LETTER

doi:10.1038/nature14263

Structural imprints *in vivo* decode RNA regulatory mechanisms

Robert C. Spitale[1]*, Ryan A. Flynn[1]*, Qiangfeng Cliff Zhang[1]*, Pete Crisalli[2], Byron Lee[1], Jong-Wha Jung[2], Hannes Y. Kuchelmeister[2], Pedro J. Batista[1], Eduardo A. Torre[1], Eric T. Kool[2] & Howard Y. Chang[1]

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[1]. 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[2,3]. 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 N^6-methyladenosine (m^6A) 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-positive rates due to spurious reverse transcription stops[4]. 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* RNA structure profiling genome wide (NAI-N_3; Fig. 1a, b and Extended Data Fig. 1). NAI-N_3 adds an azide group to NAI (2-methylnicotinic acid imidazolide), a cell-permeable SHAPE reagent[5]. By using copper-free click chemistry, a biotin moiety is selectively and efficiently added to NAI-N_3-modified RNA, providing a stringent purification handle with streptavidin-conjugated beads (Fig. 1c and Extended Data Fig. 2). NAI-N_3 generated identical profiles of reverse transcription stops to those obtained using our previously designed SHAPE reagent[5]. The fidelity of structural measurements was not affected by 'clicking' biotin onto NAI-N_3, or by molecular crowding of proteins, and NAI-N_3 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[6] 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

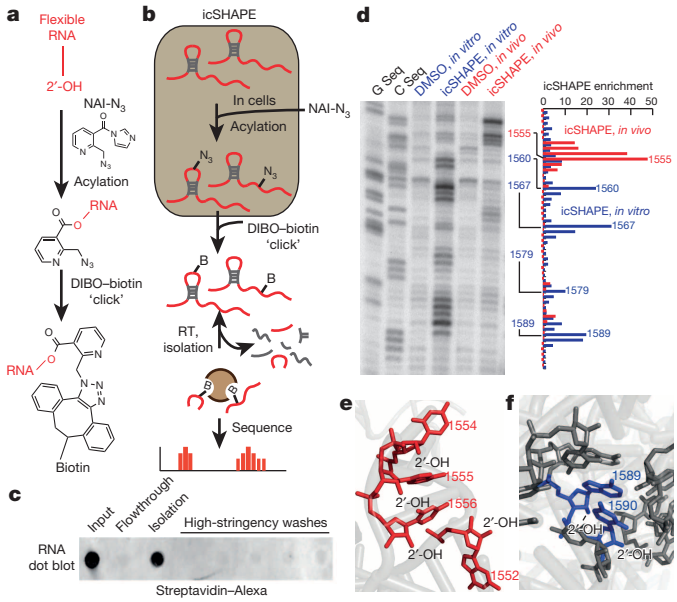


Figure 1 | icSHAPE is a novel and robust method for measuring RNA structure. **a**, Chemical scheme for the preparation of acylated RNA, which can be purified by biotin–streptavidin purification. DIBO, dibenzocyclooxtyne. **b**, Schematic of icSHAPE modification and purification steps to generate a sequencing library. RT, reverse transcription. **c**, Dot blot of biotin-modified RNA from icSHAPE through streptavidin affinity isolation. **d**, Denaturing gel electrophoresis of icSHAPE on the 18S rRNA from mouse ES cells. The corresponding icSHAPE profile, generated from deep sequencing and annotated for nucleotide position, is to the right. G Seq and C Seq indicate reverse transcription with dideoxyguanosine triphosphate or dideoxycytidine triphosphate, respectively. DMSO, dimethylsulfoxide. **e**, PyMol representation of rRNA, corresponding to regions of icSHAPE that are more reactive *in vivo* (Protein Data Bank (PDB) accession 3J3D). **f**, PyMol representation of rRNA, corresponding to regions of icSHAPE that are more reactive *in vitro*.

[1]Howard Hughes Medical Institute and Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California 94305, USA. [2]Department of Chemistry, Stanford University, Stanford, California 94305, USA.
*These authors contributed equally to this work.

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 vivo* and *in vitro*, 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 NAI-N_3 samples for As and Us (Fig. 2a). This enrichment is expected given their bias for being located in single-stranded or loop regions[7]. 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 environment[8]. Recent transcriptome-wide dimethylsulfate probing (DMS-seq), which interrogates two bases with strong bias towards adenosines (68% As and 24% Cs)[2,4], suggested that RNA structures are largely unfolded *in vivo*[2]; 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 *in vivo* and *in vitro* icSHAPE measurements, termed the '*vivo–vitro* difference' (VTD; Fig. 2b and Methods). Adenosine residues have the largest VTD, whereas guanosine and cytidine residues are less variable between environments (Fig. 2c). These observations suggest that using probes that have a broader reactivity profile, such as NAI-N_3, will give a more complete representation of RNA structure.

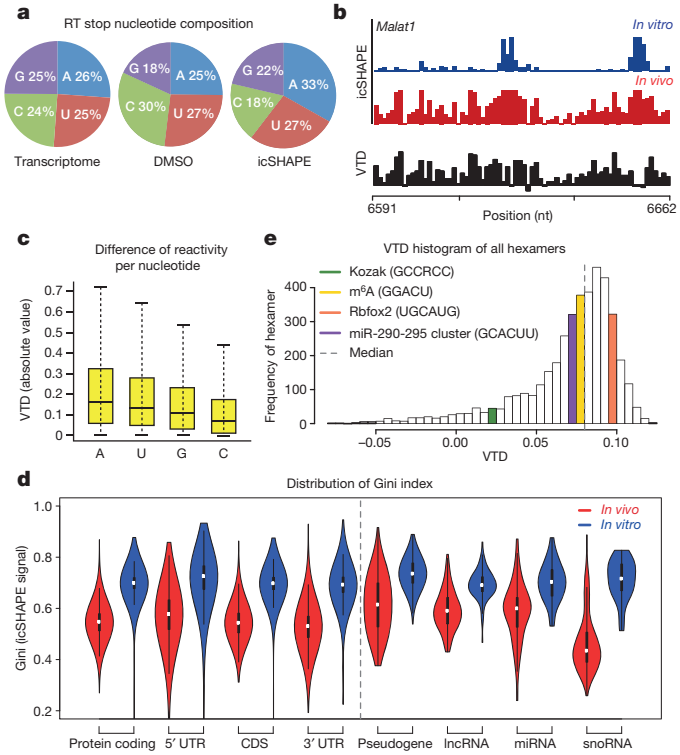


Figure 2 | icSHAPE reveals unique structural profiles for nucleobase reactivity and post-transcriptional interactions. **a**, Reverse transcription (RT) stop distribution for the transcriptome, DMSO control or icSHAPE libraries. **b**, icSHAPE data track and VTD calculation of the *Malat1* RNA (chromosome 19: 5,796,010–5,796,081). icSHAPE data are scaled from 0 (no reactivity) to 1 (maximum reactivity). nt, nucleotide. **c**, The VTD distribution for icSHAPE libraries. **d**, Gini index of icSHAPE data *in vivo* versus *in vitro*. **e**, The distribution of VTD profiles for all hexamer motifs across the transcriptome. Bin locations for several motifs for post-transcriptional regulation are highlighted.

Second, we used the Gini index[2] 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 report[2], 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* ($P < 2.2 \times 10^{-16}$, noncoding versus coding, Student's *t*-test). One exception to this rule are small nucleolar RNAs (snoRNAs), which exhibit the greatest level of increased reactivity *in vivo* among all classes of transcripts and may result from extensive rearrangements due to small nucleolar ribonucleoprotein (snoRNP) binding. Thus, most RNAs *in vivo* possess a substantial level of RNA structure beyond previous expectations based on DMS-seq[2]. Our data further suggest that RNA structural signatures *in vivo* can distinguish coding versus structural or regulatory RNAs, consistent with previous *in vitro* studies[9–12].

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, Rbfox2), RNA modification (m^6A) and miRNA seed matches[13–15] (Supplementary Tables 1 and 2). These results show that the VTD may classify RNA regulatory elements as pre-programmed 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 initiation[16], 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 output[10,17]. We used translation initiation[18] and pause sites[18], 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-site 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 kinetics[18]. 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

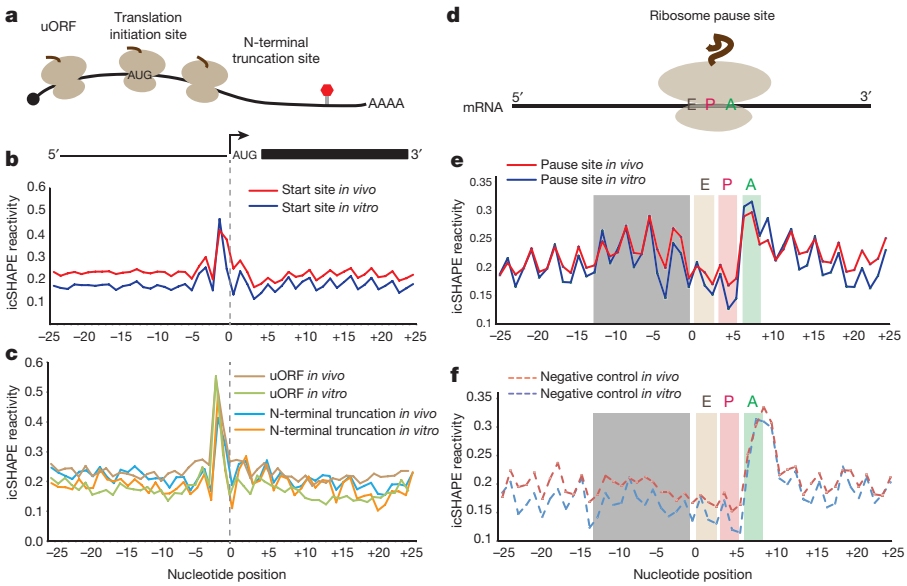


Figure 3 | icSHAPE reveals structural profiles associated with translation. **a**, Cartoon representation of ribosomes translating an mRNA. The uORF initiation site is represented by ribosome initiation upstream of the canonical start codon. The canonical start position is demarcated by 'AUG'. The N-terminal truncation is represented as a ribosome initiating to the 3′-end of the canonical start AUG. **b**, icSHAPE profile at canonical start codon position. **c**, icSHAPE profiles at uORF and N-terminal truncation sites. **d**, Cartoon representation of a paused ribosome and its corresponding A, P and E sites. **e**, The icSHAPE profile at ribosome pause sites. **f**, icSHAPE profile at negative control sites for pause sequences. Grey box highlights a region of structural difference upstream of true pause sites versus controls.

