pipetting up and down.

10| Incubate the mixture at 95 °C for 1 min to denature the RNA.

11| Place 1 µl of 100 mM 1M7 in a clean 0.65-ml reaction tube. ▲ **CRITICAL STEP** Do not pre-incubate the SHAPE reagent at 95 °C. At elevated temperatures, the competing hydrolysis reaction proceeds quickly; moisture in the tube can reduce the effective concentration of SHAPE reagent.

12| Add 9 µl of denatured RNA from Step 10 to the DC reaction tube from Step 11, mix well, and incubate at 95 °C for 1 min.

13| Place the DC reaction tube on ice while purifying the '+' and '–' samples.

RNA purification ● TIMING 1.5 h

14| Add 200 µl of chloroform to each '+' and '–' tube from Step 8, cap tightly, and shake vigorously for 15–30 s.

15| Incubate for 2 min at room temperature, and then centrifuge the tubes at 12,000*g* for 15 min at 4 °C.

16| Remove the upper aqueous phase from each tube and place it in a clean 1.7-ml microcentrifuge tube, being careful not to disturb the interphase.

17| For each sample, add 500 µl of 100% isopropanol to the aqueous phase. Mix well and incubate at –20 °C for 20–30 min.

18| Centrifuge at 12,000*g* for 15 min at 4 °C. Remove the supernatant. A small RNA pellet should be visible on the side of the tube.

19| Wash the pellet with 75% (vol/vol) ethanol, and then centrifuge at 12,000*g* for 5 min at 4 °C.
■ **PAUSE POINT** The RNA pellet can be stored in ethanol for up to a year at –20 °C.

20| Dry the pellet in air or under vacuum, and then resuspend the RNA in 88 µl of RNase-free water.

DNase treatment ● **TIMING 1 h**
21| For each sample, using both cell-free modified RNA from **Box 1** and in-cell modified RNA from Step 20, assemble the DNase reaction as follows:

Component	Amount (µl)	Final concentration
Modified RNA	88	
Turbo DNase buffer (10×)	10	1×
Turbo DNase (2 U/µl)	2	0.04 U/µl

Box 2 | Parameters for automated ΔSHAPE analysis with deltaSHAPE.py

The default parameters for ΔSHAPE analysis are set such that SHAPE reactivities must differ by at least 1.96 s.d. and ΔSHAPE values must be at least 1 s.d. away from the mean ΔSHAPE value; at least 3 nucleotides (nt) meeting these criteria are required to occur in a 5-nt window in order to be highlighted by the program. To run deltaSHAPE.py with these default parameters, enter the following command:

```
python deltaSHAPE.py file1.map file2.map
```

where 'file1.map' and 'file2.map' are the names of .map files corresponding to each experiment being compared. deltaSHAPE.py automatically generates a file named 'differences.txt', which is a tab-delimited file providing the position, sequence, and statistical data for nucleotides identified as showing significant changes between experiments. The behavior of deltaSHAPE.py can be changed using the following flags:

`--help`	Displays available flags with descriptions.
`--out` [.txt file]	Name and location of output file. Default: `./differences.txt`.
`--mask5` [int]	Specifies the number of nucleotides at the 5′ end to ignore. Default: `0`.
`--mask3` [int]	Specifies the number of nucleotides at the 3′ end to ignore. Default: `0`.
`--pad` [int]	Indicates the smoothing window size. `Window = 2*pad+1`. To turn off smoothing, set `PAD = 0`. Default: `1`.
`--Zcoeff` [float]	Adjusts the *Z*-factor stringency by changing the equation coefficient. Default: `1.96`.
`--Zthresh` [float]	Adjusts the *Z*-factor stringency by changing the cutoff threshold. Default: `0.0`.
`--SSthresh` [float]	Sets the cutoff threshold for standard score filtering. Default: `1.0`.
`--FindSite` [int,int]	Comma-separated pair of numbers indicating the window pad size and the number of required hits when finding binding sites. Default settings look for 3+ nt within a 5-nt window. Default: `2,3`.
`--magrank`	Sorts the output file by decreasing deltaSHAPE magnitude instead of nucleotide position. Default: `OFF`.
`--all`	Outputs the data for all nucleotides. Insignificant changes are listed as zero. Default: `OFF`.
`--pdf`	Saves the plot as a PDF. If the output file is given, the PDF will have the same prefix. Default: `OFF`.
`--noshow`	Generates the plot but does not display automatically. It is often used with `--pdf` or when running many comparisons in series. Default: `display plot`.
`--noplot`	Skips plotting completely. Default: `OFF`.
`--dots`	Plot markers indicating nucleotides that pass *Z*-factor and standard score filtering. This can become unwieldy for large RNAs (>1,000 nt). Standard score (open) dots are plotted above *Z*-factor (filled) dots. Default: `OFF`.
`--Zdots`	Plot markers indicating only nucleotides that pass *Z*-factor filtering. Default: `OFF`.
`--SSdots`	Plot markers indicating only nucleotides that pass standard score filtering. Default: `OFF`.
`--colorfill`	Highlights ΔSHAPE sites with coloration beneath the plot line. Default: `OFF`.
`--ymin` [float]	Sets the plot *y*-axis minimum. Default: determined automatically.
`--ymax` [float]	Sets the plot *y*-axis maximum. Default: determined automatically.
`--xmin` [float]	Sets the plot *x*-axis minimum. Default: determined automatically.
`--xmax` [float]	Sets the plot *x*-axis maximum. Default: determined automatically.

