

# RIC-seq for global *in situ* profiling of RNA–RNA spatial interactions

<https://doi.org/10.1038/s41586-020-2249-1>

Received: 3 September 2019

Accepted: 2 March 2020

Published online: 06 May 2020

 Check for updates

Zhaokui Cai<sup>1,2,5</sup>, Changchang Cao<sup>1,5</sup>, Lei Ji<sup>1,5</sup>, Rong Ye<sup>1,2</sup>, Di Wang<sup>1,2</sup>, Cong Xia<sup>3</sup>, Sui Wang<sup>1,2</sup>, Zongchang Du<sup>1,2</sup>, Naijing Hu<sup>1,2</sup>, Xiaohua Yu<sup>1</sup>, Juan Chen<sup>1</sup>, Lei Wang<sup>1,4</sup>, Xianguang Yang<sup>3</sup>, Shunmin He<sup>1</sup> & Yuanchao Xue<sup>1,2,✉</sup>

Highly structured RNA molecules usually interact with each other, and associate with various RNA-binding proteins, to regulate critical biological processes. However, RNA structures and interactions in intact cells remain largely unknown. Here, by coupling proximity ligation mediated by RNA-binding proteins with deep sequencing, we report an RNA *in situ* conformation sequencing (RIC-seq) technology for the global profiling of intra- and intermolecular RNA–RNA interactions. This technique not only recapitulates known RNA secondary structures and tertiary interactions, but also facilitates the generation of three-dimensional (3D) interaction maps of RNA in human cells. Using these maps, we identify noncoding RNA targets globally, and discern RNA topological domains and *trans*-interacting hubs. We reveal that the functional connectivity of enhancers and promoters can be assigned using their pairwise-interacting RNAs. Furthermore, we show that *CCAT1-5L*—a super-enhancer hub RNA—interacts with the RNA-binding protein hnRNP K, as well as RNA derived from the *MYC* promoter and enhancer, to boost *MYC* transcription by modulating chromatin looping. Our study demonstrates the power and applicability of RIC-seq in discovering the 3D structures, interactions and regulatory roles of RNA.

RNA structures and interactions are usually mediated by RNA-binding proteins (RBPs)<sup>1,2</sup>. In mammalian cells, there are about 1,500 RBPs that bind, guide and modify diverse RNAs<sup>2</sup>. Theoretically, global profiling of all of the RBP-mediated intra- and intermolecular RNA–RNA interactions may enable us to simultaneously deduce the structure and target of any mRNA and long noncoding RNA (lncRNA). Some RBP- or chemical-based transcriptome-wide approaches can detect RNA duplexes globally<sup>3–9</sup>, but several challenges arise from proximity ligation in dilute solutions and the biotinylation of soluble proteins in cell lysates (Supplementary Note). To address these challenges, we have developed the RIC-seq technology, which enables unbiased mapping of RBP-mediated RNA–RNA interactions at single-nucleotide resolution.

## Overview of RIC-seq technology

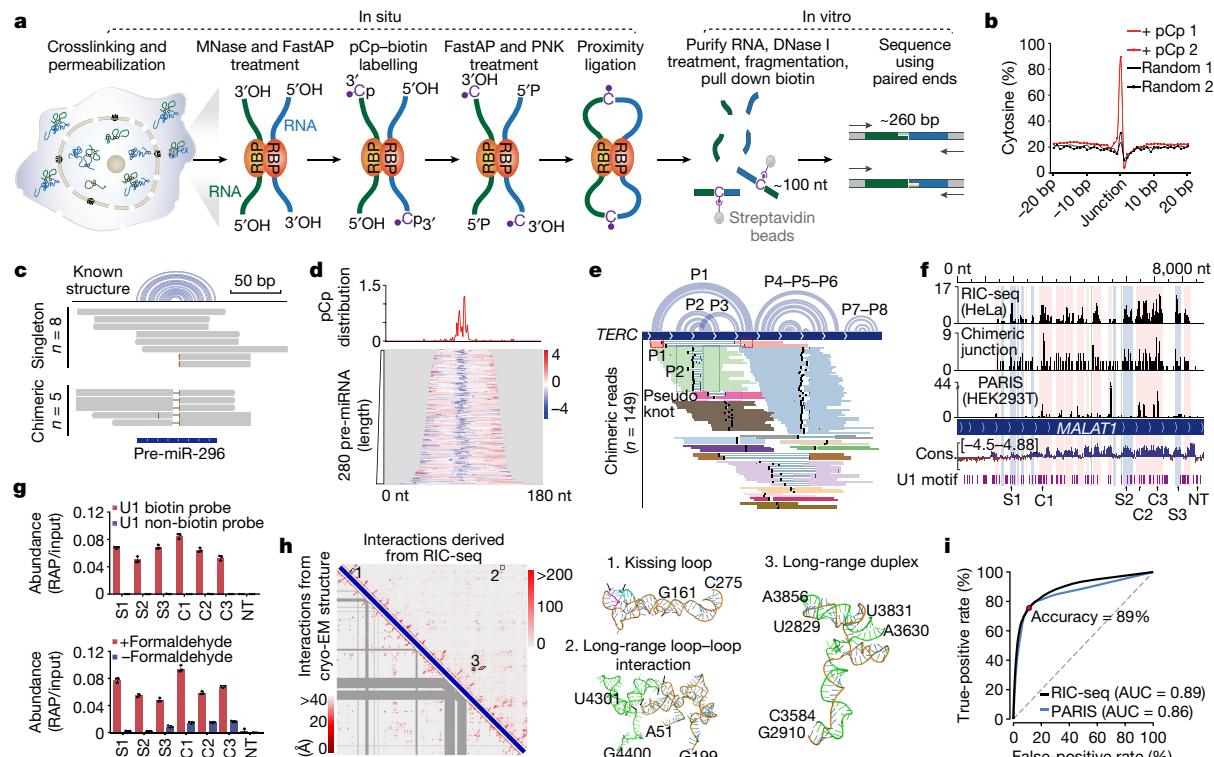
RIC-seq performs RNA proximity ligation *in situ*, and enriches chimeric reads using a biotinylated cytidine (bis) phosphate (pCp-biotin). In brief, the process involves the following steps: (1) HeLa cells are crosslinked with formaldehyde and permeabilized with detergents; (2) RNAs are randomly cut with micrococcal nuclease and dephosphorylated at their 3' overhangs; (3) the RNA 3' end is labelled with pCp-biotin and subsequently treated with FastAP alkaline phosphatase to remove the 3' phosphate group from 'Cp-biotin', and the 5' overhangs are phosphorylated with T4 polynucleotide kinase (PNK); and (4) all

