

## Dynamic Regulation of RNA Structure in Mammalian Cells

Lei Sun<sup>1,\*</sup>, Furqan Fazal<sup>2,\*</sup>, Pan Li<sup>1,\*</sup>, James P. Broughton<sup>2</sup>, Byron Lee<sup>2</sup>, Lei Tang<sup>1</sup>, Wenzhe Huang<sup>1</sup>, Howard Y. Chang<sup>2,3,#</sup>, Qiangfeng Cliff Zhang<sup>1,#</sup>

1. MOE Key Laboratory of Bioinformatics, Center for Synthetic and Systems Biology, Beijing Advanced Innovation Center for Structural Biology, Tsinghua-Peking Joint Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China

2. Center for Personal Dynamic Regulomes, Stanford University, Stanford, CA 94305, USA

3. Howard Hughes Medical Institute

\*These authors contributed equally to this work

#Jointly supervised.

Correspondence: [gczhang@tsinghua.edu.cn](mailto:gczhang@tsinghua.edu.cn) (Q.C.Z.) and [howchang@stanford.edu](mailto:howchang@stanford.edu) (H.Y.C.)

**RNA structure is intimately connected to each step of gene expression. Recent advances have enabled transcriptome-wide maps of RNA secondary structure, termed RNA structuromes. However, previous whole-cell analyses lacked the resolution to unravel the dynamic regulation of RNA structure across subcellular states. Here we reveal the RNA structuromes in three compartments — chromatin, nucleoplasm and cytoplasm. The cytotopic structuromes substantially expand RNA structural information, and enable detailed investigation of the central role of RNA structure in linking transcription, translation, and RNA decay. Through comparative structure analysis, we develop a resource to visualize the interplay of RNA-protein interactions, RNA chemical modifications, and RNA structure, and predict both direct and indirect reader proteins of RNA modifications. We validate the novel role of the RNA binding protein LIN28A as an N6-methyladenosine (m<sup>6</sup>A) modification “anti-reader”. Our results highlight the dynamic nature of RNA structures and its functional significance in gene regulation.**

RNAs fold into complex structures that are crucial for their functions and regulations including transcription, processing, localization, translation and decay<sup>1-6</sup>. Over the last few decades RNA structure has been studied extensively *in vitro* and *in silico*, and crystallography and cryo-EM structures of molecular machines such as the spliceosome and ribosome, containing RNAs at their core, have become available<sup>7,8</sup>. In recent years technologies have been developed to map RNA secondary structures for the whole transcriptome, *i.e.*, RNA structuromes, by combining biochemical probing with deep sequencing<sup>9-16</sup>. These systems biology studies have revealed

many novel insights on the RNA structure basis of gene regulation<sup>17-20</sup>. However, so far existing genome-wide structure probing studies have focused on whole-cell data, which only represents an ensemble average of RNA molecules in different subcellular compartments.

In fact, RNA undergoes a complex life cycle in eukaryotic cells, mirrored by its movement into distinct cytotopic locales<sup>4,21</sup>. RNA structure is thought to form co-transcriptionally on the chromatin template, undergo conformational changes resulting from RNA chemical modification and processing in the nucleus, and experience further changes in the cytoplasm during translation and RNA decay. Averaging the RNA structure signal in the entire cell may obscures these critical dynamic features. More importantly, detailed mapping of RNA structures *in situ* will help to elucidate how they are regulated, which is essential to understanding the RNA structure basis for gene expression regulation.

An important driving force that regulates the dynamic landscape of RNA structures in post-transcription regulation is the binding of RNA-binding proteins (RBPs). A study in Arabidopsis revealed that RNA secondary structure is anti-correlated with protein-binding density<sup>22</sup>. We recently used icSHAPE to probe RNA structuromes in mouse ES cells and examined the *in vivo* and *in vitro* structure profiles of RBFOX2, a splicing factor of the “feminizing locus on X” (Fox) family proteins; and HuR, an RBP that regulates transcript stability<sup>12</sup>. We implemented a machine learning algorithm and found that using structure signals significantly improved the prediction of RNA-binding sites of both RBPs, suggesting that RNA structure signature analysis is a powerful tool to investigate RNA–RBP interactions. However, in spite of these recent advances in our understanding of the association between RNA structure and RBP-binding, a compendium of the RNA structural basis of RBP binding is not available.

In addition to RBP binding, the modification and editing of RNAs are also an important mechanism for RNA structure regulation. RNA modification can regulate almost all RNA processes including RNA maturation, nuclear retention and exportation, translation, decay, and cell differentiation and reprogramming as well<sup>23,24</sup>. As one of the most abundant and important types of mRNA modification, N6-methyladenosine (m<sup>6</sup>A) has been showed to be able to weaken duplex RNA stability by conformational switching<sup>12,25,26</sup>. The impact of structure destabilizing effect of m<sup>6</sup>A is exemplified by a study that investigated HNRNPC, a splicing factor that preferentially binds to single-stranded polyU tracts<sup>27</sup>. Biochemical studies showed that m<sup>6</sup>A modification can disrupt the local RNA structures and promote HNRNPC binding in nearby regions<sup>28</sup>. The study defined these m<sup>6</sup>A sites as “m<sup>6</sup>A-switches”, and identified an enrichment of tens of thousands of m<sup>6</sup>A-switches in the vicinity of HNRNPC binding sites. The authors further showed these m<sup>6</sup>A-switches could regulate HNRNPC-binding and the splicing of the target

RNAs. This study demonstrated a very nice example that m<sup>6</sup>A functions as an RNA structure modulator to affect RNA splicing through interfering with protein binding. However, whether the structure context is a generic mechanism for the recognition of other “reader” proteins of m<sup>6</sup>A and other RNA modifications, is still unclear<sup>29</sup>.

Here we use *in vivo* click selective 2-hydroxyl acylation and profiling experiments (icSHAPE)<sup>12</sup>, a technique we developed to map RNA structure *in vivo*, in three compartments – chromatin, nucleoplasm and cytoplasm – in both mouse and human cells. Consequently, we were able to determine the precise relationship of RNA structure with cellular processes including transcription, translation and RNA decay in the compartment where they occur. Separately, we could quantify how RNA adopts different conformations across different cellular compartments, which we termed “structural dynamics”, and investigate the sophisticated interplay of RNA structured dynamics, RNA modification and RBP binding.

## Results

**Cytotopic RNA structure maps substantially expand the scope and comprehensiveness of RNA structures.** To investigate the dynamic regulation of RNA structure in the cell, we performed icSHAPE to measure RNA secondary structure for transcripts isolated from three subcellular compartments and in two species (Figure 1a). After performing the icSHAPE reaction of living cells *in situ*, RNA fractionation<sup>30,31</sup> enabled the study of RNA structural changes in distinct subcellular locations. Separately, we fractionated the three subcellular compartments, isolated and refolded naked RNA from each, and performed icSHAPE *in vitro*. This *in vitro* dataset served as a control for the RNA contents in each compartment. The use of both v6.5 mouse embryonic stem (mES) cells and human embryonic kidney (HEK293) cells allowed us to examine whether the structural patterns we observed are conserved across the two species and cell types.

We determined RNA structure, as previously described<sup>12,32</sup>, after enriching for messenger RNAs (mRNAs) and long noncoding RNAs (lncRNAs) by ribosome depletion, and sequencing the resulting icSHAPE libraries at high depth (~200 million reads per replicate, Table S1). We first confirmed the quality of fractionation using RT-qPCR for landmark RNAs, and Western blots for specific proteins (Figure S1). We used the icSHAPE pipeline<sup>12</sup> to calculate a score to represent the structural flexibility (indicative of unpaired RNA bases) of every nucleotide, and found good correlation across replicates (Pearson correlation  $r > 0.75$  for the top 60% most-abundant transcripts in all replicates, Figure S2). To further validate our structural data, we examined its agreement with known structures – two such RNAs are

*Ribonuclease P* RNA and Signal recognition particle (SRP) RNA (Figure S3). Both RNAs are enriched in nucleoplasm, and indeed our nucleoplasm icSHAPE data closely match the existing structural models.

