RNA structure maps across mammalian cellular compartments

Lei Sun1,2,3,4,10, Furqan M. Fazal5,6,7,10, Pan Li1,2,3,4,10, James P. Broughton5,6,7, Byron Lee5,6,7, Lei Tang1,2,3,4,11*, Wenze Huang1,2,3,4, Eric T. Kool8, Howard Y. Chang5,6,7,9,11* and Qiangfeng Cliff Zhang1,2,3,4,11*

RNA structure is intimately connected to each step of gene expression. Recent advances have enabled transcriptome-wide maps of RNA secondary structure, called ‘RNA structuromes’. However, previous whole-cell analyses lacked the resolution to unravel the landscape and also the regulatory mechanisms of RNA structural changes across subcellular compartments. Here we reveal the RNA structuromes in three compartments, chromatin, nucleoplasm and cytoplasm, in human and mouse cells. The cytotic structuromes substantially expand RNA structural information and enable detailed investigation of the central role of RNA structure in linking transcription, translation and RNA decay. We develop a resource with which to visualize the interplay of RNA–protein interactions, RNA modifications and RNA structure and predict both direct and indirect reader proteins of RNA modifications. We also validate a novel role for the RNA-binding protein LIN28A as an N6-methyladenosine modification ‘anti-reader’. Our results highlight the dynamic nature of RNA structures and its functional importance in gene regulation.

RNAs fold into complex structures that are crucial for their functions and regulations including transcription, processing, localization, translation and decay1–6. Over the past 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 available7,8. In recent years technologies have been developed to map RNA secondary structures for the whole transcriptome, that is, RNA structuromes, by combining biochemical probing with deep sequencing9–16. These systems biology studies have revealed many novel insights on the RNA structure basis of gene regulation17–20. However, so far, existing genome-wide structure probing studies have focused on whole-cell data, which only represent 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 cytotic locales41. 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 obscure these critical features. More importantly, detailed mapping of RNA structures in vivo will help to elucidate how they are regulated, which is essential to understanding the RNA structure basis for the regulation of gene expression.

An important driving force that regulates the landscape of RNA structural changes in post-transcription regulation are the RNA-binding proteins (RBPs). A study in Arabidopsis revealed that RNA secondary structure is anti-correlated with protein-binding density22. We recently used icSHAPE to probe RNA structuromes in mouse embryonic stem cells (ESCs) 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 stability23. 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 the regulation of RNA structure. RNA modification can regulate almost all RNA processes including RNA maturation, nuclear retention and exportation, translation, decay, and cell differentiation and reprogramming as well23,24. As one of the most abundant and important types of messenger RNA modification, N6-methyladenosine (m6A) has been shown to favor the unwinding of duplex RNAs by conformational switching25. The impact of the structure-destabilizing effect of m6A is exemplified by a study that investigated heterogeneous nuclear ribonucleoprotein C (HNRNPC), a splicing factor that preferentially binds to single-stranded polyU tracts26. Biochemical studies showed that m6A modification can disrupt the local RNA structures and promote HNRNPC binding in nearby regions27. The study defined these m6A sites as ‘m6A-switches’, and identified the enrichment of tens of thousands of m6A-switches in the vicinity of HNRNPC-binding sites, thereby altering HNRNPC binding and splicing of the target mRNAs. However, whether RNA structural

1MOE Key Laboratory of Bioinformatics, School of Life Sciences, Tsinghua University, Beijing 100084, China. 2Beijing Advanced Innovation Center for Structural Biology, Beijing Frontier Research Center for Biological Structures, Tsinghua University, Beijing 100084, China. 3Center for Synthetic and Systems Biology, Tsinghua University, Beijing 100084, China. 4Tsinghua-Peking Joint Center for Life Sciences, Tsinghua University, Beijing 100084, China. 5Center for Personal Dynamic Regulomes, Stanford University, Stanford, CA, USA. 6Department of Dermatology, Stanford University, Stanford, CA, USA. 7Department of Genetics, Stanford University, Stanford, CA, USA. 8Department of Chemistry, Stanford University, Stanford, CA, USA. 9Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA. 10These authors contributed equally: Lei Sun, Furqan M. Fazal, Pan Li. 11These authors jointly supervised this work: Howard Y. Chang, Qiangfeng Cliff Zhang. *e-mail: howchang@stanford.edu; qczhang@tsinghua.edu.cn
context is a general mechanism for the recognition of other ‘reader’ proteins of m6A and other RNA modifications, is still unclear29.

Here we use in vivo click-selective 2-hydroxyl acylation and profiling experiments (icSHAPE)23, 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 between RNA structure and cellular processes including transcription, translation and RNA decay in the compartment in which they occur. Separately, we could quantify how RNA adopts different conformations across different cellular compartments, which we termed ‘structural change’, and investigate the sophisticated interplay of RNA structural changes, RNA modification and RBP binding.