the resulting RNA fragments in close proximity are ligated under *in situ* and non-denaturing conditions (Fig. 1a). In the later *in vitro* stages, total RNA is extracted and fragmented (into fragments approximately 90-nt in length), the RNAs containing C-biotin (about 100 nt) are enriched and then converted into strand-specific libraries for sequencing (about 260 bp in length) (Extended Data Fig. 1a–c).

We generated 170 million and 155 million mappable reads for rRNA<sup>+</sup> and rRNA<sup>-</sup> samples, respectively. Chimeric reads constituted about 9% (rRNA<sup>-</sup>) or about 16% (rRNA<sup>+</sup>) of the uniquely mapped reads (Supplementary Table 1). In rRNA<sup>-</sup> libraries, we obtained 12.8 million chimeric reads, the arms of which mapped to different RNA fragments (Extended Data Fig. 1d–f). The intra- and intermolecular chimeric reads showed a clear enrichment for nascent transcripts, compared with RNA-sequencing (RNA-seq) data and with the signals of these reads after branch points (Extended Data Fig. 1g, h). This enrichment was not due to the loss of cytoplasmic RNA–protein complexes at washing steps, because RIC-seq captured 78.7% of the cytoplasmic RNAs (Extended Data Fig. 1i, j, Supplementary Table 2).

Each chimeric read represented one interaction between two different RNA fragments (Extended Data Fig. 1k). To identify high-confidence intramolecular interactions, we assembled multiple unique chimeric reads into a cluster for each transcript. RIC-seq was highly reproducible between replicates in pairwise-interacting RNAs ( $R=0.929$ ) (Extended Data Fig. 1l). The random ligation rate was assessed by mixing HeLa cells with *Drosophila* S2 cells and then constructed a RIC-seq library

<sup>1</sup>Key Laboratory of RNA Biology, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China. <sup>2</sup>University of Chinese Academy of Sciences, Beijing, China. <sup>3</sup>School of Life Sciences, Henan Normal University, Xinxiang, China. <sup>4</sup>College of Life Sciences, Xinyang Normal University, Xinyang, China. <sup>5</sup>These authors contributed equally: Zhaokui Cai, Changchang Cao, Lei Ji. <sup>✉</sup>e-mail: ycxue@ibp.ac.cn



**Fig. 1 | RIC-seq faithfully captures RNA secondary structures and tertiary interactions.** **a**, Schematic of RIC-seq technology. An RBP-mediated RNA–RNA contact is zoomed in. MNase, micrococcal nuclease. **b**, The percentages of cytosine around the junction of chimeric reads. **c**, Comparison of known structures (blue arcs) of pre-miR-296 to RIC-seq chimeric reads (grey boxes). Blue line, A>C mismatch; red line, pCp. **d**, pCp distribution around the apical loop of 280 pre-miRNAs in HeLa cells. Colour scale, normalized RIC-seq coverage. **e**, RIC-seq recapitulates known structures of *TERC*. The reads shown in boxes support P1, P2 or P2–P3 pseudoknot structures. **f**, RIC-seq identifies U1 binding sites in *MALAT1*. Light pink shading denotes regions detected by both RIC-seq and PARIS. Light-blue-shaded and light-green-shaded regions are revealed either by RIC-seq or by PARIS, respectively. Purple lines, U1 motif.

Arrowheads mark fragments that were chosen for RAP-qPCR. Cons., conservation; NT, non-target region. **g**, RAP-qPCR validates U1-specific binding sites on *MALAT1*. Top, 5'-end biotin-labelled probes can pull-down U1 interacting sites (C1–C3 and S1–S3) on *MALAT1*. Bottom, the U1 biotin probes can pull-down *MALAT1* upon formaldehyde treatment. Data are mean  $\pm$  s.d.;  $n=3$  biological replicates. **h**, Comparing the RNA physical interaction map and in vivo RNA 3D interaction map of 28S rRNA. The dark grey marked regions are absent from the cryo-EM structure. Boxes 1, 2 and 3 illustrate different types of RNA–RNA interaction. **i**, ROC analysis showing the performance of RIC-seq (black line) in detecting the 3D structure of 28S rRNA. Dashed line, random classifier; blue line, ROC curve of PARIS.

for sequencing. We found that the cross-species RNA–RNA intermixing ligation rate was 0.6% (Extended Data Fig. 1m).

## Validation of RNA–RNA interactions

The pCp–biotin labelling and selection were successful because around 90% of the chimeric junctions contained an additional cytosine (Fig. 1b). RIC-seq captured the stem–loop structure of pre-miRNA at single-nucleotide resolution and across a wide range of expression levels (Fig. 1c, d). Thirty-four per cent of the expressed pre-miRNAs ( $n=825$ ) in HeLa cells contained junctions marked with pCp–biotin, of which 73% were localized at  $\pm 5$  nt of the apical loop (Fig. 1d). Additionally, we captured known intra- and/or intermolecular interactions of transfer RNA, small nuclear (sn)RNA, small nucleolar (sno)RNA, *RPPH1* and *TERC* (Fig. 1e, Extended Data Fig. 2a–e).