22| Incubate the DNase reactions at 37 °C for 30 min.

23| Purify RNA from the DNase reactions and from the DC reaction using individual RNeasy Mini spin columns for each sample, according to the manufacturer's instructions.

Library preparation, sequencing, and processing of raw data ● TIMING 2–9 d, depending on the sequencing platform

24| Details for preparing, sequencing, and processing SHAPE-MaP libraries are described in Steps 16–38 of the original protocol[23].

? TROUBLESHOOTING

(Optional) ΔSHAPE analysis ● TIMING 1 h

25| To calculate regions of significant difference between experiments using ΔSHAPE, use two .map files generated by ShapeMapper[48] (see the original protocol for a description of file formats) with the deltaSHAPE.py script by typing the following command. Note that the per-nucleotide reactivity values in 'file2.map' will be subtracted from those in 'file1.map'.

```
python deltaSHAPE.py file1.map file2.map
```

Expected outcome: using default parameters, a plot of calculated ΔSHAPE values will be displayed. Regions of significant SHAPE reactivity increases and decreases will be colored purple and green, respectively. In addition, a text file, 'differences.txt' will be created that contains the ΔSHAPE values and additional statistical information about the highlighted nucleotides. The output and analysis parameters of the software can easily be adjusted by providing additional command-line flags; see **Box 2** for a complete description of available flags.

▲ CRITICAL STEP When analyzing data sets that include nucleotides masked by primer-binding sites at the 5′ or 3′ ends, it is important to specify the lengths of these masked regions using the `--mask5` and/or `--mask3` flags. Failure to do so will artificially skew the statistical calculations used by the ΔSHAPE analysis.

? TROUBLESHOOTING

The approaches for troubleshooting an in-cell SHAPE-MaP experiment are nearly identical to those for a conventional cell-free experiment. For troubleshooting advice pertaining to in-cell SHAPE modification, refer to **Table 1**. For troubleshooting advice pertaining to mutational profiling, library generation, and processing of raw data, refer to the original protocol[23].

TABLE 1 | Troubleshooting table.

Step	Problem	Possible reason	Solution
24	Low SHAPE reactivity values throughout the transcript(s) of interest	Clumped cells may be preventing the SHAPE reagent from uniformly probing the cells	Ensure that the cells are monodispersed, if grown in suspension, or that adherent cells are grown in a monolayer without clumping. Optionally, trypsinize the cells before probing
	Low SHAPE reactivity values throughout the transcript(s) of interest	pH of the growth medium may be outside the optimal range for SHAPE probing	Ensure that the pH is within the range of 7.4–8.3. If necessary and possible, replace the growth medium with PBS at a suitable pH immediately before SHAPE probing
	Low SHAPE reactivity values throughout the transcript(s) of interest	SHAPE reagent may not be permeating the cellular membrane	Perform a control experiment using cells known to be compatible with in-cell SHAPE-MaP, such as mouse embryonic stem cells or Jurkat cells Alternatively, select a different SHAPE reagent

● TIMING

Steps 1–13, in-cell RNA modification: 30 min
Steps 14–20, RNA purification: 1.5 h
Steps 21–23, DNase treatment: 1 h
Step 24, library preparation, sequencing, and data processing: 2–9 d, depending on the sequencing platform
Step 25, ΔSHAPE analysis: 1 h
Box 1, creation of a cell-free SHAPE reference sample: 6 h

ANTICIPATED RESULTS

In-cell SHAPE-MaP enables measurement of RNA flexibility at single-nucleotide resolution in living cells and can be adapted for use with RNAs of varying abundance and length. When paired with cell-free structure-probing data, in-cell SHAPE reactivities can be used to highlight possible areas of cellular interaction at both local per-nucleotide and global scales (**Figs. 2** and **3**). These data are also suitable for ΔSHAPE analysis, which identifies meaningful local differences between data sets and can be used to identify RNA–protein interaction sites with high resolution. The examples provided here highlight instructive applications of in-cell SHAPE-MaP as compared with cell-free data using the 1M7 SHAPE reagent, but, essentially, any informative set of experimental conditions and nearly any RNA-modifying reagent can be considered.