The chromatin structurome is enriched for lncRNAs (Figure 1b, S4a). As an example, we examined the structure of the human growth arrest-specific 5 (GAS5) noncoding RNA, which acts as a decoy glucocorticoid response element (GRE) by binding to the DNA-binding domain of the glucocorticoid receptor (GR)<sup>33</sup>. Indeed, the expected GAS5 RNA structure is accurately recovered in the chromatin fraction, showing low icSHAPE scores for the double-stranded GR binding motif of the GAS5 RNA, and high reactivity score for the loop region (Figure 1c). Similar to lncRNAs, snoRNAs and snRNAs are also enriched in the chromatin fraction, and to a smaller extent in the nucleoplasm fraction (both relative to the cytoplasm). Furthermore, intronic reads constitute the majority of the sequencing data in the chromatin fraction, but only ~15-20% of reads in the cytoplasmic fraction (Figure 1b, Figure S4b). For example, we obtained intron structures for the transcript heterogeneous nuclear ribonucleoprotein A2/B1 (*HNRNPA2B1*) in the chromatin fraction, but these sequences were largely absent in the nucleoplasmic and cytoplasmic fractions (Figure 1d). Interestingly, we found that RNAs *in vivo* are much more folded in intron regions than in exon regions (average Gini index of 0.7 versus 0.5. A higher Gini index indicates a more structured region<sup>11</sup>), in contrast to *in vitro* conditions (both with average Gini index 0.6); this result holds true for both human (Figure 1e) and in mouse (Figure S5a). The finding that intronic regions are more folded *in vivo* is not likely due to differential RNA-binding-protein (RBP) binding in introns versus exons, as similar trends were observed when all known RBP-binding sites were excluded in the structural comparison (Figure S5b). Instead, these results may suggest distinct interplays between RNA structures, and transcriptional or splicing regulation in introns and exons. In summary, the RNA-structural profiles of the chromatin fraction provide a rich resource to interrogate structures of lncRNAs, pre-mRNAs including introns, and other chromatin-associated RNAs, expanding the scope of the RNA structurome.

**RNA structure plays a central role in connecting many cellular events.** The cytotopic RNA structuromes allowed us to assess the roles of RNA structure (or lack thereof) in regulating each step of the gene-expression life cycle, which takes place in distinct subcellular compartments. We obtained data on transcriptional rate, translational efficiency, and RNA half-life from previous studies in human and mouse<sup>34-36</sup>, and correlated data with the Gini index of icSHAPE reactivity. RNA structure in nascent RNA has been suggested to propel or impede RNA polymerase pausing at individual genes<sup>37</sup>. We therefore analyzed the relationship between

transcription and 5'UTR (untranslated region) RNA structures of the chromatin-associated fraction, and found that lower transcriptional rate correlates modestly with more structure ( $r = -0.19$ ,  $p = 1.5 \times 10^{-6}$ , [Figure 2a](#)). Next, many studies have found RNA secondary structure upstream of or at ribosome binding sites is important for translation<sup>15,17-19</sup>. We did observe that more 5'UTR RNA structure correlates with decreased translational efficiency in the cytoplasmic fraction ( $r = -0.31$ ,  $p = 1.7 \times 10^{-48}$ , [Figure 2b](#)). Finally, as RNA degradation occurs in both nucleoplasm and cytoplasm via different pathways, we analyzed the dependence of RNA half-life on RNA structure in both fractions. We found that more-structured RNAs tended to have shorter half-lives in both the nucleus and cytoplasm ([Figure 2c-d](#)), highlighting the emerging and under-appreciated role of double-strand ribonucleases in transcript turnover<sup>38</sup>.

Quantitative correlation analysis showed that the relationships among RNA structure, transcription and translation are not binary, as there is a general trend that an RNA with lower transcriptional rate tends to simultaneously be more structured and translated less efficiently ([Figure 2e-f](#)). The positive link between transcription and translation, two major events in gene expression, has been previously appreciated<sup>39</sup>([Figure S6](#)). Recent studies have suggested different mechanisms, including m<sup>6</sup>A modification, that could account for this linkage by imprinting an mRNA transcript during its synthesis and later regulating its translation<sup>39-41</sup>. Our data suggest that genome-wide RNA structures formed at chromatin during transcription remain largely unchanged in the nucleoplasm and cytoplasm fractions ([Figure S7](#)), and might thus serve as a link between transcription and translation efficiencies. We therefore considered two models to explain our observations – in the first model RNA structure is a mediation factor that is affected by transcription, and it in turn affects translation; and in the second model RNA structure is a confounding factor that has an effect on both transcription and translation ([Figure 2g](#)). Statistical analysis suggests that while both models could be true, the first (mediation) model can account for a larger fraction of the positive correlation between transcription and translation, and is statistically more significant. In summary, RNA structure plays a general role that connects many cellular events including transcription, translation and RNA degradation ([Figure 2h](#)).

**Pervasive RNA structure dynamics across different cellular compartments.** More importantly, cytotopic RNA structuromes also enabled us to examine how RNA adopts different conformations across different cellular compartments, which we term “structural dynamics”. Overall, we saw that RNA structures were more unfolded in the chromatin fraction relative to the cytoplasmic fraction, even at sites not targeted by RBPs ([Figure S8](#)). As specific regions of an

individual RNA can be regulated differently and display different patterns of structural dynamics, we implemented a statistical method to discover structurally-dynamic regions. As an example, we show that U12 snRNA displayed structurally-dynamic regions between compartments (Figure 3a, black bars). In addition, despite high evolutionary conservation of U12, the RNA structures showed shared and unique conformational changes in human and mouse. These findings suggest that both species-specific and conserved mechanisms may regulate RNA structures and structural dynamics.

On a genome-wide scale, we found that different RNA categories showed different levels of structural dynamics *in vivo* (Figure 3b). To begin to dissect the factors that regulate RNA structural dynamics in cells, we used the same analysis pipeline to evaluate data obtained from fractionated, purified RNA that was refolded *in vitro* (Figure 3c), and compared RNA conformational changes observed between compartments *in vivo* and *in vitro*. Conceptually, structural differences between RNAs isolated from different compartments and refolded *in vitro* do not represent structural dynamics, but we will use this term for simplicity. In general, as expected, RNA structures vary less between the compartments *in vitro* relative to *in vivo* (comparing Figure 3b to Figure 3c), suggesting that fewer factors influence RNA folding *in vitro* versus *in vivo*. This finding is particularly true for highly-conserved small RNAs such as snoRNAs, miRNAs and snRNAs, suggesting that these functional RNAs adopt stable structures *in vitro* but are subjected to extensive regulation *in vivo*. The structural differences are magnified when directly comparing *in vivo* to *in vitro* icSHAPE data for each compartment (Figure 3d), and different RNA categories displayed varying levels of structural differences *in vivo* and *in vitro*, consistent with previous findings from whole-cell data<sup>12</sup>. Finally, we compared the levels of structural divergence between mouse and human for sequence-conserved regions. We used the same pipeline used above to call for regions of structural changes, and found even larger fractions of structural differences, suggesting substantial species-specific regulation of RNA structure (Figure 3e). Taken together, our analyses suggest that structural changes are pervasive, reflecting that many different factors may contribute to their regulation in different circumstances.