Besides known interactions between U1 snRNAs and *MALAT1* revealed by psoralen analysis of RNA interactions and structures (PARIS)<sup>5</sup> (referred to as common (C)1, C2 and C3 sites), we identified several additional contact sites (labelled S1, S2 and S3 in Fig. 1f) that are conserved and were validated by RNA antisense purification (RAP) coupled with quantitative (q)PCR (RAP–qPCR) (Fig. 1g). The interactions between U1 snRNAs and *MALAT1* were RBP-dependent, as samples without formaldehyde showed largely diminished RAP–qPCR signals (Fig. 1g). Notably, the additional sites found using RIC-seq showed relatively

lower U1 motif density and weaker PARIS signals than common and/or PARIS-specific sites (Extended Data Fig. 2f–h). Beyond targeting sites at 28S and 18S rRNA, we uncovered 2,549 snoRNA targets on mRNA and confirmed that blocking *SNORD22* led to increased mRNA levels of *BCL2L2* and *SPHK2* (Extended Data Fig. 2i–k).

RIC-seq showed a higher sensitivity and accuracy than previously published methods<sup>3,5–8</sup> for detecting pre-miRNA structures and interactions between U1 snRNAs and *MALAT1* (Extended Data Fig. 3a, b, Supplementary Table 3). Notably, 13.6% of the RIC-seq clusters directly overlapped with PARIS clusters (Extended Data Fig. 3c). The spanning distance and base-pairing energy of the clusters specific to RIC-seq were higher than the overlapping clusters (Extended Data Fig. 3d, e), which suggests that the clusters specific to RIC-seq may preferentially represent RBP-mediated indirect RNA contacts. Furthermore, RIC-seq detected pairwise RNAs that were often colocalized in previously published proximity RNA-seq<sup>10</sup> and APEX-seq<sup>11</sup> datasets, and the intra-compartment RNA–RNA interactions showed stronger RIC-seq signals than those of inter-compartment interactions (Extended Data Fig. 3f–i).

## Capturing 3D structure of rRNA and lncRNA

To assess whether chimeric reads in rRNAs could capture their tertiary interactions (Extended Data Fig. 4a), we built a physical interaction

map on the basis of the spatial distance between every pairwise 5-nt bins in the cryo-electron microscopy (cryo-EM) model of 28S rRNA<sup>12</sup> (Fig. 1h). We also generated a 3D interaction map on the basis of the RIC-seq data (Fig. 1h, Extended Data Fig. 4b). RIC-seq detected spatial interactions mainly within 25 Å, and captured the proximal sequence within 10 nt of known interaction sites (Extended Data Fig. 4c), which agrees well with the similarity of our two maps. Using the 3D map, we identified 95 long-range loop-loop interactions without base-pairing potential, six base-paired kissing loops, 179 non-Watson-Crick base pairs and many long-range duplexes (Fig. 1h).

We next generated a receiver operating characteristic (ROC) curve using true-positive and true-negative datasets to quantify the performance of RIC-seq for detecting 28S rRNA structures (Extended Data Fig. 4d). At a threshold of 0.07, the true-positive rate, false-positive rate and accuracy were 75.5%, 11.2% and 89%, respectively (Fig. 1i). PARIS yielded a comparable value for the area under the curve (AUC) (Fig. 1i). Notably, 28S rRNA preferentially formed local duplexes, and RBPs showed marked enrichment at the nearby regions (for example, G4000 duplex) (Extended Data Fig. 4e, f). Besides detecting known contacts between 5.8S, 28S and 18S rRNAs (Extended Data Fig. 4g), RIC-seq data also enabled us to reconstruct 28S rRNA structures in the unclear regions of the cryo-EM model, such as nucleotides 2951–3246 and 3301–3561. We found that these two regions form long and stable hairpins (Extended Data Fig. 4h, i). Furthermore, RIC-seq recapitulated the 3D structure of lncRNAs (the 7SK and SRP lncRNAs, as well as *RPPH1*) and showed AUC values comparable to those generated using PARIS (Extended Data Fig. 4j–m).

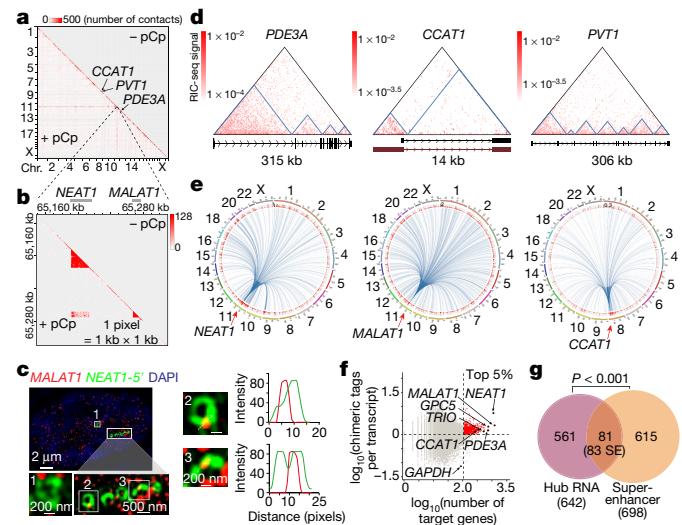
## Global view of RNA–RNA interactions

We next constructed a transcriptome-wide RNA 3D interaction map using Juicebox<sup>13</sup>. For this purpose, we first defined reliable intermolecular interactions following a Monte Carlo simulation approach<sup>10</sup>. The observed pairwise counts were compared with simulated counts to identify significant proximal interactions (Extended Data Fig. 5a–d). Compared with the pCp<sup>−</sup> control (that is, samples without pCp-biotin labelling), the pCp<sup>+</sup> dataset contained 0.8 million additional high-confidence RNA–RNA interactions (Fig. 2a, Supplementary Table 4). The three most enriched types were mRNA-to-mRNA, non-coding (nc)RNA-to-ncRNA and mRNA-to-ncRNA (Extended Data Fig. 5e).