As an example of ΔSHAPE analysis, cell-free and in-cell SHAPE-MaP profiles were generated for the mouse U1 snRNA[18]. This RNA interacts with several proteins (U1A, U1C, U1-70K, and the heteroheptameric Sm ring) to form the U1 snRNP complex. Qualitative analysis of the two SHAPE profiles indicates many differences—possibly indicative of protein interactions—throughout the RNA (**Fig. 4a**). However, when these data are analyzed using the ΔSHAPE framework, only some of the observed qualitative differences are significant (**Fig. 4b**, in green). These ΔSHAPE sites correspond well to RNA–protein interactions when superimposed on a crystallographic model of the U1 snRNP complex (**Fig. 4c**). These include nucleotides 30–33, which are deeply buried in the U1-70K binding pocket; nucleotides 73–75, which interact with the U1A protein; and nucleotides 124–129, which are single-stranded and sequestered by the Sm ring complex.

In-cell SHAPE-MaP can also be applied to large, rare transcripts. The previously described amplicon workflow uses gene-specific primers to amplify and detect chemical adducts in a specific region of interest (see Fig. 4 in ref. 23). In one example, the SHAPE-MaP amplicon strategy makes it possible to generate very-high-quality in-cell and cell-free SHAPE-MaP profiles over the entire 18-kb mouse *Xist* lncRNA, even in the background of total transcriptome RNA. In conjunction with secondary structure modeling, a global difference analysis of cell-free and in-cell SHAPE reactivities revealed locations throughout the transcript that exhibit strong regional changes (**Fig. 5a**)[24]. These large-scale changes were broken down into contributions reflecting positive and negative differences (**Fig. 5b**) to identify regions that likely transition between structured and unstructured states as a result of cellular interactions (**Fig. 5c**). This global analysis of

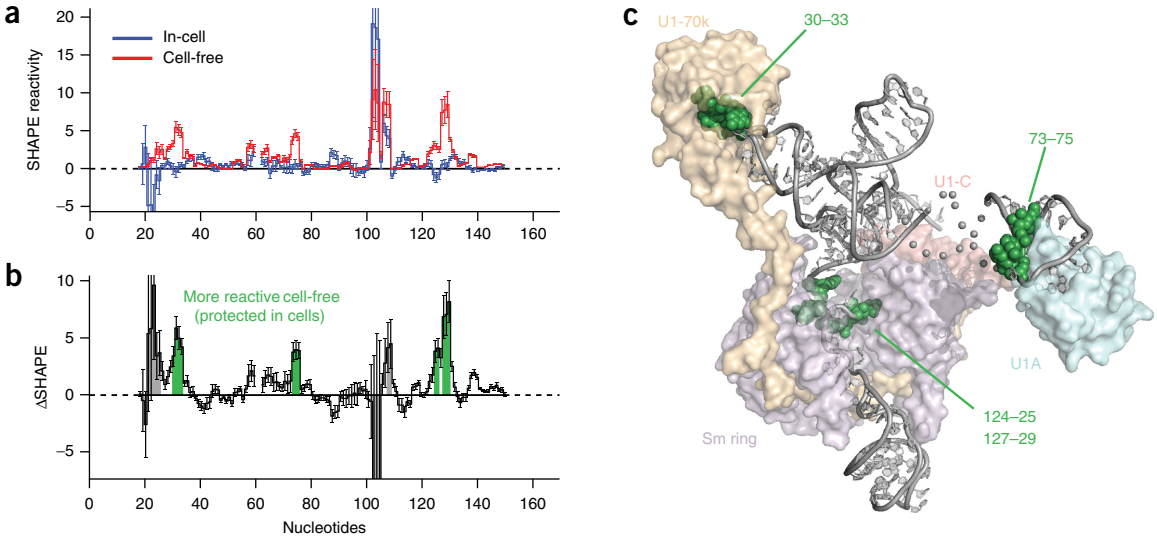

Figure 4 | ΔSHAPE analysis for a small RNP complex. (**a**) In-cell and cell-free SHAPE reactivities are shown for the mouse U1 snRNA. Error bars represent the estimated standard error, calculated by ShapeMapper. (**b**) ΔSHAPE reactivity profile for the U1 snRNA. Regions of the RNA exhibiting significant changes between in-cell and cell-free data sets are highlighted in green. Other regions exhibiting large but statistically insignificant changes are colored gray. (**c**) Crystallographic model of the U1 snRNP complex[55] highlighting the locations of ΔSHAPE sites (green spheres) and their proximities to known RNA–protein interactions. Image adapted with permission from ref. 18, American Chemical Society.