**RNA modification and RBP binding underlie RNA structure dynamics.** RNA modification and RBP-binding are important factors that are known to influence RNA structure. To disambiguate their contributions to RNA-structural dynamics, we overlaid compartment-specific RNA structuromes with RNA modifications and RBP-binding sites. Figure 4a-c shows examples



of focal conformational changes around known locations of m<sup>6</sup>A modification, pseudouridylation ( $\Psi$ ) and heterogeneous nuclear ribonucleoprotein C (HNRNPC) binding.

As m<sup>6</sup>A is well known as an RNA structure switch, favoring unpairing of dsRNA<sup>12,28</sup> we compared the genome-wide structures for m<sup>6</sup>A methylated versus non-methylated sites with the same underlying sequence motif, and confirmed similar patterns of structure destabilization in all three fractions ([Figure S9a](#)). Furthermore, the structural differences are largest in the nucleoplasm fraction, consistent with the finding that METTL3-METTL14 complex deposits m<sup>6</sup>A on nuclear RNA<sup>42</sup>. Following the structural dynamics of the same set of m<sup>6</sup>A sites from nucleoplasm and cytoplasm, we observed that RNA structure appears more open upon RNA migrating from the chromatin to the nucleoplasm, and thereafter remains the same ([Figure 4a](#), [S9a](#)). Both analyses suggest that the majority of m<sup>6</sup>A happens in nucleus. We repeated the analysis for pseudouridylation, another abundant RNA modification generated by the isomerization of uridine, which permits hydrogen bonding to the adjacent phosphate backbone. The extra hydrogen bond can rigidify RNA structure of  $\Psi$ -modified regions<sup>24</sup>. We found that in general these regions have higher icSHAPE reactivity (*i.e.* less structured), suggesting that modification hinders RNA structure folding freely, which again occurs predominantly in nucleus. ([Figure 4b](#), [S9b](#)).

All RNAs associate extensively with proteins in cells, and RNA binding protein (RBP) interactions are both sensitive to and profoundly impacts RNA structure. Taking HNRNPC as an example, we first confirmed that it bound to a stem-loop structure, inferred from more single-stranded nucleotides with flanking dsRNA ([Figure 4c](#)). We also followed the structure dynamics of the binding sites from chromatin to nucleoplasm and cytoplasm. We found that HNRNPC binding sites are more open in chromatin, also consistent with its major localization in chromatin-associated pre-RNA ([Figure S9c](#)). Our findings also suggest that HNRNPC binding could be a factor that accounts for the structural dynamics around the binding sites. Indeed, there is a significant overlap between HNRNPC binding and structurally-dynamics sites ([Figure S10](#)).

We extended the analysis to all RBPs with binding site information available from published RBP CLIP-seq experiments<sup>43</sup>. As shown in [Figure 4d](#), occupancy of many RBPs are linked with RNA structural changes, while others preferentially bind to structurally-stable regions of RNA. For example, many chromatin-associated proteins (e.g. HNRNPD and others shown in red in [Figure 4d](#)) bind to more open RNA regions; these regions become more structured after dissociating from the chromatin and the proteins. In contrast, the double-stranded-binding RBP Staufen homolog 1 (STAU1), a protein that shuttles between the nucleus and the cytoplasm,

appears to stabilize RNA structures upon its binding after RNA leaves chromatin. Thus, by determining the structuromes of multiple cytoplasmic localizations, our study provides an estimate of the relative contributions of known modification mechanisms and protein binding to RNA structure dynamics (Figure 4e). Protein binding using existing CLIP-seq data can explain most of the RNA structure dynamic sites (3392 of 5903), and many RNA modification sites with RNA structure changes overlap with protein binding sites. Our results thus suggest a complex interplay among RNA modification, protein binding and RNA structure dynamics.

**Structural analysis dissects different types of m<sup>6</sup>A readers.** Identifying RBPs that can read RNA modifications is of fundamental significance in the study of epitranscriptomics<sup>42</sup>. Using our dynamic RNA structurome data to filter published CLIP-seq data, we computed the effect that m<sup>6</sup>A modification has on protein binding (Methods). Our analysis identified most of the known m<sup>6</sup>A readers, including the canonical YTH domain proteins, and the newly identified HNRNPC<sup>28</sup> and the IGF2BP proteins<sup>44</sup>. All these readers bind to a region that contains one or more m<sup>6</sup>A sites stronger than a control (unmodified) site with the same m<sup>6</sup>A sequence motif (Figure 5a). Interestingly, the analysis also revealed several proteins with decreased bindings on modified m<sup>6</sup>A sites (termed “anti-readers”)<sup>42,45</sup>, including LIN-28 homolog A (LIN28A) and EW RNA binding protein 1 (EWSR1).

The precise pattern of RBP binding peaks and RNA structure at m<sup>6</sup>A sites can further reveal the biochemical mechanism of the m<sup>6</sup>A readers (Figure 5b). While the canonical readers bind most strongly directly at the m<sup>6</sup>A sites, the binding of HNRNPC and IGF2BP readers peaks at a distance. Our icSHAPE data supports a previous study that suggested that HNRNPC acts as m<sup>6</sup>A reader not by recognizing the N<sup>6</sup>-methyl group, but rather by binding a purine-rich motif that becomes unpaired and accessible upon nearby m<sup>6</sup>A modification<sup>28</sup> (Figure 5c). Similarly, our RNA-structural data suggest that IGF2BP proteins (here IGF2BP3) may also read the structural changes induced by the so-called m<sup>6</sup>A-switch<sup>44</sup> (Figure 5c). Furthermore, both HNRNPC and IGF2BP3 bind more tightly to flexible regions (Figure 5d).

To validate the “indirect reader” IGF2BP3 and the “anti-reader” LIN28A, we selected four endogenous m<sup>6</sup>A sites as targets. Each of the four targets contained three variants for the m<sup>6</sup>A site — an unmodified nucleotide, an m<sup>6</sup>A modification, and an adenosine-to-uracil mutation that mimics the disruption of base pairing (for IGF2BP3) or RBP binding (for LIN28A) (Figure 6a-b, S10a-b). We synthesized RNA constructs and used these RNA probes to retrieve RBPs from cell lysates. RNA pulldown analyses revealed that IGF2BP3 displays enhanced binding to the m<sup>6</sup>A-modified RNAs and uracil mutations relative to the unmethylated controls, confirming



IGF2BP3 to be a m<sup>6</sup>A-switch reader (Figure 6a, S11a). Conversely LIN28A displayed reduced binding to the m<sup>6</sup>A-modified and uracil-mutant target RNAs, supporting the hypothesis that LIN28A is an anti-reader that requires an unmethylated adenosine for binding (Figure 6b, S11b).

To confirm the anti-reader role of LIN28A, we performed LIN28A CLIP-seq experiments in the wild type and the m<sup>6</sup>A-methyltransferase *Mettl3*-knockout mES cells<sup>46</sup>. Many mRNAs containing one or more known m<sup>6</sup>A site showed increased binding to LIN28A when m<sup>6</sup>A deposition is abrogated, relative to the negative controls ( $p = 0.034$ , t test, Figure 6c-e, S11c-d). Increased LIN28A binding is not due to increased mRNA accumulation in *Mettl3* KO ES cells (Figure 6c-e, Figure S11c-d). LIN28A is an RBP known to enforce ES cell pluripotency and suppress ES cell differentiation<sup>47</sup>, while m<sup>6</sup>A is required for stem cell differentiation<sup>46</sup>. The negative regulation of m<sup>6</sup>A on LIN28A binding is consistent with the protein's functional roles. For example, LIN28A is a well-studied inhibitor of primary microRNA processing<sup>48</sup>, and m<sup>6</sup>A was recently shown to promote pri-miRNA processing<sup>29</sup>. Thus, the discovery of LIN28A as an m<sup>6</sup>A anti-reader potentially unifies their functional and molecular mechanisms in pluripotency, microRNA biogenesis, and post-transcriptional gene regulation.