Results

Cytotopic RNA structure maps substantially expand the scope and comprehensiveness of RNA structures. To investigate the regulation of RNA structural changes in the cell, we performed icSHAPE to measure RNA secondary structure for transcripts isolated from three subcellular compartments and in two species (Fig. 1a). After performing the icSHAPE reaction of living cells (hereafter ‘in vivo’), RNA fractionation20,21 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 ESCs 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 described12,32, after enriching for mRNAs and long noncoding RNAs (lncRNAs) by ribosome depletion, and sequencing the resulting icSHAPE libraries at high depth (approximately 200 million reads per replicate, Supplementary Table 1). We first confirmed the quality of fractionation using quantitative PCR with reverse transcription (RT–qPCR) for landmark RNAs, and immunoblots for specific proteins (Supplementary Fig. 1). We used the icSHAPE pipeline12 (read depth = 100 as threshold) 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 coefficient $r > 0.75$ for the top 60% of the most abundant transcripts in all replicates, Supplementary Fig. 2). As expected, the correlations between replicates are higher than those across fractions (Supplementary Fig. 3a,b), and they are also higher for RNAs that are more abundant (Supplementary Fig. 3c). We also noticed an even distribution of RNA populations that mapped across compartments (Supplementary Fig. 3d and Supplementary Table 2). 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 (Supplementary Fig. 4a,b). Both RNAs are enriched in nucleoplasm, and indeed our nucleoplasmic icSHAPE data closely match the existing structural models.

The chromatin RNA structurome is enriched for lncRNAs (Fig. 1b) and Supplementary Fig. 4c). 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 (34). Indeed, the expected GAS5 RNA structure is accurately recovered in the chromatin fraction, showing low icSHAPE scores for the double-stranded glucocorticoid-receptor-binding motif of the GAS5 RNA, and high reactivity score for the loop region (Fig. 1c). Similar to IncRNAs, small nucleolar RNAs (snoRNAs) and small nuclear RNAs (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 approximately 15–20% of reads in the cytoplasmic fraction (Fig. 1b and Supplementary Fig. 4d). 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 (Fig. 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 (35)), in contrast to in vitro conditions (both with average Gini index 0.6); this result holds true for both human (Fig. 1e) and mouse (Supplementary Fig. 4e). The finding that intronic regions are more folded in vivo is likely not because of differential RBP binding in introns versus exons, as similar trends were observed when all known RBP-binding sites were excluded in the structural comparison (Supplementary Fig. 4f,g), and RBP-binding sites were found to have no bias for intron or exon regions (7.9% of exon regions and 7.2% of intron regions are bound by RBPs). 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 IncRNAs, 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 cytotropic RNA structuromes allowed us to assess the roles of RNA structure (or lack thereof) in association with 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 samples (36-38, 41), 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 (39). We therefore analyzed the relationship between transcription and 5′-untranslated-region (UTR) RNA structures of the chromatin-associated fraction, and found that lower transcriptional rate correlates modestly with more structure (r = -0.19, P = 1.5 x 10^-4; Fig. 2a and Supplementary Fig. 5a,b). Next, many studies have found RNA secondary structure upstream of or at ribosome-binding sites may affect translation differently (37-39). Indeed, we did observe that more 5′-UTR RNA structure correlates with decreased translational efficiency in the cytoplasmic fraction (r = -0.31, P = 1.7 x 10^-46; Fig. 2b and Supplementary Fig. 5c,d). 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 (r = -0.23, P = 4.6 x 10^-46 in nucleoplasm and r = -0.18, P = 1.1 x 10^-40 in cytoplasm; Fig. 2c,d and Supplementary Fig. 5e-h). To further confirm our conclusion, we repeated the above analysis with a higher read depth cutoff (read depth = 200, Supplementary Fig. 5i) and three other datasets (Supplementary Fig. 5j-l). We also observed the same trends in mRNA 5′UTR, CDS and 3′-UTR, suggesting that the degradation is not RNA-region specific and could possibly be targeted by double-stranded ribonuclease (38) (Supplementary Fig. 6a). However, more direct evidence is needed to establish a widespread role of double-stranded ribonuclease-dependent cleavage in transcript turnover.

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 (Fig. 2e,f). The positive link between transcription and translation, two major events in gene expression, has been previously appreciated (39) (Supplementary Fig. 6b). Recent studies have suggested different mechanisms, including m6A modification, that could account for this linkage by imprinting an mRNA transcript during its synthesis and later regulating its translation (34-36). Our data suggest that genome-wide RNA structures formed at chromatin during transcription remain largely unchanged in the nucleoplasm and cytoplasm fractions, and might thus serve as a link between transcription and translation efficiencies. We therefore considered two models to explain our observations—i) 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 (Fig. 2g). Statistical analysis suggests that although 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 (Fig. 2h).

Pervasive RNA structural changes across different cellular compartments. More importantly, cytotropic RNA structuromes also enabled us to examine how RNA adopts different conformations across different cellular compartments, which we term ‘structural change’. Overall, RNA structures seemed slightly more unfolded in the chromatin fraction (Supplementary Fig. 7). As specific regions of an individual RNA can be regulated differently and display different patterns of structural changes, we implemented a statistical method to discover regions of structural variation (Methods and Supplementary Table 3). As an example, we show that U12 snRNA displayed structural-change regions between compartments (Fig. 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 change.

