Visualizing the physical basis for molecular behaviour inside living cells is a great challenge for biology. RNAs are central to biological regulation, and the ability of RNA to adopt specific structures intimately controls every step of the gene expression program. However, our understanding of physiological RNA structures is limited; current in vivo RNA structure profiles include only two of the four nucleotides that make up RNA. Here we present a novel biochemical approach, in vivo click selective 2′-hydroxyl acylation and profiling experiment (icSHAPE), which enables the first global view, to our knowledge, of RNA secondary structures in living cells for all four bases. icSHAPE of the mouse embryonic stem cell transcriptome versus purified RNA folded in vitro shows that the structural dynamics of RNA in the cellular environment distinguish different classes of RNAs and regulatory elements. Structural signatures at translational start sites and ribosome pause sites are conserved from in vitro conditions, suggesting that these RNA elements are programmed by sequence. In contrast, focal structural rearrangements in vivo reveal precise interfaces of RNA with RNA-binding proteins or RNA-modification sites that are consistent with atomic-resolution structural data. Such dynamic structural footprints enable accurate prediction of RNA–protein interactions and N6-methyladenosine (m6A) modification genome wide. These results open the door for structural genomics of RNA in living cells and reveal key physiological structures controlling gene expression.

SHAPE accurately identifies flexible (single-stranded) bases in RNA for all four nucleotides. However, current methods are potentially limited by high background rates (>70% of RNA molecules have no modification due to single-hit kinetics) and high false-positives rates due to spurious reverse transcription stops. We overcome these problems by developing a new SHAPE probe that permits in vivo SHAPE modification and subsequent selective purification of the modified RNAs. We designed, synthesized and tested a novel bifunctional chemical probe for in vivo SHAPE structure profiling genome wide (NAI-N3; Fig. 1a, b and Extended Data Fig. 1). NAI-N3 adds an azide group to NAI (2-methyl nicotinic acid imidazole), a cell-permeable SHAPE reagent. By using copper-free click chemistry, a biotin moiety is selectively and efficiently added to NAI-N3-modified RNA, providing a stringent purification handle with streptavidin-conjugated beads (Fig. 1c and Extended Data Fig. 2). NAI-N3 generated identical profiles of reverse transcription stops to those obtained using our previously designed SHAPE reagent. The fidelity of structural measurements was not affected by ‘clicking’ biotin onto NAI-N3, or by molecular crowding of proteins, and NAI-N3 showed uniform modification of all bases in denatured RNAs (Extended Data Fig. 3). We term this new chemoaffinity structure probing methodology icSHAPE; this method can also be applied to any ex vivo preparation of RNA, with slight modifications.

icSHAPE of ribosomal RNAs in mouse embryonic stem (ES) cells indicated that the method is quantitative and accurate, reporting the known structures of 18S and 28S ribosomal RNAs (Fig. 1d–f and Extended Data Fig. 4). Deep-sequencing results from icSHAPE showed strong correspondence with manual structure-probing gels (Pearson correlation r = 0.93, in vivo; Fig. 1d and Extended Data Fig. 4). rRNA is known to require the cellular environment for proper folding, and differences between in vivo and in vitro icSHAPE measurements highlighted important structural elements in the intact ribosome. We mapped our icSHAPE profiles onto the cryo-electron microscopy structure of the human 80S ribosome and searched for differences between the in vivo and in vitro conditions. Conserved (mouse to human) nucleotides of high icSHAPE signal in vivo were unpaired in the cryo-electron microscopy structure (Fig. 1e); conversely, residues lacking icSHAPE reactivity in vivo were base-paired or engaged in extensive interactions that may stabilize the RNA backbone in the mature ribosome (Fig. 1f). Overall, these data
demonstrate that icSHAPE accurately measures RNA structures, both inside and outside of living cells.

We next used icSHAPE to measure RNA structural profiles of polyadenylated transcripts in mouse ES cells and generated ~2.1 billion measurements for over 13,200 RNAs in vitro and in vivo, with high reproducibility (Extended Data Figs 5 and 6). The nucleotide composition in the transcriptome, mock-treated RNA and icSHAPE-treated RNAs are highly concordant, with a slight enrichment in NAIP-Ns for As and Us (Fig. 2a). This enrichment is expected given their bias for being located in single-stranded or loop regions1. icSHAPE thus affords the first complete RNA structurome of all four nucleotides in vivo.

icSHAPE data revealed the scale and distribution of RNA structural dynamics between in vitro conditions, in which folding is programmed entirely by sequence, versus in vivo conditions, in which folding occurs in the context of the intracellular environment2. Recent transcriptome-wide dimethylsulfate probing (DMS-seq), which interrogates two bases with strong bias towards adenosines (68% As and 24% Cs)3,4, suggested that RNA structures are largely unfolded in vitro; however, sampling only two of four nucleotides could result in an incomplete picture. We quantified RNA structural dynamics using two metrics. First, we calculated the difference in reactivity between our measurements (Fig. 2c). These observations suggest that using probes that have a broader reactivity profile, such as NAIP-Ns, will give a more complete representation of RNA structure.