## Discussion

Our analysis of RNA structuromes in different subcellular locations illuminated distinct RNA structural states in chromatin, nucleoplasm and cytosol. Fractionation enriched specific pools of RNAs, such as nuclear-enriched lncRNAs and pre-mRNAs including introns, thus substantially expand the scope and comprehensiveness of the RNA structuromes. Cytotoxic RNA structuromes revealed the intimate connection between RNA structure and RNA processes such as transcription, translation, RNA degradation, RBP interaction and RNA modification. Through comparative analysis, we were able to dissect the role of RNA modifications and RNA-binding proteins in influencing structure, and resolved the different sets of direct and indirect RNA modification readers. We further found and validated a novel role of the pluripotency regulator LIN28A as an anti-reader for m<sup>6</sup>A modification.

How RNA structure is regulated *in vivo* had remained elusive, although this information is essential to revealing hidden roles of RNA structures in gene expression regulation. Our study presents the first dynamic landscape and regulation of RNA structuromes in mammalian cells. By comparative analysis we showed that the majority of the RNA structures are stable across three locations, suggesting that they have been largely determined since their biogenesis (Figure 3a-b). This structure stability could partially explain the correlations between different RNA events including transcription, translation and RNA decay (Figure 2h). Future studies

involving structure perturbations that uncouple those functional correlations are required to test this hypothesis.

Nevertheless, our analysis has also revealed a large number of dynamic structural sites, which undergo conformational changes as RNAs transit from their sites of transcription on chromatin, are processed in the nucleus, and ultimately decoded in the cytoplasm. A recent study examined mRNA structure changes during zebrafish early embryogenesis and found translation to be a major driving force that shapes the dynamic mRNA structural landscape<sup>49</sup>. Our cytotopic data offers an opportunity to validate the finding in mammalian cells. We found that the structural dynamics between mRNAs in the chromatin fraction and the nucleoplasmic fraction is about the same as that between mRNAs in the nucleoplasmic fraction and the cytoplasmic fraction. Furthermore, the structural dynamics for mRNAs are similar to those for lncRNAs (Figure 3b). As translation remains a possible important biological process that helps to shape RNA structuromes, our observations suggest that other factors may play crucial roles that regulate RNA structure for both mRNAs and lncRNAs, in a similar fashion in mammalian cells.

Among many factors known to influence RNA structure, RNA modification and RBP-binding are important *cis*- and *trans*- regulators. Our comparative analysis illuminates their relative contributions to the observed RNA structure differences in different aspects. *In vivo* (Figure 3b) both RNA modification and RBP-binding are likely different in different compartments, whereas *in vitro* (Figure 3c) there are no RBP-binding to contribute to the structure changes. This difference in regulators may explain why RNA structures are more dynamic *in vivo*. When comparing *in vivo* to *in vitro* structure (Figure 3d) RNA modification should remain unchanged, but there are no RBP-binding to contribute to the structural differences *in vitro*, thus suggesting that RBP-binding as a whole is an important regulator of RNA structure changes. And finally, both of RBP-binding and RNA modification are likely very different in mouse and human, which may account for the big structural divergences in the two species (Figure 3e).

Finally, the specific RNA regions that undergo structural transition at each subcellular location provide direct readouts of the molecular mechanisms that shape the gene expression program. The finding of LIN28A as a m<sup>6</sup>A anti-reader may have implications for human disease, as both LIN28A and m<sup>6</sup>A have been implicated in cancer progression, germ cell development, and metabolism<sup>50</sup>. In the future, studying RNA structure dynamics together with RNA modifications and RBP binding in physiological states, and in the context of biological and

structural perturbations, will help to elucidate the complex regulatory role of RNA structures in biology and medicine.

**Acknowledgments:**

We thank members of the Chang and Zhang labs for discussion. We thank Ryan Flynn for experimental advice, and Chao Dai for computational advice. This work is supported by NIH Grants No. R01-HG004361 and R35-CA209919 (to H.Y.C.), by the National Natural Science Foundation of China (Grants No. 31671355, 91740204, and 31761163007), and the National Thousand Young Talents Program of China (to Q.C.Z.). F.M.F. was supported by a NIH T32 Genomics Training Fellowship and the Arnold O. Beckman Postdoctoral Fellowship. H.Y.C. is co-founder and serves on the SAB of Epinomics and Accent Therapeutics. Some sequencing data was generated on an Illumina HiSeq 4000 that was purchased with funds from NIH (award number S10OD018220). H.Y.C. is an Investigator of the Howard Hughes Medical Institute.

**Author contributions:**

H.Y.C and Q.C.Z. conceived the project. L.S. and F.M.F. performed icSHAPE experiments in human and mouse cell lines respectively. P.L. and L.S. analyzed all the results. J.P.B. performed the Lin28A CLIP-seq experiments. B.L. and L.T. assisted with experiments, and W.H. assisted with analysis. Q.C.Z. and H.Y.C. supervised the project. Q.C.Z., L.S., F.M.F. and H.Y.C. wrote the manuscript with inputs from all authors.

## Figure Legends

**Fig 1 | Chromatin fractions are enriched for pre-mRNA and lncRNA structures.** **a**, Experimental overview of the icSHAPE protocol. **b**, Donut charts showing read distributions of different RNA types in the three cellular compartments. The outer circles represent exon coverage while the inner ones represent intron coverage. **c**, GAS5 RNA secondary structure with icSHAPE reactivity scores shown in color. The nucleotides outlined in red interact with the GR amino acids, as shown in blue. **d**, UCSC tracks showing icSHAPE reactivity scores (y-axis), along the RNA sequence. Here 1 denotes unstructured (single-stranded) regions, and 0 denotes fully-structured regions. **e**, Violin plot of Gini index of icSHAPE data in exon versus in intron.

**Fig 2 | RNA structure plays a central role in connecting transcription, translation and RNA degradation.** **a-d**, Scatter plots of (a) transcription rate versus 5'UTR RNA structure in chromatin, (b) translational efficiency versus 5'UTR RNA structure in cytoplasm, (c) RNA half-life versus full-length-transcript RNA structure in nucleoplasm, and (d) RNA half-life versus RNA structure in cytoplasm. The 2-tailed p-value was calculated by python package function *scipy.stats.pearsonr*. **e**, Radar diagram showing 5'UTR RNA structure in chromatin, 5'UTR RNA structure in cytoplasm, transcription rate, and translational efficiency. Grey lines show all genes, and the colored lines highlight representative transcripts. **f**, Heatmap of 5'UTR RNA structure in chromatin, 5'UTR RNA structure in cytoplasm, transcription rate, and translational efficiency. Each strip represents an average of a bin comprising 5% data, ranked by RNA-structure reactivity in the chromatin fraction, 477 common transcripts are showed here. **g**, Mediator model (above) and confounding model (bottom) of RNA structure in connecting transcription rate and translation efficiency. P-values were calculated by double-ended t-test. **h**, Schematic showing RNA structure connects transcription, translation and RNA degradation.

**Fig 3 | RNA structure differences in cellular context.** **a**, U12 small nuclear RNA (snRNA) structure dynamics across cellular compartments, and the structural divergence in two species. Tracks show the icSHAPE score plotted along the RNA sequence. The black bars highlight RNA structurally-dynamic regions. **b-e**, Heatmaps showing fractions of structurally-different regions across cellular compartments (b) *in vivo*, (c) *in vitro*, (d) between *in vivo* and *in vitro*, and (e), between human and mouse. Dashed lines represent insufficient data.