Some lncRNAs, such as *NEAT1* and *MALAT1*, showed extensive *trans* interactions (Fig. 2a, b, Supplementary Table 5); these lncRNAs preferentially interacted with highly expressed transcripts, and their binding motifs were similar (Extended Data Fig. 5f, g). Moreover, the RNA targets identified by RIC-seq significantly overlap with their DNA targets, as previously revealed by capture hybridization analysis of RNA targets with sequencing (CHART-seq) in MCF-7 cells<sup>14</sup> ( $P < 5.0 \times 10^{-19}$ ) (Extended Data Fig. 5h). We found that *MALAT1* mainly bound the 5' part of *NEAT1* (Fig. 2b). Single-molecule fluorescence *in situ* hybridization (smFISH) analysis with probes targeting distinct regions of *NEAT1* (5', middle and 3') revealed that about 63.7% of the ring-like structures in the 5' part of *NEAT1* colocalized with *MALAT1* (Fig. 2c, Extended Data Fig. 5i). The middle and 3' regions of *NEAT1* showed greater spatial distance from *MALAT1* than did the 5' part of *NEAT1* (Extended Data Fig. 5j). Beyond illustrating how RIC-seq can globally identify lncRNA targets, these results also demonstrate the value of the positional information gained by RIC-seq.

## Features of RNA spatial interaction

From the RNA 3D map in HeLa cells (Fig. 2a), we identified 2,307 RNA topological domains (dense self-interacting regions) with a median size of 50 kb, which were present in 8% of the expressed genes (Fig. 2d, Extended Data Fig. 6a, b, Supplementary Table 6). Although their

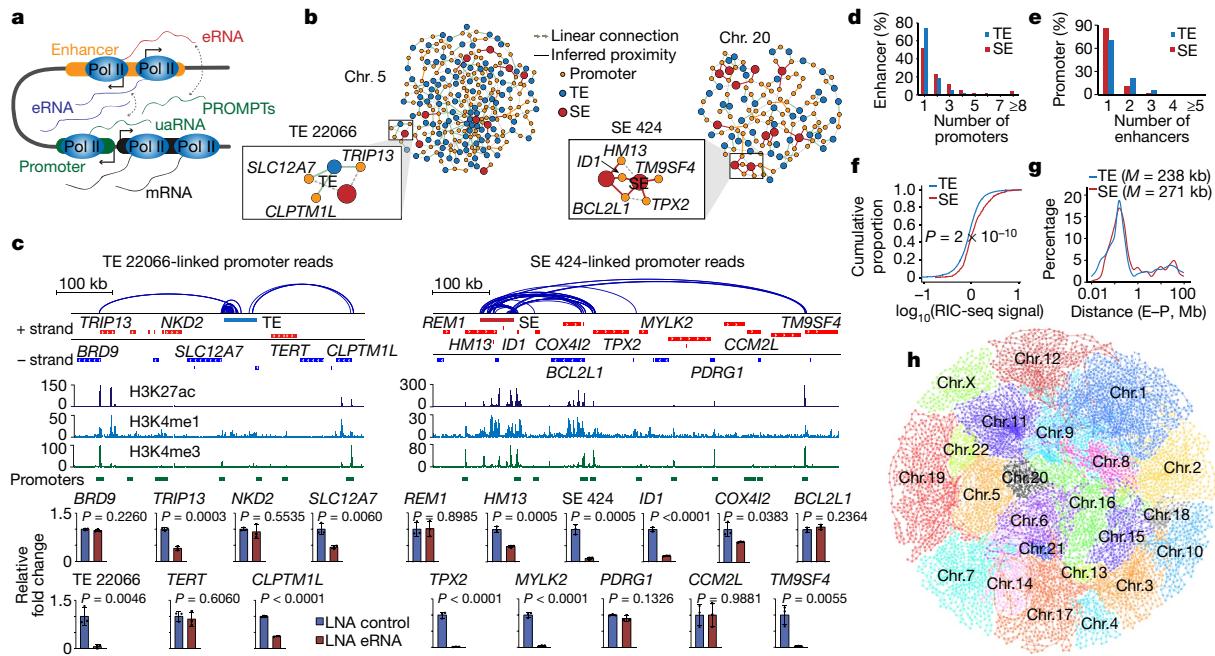


**Fig. 2 | Global view of protein-mediated RNA–RNA interactions.** **a**, RNA 3D map showing RNA–RNA interactions across all chromosomes. **b**, *NEAT1* and *MALAT1* interactions magnified from **a**. **c**, Structured illumination microscopy analysis showing the co-localization of *MALAT1* and 5' region of *NEAT1* (*NEAT1-5'*). The white boxes mark regions that are shown at higher magnification in the insets. The relative distances for overlapping foci (yellow) are quantified on the right.  $n = 15$  cells. The experiment was independently repeated twice with similar results. **d**, The topological domains (blue triangles) observed in the precursor RNA of *PDE3A*, *CCAT1* and *PVT1* based on RIC-seq data. The gene lengths are labelled at the bottom. A new isoform of *CCAT1* is displayed in maroon. **e**, Circos plot showing the *MALAT1*-, *NEAT1*- and *CCAT1*-interacting RNAs. The outer circles show different chromosomes, and the inner circles show unique contacts after normalizing to RNA abundance. **f**, Each RNA in HeLa cells is sorted on the basis of its chimeric tag intensity and the number of interacting genes. **g**, The overlap between hub RNAs and super-enhancers in HeLa cells; 81 hub RNAs overlap with 83 super-enhancers (SE). A one-sided permutation test was used to calculate the *P* value.

functions are unclear, around 20% of the RNA topological domains may reflect long-range RNA–RNA interactions organized by the spliceosome, as evidenced by the enriched exons at the boundary regions (Extended Data Fig. 6c). Notably, the binding pattern of PTBP1, HuR, hnRNPC, NONO and PSF largely demarcated RNA topological domains, whereas U2AF65, RBM10 and eIF4AIII preferentially bind the boundary region (Extended Data Fig. 6d, e).