Second, we used the Gini index2 to quantify the distribution of icSHAPE reactivity profiles. Structured RNAs have some bases that are reactive and some not, leading to unequal distribution and a high Gini index, whereas unfolded RNAs have most bases in a uniformly reactive conformation (low Gini). We found that RNAs are less folded in vivo, consistent with a previous report5, but the extent of unfolding varies in degrees that distinguish different classes of RNAs (Fig. 2d). Protein-coding messenger RNAs exhibited noticeable but partial unfolding (average Gini of 0.7 in vitro to 0.5 in vivo), with the largest variation noted at 3’ untranslated regions (UTRs) compared to coding sequences (CDSs) or 5’ UTRs. In contrast, noncoding RNAs, such as pseudogenes, long noncoding RNAs (lncRNAs) and primary microRNA (miRNA) precursors, retain substantially more of their RNA structure in vivo (P < 2.2 × 10^-16, noncoding versus coding, Student’s t-test). One exception to this rule are small nuclear RNAs (snRNAs), which exhibit the greatest level of increased reactivity in vivo among all classes of transcripts and may result from extensive rearrangements due to small nuclear ribonucleoprotein (snRNP) binding. Thus, most RNAs in vivo possess a substantial level of RNA structure beyond previous expectations based on DMS-seq2. Our data further suggest that RNA structural signatures in vivo can distinguish coding versus structural or regulatory RNAs, consistent with previous in vitro studies2-12.

The dramatically different environments that RNA experiences when inside a cell compared to in vitro suggests that our VTD parameter could provide insight into functionally important RNA regulatory elements. To assess this possibility, we measured the VTD for all hexamer sequences (Fig. 2e and Supplementary Table 1). We observed unique VTD profiles for sequence motifs driving diverse post-transcriptional processes, including translation initiation, interaction with RNA-binding proteins (RBPs; for example, Rbp2), RNA modification (m^6A) and miRNA seed matches13-15 (Supplementary Tables 1 and 2). These results show that the VTD may classify RNA regulatory elements as preprogrammed or sensitive to in vivo remodelling. Furthermore, distinctive VTD profiles precisely at sites of post-transcriptional regulatory motifs suggest that RNA structural dynamics may be used to monitor these regulatory events in cells.

We hypothesized that translational regulatory elements may have conserved icSHAPE profiles between in vivo and in vitro conditions because the Kozak sequence, important for translation initiation16, is among the most stable (low VTD) regions within mRNAs (Fig. 2d). RNA accessibility from -1 to -5 nucleotides upstream of the start codon has a major role in regulating translational output10,11. We used translation initiation18 and pause sites18, defined by ribosome profiling, to centre our structural reactivity analysis across the transcriptome (Fig. 3). Canonical initiation AUG sites are indeed preceded by ~5 nucleotides of increased accessibility, and this pattern is nearly identical to in vitro folded RNA (Fig. 3a, b). A similar pattern of conserved upstream accessibility also precedes noncanonical start sites at upstream open reading frames (uORFs) and amino-terminal truncations (Fig. 3c). Non-start-sites AUG codons are also associated with increased preceding reactivity, whereas noncanonical CUG start codons have a different profile, suggesting that RNA accessibility alone is not sufficient to dictate translational start sites (Extended Data Fig. 7). Ribosome profiling also defined ribosome pause sites as having a strong preference for glutamate or aspartate in the acceptor (A) site, where transfer RNA (tRNA) identity and the nascent peptide sequence are believed to strongly influence translation kinetics18. icSHAPE data at ribosome pause sites revealed a distinctive signature: loss of reactivity at the exit (E) and peptidyl-transferase (P) sites, whereas the A site is more reactive, preceded by a strong 3-nucleotide periodic reactivity pattern 5’ to the pause site for ~12 nucleotides (Fig. 3d, e). Furthermore, a very similar pattern was observed in vitro under conditions that do not maintain mRNA interactions with the ribosome or tRNAs, suggesting that these structural profiles are programmed by mRNA sequence. Analysis of negative control sites—defined as sites on the same transcripts that match the codon composition, are in frame, and are at least 20 nucleotides away from...
true pause sites—showed a very similar icSHAPE signature at the presumed ribosome E, P and A sites, but negative controls lacked the 5' periodic signal (grey box in Fig. 3e, f). This observation suggests that the icSHAPE signature at ribosome pause sites is probably due to the codon bias at such sites, but sequences 5' to the pause site may influence pausing. These results identify several physiological structural signatures of translational control elements, and suggest that they may be largely pre-programmed by the mRNA sequence.

In contrast, focal RNA structural rearrangements in vivo can identify sites of RBP interactions, which regulate RNA splicing, localization and stability19 (Fig. 2d). The feminizing locus on X (Fox) family of RBPs identify sites of RBP interactions, which regulate RNA splicing, localization largely pre-programmed by the mRNA sequence.