**Fig 4 | RNA modification and RBP binding underlie RNA structure dynamics.** **a-c**, RNA structure dynamics at (a) an m<sup>6</sup>A-modified site, (b) a Ψ-modified site, and (c) an HNRNPC-binding site. Tracks show the icSHAPE score plotted along the RNA sequence. **d**, Heatmap of average icSHAPE scores in RBP binding regions in different cellular compartments, ranked by increasing structure dynamics (from left to right) between the chromatin and the nucleoplasmic fractions. Proteins are annotated by their known localizations, with chromatin-associated RBPs shown in red. P-values were calculated by single-sided Mann-Whitney U test and corrected by bonferroni method. **e**, The number and overlaps of different types of RNA modification sites and RBP binding sites in RNA structurally-dynamic regions. P-values were calculated by shuffling 1000 times. \* p-value < 0.05; \*\* p-value < 1e-3; \*\*\* p-value < 1e-5

**Fig 5 | Structural analysis dissects different types of m<sup>6</sup>A readers.** **a**, Differential RBP binding to m<sup>6</sup>A sites and control sites containing an m<sup>6</sup>A motif. P-values are calculated to show the statistical significance of the binding differences by single-ended Mann-Whitney U test and corrected by Benjamini/Hochberg method. **b**, Metagene profiles of protein binding in m<sup>6</sup>A-flanking regions. **c**, Metagene profiles showing RNA structures are different between known m<sup>6</sup>A-modified sites and control (unmodified) sites, with the m<sup>6</sup>A motif around the binding sites of IGF2BP3 and HNRNPC. P-values were calculated by single-ended Mann-Whitney U test, red stars mean p-values are less than 0.01. The error bars represent the standard error of the mean. **d**, Violin plots of RBP-binding strengths of HNRNPC and IGF2BP3 in structured and flexible regions containing a m<sup>6</sup>A motif. Structured and flexible regions are defined as the RBP-binding regions at the bottom 30% and top 30% of average icSHAPE scores respectively. P-values were calculated by single-ended Mann-Whitney U test, black stars mean p-values are less than 0.05.

**Fig 6 | Validation of IGF2BP3 as an indirect m<sup>6</sup>A reader and LIN28A as an anti-reader respectively.** **a-b**, RNA pull-down assays and RBP Western blots for (a) IGF2BP3 and (b) LIN28A, using RNA probes that contain unmodified A, m<sup>6</sup>A, and U respectively derived from the indicated positions in the transcripts. M<sup>6</sup>A sites are marked with a red “m”. Histograms show RNA pull-down from three replicates. Western blots are done with anti-IGF2BP3 or anti-LIN28A antibodies respectively after RNA pull down. The error bars represent standard deviation of replicates. **c**, Density plot of LIN28A binding strength (log ratio) at m<sup>6</sup>A sites in *Mettl3* knockout (KO) versus wild-type mES cells. P-value is calculated by single-ended t-test. **d-e**, Signal

tracks of Sox2 (sex determining region Y box 2) and *Nanog* (nanog homeobox) showing LIN28A binding at specific loci in *Mettl3* KO and wildtype mES cells.

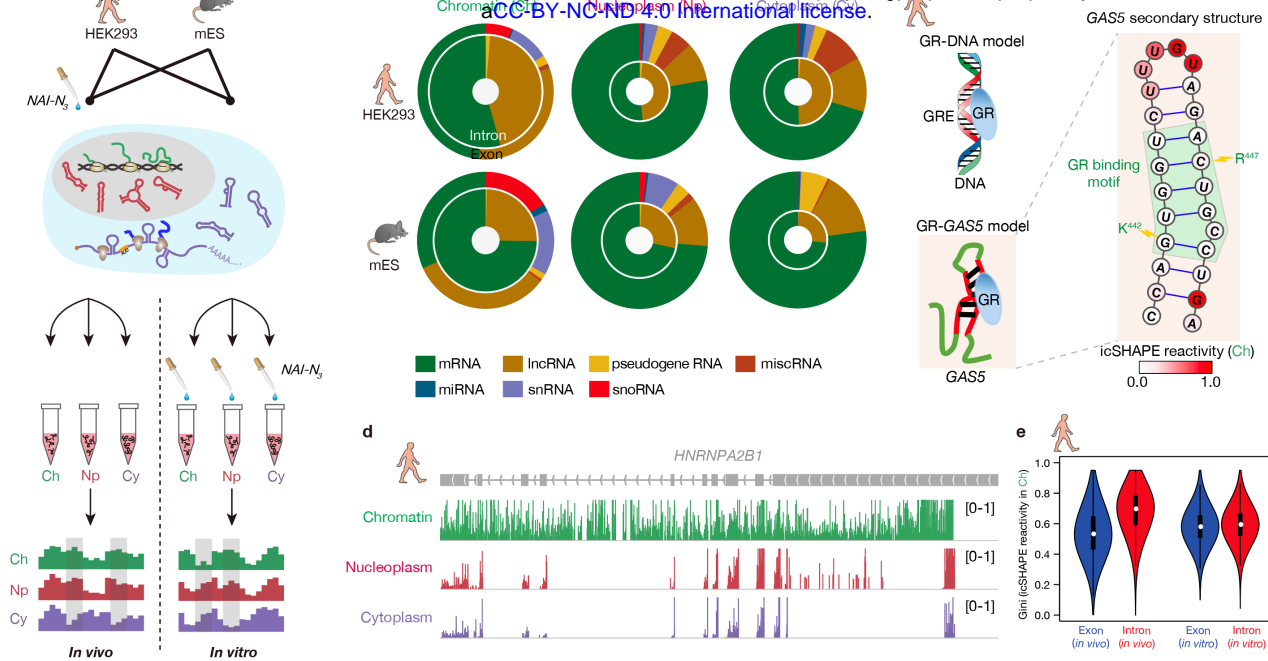
## References

- 1 Sharp, P. A. The centrality of RNA. *Cell* **136**, 577-580, doi:10.1016/j.cell.2009.02.007 (2009).
- 2 Pan, T. & Sosnick, T. RNA folding during transcription. *Annu Rev Biophys Biomol Struct* **35**, 161-175, doi:10.1146/annurev.biophys.35.040405.102053 (2006).
- 3 Warf, M. B. & Berglund, J. A. Role of RNA structure in regulating pre-mRNA splicing. *Trends in biochemical sciences* **35**, 169-178, doi:10.1016/j.tibs.2009.10.004 (2010).
- 4 Martin, K. C. & Ephrussi, A. mRNA localization: gene expression in the spatial dimension. *Cell* **136**, 719-730, doi:10.1016/j.cell.2009.01.044 (2009).
- 5 Kozak, M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361**, 13-37, doi:10.1016/j.gene.2005.06.037 (2005).
- 6 Garneau, N. L., Wilusz, J. & Wilusz, C. J. The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol* **8**, 113-126, doi:10.1038/nrm2104 (2007).
- 7 Ramakrishnan, V. Ribosome structure and the mechanism of translation. *Cell* **108**, 557-572 (2002).
- 8 Yan, C. *et al.* Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* **349**, 1182-1191, doi:10.1126/science.aac7629 (2015).
- 9 Wan, Y. *et al.* Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* **505**, 706-709, doi:10.1038/nature12946 (2014).
- 10 Ding, Y. *et al.* In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**, 696-700, doi:10.1038/nature12756 (2014).
- 11 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, doi:10.1038/nature12894 (2014).
- 12 Spitale, R. C. *et al.* Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**, 486-490, doi:10.1038/nature14263 (2015).
- 13 Lu, Z. *et al.* RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* **165**, 1267-1279, doi:10.1016/j.cell.2016.04.028 (2016).
- 14 Zubradt, M. *et al.* DMS-MapSeq for genome-wide or targeted RNA structure probing in vivo. *Nature methods* **14**, 75-82, doi:10.1038/nmeth.4057 (2017).
- 15 Mustoe, A. M. *et al.* Pervasive Regulatory Functions of mRNA Structure Revealed by High-Resolution SHAPE Probing. *Cell* **173**, 181-195 e118, doi:10.1016/j.cell.2018.02.034 (2018).
- 16 Strobel, E. J., Yu, A. M. & Lucks, J. B. High-throughput determination of RNA structures. *Nature reviews. Genetics*, doi:10.1038/s41576-018-0034-x (2018).
- 17 Mortimer, S. A., Kidwell, M. A. & Doudna, J. A. Insights into RNA structure and function from genome-wide studies. *Nature reviews. Genetics* **15**, 469-479, doi:10.1038/nrg3681 (2014).