By sorting the number of target genes and interaction intensity (chimeric reads divided by abundance), we identified 642 RNA interaction hubs, including *MALAT1*, *NEAT1* and *CCAT1* (Fig. 2e, f, Supplementary Table 7). As RNAs transcribed from these loci resemble a hub that interacts with different RNA fragments globally, we termed them hub RNAs. Seventy per cent of the hub RNAs were derived from coding genes with large introns (>50 kb), such as *PDE3A* (Fig. 2f); only 5% of the hub RNAs originated from lncRNAs and pseudogenes; and hub lncRNAs showed stronger *trans* interactions than hub pre-mRNAs (Extended Data Fig. 7a, b).

Hub RNAs had stronger *trans* interactions and were more conserved than non-hub RNAs (Extended Data Fig. 7c–e). Signals for RNA polymerase II (Pol II), trimethylation at lysine 4 of histone H3 (H3K4me3), and acetylation of lysine 27 of histone H3 (H3K27ac) around the transcription start sites of hub RNAs were slightly lower than for non-hub RNAs, whereas trimethylation of lysine 27 of histone H3 (H3K27me3) signals showed the opposite trend (Extended Data Fig. 7f–i). Additionally, hub RNAs possess the following features: (1) they usually interact with target RNAs from the same chromosome and subcellular fraction (Extended Data Fig. 7j–m); (2) hub RNAs and their targets can be classified into three categories based on 187 known binding



**Fig. 3 | Enhancer–promoter connectivity map based on pairwise-interacting RNAs.** **a**, Cartoon depicting the bidirectional transcription at promoters and enhancers in mammalian cells. uRNA, upstream antisense RNA. **b**, The enhancer–promoter connectivity networks on chromosomes 5 and 20. TE, typical enhancer. The boxed subnetworks are validated in **c**. **c**, qPCR showing changed expression of enhancer-linked promoter genes upon eRNA knockdown with LNA ASOs. The nearby non-linked genes serve as negative controls. Specific enhancer-linked promoter reads are shown as blue arc lines above the genes at positive and negative strands. The blue, red and green boxes represent typical enhancers, super-enhancers and

promoters, respectively. Data are mean  $\pm$  s.d.;  $n = 3$  biological replicates, two-tailed, unpaired *t*-test. **d**, The number of promoters paired with a typical enhancer (blue) or a super-enhancer (red). **e**, The number of typical enhancers or super-enhancers paired with promoters. **f**, Super-enhancers show stronger RIC-seq signals than typical enhancers. Two-sided Kolmogorov–Smirnov test was used to calculate the *P* value. **g**, The distance of a typical enhancer or super-enhancer (E) from paired promoters (P). *M*, median. **h**, Whole-genome enhancer–promoter and promoter–promoter connectivity map in HeLa cells. The individual chromosomes are shown in different colours.

motifs of RBPs<sup>15</sup>, and hub RNAs from the same category often interact with each other (Extended Data Fig. 7n–p); (3) hub RNAs within the same Gene Ontology term do not exhibit preferential interactions (Extended Data Fig. 7q); and (4) 13% of the hub RNAs overlapped with super-enhancers<sup>16</sup> (Fig. 2g). These super-enhancer hub RNAs were enriched for the CCUUCCCC motif, and were occupied by many RBPs (Extended Data Fig. 7r–t).

### Enhancer–promoter RNA interaction map

Long-range enhancer–promoter looping is critical for transcriptional activation<sup>17</sup>; however, correctly assigning enhancer and promoter pairing remains a major challenge in modern biology. Enhancers and promoters are bidirectionally transcribed, generating enhancer (e) RNAs and promoter upstream antisense RNAs (also known as promoter upstream transcripts (PROMPTS))<sup>17</sup>. We asked whether their pairwise-interacting RNAs could be used to infer enhancer–promoter connectivity (Fig. 3a). We obtained 27,320 chimeric reads that supported 7,996 significant ( $P \leq 0.05$ ) enhancer–promoter interactions, and 8,268 chimeric reads that supported 2,622 promoter–promoter interactions. Most of the eRNAs interacted with promoter RNAs via one strand, and were predominantly confined within topologically associating domains<sup>18,19</sup> (Extended Data Fig. 8a, b). These pairwise-interacting RNAs enabled us to construct an enhancer–promoter interaction map for HeLa cells (Extended Data Fig. 8c).