On a genome-wide scale, we found that different RNA categories showed different levels of structural change in vivo (Fig. 3b). To begin to dissect the factors that regulate RNA structural change in cells, we used the same analysis pipeline to evaluate data obtained from fractionated, purified RNA that was refolded in vitro (Fig. 3c), and compared RNA conformational changes observed between compartments in vivo and in vitro. In general, as expected, RNA structures vary less between the compartments in vitro relative to in vivo (comparing Fig. 3b to Fig. 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, micro RNAs (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 (Fig. 3d), and different RNA categories displayed varying levels of structural differences in vivo and in vitro, consistent with previous findings from whole-cell data (36). 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 (Fig. 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 structural changes. RNA modification and RBP binding are important factors that are known to influence RNA structure. To disambiguate their contributions to RNA structural change, we overlaid compartment-specific RNA structures with RNA modifications and RBPs binding sites. Figure 4a–c shows examples of focal conformational changes around known locations of m6A modification, pseudouridylates (Ψ) and HNRNPC binding.

As m6A is well known as an RNA-structure switch favoring unpairing of double-stranded RNA12,28, we compared the genome-wide structures for m6A methylated versus non-methylated sites with the same underlying sequence motif, and confirmed similar patterns of structure destabilization in all three fractions (Supplementary Fig. 8a). Furthermore, the structural differences are largest in the nucleoplasm fraction, consistent with the finding that METTL3–METTL14 complex deposits m6A on nuclear RNA42. Following the structural changes of the same set of m6A 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 (Fig. 4a and Supplementary Fig. 8a). This analysis agrees with that fact that the vast majority of m6A is deposited within the nucleus42. 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 Ψ-modified regions24. We found that in general these regions have higher icSHAPE reactivity (Ψ), suggesting that modification hinders RNA structure folding freely, which again occurs predominantly in the nucleus (Fig. 4b and Supplementary Fig. 8b). We note that our analysis is most powered to detect RNA modification effects where the modification events occur in a more homogenous fashion.
RNA modification and RBP binding underlie RNA structural changes. **Fig. 4** | RNA structural change at an m6A-modified site (a, d), RNA structural change across cellular compartments in vivo (b), in vitro (c), between in vivo and in vitro (d), and between human and mouse (e). Dashed lines represent insufficient data.
that is, the majority of the transcript copy have modifications at the site that occur in a specific cellular location). We may not be able to identify the structure-changing modifications that occur in a highly variable manner for each transcript copy.

All RNAs associate extensively with proteins in cells, and RBP interactions are both sensitive to, and profoundly impact 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 double-stranded RNA (Fig. 4c). We also followed the structural transition 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 (Supplementary Fig. 8c). Our findings also suggest that HNRNPC binding could be a factor that accounts for the structural change around the binding sites. Indeed, there is a significant overlap between HNRNPC-binding and structural-variation sites (Supplementary Fig. 9).

We extended the analysis to all RBPs with binding-site information available from published RBP high-throughput sequencing of RNA isolated by cross-linking immunoprecipitation followed by sequencing (CLIP-seq) experiments. As shown in Fig. 4d, occupancy of many RBPs are linked with RNA-structural changes, whereas others preferentially bind to structurally stable regions of RNA. For example, many chromatin-associated proteins (for example HNRNPD and others shown in red in Fig. 4d) bind to more open RNA regions; these regions become more structured after dissociating from the chromatin and the proteins. By 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 binding after RNA leaves chromatin.

**Fig. 5 | Structural analysis dissects different types of m^6^A readers.** a, Differential RBP binding to m^6^A sites and control sites containing an m^6^A motif. P values are calculated to show the statistical significance of the binding differences by single-sided Mann–Whitney U test and corrected by the Benjamini–Hochberg method. b, Metagene profiles of protein binding in m^6^A-flanking regions. c, Metagene profiles showing that RNA structures are different between known m^6^A-modified sites and unmodified sites (negative control), at m^6^A motifs overlapping a binding site of IGF2BP3 and HNRNPC. P values were calculated by single-sided Mann–Whitney U test; red asterisks indicate P < 0.01. The error bars represent the s.e.m. The numbers of HNRNPC- and IGF2BP3-binding regions are 86 and 56, respectively. d, Violin plots of RBP-binding strengths of HNRNPC and IGF2BP3 in structured and flexible regions containing a m^6^A motif. Structured and flexible regions are defined as the RBP-binding regions at the bottom 30% or top 30% of average icSHAPE scores, respectively. P values were calculated by single-sided Mann–Whitney U test. The numbers of HNRNPC- and IGF2BP3-binding regions in comparison are 137 and 320, respectively.
Thus, by determining the structuromes of multiple cytotoxic localizations, our study provides an estimate of the relative contributions of known modification mechanisms and protein binding to RNA-structural rearrangement (Fig. 4c). Protein binding using existing CLIP-seq data can explain most of the RNA structural change sites (3,392 of 5,903), 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 structural change.

Structural analysis dissects different types of m^6A readers. Identifying RBPs that can read RNA modifications is of fundamental significance in the study of epitranscriptomics. Using our cytotoxic RNA structurome data to filter published CLIP-seq data, we computed the effect that m^6A modification has on protein binding (Methods). Our analysis identified most of the known m^6A readers, including the canonical YTH domain proteins, and the newly identified HNRNPC and LIN28A as an ’anti-reader’ of m^6A modification (Fig. 5c). Similarly, our RNA-structural data suggest that IGF2BP proteins (here IGF2BP3) may also be able to read the structural changes induced by the so-called m^6A-switch (Fig. 5c).