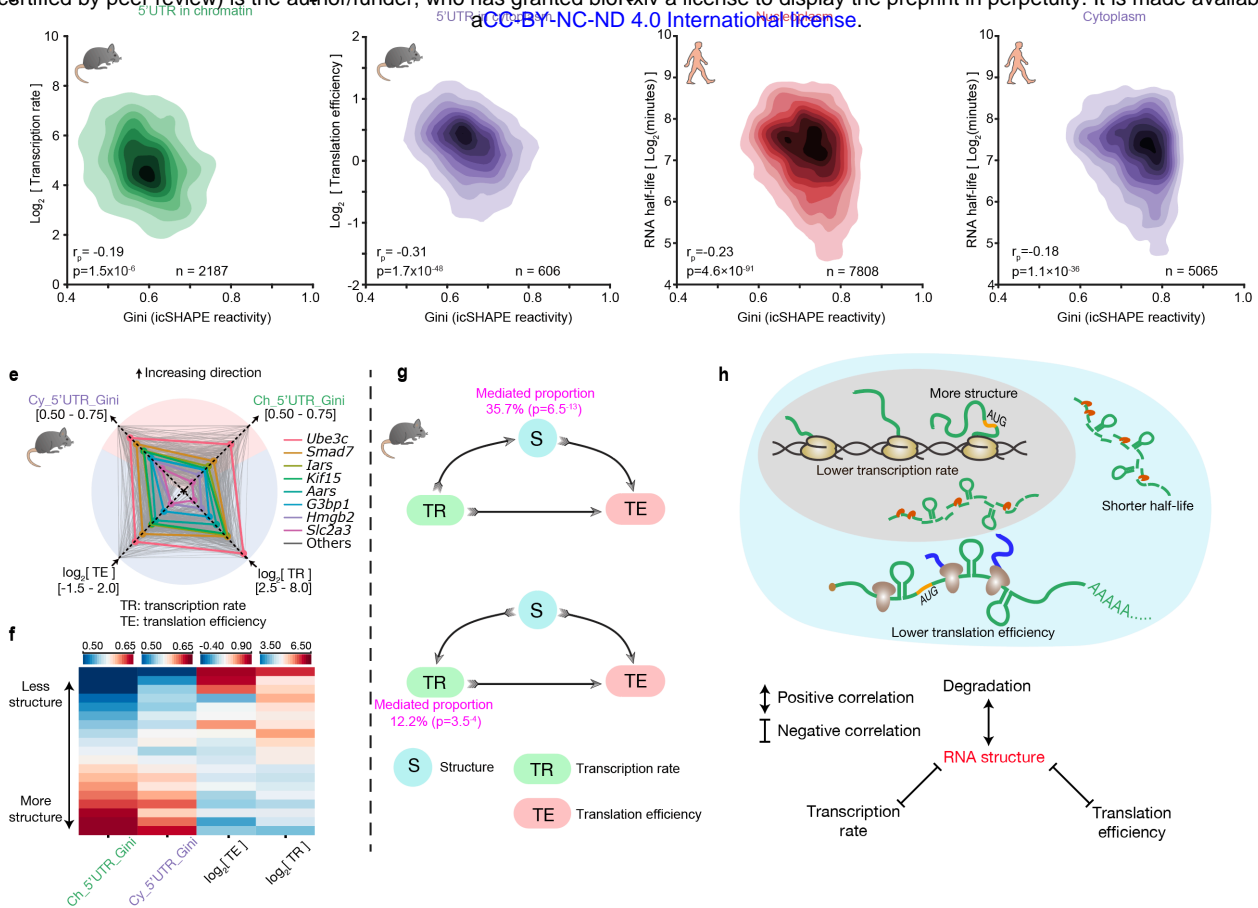


- 18 Bevilacqua, P. C., Ritchey, L. E., Su, Z. & Assmann, S. M. Genome-Wide Analysis of RNA Secondary Structure. *Annual review of genetics* **50**, 235-266, doi:10.1146/annurev-genet-120215-035034 (2016).
- 19 Piao, M., Sun, L. & Zhang, Q. C. RNA Regulations and Functions Decoded by Transcriptome-wide RNA Structure Probing. *Genomics Proteomics Bioinformatics* **15**, 267-278, doi:10.1016/j.gpb.2017.05.002 (2017).
- 20 Wan, Y., Kertesz, M., Spitale, R. C., Segal, E. & Chang, H. Y. Understanding the transcriptome through RNA structure. *Nature reviews. Genetics* **12**, 641-655, doi:10.1038/nrg3049 (2011).
- 21 Buxbaum, A. R., Haimovich, G. & Singer, R. H. In the right place at the right time: visualizing and understanding mRNA localization. *Nat Rev Mol Cell Biol* **16**, 95-109, doi:10.1038/nrm3918 (2015).
- 22 Gosai, S. J. *et al.* Global analysis of the RNA-protein interaction and RNA secondary structure landscapes of the Arabidopsis nucleus. *Molecular cell* **57**, 376-388, doi:10.1016/j.molcel.2014.12.004 (2015).
- 23 Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **169**, 1187-1200, doi:10.1016/j.cell.2017.05.045 (2017).
- 24 Zhao, B. S., Roundtree, I. A. & He, C. Post-transcriptional gene regulation by mRNA modifications. *Nat Rev Mol Cell Biol* **18**, 31-42, doi:10.1038/nrm.2016.132 (2017).
- 25 Kierzek, E. & Kierzek, R. The thermodynamic stability of RNA duplexes and hairpins containing N6-alkyladenosines and 2-methylthio-N6-alkyladenosines. *Nucleic Acids Res* **31**, 4472-4480 (2003).
- 26 Roost, C. *et al.* Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. *J Am Chem Soc* **137**, 2107-2115, doi:10.1021/ja513080v (2015).
- 27 Konig, J. *et al.* iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature structural & molecular biology* **17**, 909-915, doi:10.1038/nsmb.1838 (2010).
- 28 Liu, N. *et al.* N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518**, 560-564, doi:10.1038/nature14234 (2015).
- 29 Alarcon, C. R., Lee, H., Goodarzi, H., Halberg, N. & Tavazoie, S. F. N6-methyladenosine marks primary microRNAs for processing. *Nature* **519**, 482-485, doi:10.1038/nature14281 (2015).
- 30 Gagnon, K. T., Li, L., Janowski, B. A. & Corey, D. R. Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading. *Nat. Protoc.* **9**, 2045-2060, doi:10.1038/nprot.2014.135 (2014).
- 31 Bhatt, D. M. *et al.* Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**, 279-290, doi:10.1016/j.cell.2012.05.043 (2012).
- 32 Flynn, R. A. *et al.* Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. *Nature protocols* **11**, 273-290, doi:10.1038/nprot.2016.011 (2016).
- 33 Kino, T., Hurt, D. E., Ichijo, T., Nader, N. & Chrousos, G. P. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal* **3**, ra8, doi:10.1126/scisignal.2000568 (2010).