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 stability[19] (Fig. 2d). The feminizing locus on X (Fox) family of RBPs are important for tissue-specific control of alternative splicing, with Rbfox2 having key roles in ES cells[14,20]. High VTD at the known Rbfox2-binding motif (UGCAUG, Fig. 2e)[14,20] indicates a strong structural rearrangement *in vivo*. Alignment with the Rbfox–RNA NMR structure[21] and Rbfox2-binding sites identified by individual nucleotide crosslinking immunoprecipitation (iCLIP) in mouse ES cells[20] 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 interaction[21], 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 RNA 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 stability[15], 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 control[13]. The most prevalent modification in mRNAs, m^6A, occurs at GGm^6ACU motifs near stop codons, and acts in part to control RNA splicing and stability[22,23]. It has been hypothesized that m^6A methylation occurs at sites that contain unpaired motifs[24], but limited structural evidence *in vivo* has been presented to support this model. Comparison of icSHAPE signals at m^6A-modified versus unmodified instances of the GGACU motif in mouse ES cells[25] 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). m^6A 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 reported[24], 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 m^6A sites may arise from the ability of m^6A to destabilize RNA helices[26] (depicted in Fig. 4c) or the structural selectivity of the m^6A modification machinery for unpaired bases. In the former scenario, removal of m^6A 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 m^6A methyltransferase that is required for ES cell differentiation. We observed that in *Mettl3*-knockout cells, canonical motif sites that lost m^6A modification also substantially lost icSHAPE signal transcriptome wide (Fig. 4d), as exemplified by key m^6A target sites in *Nanog* mRNA (Fig. 4f). These results suggest that m^6A 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

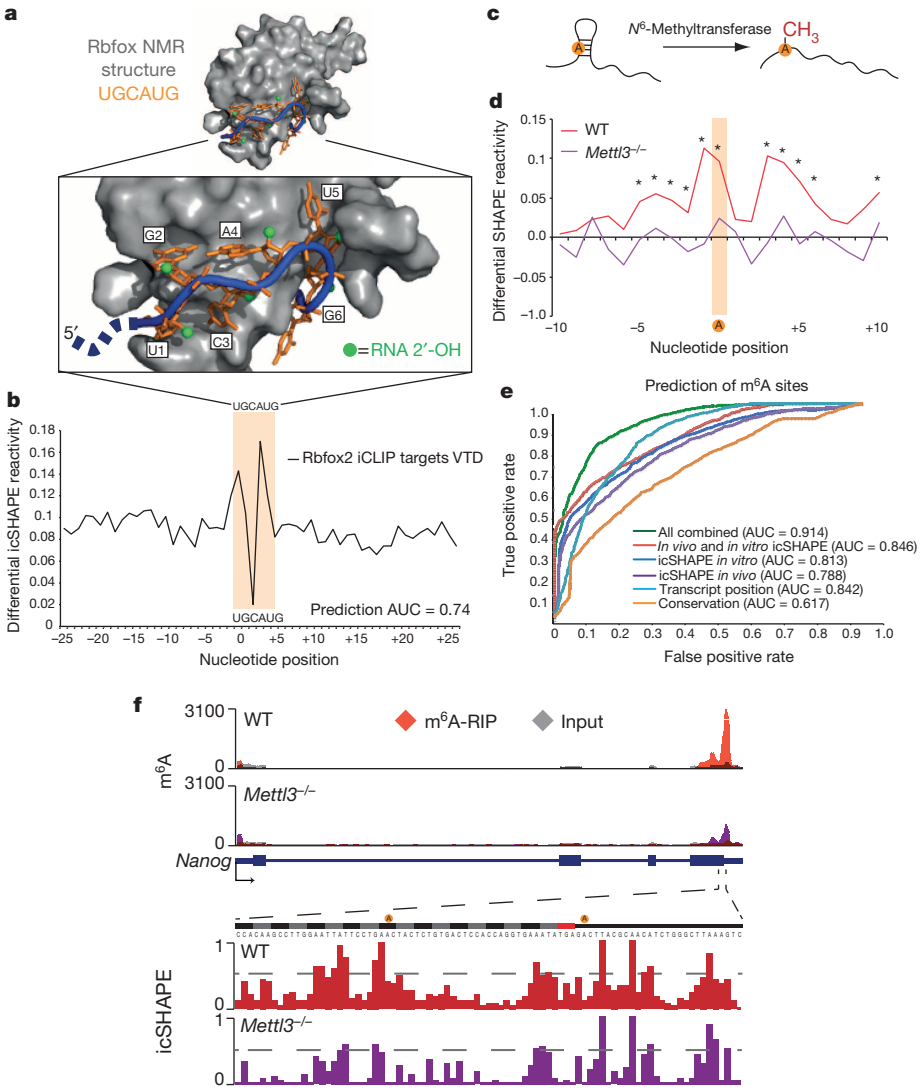


Figure 4 | icSHAPE dynamics reveal and predict post-transcriptional interactions. **a**, Structure of Rbfox1–RNA interaction, highlighting the RNA–protein interface. The RNA is shown with a blue backbone and orange bases; each 2′-hydroxyl is green (PDB accession 2ERR). **b**, The differential icSHAPE profile at Rbfox2 target mRNAs measured *in vivo* versus *in vitro* maps precisely to the Rbfox-binding sites. **c**, Model of interplay between m^{6}A and RNA structure. **d**, Differential icSHAPE signal for m^{6}A methylated versus non-methylated sites with the same underlying sequence motif, both *in vivo*. icSHAPE signals from unmodified sites are subtracted from m^{6}A-modified sites. $*P < 0.05$, false discovery rate (FDR) < 0.05, positions with significant differences. Data from wild-type (WT) and *Mettl3*-knockout mouse ES cells are plotted for comparison. **e**, Receiver operating characteristic (ROC) curve for prediction of m^{6}A sites, incorporating icSHAPE profiles. **f**, Effect of m^{6}A on RNA structure of *Nanog* mRNA. Top, location of *Mettl3*-dependent m^{6}A sites (highlight in yellow); m^{6}A-RNA immunoprecipitation (m^{6}A-RIP) data from ref. 25. Bottom, icSHAPE profile of wild-type and *Mettl3*-knockout cells.

underlying different modes of post-transcriptional regulation[27], 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 structurome 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., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures *in vivo*. *Nature* **505,** 701–705 (2014).
3. Ding, Y. *et al. In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505,** 696–700 (2014).
4. 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).
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. & Waldsich, 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. *Genes Dev.* **28,** 637–651 (2014).
15. Dai, W., Zhang, G. & Makeyev, E. V. RNA-binding protein HuR autoregulates its expression by promoting alternative polyadenylation site usage. *Nucleic Acids Res.* **40,** 787–800 (2012).
16. Kozak, M. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15,** 8125–8148 (1987).
17. Dvir, S. *et al.* Deciphering the rules by which 5′-UTR sequences affect protein expression in yeast. *Proc. Natl Acad. Sci. USA* **110,** E2792–E2801 (2013).
18. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147,** 789–802 (2011).
19. Lunde, B. M., Moore, C. & Varani, G. RNA-binding proteins: modular design for efficient function. *Nature Rev. Mol. Cell Biol.* **8,** 479–490 (2007).
20. Lovci, M. T. *et al.* Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nature Struct. Mol. Biol.* **20,** 1434–1442 (2013).
21. Auweter, S. D. *et al.* Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J.* **25,** 163–173 (2006).
22. 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^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^6A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **15,** 707–719 (2014).
26. Kierzek, E. & Kierzek, R. The thermodynamic stability of RNA duplexes and hairpins containing N^6-alkyladenosines and 2-methylthio-N^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).

METHODS

Methyl 2-(azidomethyl)nicotinate. Methyl 2-methylnicotinate (1.00 g) was dissolved in 5 ml anhydrous dichloromethane. Trichloroisocyanuric acid (2.30 g) was added and the resulting suspension stirred overnight at room temperature. The reaction was diluted with dichloromethane and quenched by the addition of saturated sodium bicarbonate solution. The phases were separated and the organic phase was washed once with brine, dried over magnesium sulfate, filtered, and concentrated to afford a yellow oil. NMR data were consistent with literature reports.

The crude product of the above reaction (1.09 g) was dissolved in 12 ml anhydrous N,N-dimethylformamide and 0.77 g sodium azide was added. The reaction was stirred overnight at room temperature then quenched with saturated sodium bicarbonate solution. The aqueous layer was extracted with ethyl acetate, and the organic layer washed three times with water and three times with brine. The organic layer was dried over magnesium sulfate, filtered, and concentrated to afford 0.91 g (71%, two steps) of a yellow oil that solidified upon standing.

NMR data were as follows: ^{1}H NMR (400 MHz, $CDCl_3$): 3.94 (3H, s), 4.88 (2H, S), 7.37 (1H, m), 8.29 (1H, dd, J = 8 Hz, 1.6 Hz), 8.76 (1H, dd, J = 4.6 Hz, 1.6 Hz); ^{13}C NMR (100 MHz, $CDCl_3$): 52.8, 54.4, 122.9, 125.0, 139.1, 152.4, 156.7, 166.0. Electrospray ionization mass spectrometry (ESI-MS) (Calc M-H = 191.06): 191.98.