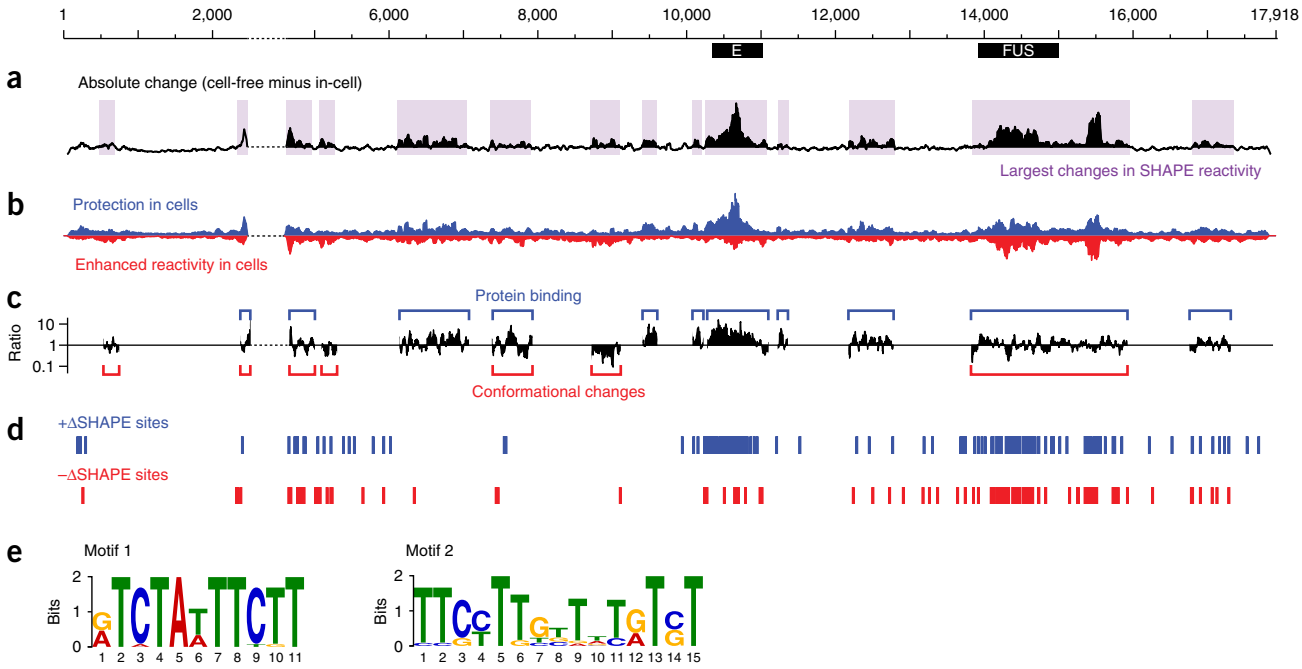

Figure 5 | Large-scale difference analyses applied to the ~18-kb *Xist* lncRNA. (**a**) Absolute difference in SHAPE reactivity between in-cell and cell-free experiments, smoothed over a 50-nt sliding window. Regions exhibiting a smoothed reactivity difference greater than the global median for at least 100 nt are highlighted in purple. (**b**) The contributions of positive (blue) and negative (red) differences to the absolute difference highlight regions where *Xist* is more or less reactive in the cellular environment than in the cell-free state, respectively. (**c**) The ratio of positive to negative differences is plotted for regions identified as undergoing large-scale reactivity changes in **a**. By examining the ratios, regions can be sorted by whether they exhibit mostly positive or mostly negative changes, suggestive of protein binding and RNA structural rearrangement, respectively. (**d**) The location of positive (blue) and negative (red) ΔSHAPE sites across the *Xist* lncRNA. In this example, ΔSHAPE sites are nonuniformly distributed and tend to overlap with regions of large-scale change identified in **a**. (**e**) Sequence motifs enriched in +ΔSHAPE sites. These U-rich sequences were identified by searching for motifs in the 20 nt surrounding the center of +ΔSHAPE sites (corresponding to protection in cells) and generally map to an unstructured region spanning nucleotides 10,000–11,000. Image adapted with permission from ref. 24, National Academy of Sciences.