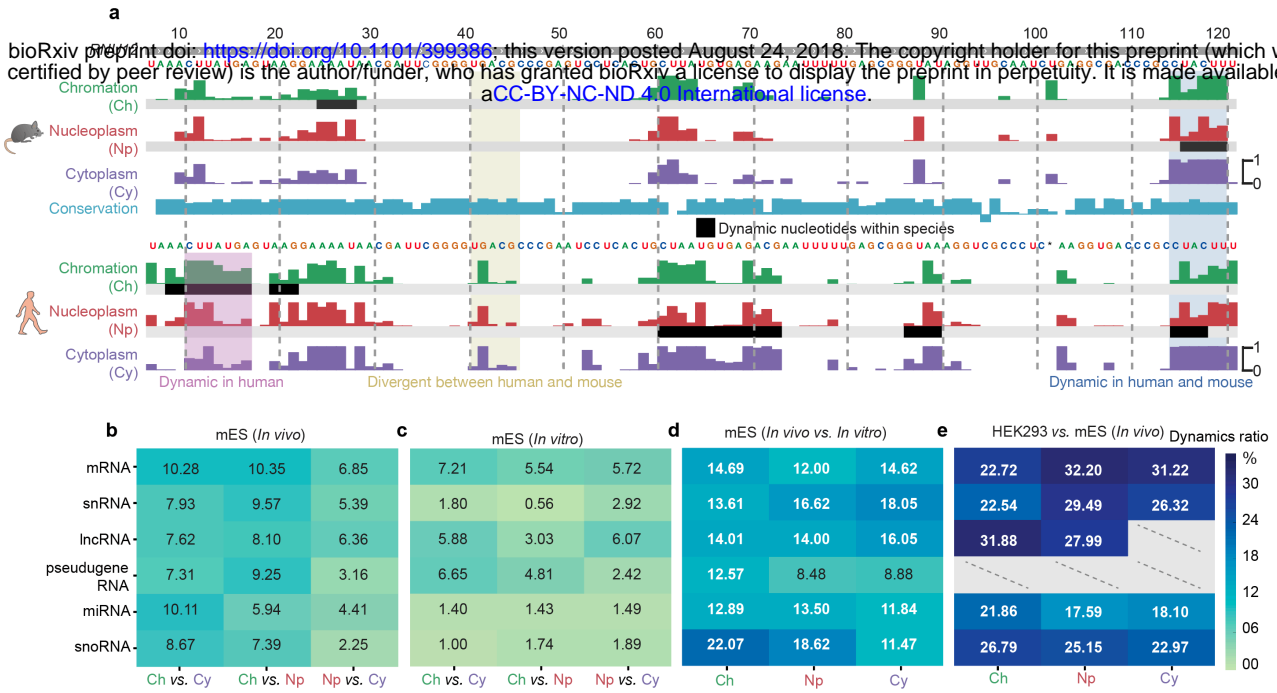
- 34 Schueler, M. *et al.* Differential protein occupancy profiling of the mRNA transcriptome. *Genome Biol* **15**, R15, doi:10.1186/gb-2014-15-1-r15 (2014).
- 35 Jonkers, I., Kwak, H. & Lis, J. T. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* **3**, e02407, doi:10.7554/eLife.02407 (2014).
- 36 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, doi:10.1016/j.cell.2011.10.002 (2011).
- 37 Zhang, J. W. & Landick, R. A two-way street: regulatory interplay between RNA polymerase and nascent RNA structure. *Trends Biochem. Sci.* **41**, 293-310, doi:10.1016/j.tibs.2015.12.009 (2016).
- 38 Houseley, J. & Tollervey, D. The many pathways of RNA degradation. *Cell* **136**, 763-776, doi:10.1016/j.cell.2009.01.019 (2009).
- 39 Harel-Sharvit, L. *et al.* RNA polymerase II subunits link transcription and mRNA decay to translation. *Cell* **143**, 552-563, doi:10.1016/j.cell.2010.10.033 (2010).
- 40 Zid, B. M. & O'Shea, E. K. Promoter sequences direct cytoplasmic localization and translation of mRNAs during starvation in yeast. *Nature* **514**, 117-121, doi:10.1038/nature13578 (2014).
- 41 Slobodin, B. *et al.* Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. *Cell* **169**, 326-337 e312, doi:10.1016/j.cell.2017.03.031 (2017).
- 42 Roundtree, I. A. & He, C. RNA epigenetics--chemical messages for posttranscriptional gene regulation. *Current opinion in chemical biology* **30**, 46-51, doi:10.1016/j.cbpa.2015.10.024 (2016).
- 43 Yang, Y. C. *et al.* CLIPdb: a CLIP-seq database for protein-RNA interactions. *BMC Genomics* **16**, 51, doi:10.1186/s12864-015-1273-2 (2015).
- 44 Huang, H. *et al.* Recognition of RNA N(6)-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nat Cell Biol* **20**, 285-295, doi:10.1038/s41556-018-0045-z (2018).
- 45 Edupuganti, R. R. *et al.* N(6)-methyladenosine (m(6)A) recruits and repels proteins to regulate mRNA homeostasis. *Nature structural & molecular biology* **24**, 870-878, doi:10.1038/nsmb.3462 (2017).
- 46 Batista, P. J. *et al.* m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell stem cell* **15**, 707-719, doi:10.1016/j.stem.2014.09.019 (2014).
- 47 Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917-1920, doi:10.1126/science.1151526 (2007).
- 48 Viswanathan, S. R. & Daley, G. Q. Lin28: A microRNA regulator with a macro role. *Cell* **140**, 445-449, doi:10.1016/j.cell.2010.02.007 (2010).
- 49 Beaudoin, J. D. *et al.* Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. *Nature structural & molecular biology* **25**, 677-686, doi:10.1038/s41594-018-0091-z (2018).
- 50 Shyh-Chang, N. & Daley, G. Q. Lin28: primal regulator of growth and metabolism in stem cells. *Cell stem cell* **12**, 395-406, doi:10.1016/j.stem.2013.03.005 (2013).



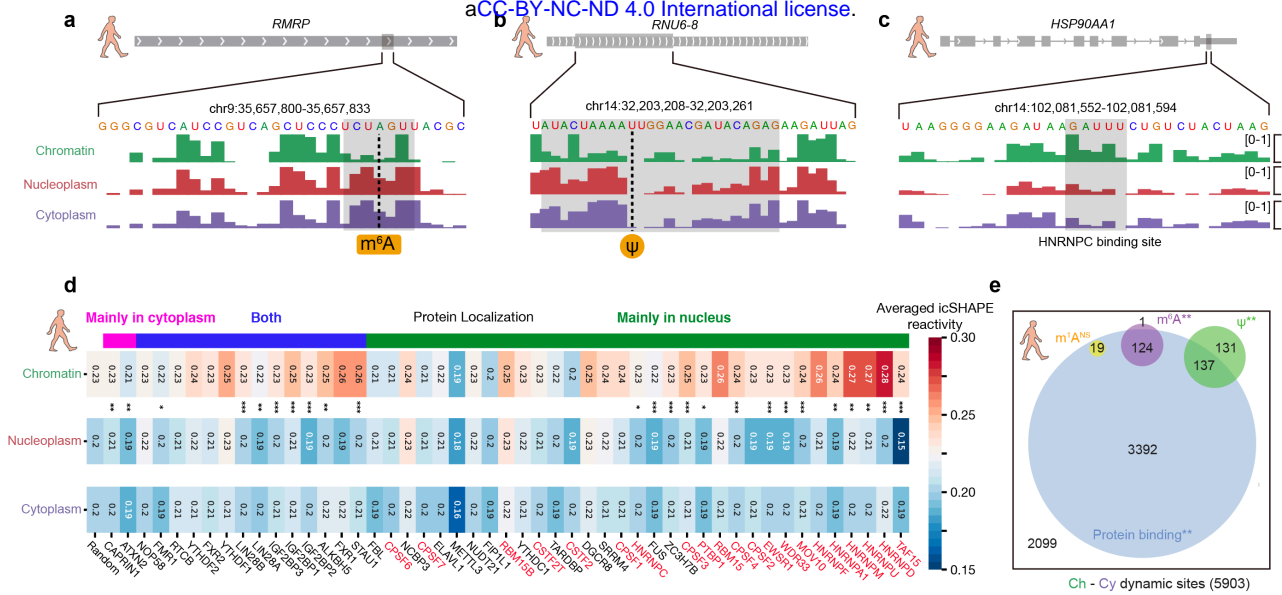
**Fig 1 | Chromatin fractions are enriched for pre-mRNA and lncRNA structures. a**, Experimental overview of the icSHAPE protocol. **b**, Donut charts showing read distributions of different RNA types in the three cellular compartments. The outer circles represent exon coverage while the inner ones represent intron coverage. **c**, GAS5 RNA secondary structure with icSHAPE reactivity scores shown in color. The nucleotides outlined in red interact with the GR amino acids, as shown in blue. **d**, Tracks showing icSHAPE reactivity scores (y-axis), along the RNA sequence. Here 1 denotes unstructured (single-stranded) regions, and 0 denotes fully-structured regions. **e**, Violin plot of Gini index of icSHAPE data in exon versus in intron.