mRNA codon bias at such sites, but sequences 5' to the motif showed strong icSHAPE VTD signals. The 2'-hydroxyl groups of these three residues are flipped outward while G2 and A4 base pair upon Rbfox interaction15, consistent with the adoption of new structural rearrangement in vivo. Alignment with the Rbfox–RNA NMR structure14 and Rbfox2-binding sites identified by individual nucleotide crosslinking immunoprecipitation (iCLIP) in mouse ES cells26 showed that the differential icSHAPE signal precisely matches the key RNA residues involved in Rbfox interaction (Fig. 4a, b). U1, G2 and A4 in the motif showed strong icSHAPE VTD signals. The 2'-hydroxyl groups of these three residues are flipped outward while G2 and A4 base pair upon Rbfox interaction15, consistent with the adoption of new structural environments in vivo that we detected at these residues. In principle, the dynamic structural footprints of RBPs may enable comprehensive readout of RNA–RBP interactions in vivo. We tested this idea by implementing a support vector machine (SVM) algorithm to learn which dynamic icSHAPE signals are best able to predict sites of mRNA regulation, using held out data for cross-validation of prediction accuracy (Extended Data Fig. 8 and Methods). Indeed, the combination of both in vivo and in vitro icSHAPE data increased the ability to predict true Rbfox2-binding sites compared to motif sequence or conservation alone, particularly at lower false-positive rates where accuracy is most important (area under the curve (AUC) = 0.74; Extended Data Fig. 8).

As an independent validation, we used icSHAPE data to predict the binding sites of HuR, an RBP that regulates transcript stability15, and also performed the first HuR iCLIP, to our knowledge, in mouse ES cells. Comparing in vivo versus in vitro icSHAPE data precisely identified peaks of structural arrangement at authentic HuR-binding sites (defined by iCLIP sites), and enabled reasonably accurate prediction of HuR binding from icSHAPE data alone (AUC = 0.841; Extended Data Fig. 8 and HuR iCLIP data in Extended Data Fig. 9). Thus, icSHAPE data can distinguish true binding sites from other sequence motif instances, collectively boosting prediction accuracy.

We also identified a critical connection between RNA structure and RNA modification, a newly appreciated and pervasive mode of post-transcriptional control13. The most prevalent modification in mRNAs, m6A, occurs at GGmAGAUU motifs near stop codons, and acts in part to control RNA splicing and stability22,23. It has been hypothesized that m6A methylation occurs at sites that contain unpaired motifs24, but limited structural evidence in vivo has been presented to support this model. Comparison of icSHAPE signals at m6A-modified versus unmodified instances of the GGAGAUU motif in mouse ES cells25 revealed a specific structural signature, with stronger icSHAPE reactivity (consistent with unpaired RNA) at positions both surrounding and including the modified A (Fig. 4d and Extended Data Fig. 8). m6A sites in different subdomains of mRNAs or in lncRNAs have nearly identical icSHAPE profiles (Extended Data Fig. 10). Evaluation of all predictive features using our SVM algorithm showed that motif conservation or motif position offers some predictive value (AUC = 0.617 or 0.824, respectively) as previously reported25, but use of icSHAPE data (AUC = 0.846) or all features together (AUC = 0.914) improved prediction rate (Fig. 4e). These results show that icSHAPE structure profiles can be used accurately to predict post-transcriptional modifications on a transcriptome-wide scale.

The strong RNA structural signature at m6A sites may arise from the ability of m6A to destabilize RNA helices26 (depicted in Fig. 4c) or the structural selectivity of the m6A modification machinery for unpaired bases. In the former scenario, removal of m6A should cause increased base-pairing (loss of icSHAPE signal) whereas the latter scenario predicts little change to RNA structural profile. To distinguish between these hypotheses, we determined the icSHAPE profile of mouse ES cells genetically ablated for Mettl3 (ref. 25), a key m6A methyltransferase that is required for ES cell differentiation. We observed that in Mettl3-knockout cells, canonical motif sites that lost m6A modification also substantially lost icSHAPE signal transcriptome wide (Fig. 4d), as exemplified by key m6A target sites in Nanog mRNA (Fig. 4f). These results suggest that m6A impacts RNA structure, favouring the transition from paired to unpaired RNA. The ability to couple genetic perturbation with comprehensive, base-resolution structural maps in vivo is a potentially powerful approach to dissect regulators of RNA structure.

Understanding how RNA structures contribute to biological regulation opens the door to understanding a physical dimension of the transcriptome. icSHAPE bridges a gap in RNA-sequencing technologies that currently lack the ability to infer a mechanistic basis of biological function. The ability to view the structural dynamics of all four RNA bases in living cells is essential to uncover specific sequence motifs.
underlying different modes of post-transcriptional regulation, and has enabled the accurate identification and de novo prediction of trans-acting factor binding and chemical modification at single-nucleotide resolution. In the future, viewing the RNA structure when cells are exposed to different stimuli or genetic perturbations should revolutionize our understanding of gene regulation in biology and medicine.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 August 2014; accepted 26 January 2015.

Published online 18 March 2015.