2-(azidomethyl)nicotinic acid. Methyl 2-(azidomethyl)nicotinate (0.50 g) was stirred vigorously in 10 ml of 1:1 MeOH:10% aqueous NaOH. After 10 min thin layer chromatography (TLC) indicated complete consumption of starting material. Twenty-five millilitres of water were added, the crude reaction mixture was washed once with ether (10 ml), then acidified to pH 4 with 10% aqueous HCl and extracted five times with 50 ml ethyl acetate. The organic layer was dried over magnesium sulfate, filtered, and concentrated to afford 0.46 g (99%) of a white solid that was pure by NMR.

NMR data were as follows: ^{1}H NMR (400 MHz, DMSO-d_6): 4.81 (2H, s), 7.50 (1H, m), 8.28 (1H, dd, J = 7.8 Hz, 1.6 Hz), 8.74 (1H, dd, J = 5 Hz, 1.6 Hz), 13.64 (1H, br. s); ^{13}C NMR (100 MHz, DMSO-d_6): 53.4, 123.4, 125.9, 139.0, 151.9, 156.0, 167.1. ESI-MS (Calc M-H = 177.04): 177.05.

2-(azidomethyl)nicotinic acid acyl imidazole. 2-(Azidomethyl)nicotinic acid (0.15 g) was dissolved in 0.21 ml anhydrous dimethylsulfoxide. A solution of 0.14 g carbonyldiimidazole in 0.21 ml anhydrous dimethylsulfoxide was added drop-wise, creating rapid gas evolution. The reaction was allowed to proceed for 1 h and the resulting solution used as a 2 M stock solution for RNA SHAPE experiments. For NMR data collection, an analytical sample was prepared in dichloromethane as described earlier. The reaction was stirred overnight and the solvent removed in vacuum. The product was then isolated by flash column chromatography on silica (ethyl acetate).

NMR data were as follows: ^{1}H NMR (400 MHz, DMSO-d_6): 4.62 (2H, s), 7.16 (1H, dd, J = 1.7 Hz, 0.8 Hz), 7.60 (1H, dd, J = 7.9 Hz, 4.9 Hz), 7.66 (1H, m), 8.15 (2H, m), 8.84 (1H, dd, J = 4.9 Hz, 1.7 Hz); ^{13}C NMR (100 MHz, DMSO-d_6): 52.8, 117.9, 123.0, 127.3, 130.9, 137.7, 138.5, 151.9, 154.6, 164.7.

***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 water for 2 min at 95 °C. The RNA was then flash-cooled on ice. The RNA 3× SHAPE buffer (333 mM HEPES, pH 8.0, 20 mM $MgCl_2$, 333 mM NaCl) was added and the RNA was allowed to equilibrate at 37 °C for 10 min. To this mixture, 1 µl of 10× electrophile stock in DMSO (+) or DMSO (−) was added. The reaction was permitted to continue until the desired time. Reactions were cleaned up using RNeasy columns (Qiagen) following the manufacturer's protocol and eluted RNase-free water.

***In vitro* manual SHAPE analysis.** ^{32}P-end-labelled DNA primer (reverse primer above) was annealed to 3 µg of total RNA by incubating at 95 °C for 2 min followed by a step-down cooling (2° per s) to 4 °C. To the reaction first-strand buffer, dithiothreitol (DTT) and dNTPs were added. The reaction was pre-incubated at 52 °C for 1 min, then Superscript III (2 U µl^{-1} final concentration) was added. Extensions were performed for 10 min. To the reaction, 1 µ of 4 M NaOH was added and allowed to react for 5 min at 95 °C. Ten microlitres of Gel Loading Buffer II (GLBII, Ambion) was then added, and cDNA extensions were resolved on 8% denaturing (7 M urea) polyacrylamide gels (29:1 acrylamide:bisacrylamide, 1× TBE). All (−) lanes are those from DMSO control-treated cells. In addition, all sequencing lanes are from DMSO control-treated cells. cDNA extensions were visualized by phosphorimaging (STORM, Molecular Dynamics). cDNA bands were integrated with SAFA[28]. SHAPE reactivities were normalized to a scale spanning 0 to 1.5, where 1.0 is defined as the mean intensity of highly reactive nucleotides[29]. RNA secondary structures were predicted using mFOLD software[30].

Characterization of manual SHAPE-enriched reverse transcription stops. *Copper-free click chemistry of acylated RNA.* In a typical reaction, acylated RNA (1 pmol) was reacted with 100 equivalents of DIBO-biotin (Life Technologies) for 2 h, at 37 °C, in 1× PBS. Reactions were extracted once with acid phenol:chloroform (pH 4.5 ± 0.2) and twice with chloroform. RNA was precipitated with 40 µl of 3 M sodium acetate buffer (pH 5.2) and 1 µl of glycogen (20 µg µl^{-1}). Pellets were washed twice with 70% ethanol and resuspended in 10 µl RNase-free water.

Enrichment of NAI-N_3-modified RNA. The following protocol was used for manual enrichment protocols used to optimize capture conditions. To 1 pmol of precipitated and biotinylated RNA (in 900 µl of binding buffer: 50 mM Tris-HCl pH 7.0 and 1 mM EDTA) was added 50 µl (slurry) of DYNAL MyOneC1 beads (Life Technologies). The reaction mixture was then incubated for 1 h at room temperature. The beads were then collected on a magnetic plate and the solution decanted. The beads were then resuspended and washed four times with Biotin Wash Buffer (10 mM Tris-HCl, pH 7.0, 1 mM EDTA, 4 M NaCl, 0.2% Tween). The beads were then washed three times with RNase-free water. To elute the purified RNA, streptavidin-conjugated beads were incubated in 1× proteinase K buffer with 20 U of proteinase K (Life Technologies), 1 mM D-biotin (Sigma-Aldrich), and 20 U of SUPERaseIn (Life Technologies). The reaction was permitted to run for 30 min at 37 °C. Beads were then collected by magnet and the supernatant removed and set on ice. This was repeated twice more and elutions were pooled. Reactions were extracted once with acid phenol:chloroform (pH 4.5 ± 0.2) and twice with chloroform. RNA was precipitated with 40 µl of 3 M sodium acetate buffer (pH 5.2) and 1 µl of glycogen (20 µg µl^{-1}). Pellets were washed twice with 70% ethanol and resuspended in 10 µl RNase-free water.

Dot blot analysis of enriched NAI-N_3-modified RNA. Hybond N+ membranes (GE) were pre-incubated in 1× PBS. Precipitated RNA was dissolved in 100 µl of 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 midi-column and protocol (Qiagen) as recommended by the manufacturer. *In vivo* modification of cellular RNAs was performed as described previously[5]. 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-N_3 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 5S rRNA as previously reported with a single-hit regime[5]. After NAI-N_3 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-N_3 (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 two times using the same poly-dT beads ('double poly-A selection'). The NAI-N_3 sample may have lower yields after purification so additional starting material could be required.

NAI-N_3 biotinylation and RNA fragmentation. All RNA samples (NAI-N_3 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 DIBO-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 RLT (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), one no-buffer spin to dry the column, and finally two 50 µl elutions 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

(Ambion) was added and samples were placed back at 95 °C for 70–90 s. Reactions were quenched by adding 1 µl of RNA Fragmentation Stop Solution (Ambion) and moved to ice. RNA was cleaned up by adding 35 µl of Buffer RLT and 100 µl of 100% EtOH and purified using RNeasy Mini columns as described earlier. Samples are then concentrated with a lyophilizer to 5 µl.

RNA end repair, RNA ligation, and RNA size selection. To resolve the 3′-end phosphate generated by the fragmentation process, T4 PNK is used. To each 5 µl sample 2 µl of 5× PNK buffer (350 mM Tris-HCl pH 6.5, 50 mM $MgCl_2$, 25 mM DTT), 1 µl SUPERaseIn, and 2 µl of T4 PNK (NEB) is added, mixed by flicking, and incubated at 37 °C for 1 h. After end-repair samples are moved directly to 3′-end ligation by adding 1 µl of 50 µM 3′ Adaptor, 1 µl of 10× T4 RNL2tr buffer (NEB), 1.5 µl of T4 RNL2tr K227Q (NEB), 1 µl of 100 mM DTT, and 8 µl of 50% PEG8000. Mix samples by flicking and incubate at 16 °C overnight.

Ligation. Note that NAI-N_3 samples must use 3′-Adaptor-3′ddc (/5rApp/AGATCGGAAGAGCGGTTCAG/3ddC/) while DMSO samples must use 3′-Adaptor-3′Biotin (/5rApp/AGATCGGAAGAGCGGTTCAG/3Bio/). The 'click' chemistry will label only the NAI-N_3-modified RNAs in the NAI-N_3 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-N_3 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 using a lyophilizer and 5 µl of GLBII is added and stored on ice. To size select the RNA samples a mini 6% TBE PAGE gel with 7 M urea is cast and pre-run to 50 W for 8 min. Samples are loaded without prior heating and the PAGE gel is imaged using a 1:10,000 dilution of SybrGold (Life Technologies). RNA is visualized on a BlueBox (Clare Chemical) and fragmented RNA ranging between 20–120 nucleotides (40–140 nucleotides with the 3′ Adaptor ligated) are excised with a scalpel. Gel slices are crushed through a 0.75 ml tube nested in a 2 ml tube by centrifugation and 300 µl of Crush Soak Buffer (500 mM NaCl, 1 mM EDTA) is added with 3 µl of SUPERaseIn. RNA is eluted overnight at 4 °C on rotation.