Xist highlighted a repetitive sequence element in the middle of the transcript (repeat E) that undergoes a marked change from being highly unstructured in the cell-free state to being highly constrained in cells, probably due, at least in part, to extensive binding by proteins.

ΔSHAPE analysis can also be applied to large transcripts such as *Xist*. When in-cell and cell-free probing data sets were compared, ~200 ΔSHAPE sites were identified across the *Xist* RNA (**Fig. 5d**)[24]. When many sites are identified within a large transcript, it is often helpful to apply additional downstream analyses to further explore the nature of these sites. For example, ΔSHAPE sites at which the in-cell reactivity is reduced relative to the cell-free state are good candidates for possible RNA–protein interaction sites[18]. These sites showing protection in cells can be searched for shared sequence motifs. In the *Xist* lncRNA, we identified strong signals corresponding to two U-rich motifs that are likely bound by proteins in the repeat E unstructured repetitive sequence element (**Fig. 5e**). Identification of these sites was made possible by focusing specifically on sequence motifs shared among in-cell protected regions[24].

Second, ΔSHAPE sites can be used to support or identify interactions with specific proteins, based on overlap with protein-binding sites identified by CLIP-seq experiments. CLIP sites mapped on the *Xist* RNA typically span much of the RNA, and we found that it was generally not possible to identify any sequence or structure consensus interaction site for a given protein (**Fig. 6a**, open symbols). We reasoned that specific protein binding should impact the RNA sufficiently to perturb the local SHAPE reactivity, as measured by a ΔSHAPE signal. Indeed, when CLIP-seq sites for different proteins were filtered by ΔSHAPE sites, the smaller set of ΔSHAPE-supported sites revealed sequence- and structure-recognition features that were otherwise obscured by the likely low resolution and low stringency of typical CLIP-seq sites (**Fig. 6a**, closed symbols; **Fig. 6b**). In the case of two proteins, CELF1 and PTBP1, CLIP sites overlapping ΔSHAPE sites were enriched for a sequence motif that could not be identified when searching over all reported CLIP sites (**Fig. 6b**).

We also examined the interactions of *Xist* with the FUS protein, which binds many RNAs and is generally considered to be a promiscuous binder[52]. We evaluated the cell-free-derived structure model of *Xist* at ΔSHAPE sites overlapping FUS CLIP sites and identified a conserved structural motif: the ΔSHAPE-supported FUS sites tend to be single-stranded but are flanked by base pairs (**Fig. 6c**). SHAPE-supported FUS sites were also overrepresented in a specific region with high levels