To validate the role of IGF2BP3 as a possible ’indirect reader’ and LIN28A as an ’anti-reader’ of m^6A modification, we selected four endogenous m^6A sites as targets. Each of the four targets contained three variants for the m^6A site—an unmodified nucleotide, an m^6A modification, and an adenosine-to-uracil mutation relative to the unmethylated controls, that IGF2BP3 displays enhanced binding to the m^6A-modified RNAs and uracil mutations relative to the unmethylated controls, confirming IGF2BP3 to be a m^6A-switch reader to the hairpin probes (Fig. 6a and Supplementary Fig. 10a). IGF2BPs contain different RNA-binding domains including two RNA-recognition motifs and four K-homology domains. A recent study suggested that the third and fourth K-homology domains of IGF2BP3 can recognize m^6A directly via a GGAC motif. Our data suggest that IGF2BP3 may also bind different RNA targets in a manner that is dependent on the m^6A-structural switch, akin to the indirect m^6A reader hnRNPC. Further experiments are necessary to validate
the observations in vivo and to dissect the different domains of IGF2BP3 in reading m^A modifications.

Conversely LIN28A displayed reduced binding to the m^A-modified and uracil-mutant target RNAs, supporting the hypothesis that LIN28A is an anti-reader that requires an unmethylated adenosine for binding (Fig. 6b and Supplementary Fig. 10b). To confirm the anti-reader role of LIN28A, we performed LIN28A CLIP-seq experiments in the wild type and the m^A-methyltransferase Mettl3-knockout mouse ESCs48. Many mRNAs containing one or more known m^A site showed increased binding to LIN28A when m^A deposition is abrogated, relative to the negative controls (P = 0.034, t-test; Fig. 6c–e and Supplementary Fig. 10c,d). Increased LIN28A binding is not due to increased mRNA accumulation in Mettl3-knockout ESCs (Fig. 6c–e and Supplementary Fig. 10c,d). LIN28A is an RBP known to enforce ESC pluripotency and suppress ESC differentiation49, while m^A is required for stem cell differentiation50. The negative regulation of m^A on LIN28A binding is consistent with the functional role of the protein. For example, LIN28A is a well-studied inhibitor of primary microRNA processing51, and m^A was recently shown to promote pri-miRNA processing52. Thus, the discovery of LIN28A as an m^A anti-reader potentially utilizes 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. Cytotopic 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^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 the regulation of gene expression. Our study presents the first landscape and regulation of RNA structuromes and their changes in mammalian cells. Using 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 (Fig. 3a,b). This structure stability could partially explain the correlations between different RNA events including transcription, translation and RNA decay (Fig. 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 sites with RNA-structure changes, 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 landscape of mRNA structural changes53. Our cytotic data offer an opportunity to validate the finding in mammalian cells. We found that the structural change between mRNAs in the chromatin fraction and the nucleoplasmic fraction is approximately the same as that between mRNAs in the nucleoplasmic fraction and the cytoplasmic fraction. Furthermore, the structural changes for mRNAs are similar to those for lncRNAs (Fig. 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 (Fig. 3b) both RNA modification and RBP-binding are probably different in different compartments, whereas in vitro (Fig. 3c) there are no RBPs binding to contribute to the structure changes. This difference in regulators may explain why RNA structures are more diverse in vivo. When comparing in vivo to in vitro structure (Fig. 3d) RNA modification should remain unchanged, but there are no RBPs 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 RBP binding and RNA modification are probably very different in mice and humans, which may account for the big structural divergences in the two species (Fig. 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^A anti-reader may have implications for human disease, as both LIN28A and m^A have been implicated in cancer progression, germ-cell development and metabolism54. In the future, studying RNA structural transitions 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.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41594-019-0200-7.

Received: 8 August 2018; Accepted: 7 February 2019; Published online: 18 March 2019