Reverse transcription, streptavidin capture, cDNA elution and cDNA size selection. RNA samples are purified away from residual PAGE using 0.45 µm Spin-X columns (Corning) and the 300 µl elutions are transferred to siliconized 1.5 ml tubes (Fisher Scientific, used in all subsequent steps). RNA is precipitated by adding 30 µl of 3 M sodium acetate buffer (pH 5.2), 0.8 µl of GlycoBlue (Ambion) and 1 ml of 100% EtOH. Samples are frozen for 1 h on dry ice, spun at maximum speed (15,000 r.p.m.) for 1 h at 4 °C, washed with 800 µl of ice-cold 80% EtOH, decanted, air-dried and then resuspended in a 0.5 ml PCR tube with 11.5 µl of water. To the RNA samples add 1 µl of 10 µM RT primer (/5phos/DDDNNAACCNNNNAGATCGGAAGAGCGTCGTGAT/iSp18/GGATCC/iSp18/TACTGAACCGC, /5phos/ = 5′ phosphate, D = A/T/G, /iSp18/ = 18carbon PEG spacer) and 1 µl of 10 mM dNTPs. Heat the samples to 70 °C for 5 min and then cool slowly to 25 °C (2° per s) and hold at 25 °C for 1 min. After primer annealing add 0.5 µl of SUPERaseIn, 1 µl 100 mM DTT, 4 µl of 5× First Strand Buffer and 1 µl of SuperScript III (Life Technologies). cDNA extension occurs for 3 min at 25 °C, 7 min at 42 °C, and finally at 52 °C for 15 min. After cDNA extension do not raise samples above 37 °C to avoid denaturing conditions.

MyOneC1 streptavidin beads for cDNA capture and NAI-N_3-modified RNA enrichment are prepared (40 µl slurry per sample) by washing three times in 1 ml of Biotin Bind Buffer (100 mM Tris-HCl pH 7.0, 10 mM EDTA, 1 M NaCl) and resuspending the beads in 40 µl Biotin Bind Buffer and 1 µl SUPERaseIn per reaction. After the reverse transcription reaction completes, 40 µl of pre-washed beads are added to each sample, mixed by flicking, and incubated at room temperature for 45 min. After streptavidin capture, samples are washed at room temperature serially with four times 100 µl of Biotin Wash Buffer, two times 100 µl 1× PBS and finally moved to 1.5 ml tubes. cDNA is eluted by adding 1 µl RNaseA/T1 cocktail (Ambion), 1 µl RNaseH (Enzymatics), 12.5 µl 50 mM D-biotin, 5 µl 10× Elution Buffer (500 mM HEPES, 750 mM NaCl, 30 mM $MgCl_2$, 1.25% Sarkosyl, 0.25% Na-deoxycholate, 50 mM DTT), 30.5 µl water and incubating at 37 °C for 30 min in a Thermomixer at 800 r.p.m. Samples are mixed with 1 µl 100% DMSO, heated to 95 °C for 3 min, placed on a magnet, and the 50 µl cDNA elution moved to a new tube. The elution is repeated once (total of two times and final of 100 µl). cDNA is processed by adding 1 ml of Buffer PNI and purifying over a MiniElute columns (Qiagen), following the manufacturer's protocol, and eluting twice in 15 µl of Buffer EB (final 30 µl). cDNAs are concentrated using a lyophilizer to 5 µl and an equal volume of GLBII is added. Size selection of cDNAs is performed as was done for the RNA size selection. 6% PAGE gel pre-running is critical to achieve denaturing conditions as well as heating the samples to 95 °C for 3 min before PAGE separation. cDNAs are selected for insert sizes of ~20–120 nucleotides (~85–205 nucletodies 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 overnight 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 10× CircLigaseII Buffer, 1 µl of CircLigaseII (Epicentre) 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 ~27 µl). Samples are initially amplified in a 60 µl qPCR reaction (27 µl cDNA, 30 µl 2× Phusion HF Master Mix, 0.75 µl of 10 µM P3_short primer (CTGAACCGCTCTTCCGATCT), 0.75 µl of 10 µM P5_short primer (ACACGACGCTCTTCCGATCT), 0.72 µl of 25× SybrGold). 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 40 µl reactions (19 µl library DNA, 0.5 µl of 10 µM P3_solexa primer (CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT), 0.5 µl of 10 µM P5_solexa primer (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT)) and cleaned up using Agencourt AMPure XP beads (Beckman) according to the manufacturer's protocol and we eluted the library in 20 µl 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 1× 100 bp cycle run.

iCLIP and data analysis. The iCLIP method was performed as described before with the specific modifications below[31]. v6.5 mouse ES cells were grown as described earlier and UV-C crosslinked to a total of 0.3 J cm^{-2}. 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% N-lauroylsarcosine) 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 Protien 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: 2× 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), 1× high salt buffer (15 mM Tris-HCl pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% Triton X-100, 1% Na-deoxycholate, 1 M NaCl), 1× NT2 buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM $MgCl_2$, 0.05% NP-40). 3′-End RNA dephosphorylation, 3′-end ssRNA ligation, 5′ labelling, SDS–PAGE separation and transfer, autoradiograph, RNP isolation, Proteinase K treatment, and overnight RNA precipitation took place as previously described[31]. The 3′-ssRNA ligation adaptor was modified to contain a 3′-biotin moiety as a blocking agent. The iCLIP library preparation was performed as described previously[31]. 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 1× 75 bp cycle run. iCLIP data analysis was performed as previously described[31].

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*). We performed single-end sequencing on Illumina's HiSeq sequencer and obtained approximately 200 million to 600 million raw reads for each replicates, totalling 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[32]. We mapped trimmed reads to the mouse transcriptome of the Ensembl annotation (build GRCm38.74)[33] by 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 first 5′-end nucleotide mapped. This represents the reverse transcription stop, which corresponds to the position adjacent to modified nucleotides in the NAI-N_3 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.

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, NAI 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_3 libraries, and then adjusted by the background base density:

$$R = (\mathrm{RT\_stop}_{\mathrm{NAI\text{-}N3}} - \alpha \mathrm{RT\_stop}_{\mathrm{DMSO}})/\mathrm{background\_base\_density}_{\mathrm{DMSO}}$$

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 α 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* polyA-selected RNA libraries, among which the majority are mRNAs (Extended Data Fig. 6).

Metagene analysis of translation, pause, m^{6}A and protein-binding sites. We calculated metagene structure profile around different functional sits by averaging all valid reactivity score R: (1) 10 nucleotides upstream and downstream of the RNA methylation m^{6}A site, as determined by our laboratory previously[25]; (2) 25 nucleotides upstream and downstream of the translation pause site, as determined in the same ribosome profiling experiment; (3) 25 nucleotides upstream and downstream of the RNA methylation m^{6}A site; (4) 25 nucleotides upstream and downstream of the binding sites of RNA-binding proteins Rbfox2 (ref. 35) and HuR (Extended Data Figs 8 and 9; see later for details).

In the analysis of differential profiles of icSHAPE reactivity scores around m^{6}A and negative control sites, we retrieved a set of target m^{6}A sites that have icSHAPE reactivity scores in both wild-type and *Mettl3*-knockout cells, and defined a set of a similar number of non-methylated m^{6}A sites with the same sequence motifs (GG ACU). For both wild-type and knockout cells, we calculated the profiles of average reactivity scores in target sites and negative controls and subtracted the latter from the former scores to define the differential icSHAPE profiles.

We calculated *in vivo* and *in vitro* metagene structure profile separately. For each transcript functional site and its flanking regions, we demand a stringent R score for thousands of transcripts being compared. We generated roughly the same number of negative controls for each set of functional sites. And whenever a sequence motif exists for a functional site, we use that motif in generating the negative control. For example, the same sequence motif GGACU is used to scan the transcriptome and negative controls are randomly selected from the hits, excluding regions that are close to a true m^{6}A site.

The HuR iCLIP experiments are performed and clusters of binding sites are determined with the pipeline as described previously[31]. Threshold 9 (at least 9 unique reverse transcription stops are each genomic coordinate) is used to filter for the true binding sites. The highest peak and its flanking 50 nucleotides in each cluster were retrieved and used to call sequence motifs by using HOMER, with random sequences of 50 nucleotides from the same set of transcripts as background. The motifs are used as the anchor point in calculating metagene profiles and also used to generate negative controls, using the same protocol as the m^{6}A negative control generation described earlier.

VTD analysis. We defined and calculated the VTD profile of a transcript by subtracting its valid *in vitro* structural profile from the *in vivo* one. We calculate the average VTD profiles for all 4,096 possible hexamers in our structurome. The overall VTD score of each hexamer is defined as the average score across the six bases of the hexamer.

We retrieved sequence motifs of important functional sites, including Kozak sequences (GCCRCC), m^{6}A sites (GGACU), miR-290 family hexamers seed matches (GCACUU, complementary to the seeds) and Rbfox2-binding sites (UGCAUG), and highlighted their VTD scores on the VTD histogram of all hexamers. For sites with ambiguity, for example, m^{6}A sites, we took the average of all hexamers that contain GGACU.

We also compiled a resource (Supplementary Tables 1 and 2) of VTD scores for all RNA protein-binding motifs studied by RNAcompete experiments[36], and all mouse microRNA hexamers seed matches from miRBase[37]. In addition to the VTD scores, for every motif or seed match, we asked three questions by using a permutation test: (1) is the absolute value of the motif (or seed match) VTD significantly less than a random hexamer, that is, represents a stable region; (2) is the motif (or seed match) VTD significantly smaller than a random hexamer, that is, represents a region that is more structured *in vivo*; and (3) is the motif (or seed match) VTD significantly bigger than a random hexamer, that is, represents a region that is more structured *in vitro*.