1. Wan, Y., Kertesz, M., Spitale, R. C., Segal, E. & Chang, H. Y. Understanding the transcriptome through RNA structure. Nature Rev. Genet. 12, 641–655 (2011).
2. Rouskin, S., Zubratt, M., Washietl, S., Kelis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature 505, 701–705 (2014).
3. Ding, Y. et al. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature 505, 696–700 (2014).
4. Lucke, 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).
5. Spitale, R. C. et al. RNA SHAPE analysis in living cells. Nature Chem. Biol. 9, 18–20 (2013).
6. Anger, A. M. et al. Structures of the human and Drosophila 80S ribosome. Nature 497, 80–85 (2013).
7. Weeks, K. M. & Mauger, D. M. Exploring RNA structural codes with SHAPE chemistry. Acc. Chem. Res. 44, 1280–1291 (2011).
8. Schroeder, R., Grossberger, R., Pichler, A. & Waldsch, C. RNA folding in vivo. Curr. Opin. Struct. Biol. 12, 296–300 (2002).
9. Wan, Y. et al. Landscape and variation of RNA secondary structure across the human transcriptome. Nature 505, 706–709 (2014).
10. Wan, Y. et al. Genome-wide measurement of RNA folding energies. Mol. Cell 48, 169–181 (2012).
11. Kertesz, M. et al. Genome-wide measurement of RNA secondary structure in yeast. Nature 467, 103–107 (2010).
12. Vandivier, L. et al. Arabidopsis mRNA secondary structure correlates with protein function and domains. Plant Signal. Behav. 8, e24301 (2013).
13. Jia, G., Fu, Y. & He, C. Reversible RNA adenosine methylation in biological regulation. Trends Genet. 29, 108–115 (2013).
14. Jangi, M., Boutz, P. L., Paul, P. & Sharp, P. A. Rbfox2 controls autoregulation in RNA-binding protein networks. Nature 467, 637–651 (2014).
15. Dai, W., Zhang, G. & Makeyev, E. V. RNA-binding protein HuR autoregulates its expression by promoting alternative polyadenylation site usage. Proc. Natl Acad. Sci. USA 110, E2792–E2801 (2013).
16. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals complexity and dynamics of mammalian proteomes. Cell 147, 789–802 (2011).
17. Kozak, M. An analysis of 5’-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. 15, 8125–8148 (1987).
18. Dvir, S. et al. Deciphering the rules by which 5’-UTR structures affect protein expression in yeast. Proc. Natl Acad. Sci. USA 110, E2792–E2801 (2013).
19. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals complexity and dynamics of mammalian proteomes. Cell 147, 789–802 (2011).
20. Lunde, B. M., Moore, C. & Varani, G. RNA-binding proteins: modular design for efficient function. Nature Rev. Mol. Cell Biol. 8, 479–490 (2007).
21. Lovci, M. T. et al. Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. Nature Struct. Mol. Biol. 20, 1434–1442 (2013).
22. Auerwer, S. D. et al. Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. EMBO J. 25, 163–173 (2006).
23. Meyer, K. D. et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3’ UTRs and near stop codons. Cell 149, 1635–1646 (2012).
23. Wang, X. et al. N\textsuperscript{6}-methyladenosine-dependent regulation of messenger RNA stability. Nature 505, 117–120 (2014).

24. Schwartz, S. et al. High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. Cell 155, 1409–1421 (2013).

25. Batista, P. J. et al. m\textsuperscript{8}A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell 14, 707–719 (2014).

26. Kierzek, E. & Kierzek, R. The thermodynamic stability of RNA duplexes and hairpins containing N\textsuperscript{6}-alkyladenosines and 2-methylthio-N\textsuperscript{6}-alkyladenosines. Nucleic Acids Res. 31, 4472–4480 (2003).

27. McGinnis, J. L. & Weeks, K. M. Ribosome RNA assembly intermediates visualized in living cells. Biochemistry 53, 3237–3247 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank members of the Chang laboratory, J. Weissman and J. Doudna for comments. This work was supported by National Institutes of Health (NIH) R01HG004361 and P50HG007735, and the California Institute for Regenerative Medicine (H.Y.C.), NIH R01068122 (E.T.K.), the A.P. Giannini Foundation (R.C.S.), a Stanford Dean’s Fellowship (Q.C.Z.), NIH T32AR007422 (P.J.B.), and the Stanford Medical Scientist Training Program and NIH F30CA189514 (R.A.F.). H.Y.C. is an Early Career Scientist of the Howard Hughes Medical Institute.

Author Contributions R.C.S., E.T.K. and H.Y.C. conceived the study. R.C.S., P.C., J.-W.J., H.Y.K. and E.T.K. performed chemical design and synthesis. R.C.S., R.A.F., B.L., E.A.T. and P.J.B. performed biological experiments. Q.C.Z., R.C.S., R.A.F. and H.Y.C. performed data analysis. R.C.S. and H.Y.C. wrote the paper with input from all authors.

Author Information All genomic data sets have been deposited in the Gene Expression Omnibus under accession number GSE64169. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.Y.C. (howchang@stanford.edu) or E.T.K. (kool@stanford.edu).

©2015 Macmillan Publishers Limited. All rights reserved
In vitro transcription and acylation of RNA. RNA was transcribed from amplified inserts using T7 Megascript kit from Ambion, following the manufacturer’s protocol. In a typical in vitro modification protocol, RNA was heated in metal-free 1× PBS. RNA was added to the Hybond membrane and crosslinked using 254 nm ultraviolet light. The Hybond membrane was washed three times with 1× PBS. To the membrane was added NorthernLights Streptavidin NL493 (in PBS-Tween-20) for visualization. After incubation, the membrane was washed three times in 1× PBS-Tween-20. The membrane was dried and imaged by phosphorimaging (STORM, Molecular Dynamics).