References

1. Sharp, P. A. The centrality of RNA. Cell 136, 577–580 (2009).
2. Pan, T. & Sosnick, T. RNA folding during transcription. Annu. Rev. Biophys. Biomol. Struct. 35, 161–175 (2006).
3. Warf, M. B. & Berglund, J. A. Role of RNA structure in regulating pre-mRNA splicing. Trends Biochem. Sci. 35, 169–178 (2010).
4. Martin, K. C. & Ephrussi, A. mRNA localization: gene expression in the spatial dimension. Cell 136, 719–730 (2009).
5. Kozak, M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene 361, 13–37 (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 (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 splicingosome at 3.6-angstrom resolution. Science 349, 1182–1191 (2015).
9. Wan, Y. et al. Landscape and variation of RNA secondary structure across the human transcriptome. Nature 505, 706–709 (2014).
10. Ding, Y. et al. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature 505, 696–700 (2014).
11. Rousskin, S., Zabraht, 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).
12. Spitale, R. C. et al. Structural imprints in vivo decode RNA regulatory mechanisms. Nature 519, 486–490 (2015).
13. Lu, Z. et al. RNA duplex map in living cells reveals higher-order transcriptome structure. Cell 165, 1267–1279 (2016).
14. Zubrhardt, M. et al. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. Nat. Methods 14, 75–82 (2017).
15. Mustoe, A. M. et al. Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. Cell 173, 181–195 e118 (2018).
16. Strobel, E. J., Yu, A. M. & Lucks, J. B. High-throughput determination of RNA structures. Nat. Rev. Genet. 19, 615–634 (2018).
17. Mortimer, S. A., Kidwell, M. A. & Doudna, J. A. Insights into RNA structure and function from genome-wide studies. Nat. Rev. Genet. 15, 469–479 (2014).
18. Bevilacqua, P. C., Ritchey, L. E., Su, Z. & Assmann, S. M. Genome-wide analysis of RNA secondary structure. Annu. Rev. Genet. 50, 235–266 (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 (2017).
20. Wan, Y., Kertesz, M., Spitale, R. C., Segal, E. & Chang, H. Y. Understanding the transcriptome through RNA structure. Nat. Rev. Genet. 12, 641–655 (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 (2015).
22. Gosai, S. J. et al. Global analysis of the RNA–protein interaction and RNA secondary structure landscape of the Arabidopsis nucleus. Mol. Cell 57, 376–388 (2015).
23. Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA modifications in gene expression regulation. Cell 169, 1187–1200 (2017).
24. Zhao, B. S., Roundtree, I. A. & He, C. Post-transcriptional gene regulation by RNA modifications. Nat. Rev. Mol. Cell Biol. 18, 31–42 (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 (2015).
27. Konig, J. et al. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat. Struct. Mol. Biol. 17, 909–915 (2010).
28. Liu, N. et al. N6-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions. Nature 518, 560–564 (2015).
29. Alarcon, C. R., Lee, H., Goodarzi, H., Halberg, N. & Tava zoie, S. F. N6-methyladenosine marks primary microRNAs for processing. Nature 519, 482–485 (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 (2014).
31. Bhatt, D. M. et al. Transcript dynamics of protamine-like genes revealed by sequence analysis of subcellular RNA fractions. Cell 150, 279–290 (2012).
32. Flynn, R. A. et al. Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. Nat. Protoc. 11, 273–290 (2016).
33. Kino, T., Huri, D. E., Ichijo, T., Nader, N. & Chrousos, G. P. Noncoding RNA: a spring-loaded base modification. Cell 169, 1187–1200 (2017).
34. Schueller, M. et al. Differential protein occupancy profiling of the mRNA transcriptome. Genome Biol. 15, 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 (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 (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 (2016).
38. Housely, J. & Tollervey, D. The many pathways of RNA degradation. Cell 136, 763–776 (2009).
39. Harel-Sharvit, L. et al. RNA polymerase II subunits link transcription and mRNA decay to translation. Cell 143, 552–563 (2010).
Methods

Cell culture and NAI-N, modification in vivo. Human HEK293 cells were bought from Cell Bank, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences. The cells were cultured in DMEM supplemented with 10% FBS at 37 °C with 5% CO₂ in 15 cm plates. V6.5 mouse ESCs were purchased from Novus Biologicals. The cells were cultured at 37 °C with 5% CO₂ in 15 cm plates coated with 0.2% gelatin, in knockout DMEM (Gibco) media supplemented with 15% FBS, 1% PenStrep, 1% MEM non-essential amino acids, 1% Glutamax, 0.2% β-mercaptoethanol, and 0.01% leukemia inhibitory factor (LIF). All cell lines were tested negative for mycoplasma contamination. Cells were cultured to approximately 80–90% confluency, then rinsed, collected and treated with NAI-N as previously described32.

Subcellular fractionation. Subcellular fractionation for HEK293 cells was performed as previously described32 with the following modifications. Cell pellets were first resuspended in cold cytoplasmic lysis buffer using wide orifice tips and incubated on ice for 6 min. The subsequent steps were as described previously32, with the exception that for each collected subcellular fraction, 5% was used for immunoblot analysis, and then 1 ml Trizol LS (Life Technologies) was added to the remaining aliquot for RNA purification using the QIAGEN RNA cleanup protocol.

Fractionation for mouse ESCs was carried out as previously described32. We confirmed the efficacy of fractionation by imaging using DAPI to stain for intact nuclei, and used endoplasmic-reticulum-tracker red (BODIPY TR Glibenclamide, Thermo Fisher Scientific) to confirm removal of endoplasmic-reticulum contaminants in the nuclear fractions. Riboblock RNase inhibitor (Thermo Fisher Scientific) was used to prevent RNA degradation.

Immunoblotting and RT–qPCR. Immunoblotting and RT–qPCR of marker proteins and transcripts for HEK293 subcellular compartments were used to verify the subcellular fractionation results. Immunoblots were performed with antibodies for three proteins—GAPDH (Abcam), SNRP70 (Abcam) and histone H3 (Abcam). Samples of subcellular fractions were boiled at 95 °C for 10 min, then spun at 14,000g for 3 min at room temperature to minimize the influence of sticky DNA (especially in the chromatin samples) on immunoblots. For every immunoblot, 1% of the sample volume was used. For RT–qPCR, the same percentage (1%) of RNA samples was used. Then the ratio of each marker gene (GAPDH, U1, ACTIN (intron)) was calculated in the respective chromatin, nucleoplasmic and cytoplasmic fractions.