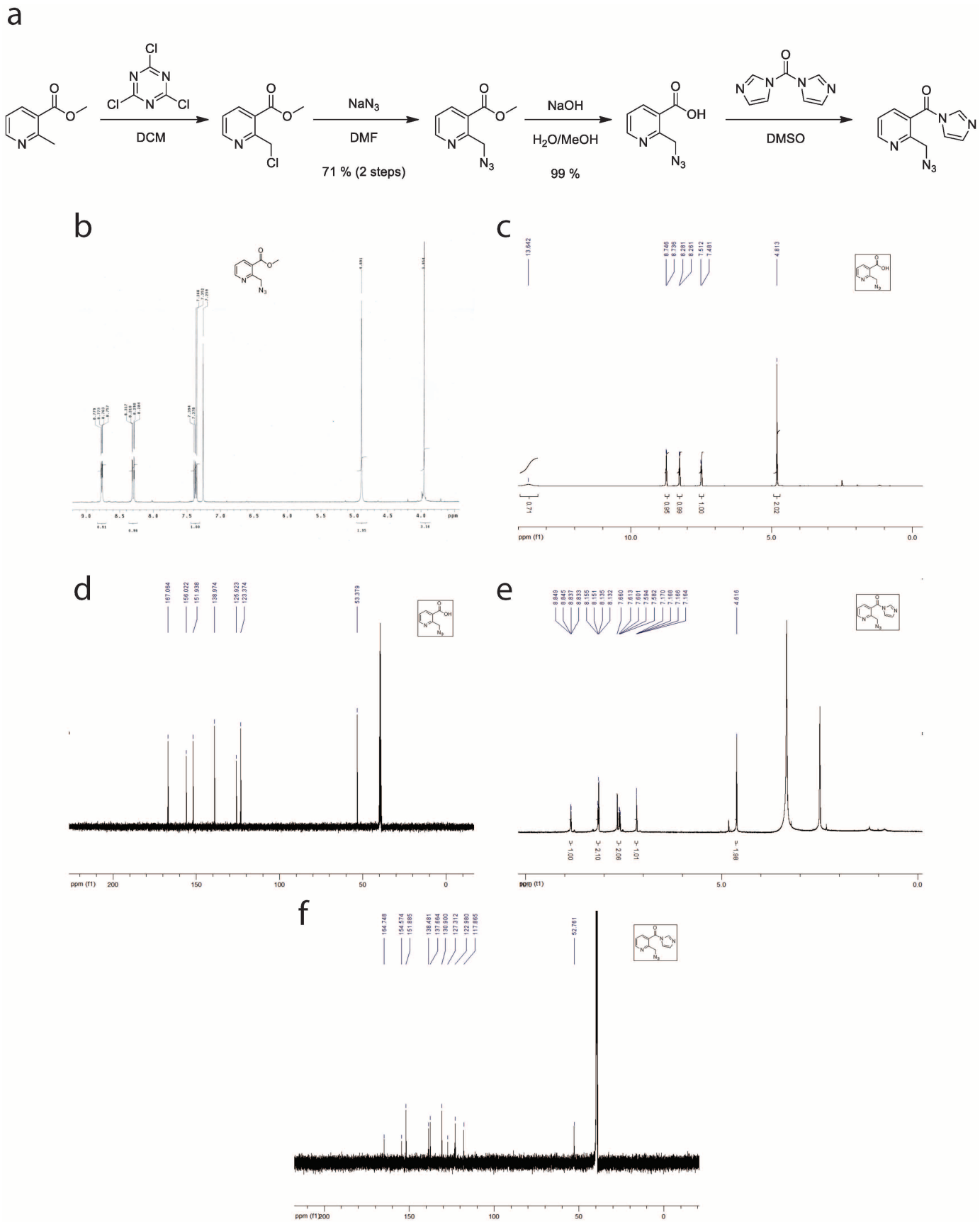
Structure-based prediction of m^{6}A sites and protein-binding sites. We constructed a set of SVM models[38] for the prediction of m^{6}A sites and protein-binding sites using structural profiles, genomic locations, conservations and their combinations.

The structural profile is limited to the range from the -10 to the $+10$ position of the m^{6}A site or the motifs 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^{6}A 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 start 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[39] for nucleotides in the range from the -10 to the $+10$ position of the m^{6}A site or the motifs of the protein-binding sites.

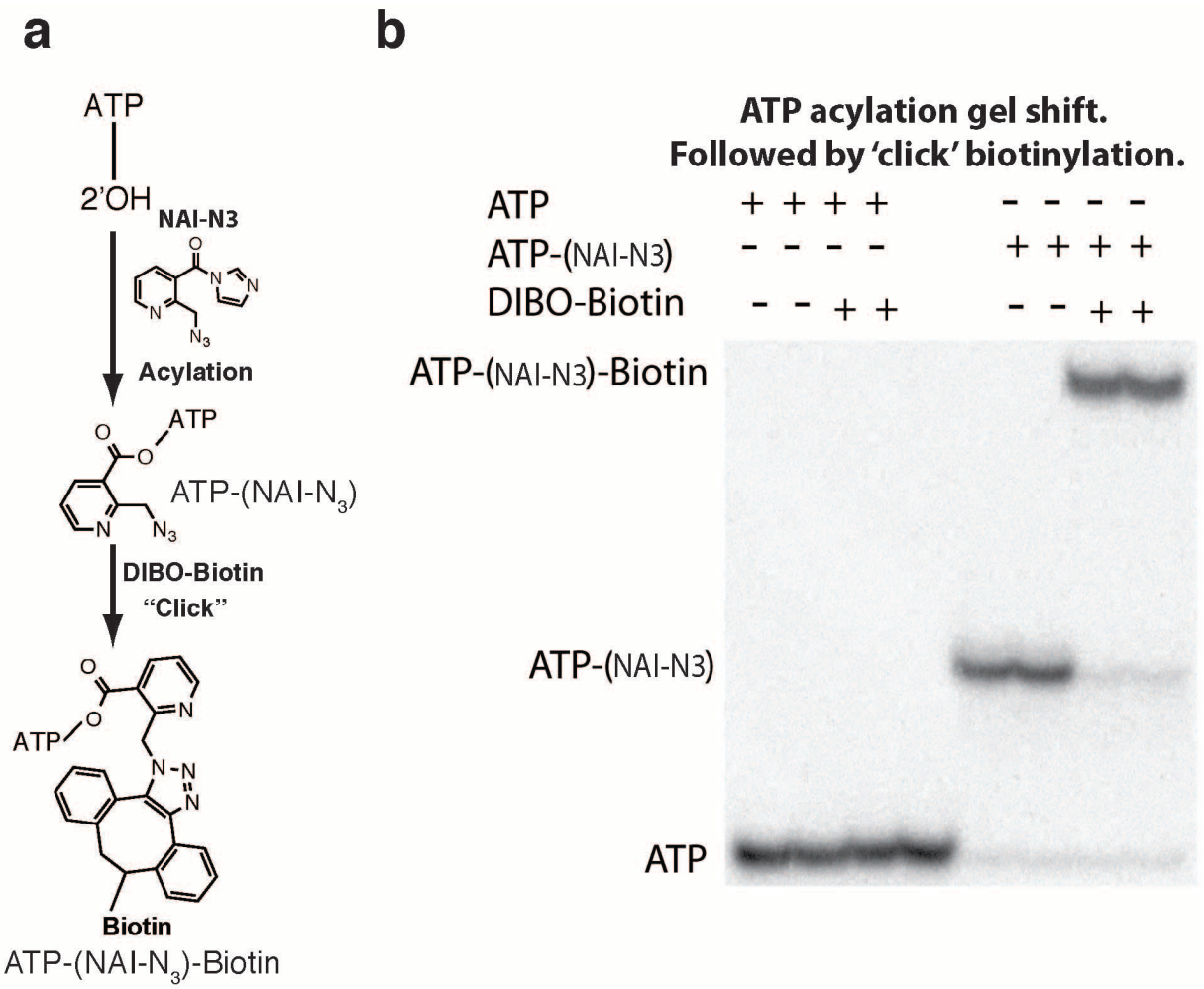
We used the same set of positive and negative controls and the best predictor is selected by using a parameter-searching tool that is included with the LIBSVM package (http://www.csie.ntu.edu.tw/~cjlin/libsvm/). We used a fivefold cross-validation and calculated the AUC of the ROC curve to evaluate the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