Tissue culture and in vivo SHAPE modification. Mouse ES cells (v6.5 line) were grown on gelatinized dishes in serum and LIF. Unmodified total RNA was extracted by removing media, washing once in room temperature 1× PBS, and adding 2 ml (10 cm dish) or 7 ml (15 cm dish) of TRIzol directly to the cells. Subsequent RNA clean up was performed using the miRNeasy mini- or mid-column and protocol (Qiagen) as recommended by the manufacturer. In vivo modification of cellular RNAs was performed as described previously. Briefly, cells were rinsed once on the plate in room temperature 1× PBS, decanted, scraped in 1× PBS, and collected into a 15 ml tube. Cells were pelleted at room temperature and resuspended in 450 μl of 1× PBS. Fifty microlitres of 10× electrophile stock in DMSO (+) or DMSO (−) was added drop-wise, immediately mixed by inversion, and incubated at 37 °C on end-over-end rotation for 20 min. Reactions were pelleted for 1 min at 4°C at 10,000 r.p.m. and resuspended in 500 μl of 1× PBS. Samples were then transferred to 15 ml tubes with 2–7 ml of pre- aliquoted TRIZol and RNA was extracted as described earlier.

Methods to ensure titrated hit kinetics of RNA modification. We titrated NAI-N3 for single-hit kinetics that are comparable to those routinely used in chemical probing of RNA structure. For example, we obtained nearly identical secondary structure for 55 RNA as previously reported with a single-hit regime. After NAI-Mold modification and biotin pulldown, we retrieved approximately 10–20% of the input RNA as modified RNA, consistent with the expected Poisson distribution of single-hit modification.

icSHAPE deep-sequencing library preparation. RNA preparation. DMSO (mock) or NAI-N3 (experimental) modified total RNA was used as input for the deep-sequencing library preparation. Before library preparation, input RNA should be modified (or mock-treated) under in vitro or in vivo conditions as described earlier. For ‘total RNA’ libraries, no additional processing was needed. For ‘poly-A selected’ samples, 200 μg of total RNA was used per poly-A purist column (Ambion), which should yield ~2 μg of enriched RNA. Poly-A selection was performed a total of 2 times using the same poly-A beads (double poly-A selection). The NAI-N3 sample may have lower yields after purification so additional starting material could be required.

NAI-N3 biotinylation and RNA fragmentation. All RNA samples (NAI-N3 and DMSO treated) are processed through a copper-free ‘click’ reaction. RNA is brought to 97 μl in 1× PBS and 1 μl of SUPERaseIn and 2 μl of 185 mM DMSO-biotin are added. Samples were mixed by brief vortexing and then incubated at 37 °C for 2 h in a Thermomixer (Eppendorf). Reactions were stopped by adding 350 μl of Buffer RLE (Qiagen) and then 900 μl of 100% ethanol (EtOH). Each RNA sample was processed by passing over a RNeasy Mini column (Qiagen), two 500 μl washes with Buffer RPE (Qiagen), two 500 μl washes with Buffer RPE (Qiagen), and two 500 μl washes with Buffer RPE (Qiagen), and finally two 500 μl washes with Buffer RPE (Qiagen). The RNA was eluted in RNase-free water (final 100 μl). Samples were then frozen for 5 min on dry ice and concentrated to 9 μl using a lyophilizer (Labconco). Concentrated RNA samples (9 μl) were then moved to 0.5 ml PCR tubes for fragmentation. Samples were heated to 95 °C for 90 s and then 1 μl of 10× RNA Fragmentation Reagent

©2015 Macmillan Publishers Limited. All rights reserved
and hold at 25°C denaturing conditions.

Samples are frozen for 1 h on dry ice, spun at maximum speed. RNA samples add 100% EtOH. Mix samples by flicking and incubate at 16°C overnight.

**Ligation.** Note that NAI-N3 samples must use 3'-Adaptor-3'dde (5'/App/AGAT CGGAAGACGGGTTAGGG/G3/ddc) while DMSO samples use 3'-Adaptor-3' Biotin (5'/App/AGATCGGAAGACGGGTTAGGG/G3/Bio). The 'click' chemistry will label only the NAI-N3-modified RNAs in the NAI-N3 pool of transcripts with a biotin moiety, thus allowing the selective purification of structurally informative molecules. The DMSO samples are not capable of 'click' chemistry and every molecule in this pool is desired for sequencing so addition of a biotin moiety must happen in an unbiased fashion. Thus, DMSO samples have a 3'-biotin modification added specifically to their 3' Adaptor to allow for downstream processing in parallel of the DMSO and NAI-N3 samples.

After the overnight ligation, 30 µl of water, 185 µl of Buffer RLT and 400 µl of 100% EtOH is added to each sample and purified using RNeasy Mini columns as described earlier. Samples are concentrated to 5 µl. RNA end repair, RNA ligation, and RNA size selection.

**cDNA size selection.** cDNAs are selected for insert sizes of 20–120 nucleotides (−85–205 nucleotides with RT primer extension) and, depending on the input material amount, the libraries may be invisible at this step. Gel slices are crushed as above, 300 µl of Crush Soak Buffer is added and cDNAs are eluted at 50°C on rotation.