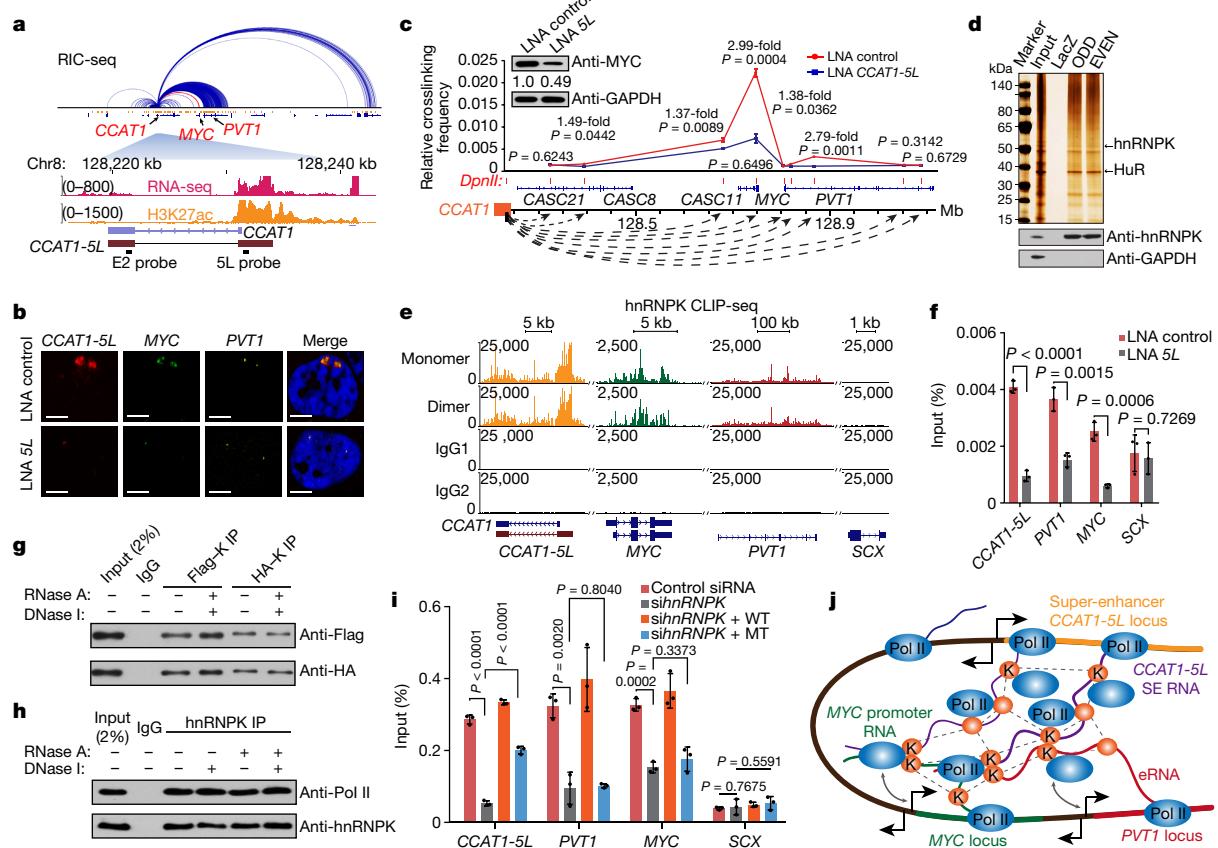
We next used Cytoscape<sup>20</sup> to visualize enhancer–promoter and promoter–promoter interaction networks, as illustrated for chromosome 5 and chromosome 20 (Fig. 3b, Supplementary Table 8). To validate the enhancer–promoter pairings that we deduced, we chose five typical enhancers and two super-enhancers for eRNA knockdown using locked

nucleic acid (LNA) antisense oligonucleotides (ASOs). In contrast to 16 randomly chosen non-target controls, among the 31 linked promoters we found that 27 linked target genes showed reduced transcription, whereas *FPCS* (linked to typical enhancer 30508) showed the opposite trend and levels of *BCL2L1*, *SAE1* and *STXBPI* were not changed (Fig. 3c, Extended Data Fig. 8d, e). Together, these data demonstrate that it is possible to infer the functional pairing of enhancers and promoters from their interacting RNAs.

Compared with typical enhancers, super-enhancers tended to pair with multiple promoters (49% versus 26%) (Fig. 3d). Conversely, most promoters contacted only a single typical enhancer or super-enhancer (Fig. 3e). Super-enhancers had stronger RIC-seq signals and spanned relatively longer distances than typical enhancers to interact with other RNAs (Fig. 3f, g). We speculated that some RBPs might account for this ability (Extended Data Fig. 7r–t). Next, we displayed the genome-wide enhancer–promoter and promoter–promoter networks detected by RIC-seq using Cytoscape (Fig. 3h). Notably, the resulting whole-genome networks revealed the organization of individual chromosomes and recapitulated the reported chromatin territories, as shown by chromosome painting<sup>21</sup>. Although the thorough verification of these complex networks awaits future work, the enhancer–promoter RNA contacts that we have elucidated lay a foundation for the study of the function of eRNAs in the nucleus.

### *CCAT1-5L* modulates chromatin looping

We next studied one hub lncRNA, *CCAT1*, because it overlapped with super-enhancer 638 and has previously been shown to be frequently upregulated in diverse cancers<sup>22</sup>. By 5' and 3' rapid amplification of cDNA ends mapping, we identified an isoform of *CCAT1* that has an extra



**Fig. 4 | The super-enhancer CCAT1-5L regulates MYC transcription via chromatin looping.** **a**, A snapshot of the RIC-seq, RNA-seq and H3K27ac signals around the CCAT1 locus. Northern blot probes, black lines. The chimeric interactions between CCAT1-5L and MYC are shown in red. **b**, CCAT1-5L, MYC promoter RNA and PVT1 are colocalized by smFISH in HeLa cells. 5L, CCAT1-5L. Scale bar, 5  $\mu$ m. **c**, 3C-qPCR analysis of long-distance interactions at the CCAT1-5L, MYC and PVT1 loci upon knockdown of CCAT1-5L RNA with LNA mix. Western blot showing reduced MYC protein levels. Arrows indicate the potential ligation products of *DpnII* restriction fragments. **d**, HnRNP K interacts with CCAT1-5L by ChIP-MS (top) and western blot (bottom). LacZ, control probe; ODD and EVEN, two different sets of antisense oligonucleotides to CCAT1-5L. **e**, HnRNP monomer and dimer binding sites at CCAT1-5L, MYC and PVT1 RNA revealed by CLIP-seq. SCX acts as a negative control. **f**, HnRNP K occupies the CCAT1-5L, MYC and PVT1 loci, as shown by ChIP-qPCR. **g**, Co-IP of Flag-tagged hnRNP K (Flag-K) and HA-tagged hnRNP K (HA-K) protein from the transfected cell lysates. **h**, Co-IP of hnRNP K and Pol II in an RNA- and DNA-independent manner. **i**, ChIP-qPCR showing the occupancy of Pol II at the CCAT1-5L, MYC and PVT1 loci before or after hnRNP K knockdown. sihnRNP K, siRNA against hnRNP K; WT, siRNA-resistant hnRNP K; MT, hnRNP K (T389A/Q391A) mutant. **j**, A model of CCAT1-5L-mediated chromatin looping on the transcriptional activation of MYC. K, hnRNP K; dashed lines, dimerization. Data in **c** are mean  $\pm$  s.e.m. Data in **f** and **i** are mean  $\pm$  s.d.;  $n = 3$  biological replicates, two-tailed, unpaired *t*-test. The experiments in **b**, **d**, **g**, **h** were independently repeated three times with similar results.