For mouse ESCs, similar immunoblot experiments were carried out using antibodies against actin (Abcam), histone H3 (Abcam) and SNRP70 (Abcam). We confirmed that NAI-N treatment did not affect fractionation. Protein was quantified using the Pierce BCA protein assay kit (Thermo Fisher Scientific). We loaded the protein from each compartment in proportion to the amount obtained (roughly 1:1:2 for chromatin:nucleoplasm:cytoplasm).

NAI-N, modification in vitro. For refolding, purified RNA was first denatured at 95 °C for 3 min, quickly cooled to 4 °C, and then incubated at 53 °C. We used the Pierce BCA protein assay kit (Thermo Fisher Scientific). We loaded the protein from each compartment in proportion to the amount obtained (roughly 1:1:2 for chromatin:nucleoplasm:cytoplasm).

icSHAPE library construction of subcellular fractions. RNA (10μg) from each subcellular fraction for HEK293 cells was depleted of ribosomal RNA using the mouse/human ribonuclease inhibitor (Invitrogen). About 500 ng RNA was recovered in the nucleoplasmic and cytoplasmic samples and about 2 μg RNA in the chromatin fractions. RNA from mouse ESC fractions were depleted using the ribonuclease eukaryotic system v2 kit (Thermo Fisher Scientific). icSHAPE sequencing libraries were then constructed from these RNA samples as previously described32. Libraries of mouse ESCs were sequenced on the Hiseq 2500 to approximately 200 million reads per replicate.

LIN28A plasmid transfection and RNA pull-down. pCMV-Flag human LIN28A vector (30μg) was transfected into 15 cm plates with 90 μl polyethylenimine (PEI) dissolved in 1 ml opti-MEM, following the standard transfection protocol. Fresh medium was added after 6 h, and cells were collected after 48 h. RT–qPCR (Takara) was performed to test the transfection efficiency.

The in vitro RNA pull-down assay was performed as described32. In summary 100 pmol of RNA oligonucleotides were refolded by heating at 90 °C for 1 min, and then incubated at 30 °C for 5 min. HEK293 cells (2×10⁵) were lysed in lysis buffer (150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.5 mM DTT, 50 mM Tris- HCl, pH 7.5, 0.5% sodium deoxycholate) with 10 μl PMSF (Amresco), 10 μl phosphatase inhibitor cocktail (Promega), 2.5 μl SUPERase In inhibitor (Life Technologies) and 2.5 μl Ribonuclease inhibitor (Promega). mRNA-bound streptavidin beads were added into the buffer and incubated at 4 °C for 45 min. The beads were washed first with high salt buffer (50 mM Tris-HCl, pH 7.5, 1 M NaCl, 1% TRITON X-100) at room temperature for 4 min and then 2 more times with low salt buffer, also at room temperature and for 4 min. Proteins were eluted in 15 μl elution buffer (1% SDS, 50 mM Tris-HCl pH 8.0, 1 M NaCl) at 95 °C for 10 min. The eluted protein samples were quantified by immunoblot with IGF2BP3 antibody (Abcam) and LIN28A antibody (Abcam). Blotting membranes were stained by ECL-prime (RPN2232, GE Healthcare) and visualized by a digital imaging system. Control samples were prepared similarly to the lysate samples, with the exception that no RNA oligonucleotides were added. The sequences of IGF2BP3 binding probes and LIN28A binding probes are in the Supplementary Note 1.

CLIP-seq experiments. To carry out CLIP-seq experiments, Metf3 wild type and knockout mouse ESCs32 were grown on gelatinized plates in Knockout DMEM media (Gibco) supplemented with 15% FBS, 1% penicillin–streptomycin (Gibco), glutamine (Gibco), non-essential amino acids, basal medium eagle (Thermo Fisher Scientific), and ESGRO LIF (EMD Millipore). To collect cells for iCLIP31, cells were initially washed with ice-cold PBS, then cross-linked on ice with UV-C (254 nm) at 0.3 cm⁻² in a Stratalinker, washed again with ice-cold PBS (+10 mM EDTA) for 5 min, and removed from the plate.

Immunoprecipitation of cross-linked protein–RNA complexes was performed for 3.5 h at 4 °C with anti-LIN28 antibody (Abcam ab46020). Washing of LIN28 RNA complexes, RNA mixing, on-bead biochemistry, and library generation was performed in triplicate as previously described32. Mouse anti-GFP (Abcam) was used to generate a negative-control library. Reverse transcription of LIN28 bound RNA was performed with SuperScript IV (Thermo Fisher Scientific) for 30 min. Library amplification was done for a total of 14 cycles using Phusion HF polymerase master mix (NEB). Libraries were gel purified and submitted for sequencing. Sequencing were run on NextSeq 500 with custom sequencing primer P6_seq2 (see Supplementary Note 1).