**cDNA circularization, library qPCR, library size selection and sequencing PCR.** Purification of eluted cDNA is performed as described earlier for RNA elution. After cDNA precipitation, samples are resuspended in 16 µl of water, 2 µl of 10X CircLigaseII Buffer, 1 µl of CircLigaseII (Epicycle) and moved to 0.5 ml PCR tubes. cDNA circularization takes place at 60°C for 120 min in a PCR machine. Circularized cDNA is purified by adding 200 µl of Buffer PNI and processing as described earlier using MiniElute columns, eluting the cDNA twice in 14 µl (final 27µl). Samples are initially amplified in a 60 µl qPCR reaction (27 µl cDNA, 30 µl 2X Phusion HF Master Mix, 0.75 µl of 10 µM P5_shorter primer (CTGACGGCTTTCTCGGATCT), 0.75 µl of 10 µM P5_shorter primer (ACACAGGAGCTCTCTCATCT), 0.75 µl of 2X SuperGold). The qPCR machine is programmed as follows: 98°C for 1 min, 98°C for 15 s, 62°C for 30 s, 72°C for 45 s. After qPCR amplification, samples are purified with 600 µl of Buffer PNI and MiniElute columns as described earlier. Library DNA is eluted twice in 15 µl (total 30 µl) and concentrated using a lyophilizer to less than 5 µl. A second 6% TBE 7 m urea PAGE gel selection is performed as described earlier to remove any PCR dimer products and all short qPCR primers. Gel slices are crushed as described earlier and eluted overnight at 50°C on rotation. Purification of library DNA is performed as described earlier, post-PAGE gel elution and after precipitation, resuspended in 19 µl of water.

A final library PCR amplification is performed for three cycles in 10 µl reactions (19 µl DNA library, 0.5 µl of 10 µM P3_solaex primer (CAAGCGAAGAGCATAACGATCTCAGCTTCGGATCT), 0.75 µl of 10 µM P5_solaex primer (AATGATACGGGACACAGCAGATCTACGACTCTTCCGA AACGCAGTCTTCTCGGGAGA), 0.75 µl of 2X SuperGold). The qPCR machine is programmed as follows: 98°C for 1 min, 98°C for 15 s, 62°C for 30 s, 72°C for 45 s. After qPCR amplification, samples are purified with 600 µl of Buffer PNI and MiniElute columns as described earlier. Library DNA is eluted twice in 15 µl (total 30 µl) and concentrated using a lyophilizer to less than 5 µl. A second 6% TBE 7 m urea PAGE gel selection is performed as described earlier to remove any PCR dimer products and all short qPCR primers.

**iCLIP and data analysis.** The iCLIP method was performed as described before with the specific modifications below. v6.5 mouse ES cells were grown as described earlier and UV-C crosslinked to a total of 0.3 cm². Whole-cell lysates were generated in CLIP lysis buffer (50 mM HEPES, 200 mM NaCl, 1 mM EDTA, 10% glycerol, 0.1% NP-40, 0.2% Triton X-100, 0.5% Na-laurylsarcosine) and briefly sonicated using a probe-tip Branson sonicator to solubilize chromatin. Each iCLIP experiment was normalized for total protein amount, typically 2 mg, and partially digested with RNaseA (Affymetrix) for 10 min at 37°C and quenched on ice. Immunoprecipitations of HuR were carried out with Proteín G Dynabeads (Life Technologies) and anti-HuR antibody (3A2, Santa Cruz) for 3 h at 4°C on rotation. Samples were washed sequentially in 1 ml for 5 min each at 4°C: 2X high stringency buffer (15 mM Tris-HCl pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% Triton X-100, 1% Na-deoxycholate, 120 mM NaCl, 25 mM KCl), 1X high salt buffer (15 mM Tris-HCl pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% Triton X-100, 1% Na-deoxycholate), 1X NT2 buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM MgCl2, 0.5% NP-40). 3' End RNA dephosphorylation, 5' end ssRNA ligation, 5' label- lour AMPPNP X beads (Beckman) according to the manufacturer's protocol and finally we eluted the library in 20 ml of water. Final library material was quantified on the BioAnalyzer High Sensitivity DNA chip (Agilent) and then sent for deep sequencing on the Illumina HiSeq2500 machine for 1X 100 bp cycle run. iCLIP data analysis was performed as described previously. Final library material was quantified on the BioAnalyzer High Sensitivity DNA chip (Agilent) and then sent for deep sequencing on the Illumina HiSeq2500 machine for 1X 75 bp cycle run. iCLIP data analysis was performed as described previously.

**RNA structure analysis.** Sequencing, reads mapping and data quality control. We generated four replicates for each library (DMSO PolyA, NAI PolyA in vivo and in vitro, DMSO PolyT and DMSO PolyC) sequencing on Illumina’s HiSeq sequencer and obtained approximately 200 million to 600 million raw reads for each replicate, totaling 3.9 billion reads. We collapsed these reads to remove PCR duplicates (only reads that have identical sequences including barcode region are regarded as duplicates). Collapsed reads were then subjected to barcode removal and primer and linker trimming by using Trimmomatic. We mapped trimmed reads to the mouse transcriptome of the Ensembl annotation (build GRCm38.p7) using Bowtie2 (ref. 34). For reads that can be mapped to multiple locations of the transcriptome, we evenly distribute them to up to ten random hits. Finally, we obtained 2.1 billion mapped reads in total. We define the ‘1 positions’ of each sequencing read as the location to which the nucleotide is mapped. This corresponds to the reverse transcription stop, which corresponds to the position adjacent to modified nucleotides in the NAI-N3 libraries, and intrinsic modified (or fragmentation) positions in the DMSO libraries. We defined reverse transcription stop coverage as the number of times a base is mapped as a reverse transcription stop.