revealed by CLIP-seq. SCX acts as a negative control. **f**, HnRNP K occupies the CCAT1-5L, MYC and PVT1 loci, as shown by ChIP-qPCR. **g**, Co-IP of Flag-tagged hnRNP K (Flag-K) and HA-tagged hnRNP K (HA-K) protein from the transfected cell lysates. **h**, Co-IP of hnRNP K and Pol II in an RNA- and DNA-independent manner. **i**, ChIP-qPCR showing the occupancy of Pol II at the CCAT1-5L, MYC and PVT1 loci before or after hnRNP K knockdown. sihnRNP K, siRNA against hnRNP K; WT, siRNA-resistant hnRNP K; MT, hnRNP K (T389A/Q391A) mutant. **j**, A model of CCAT1-5L-mediated chromatin looping on the transcriptional activation of MYC. K, hnRNP K; dashed lines, dimerization. Data in **c** are mean  $\pm$  s.e.m. Data in **f** and **i** are mean  $\pm$  s.d.;  $n = 3$  biological replicates, two-tailed, unpaired *t*-test. The experiments in **b**, **d**, **g**, **h** were independently repeated three times with similar results.

5' extension (of about 4,700 nt), which we named CCAT1-5L (Fig. 4a). Genotyping analysis excluded the integration of human papillomavirus at the 5L-extension locus (Extended Data Fig. 9a). Northern blot analysis confirmed the expression of CCAT1 in diverse cell types, whereas the 5L-specific probe detected CCAT1-5L only in HeLa cells (Fig. 4a, Extended Data Fig. 9b). CCAT1-5L was nuclear-retained and formed 2 or 3 foci per cell (Extended Data Fig. 9c–e). The extra extended region of CCAT1-5L spanned a long distance to interact with MYC promoter RNA and eRNA PVT1<sup>23</sup> (Fig. 4a). These three RNAs and their corresponding DNA loci were colocalized in vivo, as shown by smFISH or DNA FISH, respectively (Fig. 4b, Extended Data Fig. 9f). We next used four LNA ASOs that targeted the 5' extended region to knockdown CCAT1-5L. Compared with non-targeting controls, the RNA and protein levels of MYC were reduced by around 50% and around 70%, respectively (Extended Data Fig. 9g–i). The level of PVT1 was also reduced by about 64% (Extended Data Fig. 9h), which indicates that CCAT1-5L might synergize with the promoter RNA and eRNA of MYC to coordinate PVT1 and MYC expression.

To investigate whether the extra extended region of CCAT1-5L controls MYC transcription, we deleted it using CRISPR–Cas9 and observed a reduced expression of MYC by two- to fivefold (Extended Data Fig. 9j, k).

Furthermore, we used dCas9-KRAB to block the transcription of the extra extended region of CCAT1-5L without affecting the transcription of CCAT1; accordingly, MYC expression was reduced by around threefold (Extended Data Fig. 9l–n). A quantitative chromosome conformation capture assay with qPCR (3C-qPCR) revealed that the chromatin looping between the CCAT1-5L, PVT1 and MYC loci was significantly reduced upon the depletion of CCAT1-5L (Fig. 4c). These data support a model in which CCAT1-5L interacts with MYC promoter RNAs and eRNAs to modulate long-range chromatin looping.

### CCAT1-5L regulates MYC via hnRNP K

We detected an enriched CCUUCCC motif in CCAT1-5L, MYC and PVT1 RNA (Extended Data Fig. 9o, p). Next, we performed comprehensive identification of RBPs by mass spectrometry (ChIRP-MS) using biotinylated antisense probes for CCAT1-5L, which yielded two RBP candidates: HuR and hnRNP K (Fig. 4d). We focused on hnRNP K for its previously reported roles in transcriptional activation<sup>24</sup>. Crosslinking immunoprecipitation and sequencing (CLIP-seq) confirmed that hnRNP K intimately bound CCAT1-5L, MYC promoter and PVT1 RNA, but did not bind SCX (Fig. 4e). Chromatin immunoprecipitation with qPCR

# Article

(ChIP-qPCR) further revealed *CCAT1-SL*-dependent hnRNP binding at *CCAT1-SL*, *MYC* and *PVT1* loci (Fig. 4f). Notably, hnRNP has previously been shown to occur in dimeric and multimeric forms in vitro via two critical amino acids (T389 and Q391)<sup>25</sup>. We verified that hnRNP formed homodimers via co-immunoprecipitation (co-IP) assays with cotransfected Flag- and HA-tagged hnRNP (Fig. 4g). Moreover, both monomeric and dimeric hnRNP bound similar RNA repertoires via the CCUCCCC motif (Extended Data Fig. 9q, r). Consistently, knocking down *hnRNP* decreased *MYC* expression by twofold and also reduced the looping frequency between *CCAT1-SL* and *MYC* loci (Extended Data Fig. 9s).