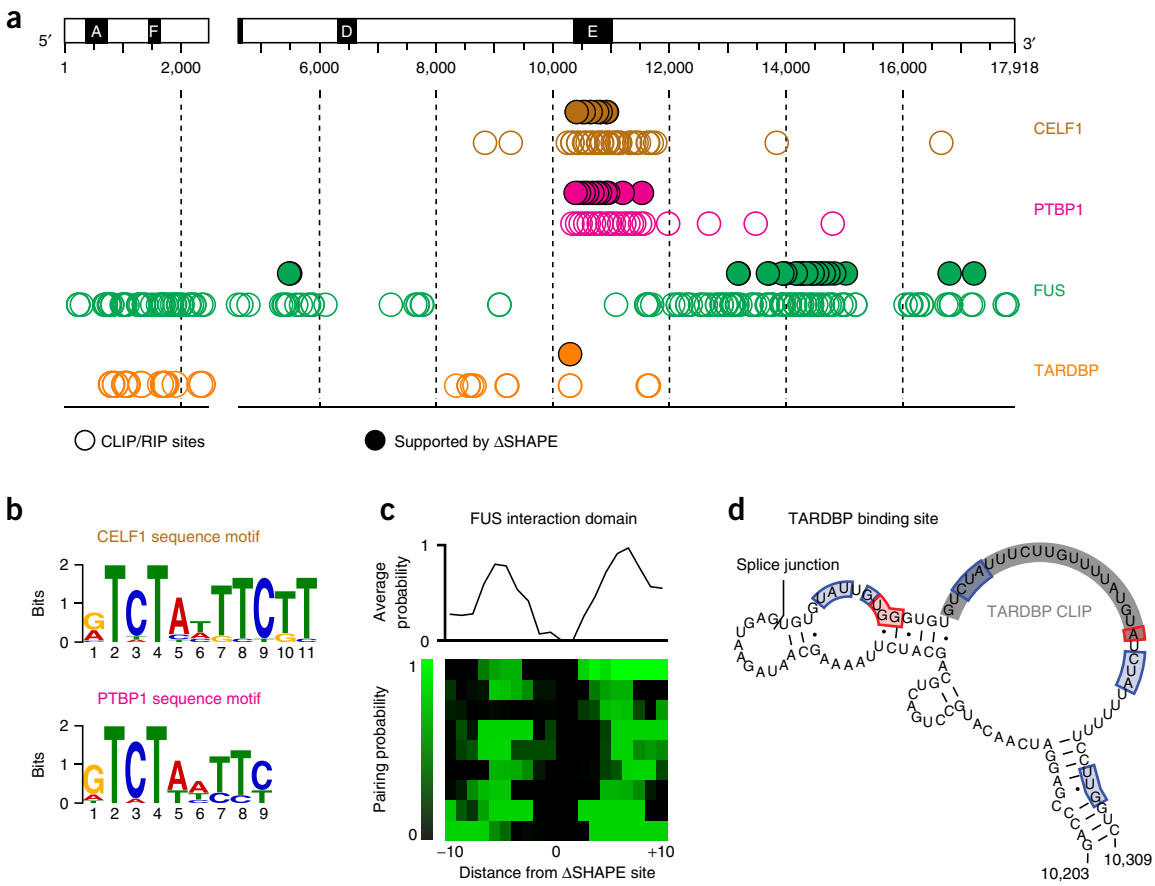

Figure 6 | Examples of downstream analyses of mouse *Xist* ΔSHAPE sites. (**a**) Locations of reported RNA-binding protein CLIP sites supported by ΔSHAPE are emphasized with filled circles; sites identified by CLIP only are shown with open symbols. CELF1 and PTBP1 sites generally occur near the repeat E region; FUS sites are situated in a low SHAPE/low-entropy region spanning nucleotides 13,500–15,000 (ref. 24). Although multiple TARDBP CLIP sites have been reported, only one is supported by ΔSHAPE. (**b**) Motif analysis of the CELF1 and PTBP1 ΔSHAPE-supported CLIP sites reveals similar sequence motifs, suggesting that these RNA-binding proteins recognize similar sequence elements in *Xist*. These motifs could not be identified when searching over all reported CLIP sites for these proteins. (**c**) Structural preferences of ΔSHAPE-supported FUS CLIP sites. The predicted pairing probabilities were calculated for the 20-nt window centered on ΔSHAPE sites overlapping FUS CLIP sites, revealing a preference for single-stranded binding sites flanked by base pairs. (**d**) Structural model of the TARDBP–*Xist* interaction site. Positive (blue) and negative (red) ΔSHAPE sites, and the reported CLIP site (gray) are shown. Image adapted with permission from ref. 24, National Academy of Sciences.

of well-determined RNA secondary structure, using the low SHAPE/low-entropy metric outlined in the original protocol[23,24]. In-cell SHAPE experiments thus revealed an underlying structural code for RNA–protein interactions and suggested that a well-determined and highly structured element in a large RNA might constitute a recognition domain for binding by multiple FUS proteins.

Finally, we examined the interactions between *Xist* and TARDBP, a splicing regulator with a reported preference for UG-rich sequences[53,54]. Based on two fully independent biological replicates, only a single TARDBP CLIP site was supported by ΔSHAPE. This site overlapped a UG-rich structural motif at a splice junction (**Fig. 6d**), suggesting that TARDBP recognizes a specific sequence and structural context to regulate *Xist* splicing. These examples of downstream ΔSHAPE analysis highlight how a comparison of two or more experimental states can lead to insightful hypotheses about RNA structure–function relationships.

In conclusion, in-cell SHAPE-MaP enables RNA structure to be examined at single-nucleotide resolution in a native cellular environment, including for RNAs of low to moderate abundance. In-cell SHAPE is simple to perform and amenable to many cell types, media, and growth conditions. Global difference analysis and the ΔSHAPE framework are proven approaches for examining how the cellular environment impacts RNA structure relative to a reference state and can complement other biochemical assays. We anticipate that in-cell SHAPE-MaP will be a broadly useful technology for understanding the role of RNA structure–function interrelationships in regulating cellular biology.

ACKNOWLEDGMENTS Work in our lab focused on developing quantitative and biophysically rigorous RNA structure-probing technologies is supported by the National Institutes of Health (NIH; R35 GM122532 and R01 AI068462). M.J.S. was a Graduate Research Fellow of the National Science Foundation (DGE-1144081) and was supported in part by an NIH training grant in molecular and cellular biophysics (T32 GM08570). We are indebted to the Calabrese laboratory at the University of North Carolina at Chapel Hill for assistance in developing in-cell probing strategies and to members of the Weeks laboratory for thoughtful feedback regarding the analysis algorithms and strategies described here.