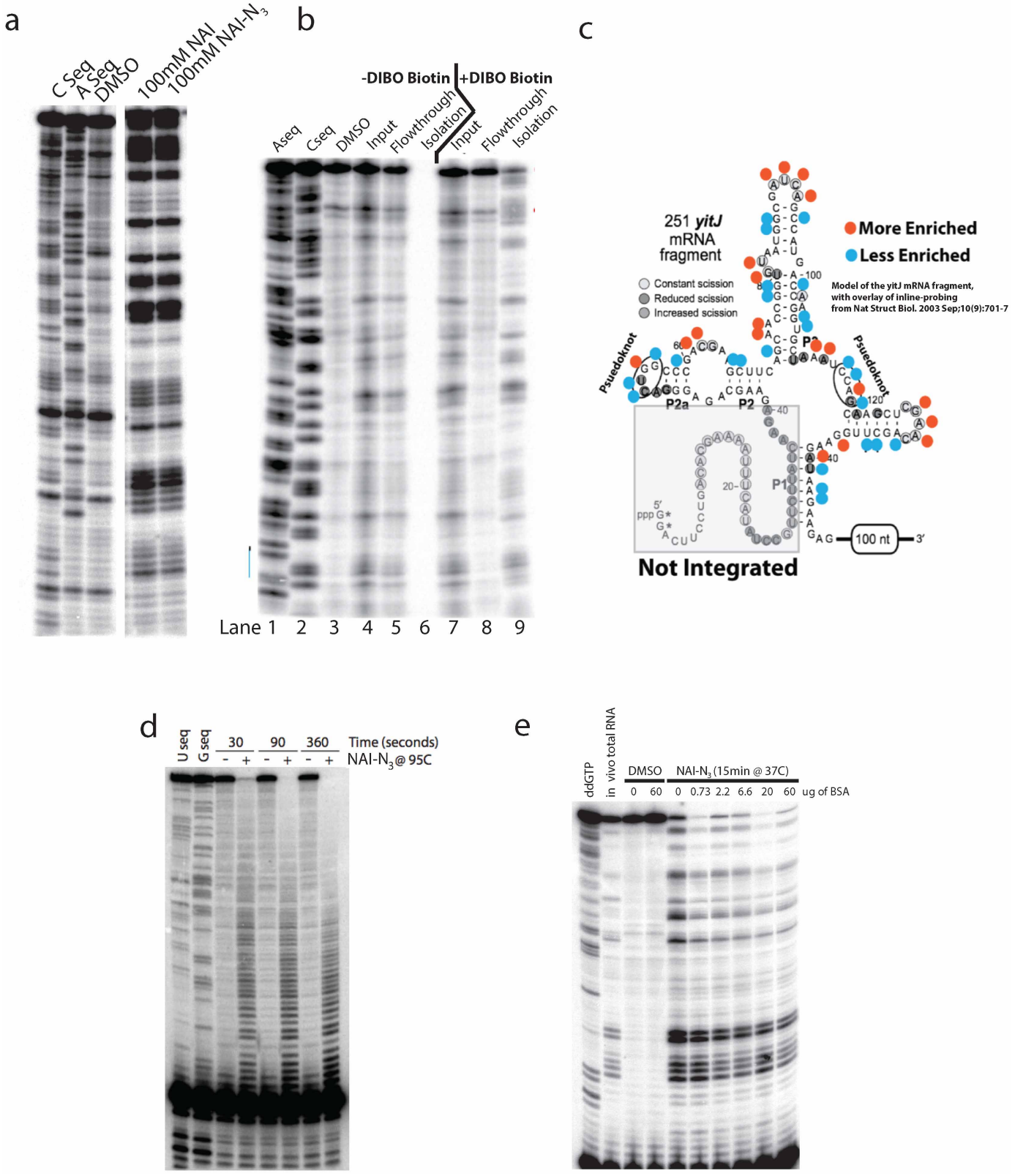
28. Das, R., Laederach, A., Pearlman, S. M., Herschlag, D. & Altman, R. B. SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA* **11,** 344–354 (2005).
29. Gherghe, C. *et al.* Definition of a high-affinity Gag recognition 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. Jangi, M., Boutz, P. L., Paul, P. & Sharp, P. A. Rbfox2 controls autoregulation in RNA-binding protein networks. *Genes Dev.* **28,** 637–651 (2014).
36. Ray, D. *et al.* A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499,** 172–177 (2013).
37. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42,** D68–D73 (2014).
38. Cortes, C. & Vapnik, V. Support–vector networks. *Mach. Learn.* **20,** 273–297 (1995).
39. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15,** 1034–1050 (2005).



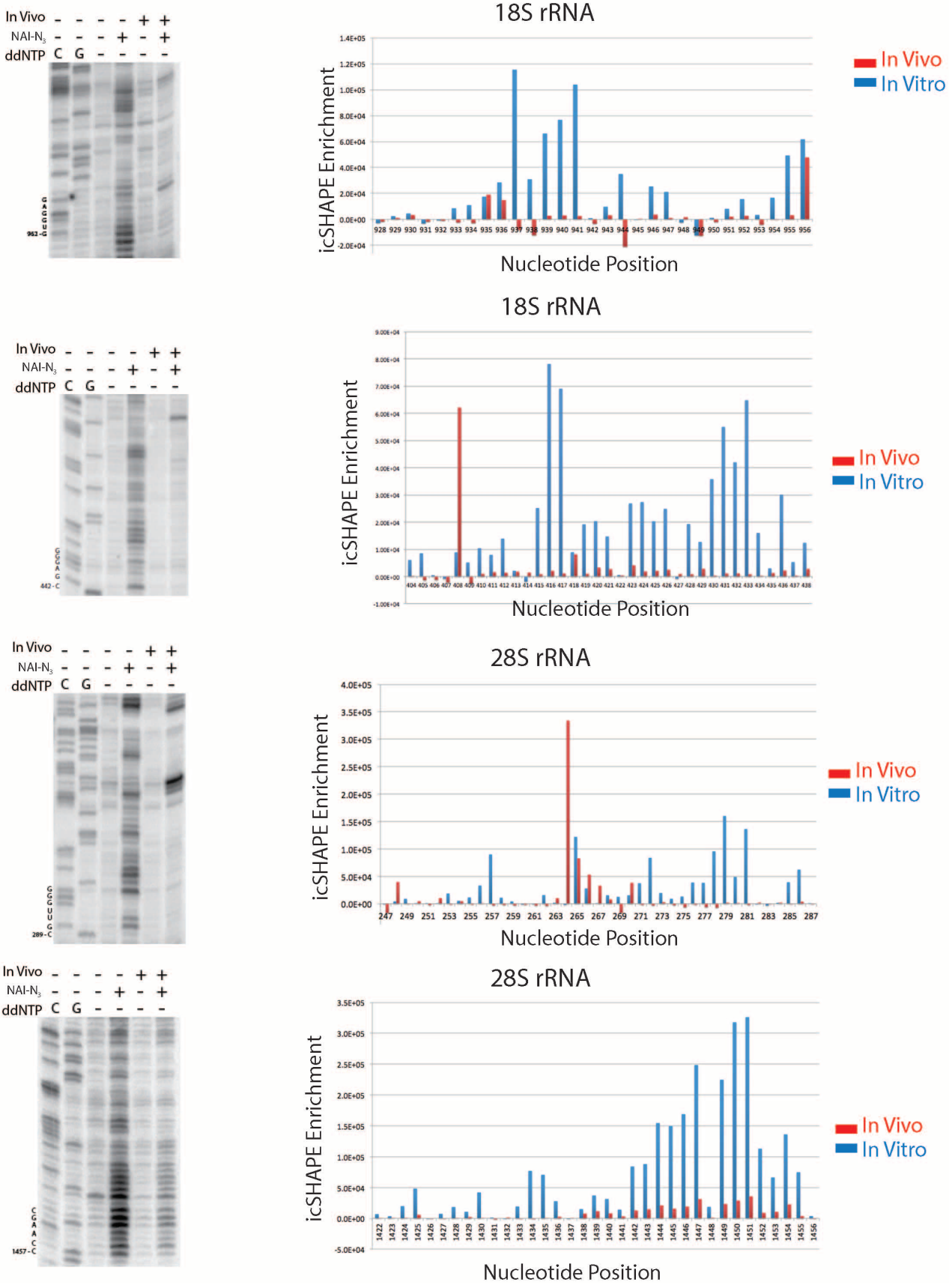
Extended Data Figure 1 | Chemical synthesis of NAI-N_3. **a**, Synthetic scheme for NAI-N_3. **b**, ^{1}HMR of methyl 2-(azidomethyl)nicotinate. **c**, ^{1}HNMR of 2-(azidomethyl)nicotinic acid. **d**, ^{13}CNMR of 2-(azidomethyl)nicotinic acid. **e**, ^{1}HNMR of 2-(azidomethyl)nicotinic acid acyl imidazole. **f**, ^{13}CNMR of 2-(azidomethyl)nicotinic acid acyl imidazole.



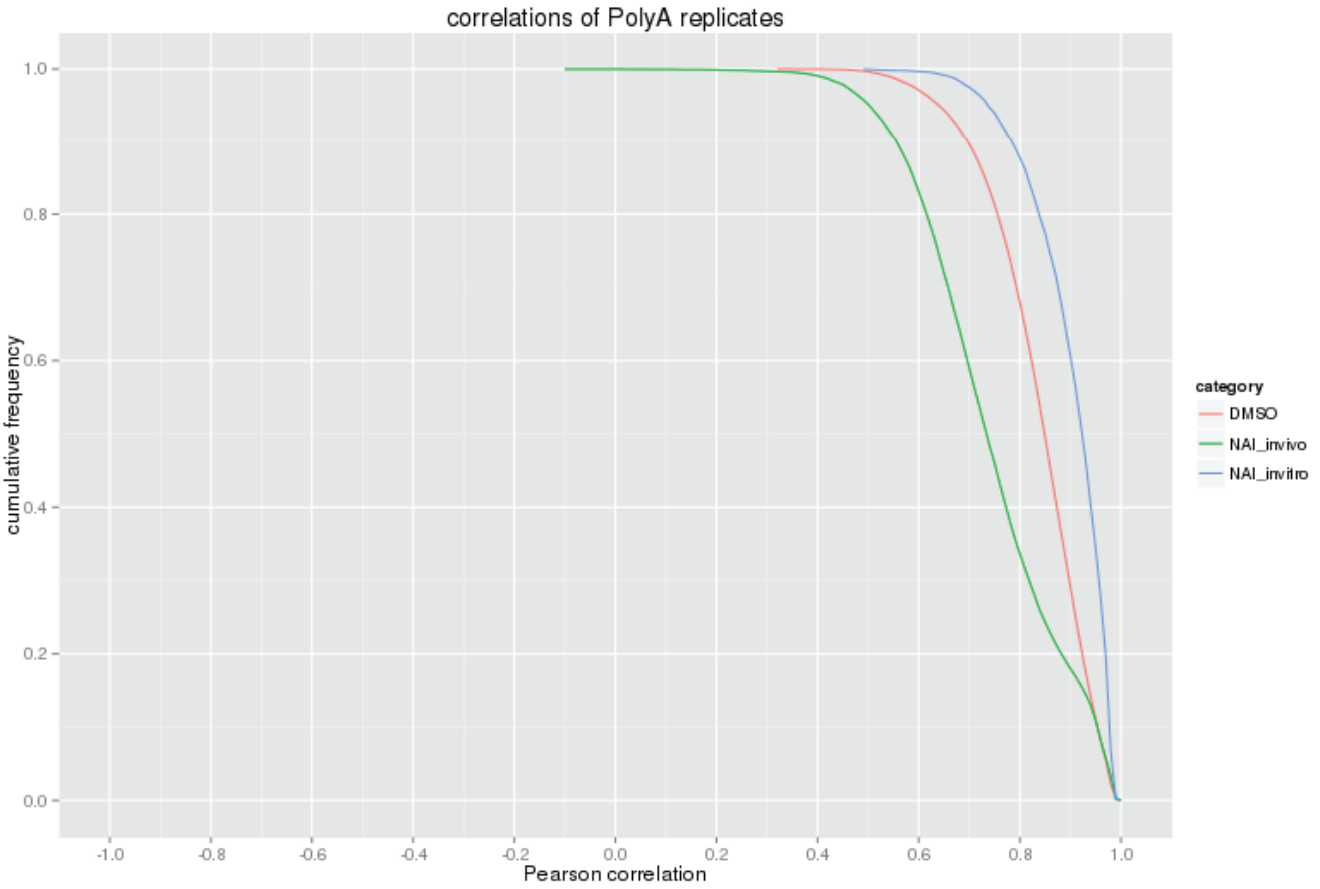
Extended Data Figure 2 | NAI-N_3 is a novel RNA acylation reagent that enables RNA purification. **a**, Chemical schematic of RNA acylation and copper-free 'click' chemistry using NAI-N_3 and dibenzocyclooxtyne (DIBO)–biotin conjugate. **b**, ATP acylation gel shift showing ATP acylation and copper-free 'click' chemistry using NAI-N_3 and DIBO–biotin conjugate.