Reads were mapped to mm10 assembly and PCR duplicates were removed using Cuﬄinks (v2.2.1). Reproducible reverse-transcription stop sites were identified using the FAST-iCLIP pipeline32. Lin28a-binding sites were called by Piranha54 with the following parameters: -g / g - genome.fa - genome_size -W -M -F -G -A -O -E -w -transcriptome.fa -d transcriptome.collapsed.info genome.gff

Processed reads were mapped to the transcriptomes by using bowtie2 with icSHAPE suggested parameters (−n non-deterministic --norc). icSHAPE scores were then calculated as previously described, with enough coverage (read depth > 100). To study intron structures, the full-gene sequences with introns were generated from the genome and annotation files. Processed reads were mapped to the full-gene sequences to calculate icSHAPE scores for both exons and introns.

Transcriptional rate, translational efficiency and half-life analysis. Publicly available Ribo-seq (SRR315623) and RNA-seq (SRR315594) datasets were used to calculate translational efficiency for each transcript in mouse ESCs36,57, and a GRO-seq (SRR935117, SRR935118, SRR9342449, SRR94450, SRR942451) dataset was used to calculate transcriptional rate, also in mouse ESCs35,58,59. All sequenced reads were aligned to mouse protein-coding transcripts (including five mouse rRNA transcripts) by using bowtie2 with default parameters. For genes with multiple isoforms, the isoform with the longest coding sequence was chosen as the reference. Translational efficiency for each transcript was calculated as the reads per kilobase of mRNA per million reads (RPKM) of the coding sequence region in the Ribo-seq library, divided by the RPKM of the whole transcript in the RNA-seq library57. Transcriptional rate for each transcript was calculated as its number of transcribed reads (mapping reads divided by the transcripts length) in the GRO-seq library, the data was normalized by sequencing depth to compare between different samples. Pre-calculated half-life data for each transcript in K290 cells were collected from a previous study62.

All correlations and P values were calculated with Python package Seaborn (https://seaborn.pydata.org).

Mediation and confounding-factor analysis. The mediation and confounding-factor analysis for the role of RNA structures in connecting transcription and translation were performed using methods described in the literature60,61. Data of translational efficiency, transcriptional rate and RNA structure (in terms of Gini index of icSHAPE scores of the cytoplasmic 5’-UTR) of 477 mouse transcripts were used for the analysis.

To test whether RNA structure is a mediation factor that is affected by transcription and then affects translation, two linear regressions were performed: \[ T = \beta_0 + \beta_1 T + \epsilon_1 \]
Here $c$ is the regression coefficient relating transcriptional rate (TR) to translational efficiency (TE), and $e$ is the regression coefficient relating transcriptional rate to translational efficiency adjusted for the mediator, that is, RNA structure (S). $i$ is the intercept, $e$ is the regression error, and $b$ is the regression coefficient relating structure (S) to TE.

The value of the mediated effect was estimated by taking the difference in the coefficients, $c - c'$. The mediated proportion $(1 - c'/c)$ is used to measure the effect size of mediated effect. The mediated effect $(c - c')$ divided by the standard error $(\sqrt{\sigma^2 + \frac{\sigma^2}{N-1}})$, here $\sigma$ is the standard error of $c$, $\sigma'$ is the standard error of $c'$, $\nu$ is the correlation of TR and S) was compared to the $t_{nu}$ distribution to determine whether the mediated effect is significant.

To test whether RNA structure is a confounding factor that has an effect on both transcription and translation, another two linear regressions were performed:

$$TE = i + c' TR + b S + e$$

$$TE = i + c S + b TR + e$$

The mediated proportion $(1 - c'/c)$ was estimated, and significance test was conducted the same as described above.

Structural change site analysis. For each compartment, two DMSO replicates and two NAI-N3 replicates were used to calculate four isSHAPE scores combinatorially. For each compartment, two DMSO replicates and two NAI-N3 replicates were used to calculate four ICSHAPE scores combinatorially. For each compartment, two DMSO replicates and two NAI-N3 replicates were used to calculate four ICSHAPE scores combinatorially.

Overlay of RNA modifications, RBP bindings and structural change sites. RBP-binding data by CLIP-seq experiments were collected from CLIPdb14, and mA, mA, m6A and pseudouridylation sites were collected from literature15, 16. Protein localization information was from the website GeneCards (https://www.genecards.org), which hosts the UniProt and COMPARTMENTS localization information. Some literatures were also referred to, especially for the protein localization in the chromatin fraction. To study whether structural-change sites are enriched in RBP-binding regions, mA sites, mA sites or pseudouridylation sites, change sites were randomly shuffled within the transcript for 1,000 times. Thereafter the number of change sites within an RBP-binding region or within the flanking region (10 nt) of a modification site was counted. The real overlapping number was compared with the overlapping numbers in all permutations to calculate the P value: $P < 0.05$, $**P < 0.001$, $***P < 0.0001$.

mA preference of RBPs. True mA modification sites with GGACU motifs were obtained from a published dataset17. To generate a control set of mA modification sites, each transcript with true GGACU mA sites was scanned by the GGACU motif to produce the same number of pseudo-mA sites with the same sequence motif, avoiding 20 nt flanking regions of a true mA site. For each RBP, the binding strength to each binding site was normalized to the computational prediction of RNA-binding residues in protein sequences.

m6A binding peak calling and analysis. Mouse ESC mA modification sites were collected from a published dataset18. Those mA sites with the ratio of m6A modification sites to m6A modification sites in Motif3 knockout versus wild type less than 0.8 were filtered out. Lin28a CLIP-seq experiments were performed by using bowtie2 with default parameters. CLIP binding peaks were called with Piranha19 with parameters “-b 50 -s”. To study the correlation between mA and Lin28a, their genome coordinates were mapped onto the transcriptome. mA sites were shuffled within the same transcripts, keeping the same mA sequence context (keeping the one base upstream and one base downstream of the mA modification site unchanged). RBP-binding sites were also shuffled within the same transcripts. A Lin28a-binding site was defined as overlapped with a mA site if their distance was less than 50 nt.