©2015 Macmillan Publishers Limited. All rights reserved.
We calculated the expression level of all transcripts in the mouse transcriptome in terms of reads per kilobase per million mapped reads (RPKM). The correlations of transcript expression value (RPKM > 0.1) in different replicates are very high (in the range of 0.96 to 1.00). We constructed the background base density profile for each transcript as the sequencing depth of each base in the DMSO libraries. We also calculated the correlation of reverse transcription stops for each transcript in different replicates. As shown in Extended Data Fig. 5, the correlation is high for most transcripts if we limit the analysis to transcripts of average reverse transcription stop coverage higher than 2 and regions of background base density higher than 200. So for each library (DMSO poly-A, NA poly-A in vivo and in vitro) we combine all four replicates into one for the following analyses.

**Reactivity score calculation and construction of structural profile.** We performed a 5–5% normalization for each transcript; that is, the mean of the reverse transcription stops of the second top 5% bases, excluding the 32 bases at the beginning and 32 bases at the end of the transcript, will be normalized to 1, and all reverse transcription stops will be normalized proportionally.

**We defined reactivity score (R)** as the subtraction of background reverse transcription stops (DMSO libraries) from reverse transcription stops of the modified NAi-N, libraries, and then adjusted by the background base density:

\[
R = (R_{\text{TRP}^{\text{NAi-N}}} - R_{\text{TRP}^{\text{DMSO}}}) / \text{background base density}_{\text{DMSO}}
\]

The score is then 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). We trained the parameter \( \alpha \) on the ribosomal RNA structures, and set it to 0.25 to maximize the correlation of reactively score \( R \) determined by deep sequencing and reactivity score measured in low-throughput gel shift experiments.

**For each transcript, we defined its structural profile as the vector of base-resolution reactivity scores from the beginning to the end.** The valid structural profile of a transcript is limited to regions of reverse transcription stop coverage higher than 2 and background base density higher than 200. Finally, we obtained valid structural profiles for, respectively, 19,347 and 13,281 transcripts from in vivo and in vitro poly-A-selected RNA libraries, among which the majority are mRNAs (Extended Data Fig. 6).

**Metagene analysis of translation, pause, m^6A and protein-binding sites.** We calculated metagene structure profile around different functional sites by averaging all valid reactivity score \( R \) in the range from 0 to 10. We defined a set of genomic features for the prediction of \( m^6A \) sites and protein-binding sites, using structural profiles, genomic locations, conservations and their combinations.

The structural profile is limited to the range from the −10 to the +10 position of the \( m^6A \) site or the motif of the protein-binding sites. We used in vivo and in vitro reactivity scores separately and jointly in making predictions. We also retrieved a set of genomic features for the prediction of \( m^6A \) sites and protein-binding sites, including whether the site is in the 5’ UTR, CDS or 3’ UTR, whether it is at the last exon, whether it is at the largest exon, the distance to stop codon, the distance to stop codon, the distance to 5’ of the splicing junction, and so on. In addition, we retrieved the UCSC 60-way phastCons conservation score for \( m^6A \) sites in the range from the −10 to the +10 position of the \( m^6A \) site or the motifs of the protein-binding sites.

We used the same set of positive and negative controls and the best predictor is selected by using a parameter-searching tool that is included with the LIBSVVM package (http://www.csie.ntu.edu.tw/~cjlin/libsvm/). We used a fivefold cross-validation and calculated the AUC of the ROC curve to evaluate the performance of the predictors (Extended Data Fig. 8).

**Source code.** Source code used for the icSHAPE analysis is freely available at https://github.com/qczhang/icSHAPE.

28. Das, R., Laederach, A., Pearman, S. M., Herschlag, D. & Altman, R. B. SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. RNA 11, 344–354 (2005).

29. Gherghie, C. et al. Definition of a high-affinity Gag recognition structure mediating packaging of a retroviral RNA genome. Proc Natl Acad Sci USA 107, 19248–19253 (2010).

30. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415 (2003).

31. Flynn, R. A. et al. Dissecting noncoding and pathogen RNA-protein interactomes. RNA 21, 135–143 (2015).

32. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014).

33. Flicek, P. et al. Ensembl 2014. Nucleic Acids Res. 42, D749–D755 (2014).

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

35. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 42, D68–D73 (2014).