AUTHOR CONTRIBUTIONS M.J.S. and K.M.W. conceived the use of SHAPE-MaP in living cells. M.J.S. developed the in-cell probing strategy and created the ΔSHAPE analysis procedure and software. Both authors wrote and edited the manuscript.

COMPETING INTERESTS
K.M.W. has equity ownership in and serves as an advisor to Ribometrix, to which SHAPE-MaP technologies have been licensed. M.J.S. is an employee of Ribometrix.

1. Sharp, P.A. The centrality of RNA. *Cell* **136**, 577–580 (2009).
2. Wang, Z. & Burge, C.B. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14**, 802–813 (2008).
3. Ghildiyal, M. & Zamore, P.D. Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* **10**, 94–108 (2009).
4. Sherwood, A.V. & Henkin, T.M. Riboswitch-mediated gene regulation: novel RNA architectures dictate gene expression responses. *Annu. Rev. Microbiol.* **70**, 361–374 (2016).
5. Lau, M.W.L. & Ferré-D'Amaré, A.R. Many activities, one structure: functional plasticity of ribozyme folds. *Molecules* **21**, 1570 (2016).
6. Rinn, J.L. & Chang, H.Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166 (2012).
7. Engreitz, J.M., Ollikainen, N. & Guttman, M. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat. Rev. Mol. Cell Biol.* **17**, 756–770 (2016).
8. Cech, T.R. & Steitz, J.A. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157**, 77–94 (2014).
9. Mortimer, S.A., Kidwell, M.A. & Doudna, J.A. Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.* **15**, 469–479 (2014).
10. Butcher, S.E. & Pyle, A.M. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.* **44**, 1302–1311 (2011).
11. Nicholson, B.L. & White, K.A. Exploring the architecture of viral RNA genomes. *Curr. Opin. Virol.* **12**, 66–74 (2015).
12. Gebhard, L.G., Filomatori, C.V. & Gamarnik, A.V. Functional RNA elements in the dengue virus genome. *Viruses* **3**, 1739–1756 (2011).
13. Licatalosi, D.D. & Darnell, R.B. RNA processing and its regulation: global insights into biological networks. *Nat. Rev. Genet.* **11**, 75–87 (2010).
14. Spitale, R.C. *et al.* RNA SHAPE analysis in living cells. *Nat. Chem. Biol.* **9**, 18–20 (2012).
15. 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).
16. Loughrey, D., Watters, K.E., Settle, A.H. & Lucks, J.B. SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Res.* **42**, e165–e165 (2014).
17. Siegfried, N.A., Busan, S., Rice, G.M., Nelson, J.A.E. & Weeks, K.M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* **11**, 959–965 (2014).
18. Smola, M.J., Calabrese, J.M. & Weeks, K.M. Detection of RNA-protein interactions in living cells with SHAPE. *Biochemistry* **54**, 6867–6875 (2015).
19. McGinnis, J.L. *et al.* In-cell SHAPE reveals that free 30S ribosome subunits are in the inactive state. *Proc. Natl. Acad. Sci. USA* **112**, 2425–2430 (2015).
20. Tyrrell, J., McGinnis, J.L., Weeks, K.M. & Pielak, G.J. The cellular environment stabilizes adenine riboswitch RNA structure. *Biochemistry* **52**, 8777–8785 (2013).
21. McGinnis, J.L. & Weeks, K.M. Ribosome RNA assembly intermediates visualized in living cells. *Biochemistry* **53**, 3237–3247 (2014).
22. Fang, R., Moss, W.N., Rutenberg-Schoenberg, M. & Simon, M.D. Probing Xist RNA structure in cells using targeted structure-Seq. *PLoS Genet* **11**, e1005668 (2015).
23. Smola, M.J., Rice, G.M., Busan, S., Siegfried, N.A. & Weeks, K.M. Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.* **10**, 1643–1669 (2015).
24. Smola, M.J. *et al.* SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. *Proc. Natl. Acad. Sci. USA* **113**, 10322–10327 (2016).
25. Weeks, K.M. & Mauger, D.M. Exploring RNA structural codes with SHAPE chemistry. *Acc. Chem. Res.* **44**, 1280–1291 (2011).
26. Spitale, R.C. *et al.* Structural imprints *in vivo* decode RNA regulatory mechanisms. *Nature* **519**, 486–490 (2015).
27. Homan, P.J. *et al.* Single-molecule correlated chemical probing of RNA. *Proc. Natl. Acad. Sci. USA* **111**, 13858–13863 (2014).
28. Larman, B.C., Dethoff, E.A. & Weeks, K.M. Packaged and free satellite tobacco mosaic virus (STMV) RNA genomes adopt distinct conformational states. *Biochemistry* **56**, 2175–2183 (2017).
29. Underwood, J.G. *et al.* FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* **7**, 995–1001 (2010).
30. 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).
31. Ding, Y. *et al.* *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**, 696–700 (2014).
32. Incarnato, D., Neri, F., Anselmi, F. & Oliviero, S. Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol.* **15**, 491 (2014).
33. 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 (2015).
34. Krokhotin, A., Mustoe, A.M., Weeks, K.M. & Dokholyan, N.V. Direct identification of base-paired RNA nucleotides by correlated chemical probing. *RNA* **23**, 6–13 (2017).
35. McGinnis, J.L., Dunkle, J.A., Cate, J.H.D. & Weeks, K.M. The mechanisms of RNA SHAPE chemistry. *J. Am. Chem. Soc.* **134**, 6617–6624 (2012).
36. Gherghe, C.M., Shajani, Z., Wilkinson, K.A., Varani, G. & Weeks, K.M. Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S2) in RNA. *J. Am. Chem. Soc.* **130**, 12244–12245 (2008).
37. Deigan, K.E., Li, T.W., Mathews, D.H. & Weeks, K.M. Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. USA* **106**, 97–102 (2009).
38. Hajdin, C.E. *et al.* Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl. Acad. Sci. USA* **110**, 5498–5503 (2013).
39. Watters, K.E., Abbott, T.R. & Lucks, J.B. Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. *Nucleic Acids Res.* **44**, e12 (2016).
40. Takahashi, M.K. *et al.* Using in-cell SHAPE-Seq and simulations to probe structure-function design principles of RNA transcriptional regulators. *RNA* **22**, 920–933 (2016).
41. Kuhlmann, M.M., Chattopadhyay, M., Stupina, V.A., Gao, F. & Simon, A.E. An RNA element that facilitates programmed ribosomal readthrough in turnip crinkle virus adopts multiple conformations. *J. Virol.* **90**, 8575–8591 (2016).
42. Sztuba-Solinska, J. *et al.* Kaposi's sarcoma-associated herpesvirus polyadenylated nuclear RNA: a structural scaffold for nuclear, cytoplasmic and viral proteins. *Nucleic Acids Res.* **45**, 6805–6821 (2017).
43. Lee, B. *et al.* Comparison of SHAPE reagents for mapping RNA structures inside living cells. *RNA* **23**, 169–174 (2017).
44. Lavender, C.A. *et al.* Model-Free RNA sequence and structure alignment informed by SHAPE probing reveals a conserved alternate secondary structure for 16S rRNA. *PLoS Comput. Biol.* **11**, e1004126 (2015).
45. Engreitz, J.M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973–1237973 (2013).