Statistics and reproducibility. Fig. 3b–e: statistical comparisons of the compartment structural changes were carried out with the single-sided t-test adjusted by the Bonferroni method. Statistical significance was set to adjusted $P \leq 0.05$.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

For null hypothesis testing, the test statistic (e.g. $F$, $t$, $r$) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted.

Give $P$ values as exact values whenever suitable.

For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

Estimates of effect sizes (e.g. Cohen’s $d$, Pearson’s $r$), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

1. bowtie2 2.3.2
2. trimmomatic-0.30.
3. icSHAPE pipeline: https://github.com/qczhang/icSHAPE
4. Python statistic library scipy 0.19.1
5. Python visualization library seaborn 0.8.0
6. Piranha 1.2.1

Data analysis

All scripts can be found on GitHub at: https://github.com/lipan6461188/RNA_Structure_Dynamics.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data generated during the study available in a public repository

mES Ribo-seq data that support the findings of this study have been deposited in GEO with the GSE30839 accession code 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). Figure 2b, S6

mES GRO-seq data that support the findings of this study have been deposited in GEO with the GSE48895 accession code Jonkers, I., Kwak, H. & Lis, J. T. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin,
Data generated during the study available in a public repository HEK293 degradation data that support the findings of this study have been deposited in GEO with the accession code GSE49831. 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). Figure 2c

Data generated during the study available in a public repository HEK293 RNA binding protein data by CLIP-seq experiments that support the findings of this study have been deposited in CLIPdb database. 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). Figure 4e, 5, S5b

Authors can confirm that all relevant data are included in the paper and/or its supplementary information files. The authors declare that the m6A data supporting the findings of this study are available within the paper and its supplementary information files. 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). Figure 6d

Data produced in this study about cellular RNA structure (icSHAPE) in HEK293 cell lines and mES cell lines, and Lin28A binding sites (irCLIP) in mES cell lines have been deposited in GEO with the accession code GSE117840.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences  ☐ Behavioural & social sciences  ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size: Samples sizes were determined according to the published icSHAPE protocol.

Data exclusions: No data were excluded.

Replication: The high-throughput icSHAPE was two replicates. The validation experiments of RNA pull down and irCLIP for m6A reader were three replicates. All the replicates are good consistent.

Randomization: Allocation was random.

Blinding: Yes, the investigator was blinded.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a  Involved in the study

☐ Antibodies

☐ Eukaryotic cell lines

☐ Palaeontology

☐ Animals and other organisms

☐ Human research participants

☐ Clinical data

Methods

n/a  Involved in the study

☐ ChIP-seq

☐ Flow cytometry

☐ MRI-based neuroimaging

Antibodies

Antibodies used

Anti-GAPDH antibody [EPR16891] (ab181602)
SNRP70 antibody (ab51266)
Anti-Histone H3 antibody - Nuclear Loading Control and ChIP Grade (ab1791)
| Validation                                                                 |                                                                 |
|---------------------------------------------------------------------------|------------------------------------------------------------------|
| Anti-GAPDH antibody: [https://www.abcam.com/gapdh-antibody-epr16891-loading-control-ab181602.html](https://www.abcam.com/gapdh-antibody-epr16891-loading-control-ab181602.html) |
| SNRP70 antibody: [https://www.abcam.com/snrp70-antibody-ab51266.html](https://www.abcam.com/snrp70-antibody-ab51266.html) |
| Anti-Histone H3 antibody: [https://www.abcam.com/histone-h3-antibody-nuclear-loading-control-and-chip-grade-ab1791.html](https://www.abcam.com/histone-h3-antibody-nuclear-loading-control-and-chip-grade-ab1791.html) |
| Anti-Lin28A: [https://www.abcam.com/lin28a-antibody-epr4640-ab124765.html](https://www.abcam.com/lin28a-antibody-epr4640-ab124765.html) |
| Anti-IMP3: [https://www.abcam.com/imp3-antibody-epr12021-ab177477.html](https://www.abcam.com/imp3-antibody-epr12021-ab177477.html) |

**Eukaryotic cell lines**

Policy information about cell lines

Cell line source(s) | Mouse ES cells (v6.5 line) from Howard Chang lab  
| human HEK293 cell line: bought from Cell Bank, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai

Authentication | human HEK293 cell line was authenticated by Cell Bank of Shanghai Institutes for Biological Sciences

Mycoplasma contamination | all cell lines are tested negative for mycoplasma contamination

Commonly misidentified lines (See ICLAC register) | *Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*