**Fig 2 | RNA structure plays a central role in connecting transcription, translation and RNA degradation.** **a-d**, Kernel density estimation (KDE) plots of (a) transcription rate versus 5'UTR RNA structure in chromatin, (b) translational efficiency versus 5'UTR RNA structure in cytoplasm, (c) RNA half-life versus full-length-transcript RNA structure in nucleoplasm, and (d) RNA half-life versus RNA structure in cytoplasm. The 2-tailed p-value was calculated by python package function *scipy.stats.pearsonr*. **e**, Radar diagram showing 5'UTR RNA structure in chromatin, 5'UTR RNA structure in cytoplasm, transcription rate, and translational efficiency. Grey lines show all genes, and the colored lines highlight representative transcripts. **f**, Heatmap of 5'UTR RNA structure in chromatin, 5'UTR RNA structure in cytoplasm, transcription rate, and translational efficiency. Each strip represents an average of a bin comprising 5% data, ranked by RNA-structure reactivity in the chromatin fraction, 477 common transcripts are showed here. **g**, Mediator model (above) and cofounding model (bottom) of RNA structure in connecting transcription rate and translation efficiency. P-values were calculated by double-ended t-test. **h**, Schematic showing RNA structure connects transcription, translation and RNA degradation.

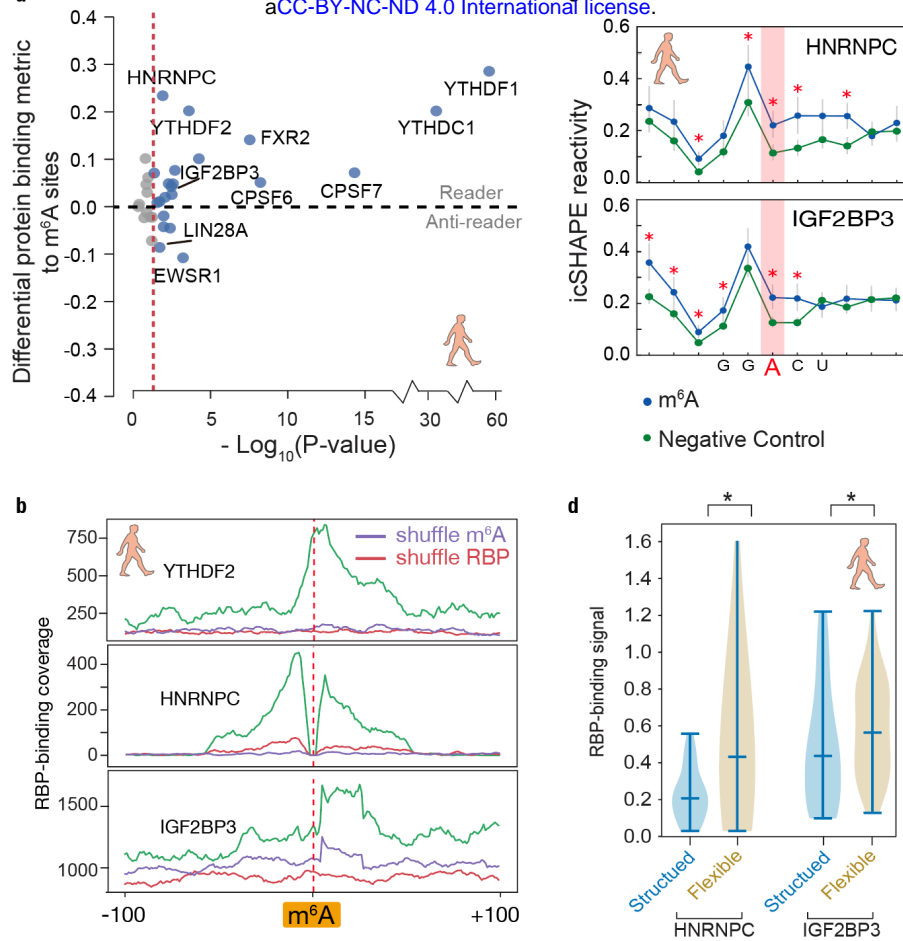


**Fig 3 | RNA structure differences in cellular context. a**, U12 small nuclear RNA (snRNA) structure dynamics across cellular compartments, and the structural divergence in two species. Tracks show the icSHAPE score plotted along the RNA sequence. The black bars highlight RNA structurally-dynamic regions. **b-e**, Heatmaps showing fractions of structurally-different regions across cellular compartments (b) *in vivo*, (c) *in vitro*, (d) between *in vivo* and *in vitro*, and (e), between human and mouse. Dashed lines represent insufficient data.



**Fig 4 | RNA modification and RBP binding underlie RNA structure dynamics.** **a-c**, RNA structure dynamics at (a) an m<sup>6</sup>A-modified site, (b) a Ψ-modified site, and (c) an HNRNPC-binding site. Tracks show the icSHAPE score plotted along the RNA sequence. **d**, Heatmap of average icSHAPE scores in RBP binding regions in different cellular compartments, ranked by increasing structure dynamics (from left to right) between the chromatin and the nucleoplasmic fractions. Proteins are annotated by their known localizations, with chromatin-associated RBPs shown in red. P-values were calculated by single-sided Mann-Whitney U test and corrected by bonferroni method. **e**, The number and overlaps of different types of RNA modification sites and RBP binding sites in RNA structurally-dynamic regions. P-values were calculated by shuffling 1000 times. \* p-value < 0.05; \*\* p-value < 1e-3; \*\*\* p-value < 1e-5





**Fig 5 | Structural analysis dissects different types of m<sup>6</sup>A readers.** **a**, Differential RBP binding to m<sup>6</sup>A sites and control sites containing an m<sup>6</sup>A motif. P-values are calculated to show the statistical significance of the binding differences by single-ended Mann-Whitney U test and corrected by Benjamini/Hochberg method. **b**, Metagene profiles of protein binding in m<sup>6</sup>A-flanking regions. **c**, Metagene profiles showing RNA structures are different between known m<sup>6</sup>A-modified sites and control (unmodified) sites, with the m<sup>6</sup>A motif around the binding sites of IGF2BP3 and HNRNPC. P-values were calculated by single-ended Mann-Whitney U test, red stars mean p-values are less than 0.01. The error bars represent the standard error of the mean. **d**, Violin plots of RBP-binding strengths of HNRNPC and IGF2BP3 in structured and flexible regions containing a m<sup>6</sup>A motif. Structured and flexible regions are defined as the RBP-binding regions at the bottom 30% and top 30% of average icSHAPE scores respectively. P-values were calculated by single-ended Mann-Whitney U test, black stars mean p-values are less than 0.05.