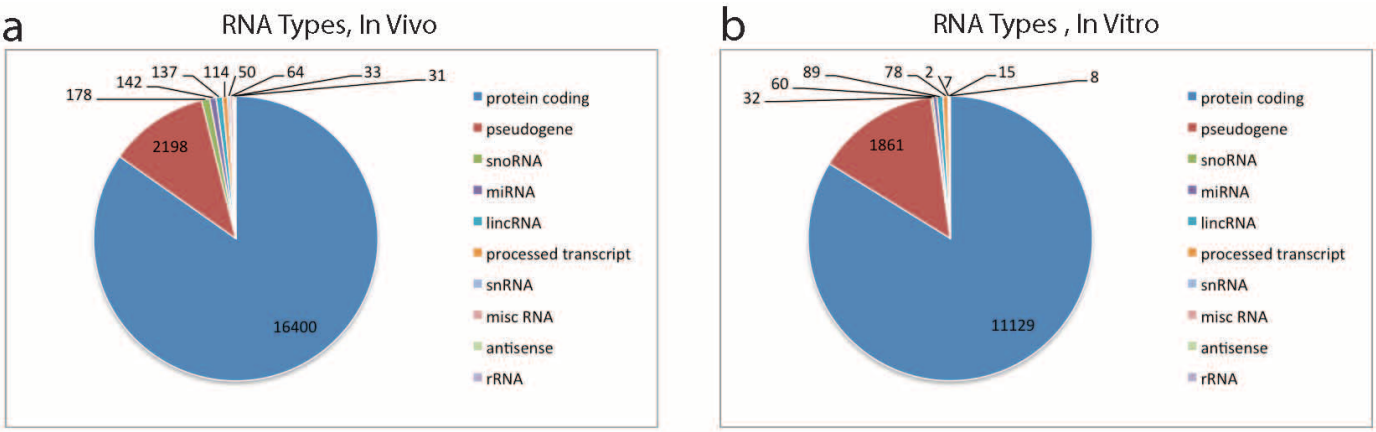
Extended Data Figure 3 | NAI-N_3 is a novel RNA acylation reagent that accurately reads out RNA structure. **a**, Comparative denaturing gel of NAI and NAI-N_3 RNA acylation. **b**, Denaturing gel analysis of cDNAs that originate from the biotin-purification protocol (Extended Data Fig. 1). **c**, Secondary 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-N_3 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.



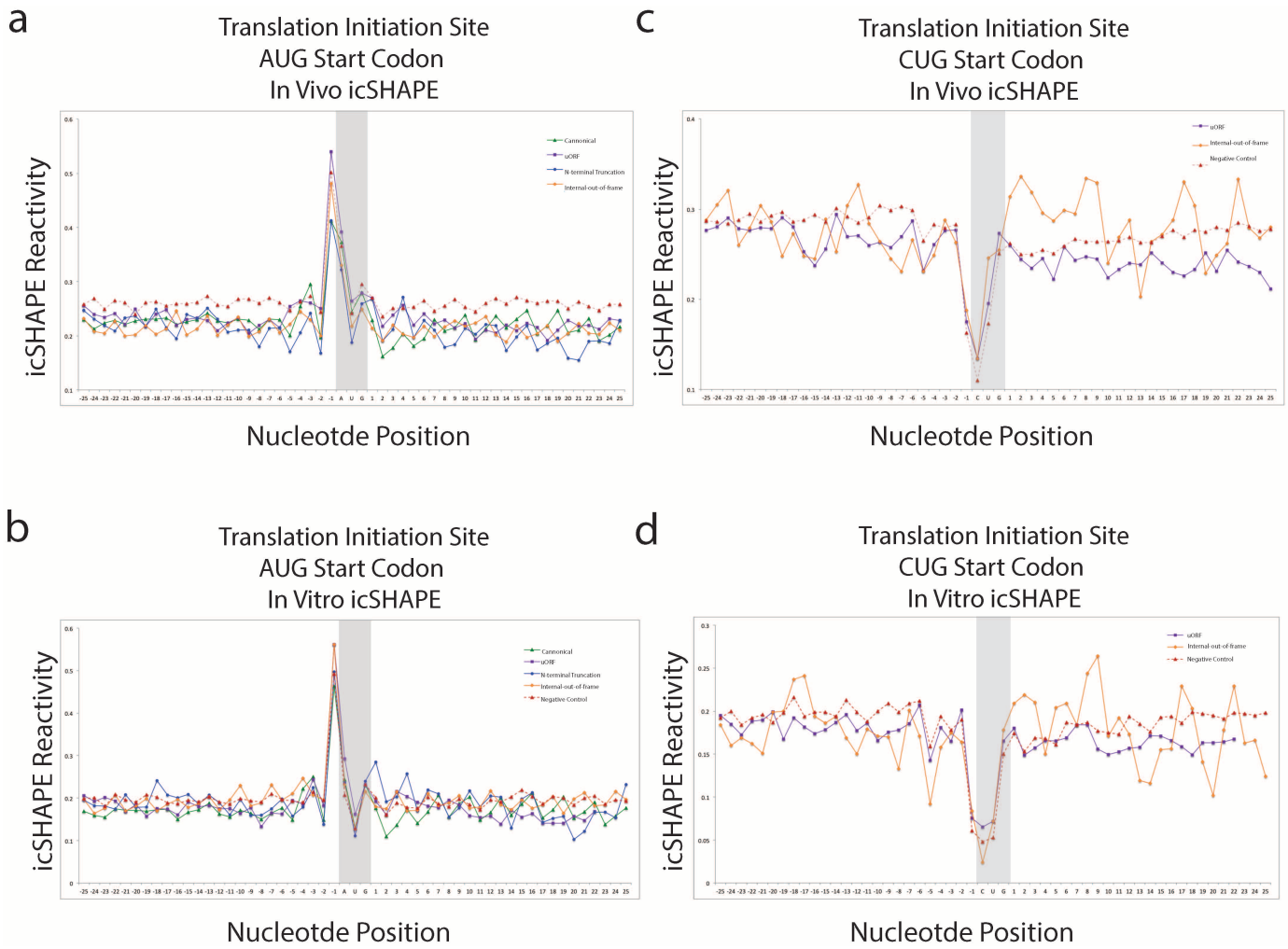
Extended Data Figure 4 | icSHAPE is capable of reproducing RNA acylation profiles obtained by manual RNA modification experiments. icSHAPE profiles (right) of rRNA, and compared to those obtained by manual SHAPE (left).



Extended Data Figure 5 | Reverse transcription stops measured by icSHAPE are very well correlated in different library replicates.