46. Simon, M.D. *et al.* High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* **504**, 465–469 (2013).
47. Jabara, C.B., Jones, C.D., Roach, J., Anderson, J.A. & Swanstrom, R. Accurate sampling and deep sequencing of the HIV-1 protease gene using a primer ID. *Proc. Natl. Acad. Sci. USA* **108**, 20166–20171 (2011).
48. Busan, S. & Weeks, K.M. Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA* **24**, 143–148 (2017).
49. Mortimer, S.A. & Weeks, K.M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* **129**, 4144–4145 (2007).
50. Steen, K.-A., Siegfried, N.A. & Weeks, K.M. Synthesis of 1-methyl-7-nitroisatoic anhydride (1M7). *Protoc. Exchange* http://dx.doi.org/10.1038/protex.2011.255 (2011).
51. Turner, R., Shefer, K. & Ares, M. Safer one-pot synthesis of the 'SHAPE' reagent 1-methyl-7-nitroisatoic anhydride (1m7). *RNA* **19**, 1857–1863 (2013).
52. Wang, X., Schwartz, J.C. & Cech, T.R. Nucleic acid-binding specificity of human FUS protein. *Nucleic Acids Res.* **43**, 7535–7543 (2015).
53. Lagier-Tourenne, C., Polymenidou, M. & Cleveland, D.W. TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. *Hum. Mol. Genet.* **19**, R46–R64 (2010).
54. Bhardwaj, A., Myers, M.P., Buratti, E. & Baralle, F.E. Characterizing TDP-43 interaction with its RNA targets. *Nucleic Acids Res.* **41**, 5062–5074 (2013).
55. Kondo, Y., Oubridge, C., van Roon, A.-M.M. & Nagai, K. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5′ splice site recognition. *Elife* **4** 04986 (2015).