36. Cortes, C. & Vapnik, V. Support–vector networks. Mach. Learn. 20, 273–297 (1995).

37. Siepel, A. et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Nature Genet. 35, 1034–1050 (2005).
Extended Data Figure 1 | Chemical synthesis of NAI-N$_3$. 

a, Synthetic scheme for NAI-N$_3$. b, $^1$HMR of methyl 2-(azidomethyl)nicotinate. c, $^1$HNMR of 2-(azidomethyl)nicotinic acid. d, $^{13}$CNR of 2-(azidomethyl)nicotinic acid. 

e, $^1$HNMR of 2-(azidomethyl)nicotinic acid acyl imidazole. f, $^{13}$CNMR of 2-(azidomethyl)nicotinic acid acyl imidazole.
Extended Data Figure 2 | NAI-N₃ is a novel RNA acylation reagent that enables RNA purification.  a, Chemical schematic of RNA acylation and copper-free ‘click’ chemistry using NAI-N₃ and dibenzocyclooctyne (DIBO)-biotin conjugate.  b, ATP acylation gel shift showing ATP acylation and copper-free ‘click’ chemistry using NAI-N₃ and DIBO-biotin conjugate.
**Extended Data Figure 3** | NAI-N₃ is a novel RNA acylation reagent that accurately reads out RNA structure. 

- **a**, Comparative denaturing gel of NAI and NAI-N₃ RNA acylation. 
- **b**, Denaturing gel analysis of cDNAs that originate from the biotin-purification protocol (Extended Data Fig. 1). 
- **c**, Secondary structure of the SAM-I Riboswitch with enriched residues highlighted in orange and depleted residues highlighted in blue. 
- **d**, Denaturing gel analysis of denatured RNA probed with NAI-N₃ shows even coverage of 2'-hydroxyl reactivity when RNA is unfolded. 
- **e**, Protein titration with bovine serum albumin (BSA), demonstrating no difference in the SHAPE pattern as a function of protein concentration.
Extended Data Figure 4 | icSHAPE is capable of reproducing RNA acylation profiles obtained by manual RNA modification experiments. icSHAPE profiles (right) of rRNA, and compared to those obtained by manual SHAPE (left).
Extended Data Figure 5 | Reverse transcription stops measured by icSHAPE are very well correlated in different library replicates.
Extended Data Figure 6 | icSHAPE is capable of measuring the RNA structure profiles of thousands of RNAs simultaneously.  

a, The RNAs represented in polyA-selected RNA, in vivo.  
b, The RNAs represented in polyA-selected RNA, in vitro.
Extended Data Figure 7 | Non-AUG start codons are associated with preceding reactivity, and non-AUG start codons have a different profile, suggesting that RNA accessibility alone is not sufficient to drive translation.

a, icSHAPE profile at AUG start codons, in vivo. b, icSHAPE profile at AUG start codons, in vitro. c, icSHAPE profile at CUG start codons, in vivo. d, icSHAPE profile at CUG start codons, in vitro.
Extended Data Figure 8 | icSHAPE can be used to predict post-transcriptional regulatory elements. a, icSHAPE profile at Rbfox2 targets, \textit{in vivo}. b, icSHAPE profile at Rbfox2 targets, \textit{in vitro}. c, ROC curve of Rbfox2 RNA–protein interactions, predicted using icSHAPE profiles. d, icSHAPE profile at m$^6$A targets, \textit{in vivo}. The negative control is the set of motif instances that are not m$^6$A modified. e, icSHAPE profile at m$^6$A targets, \textit{in vitro}. f, ROC curve of m$^6$A RNA modification sites, predicted using icSHAPE profiles. g, icSHAPE profile at HuR targets, \textit{in vivo}. h, icSHAPE profile at HuR targets, \textit{in vitro}. i, ROC curve of HuR RNA–protein interactions, predicted using icSHAPE profiles.
Extended Data Figure 9 | iCLIP analysis of HuR in mouse ES cells. a, Global binding preference of the RBP HuR in mouse ES cells as represented by reverse transcription (RT) stops across the mouse transcriptome (mm9). HuR mainly binds protein-coding, processed and ribosomal RNAs. b, Number of unique RNA transcripts bound by HuR. c, HuR reverse transcription stops distributed across protein-coding transcript functional domains. HuR prefers intronic and 3' UTR regions. d, Metagene analysis of all HuR binding sites. Each mRNA region (5' UTR, CDS or 3' UTR) was scaled to a standard width and reverse transcription stop density across all bound protein-coding genes and was plotted, revealing a clear enrichment for 3' UTR regions in mature protein-coding transcripts. e, Individual mRNA binding events of HuR to genes important for mouse ES cell biology, including Tet1, β-actin, Elav1 (encoding HuR itself) and Lin28a. Discrete binding sites are observed focused in 3' UTR and intronic regions.
Extended Data Figure 10 | m^6A-associated RNA structure features are preserved, independent of their position along the RNA transcript.
Erratum: Structural imprints in vivo decode RNA regulatory mechanisms

Robert C. Spitale, Ryan A. Flynn, Qiangfeng Cliff Zhang, Pete Crisalli, Byron Lee, Jong-Wha Jung, Hannes Y. Kuchelmeister, Pedro J. Batista, Eduardo A. Torre, Eric T. Kool & Howard Y. Chang

Nature 519, 486–490 (2015); doi:10.1038/nature14263

In Fig. 2a of this Letter, an error in the placement of the labels on the pie chart was introduced during the production process. The correct numbers were reported, but the size of the pie sectors was incorrect. In addition, on page 488, we made an incorrect attribution of the acronym 'Fox'. The Fox family of RBP is named after the gene 'feminizing locus on X' (PMID: 7821230), rather than 'forkhead box'. We apologize for this oversight. These corrections do not affect the conclusions of the paper, and both errors have now been corrected in the online versions of the paper.