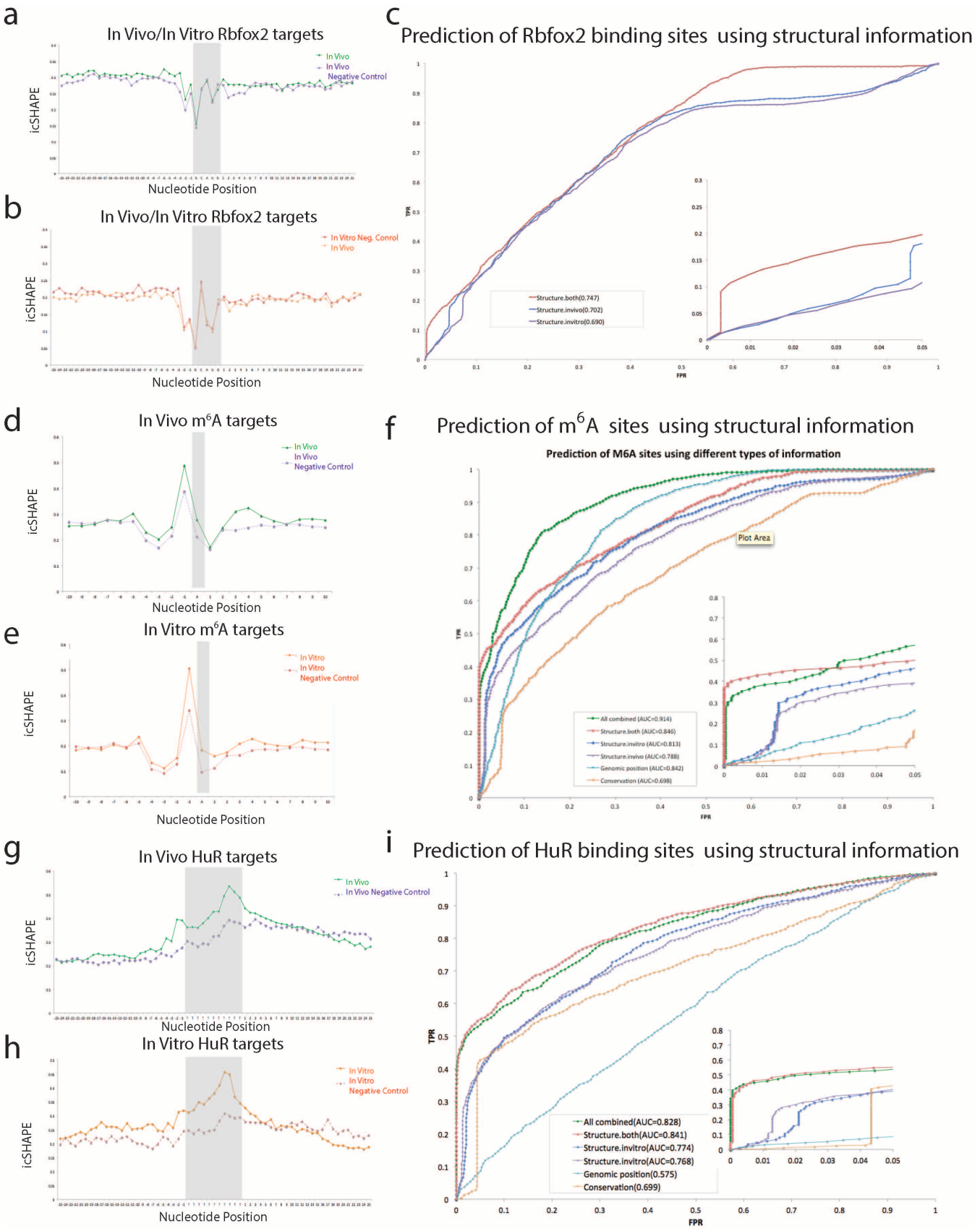


Extended Data Figure 6 | icSHAPE is capable of measuring the RNA structure profiles of thousands of RNAs simultaneously. **a**, The RNAs represented in polyA-selected RNA, *in vivo*. **b**, The RNAs represented in polyA-selected RNA, *in vitro*.

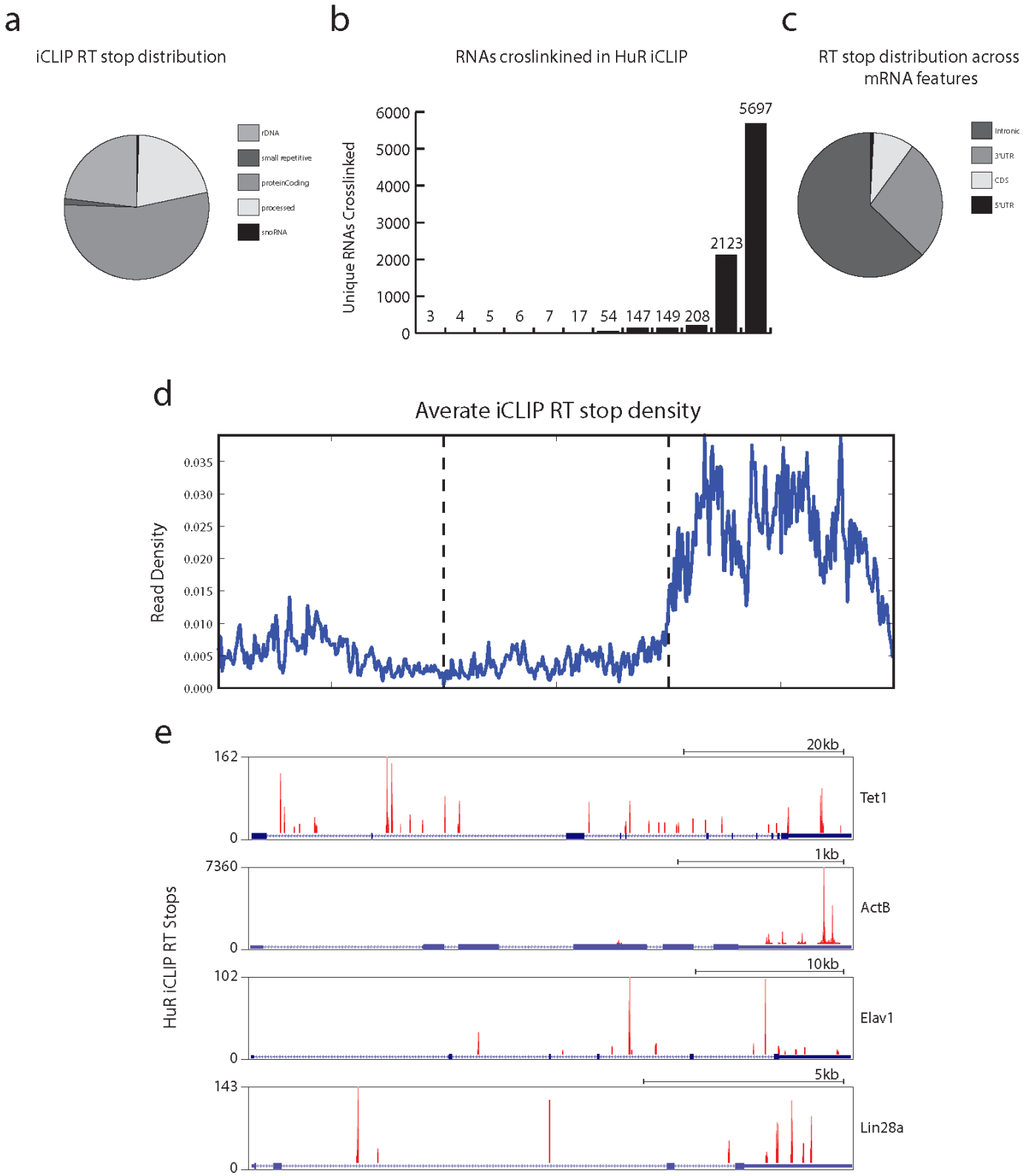


Extended Data Figure 7 | Non-AUG start codons are associated with preceding reactivity, and non-AUG start codons have a different profile, suggesting that RNA accessibility alone is not sufficient to drive translation.

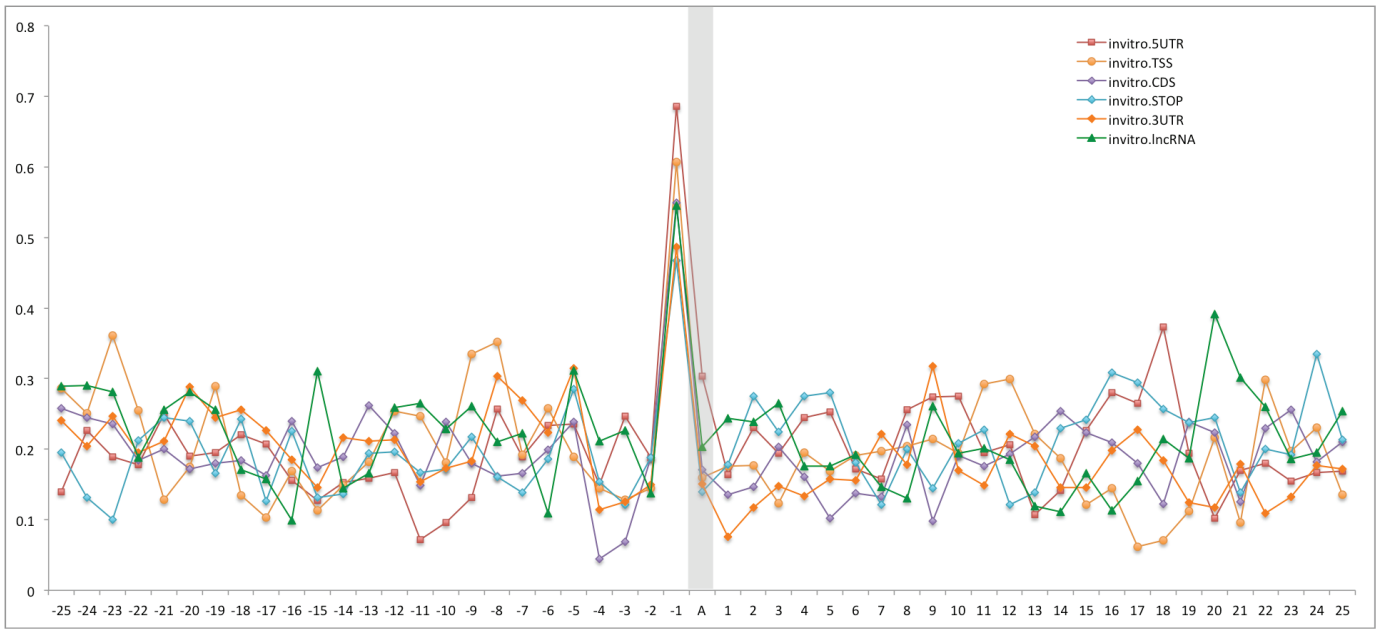
a, icSHAPE profile at AUG start codons, *in vivo*. **b**, icSHAPE profile at AUG start codons, *in vitro*. **c**, icSHAPE profile at CUG start codons, *in vivo*. **d**, icSHAPE profile at CUG start codons, *in vitro*.



Extended Data Figure 8 | icSHAPE can be used to predict post-transcriptional regulatory elements. **a**, icSHAPE profile at Rbfox2 targets, *in vivo*. **b**, icSHAPE profile at Rbfox2 targets, *in vitro*. **c**, ROC curve of Rbfox2 RNA–protein interactions, predicted using icSHAPE profiles. **d**, icSHAPE profile at m^{6}A targets, *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, *in vitro*. **f**, ROC curve of m^{6}A RNA modification sites, predicted using icSHAPE profiles. **g**, icSHAPE profile at HuR targets, *in vivo*. **h**, icSHAPE profile at HuR targets, *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^{6}A-associated RNA structure features are preserved, independent of their position along the RNA transcript.

ERRATUM

doi:10.1038/nature15717

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.