To uncover how hnRNP promotes transcription, we checked the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database and noted a potential interaction between hnRNP and Pol II. Co-IP assays confirmed that hnRNP interacted with Pol II independently on RNA and DNA (Fig. 4h). Moreover, knockdown of *hnRNP* mediated by small interfering (si)RNA reduced the occupancy of Pol II at the *CCAT1-SL*, *MYC* and *PVT1* loci by 50–75%, and these reductions were rescued by ectopic expression of a siRNA-resistant variant of hnRNP (Fig. 4i). HnRNP(T389A/Q391A), a dimerization-dead variant, rescued Pol II occupancy at the *CCAT1-SL* locus by about 70% but did not rescue the Pol II loading defects at the *MYC* and *PVT1* loci (Fig. 4i, Extended Data Fig. 9t). These data indicate that hnRNP oligomerization is required for Pol II loading at the *MYC* promoter.

Lastly, we investigated whether the oncogenic effects of *CCAT1-SL* depend on *MYC*. Ectopically expressed *CCAT1-SL* was colocalized with *MYC* promoter RNA and the DNA locus, and increased *MYC* expression by 31% (Extended Data Fig. 10a–c). Overexpression of *CCAT1-SL* increased proliferation by 32%, whereas *CCAT1-SL* knockdown reduced cell proliferation by 27% (Extended Data Fig. 10d). These phenotypes were fully rescued via knockdown or overexpression of *MYC*, respectively (Extended Data Fig. 10e). Additionally, *CCAT1-SL* also regulated the metastatic ability of HeLa cells via *MYC* (Extended Data Fig. 10f).

## Discussion

Beyond detecting duplexes and long-range loop–loop interactions, we have shown that RIC-seq can unbiasedly map ncRNA targets, including lncRNAs, snoRNAs and eRNAs. Our study reveals that super-enhancer hub lncRNA *CCAT1-SL* physically interacts with hnRNP and binds *MYC* promoter RNA and eRNA; these interactions are necessary for chromatin looping between the *CCAT1-SL*, *MYC* and *PVT1* loci. Because the oligomerization of hnRNP is required for Pol II occupancy at the *MYC* and *PVT1* loci, we speculate that the hnRNP oligomerization process may function to deliver Pol II from super-enhancer loci to the *MYC* promoter to activate its transcription (Fig. 4j). In summary, this application highlights the value of RIC-seq technology to expedite the discovery of RNA-mediated transcriptional regulation and nuclear organization in the 3D space of the nucleus.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2249-1>.

1. Butcher, S. E. & Pyle, A. M. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.* **44**, 1302–1311 (2011).
2. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829–845 (2014).
3. Aw, J. G. et al. In vivo mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol. Cell* **62**, 603–617 (2016).
4. Kudla, G., Granneman, S., Hahn, D., Beggs, J. D. & Tollervey, D. Cross-linking, ligation, and sequencing of hybrids reveals RNA–RNA interactions in yeast. *Proc. Natl Acad. Sci. USA* **108**, 10010–10015 (2011).
5. Lu, Z. et al. RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* **165**, 1267–1279 (2016).
6. Nguyen, T. C. et al. Mapping RNA–RNA interactome and RNA structure in vivo by MARIO. *Nat. Commun.* **7**, 12023 (2016).
7. Raman, V., Qiu, R. & Shendure, J. High-throughput determination of RNA structure by proximity ligation. *Nat. Biotechnol.* **33**, 980–984 (2015).
8. Sharma, E., Sterne-Weiler, T., O'Hanlon, D. & Blencowe, B. J. Global mapping of human RNA–RNA interactions. *Mol. Cell* **62**, 618–626 (2016).
9. Sugimoto, Y. et al. hiCLIP reveals the *in vivo* atlas of mRNA secondary structures recognized by Staufen 1. *Nature* **519**, 491–494 (2015).
10. Morf, J. et al. RNA proximity sequencing reveals the spatial organization of the transcriptome in the nucleus. *Nat. Biotechnol.* **37**, 793–802 (2019).
11. Fazal, F. M. et al. Atlas of subcellular RNA localization revealed by APEX-seq. *Cell* **178**, 473–490.e426 (2019).
12. Anger, A. M. et al. Structures of the human and *Drosophila* 80S ribosome. *Nature* **497**, 80–85 (2013).
13. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
14. West, J. A. et al. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell* **55**, 791–802 (2014).
15. Ray, D. et al. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172–177 (2013).
16. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
17. Kim, T. K. & Shiekhattar, R. Architectural and functional commonalities between enhancers and promoters. *Cell* **162**, 948–959 (2015).
18. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
19. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
20. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
21. Bolzer, A. et al. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* **3**, e157 (2005).
22. Xiang, J. F. et al. Human colorectal cancer-specific *CCAT1-L* lncRNA regulates long-range chromatin interactions at the *MYC* locus. *Cell Res.* **24**, 513–531 (2014).
23. Tseng, Y. Y. et al. *PVT1* dependence in cancer with *MYC* copy-number increase. *Nature* **512**, 82–86 (2014).
24. Michelotti, E. F., Michelotti, G. A., Aronsohn, A. I. & Levens, D. Heterogeneous nuclear ribonucleoprotein K is a transcription factor. *Mol. Cell. Biol.* **16**, 2350–2360 (1996).
25. Backe, P. H., Messias, A. C., Ravelli, R. B., Sattler, M. & Cusack, S. X-ray crystallographic and NMR studies of the third KH domain of hnRNP K in complex with single-stranded nucleic acids. *Structure* **13**, 1055–1067 (2005).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

## Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

### Construction of RIC-seq libraries

Construction of the RIC-seq library for sequencing includes crosslinking, permeabilization, MNase digestion, pCp–biotin labelling, proximity ligation, RNA purification, DNase I treatment, RNA fragmentation, pCp-biotin selection and strand-specific library preparation.

**Crosslinking, permeabilization and MNase digestion.** HeLa (ATCC, CCL-2) cells were grown in DMEM (Thermo Fisher, C11965500BT) containing 10% fetal bovine serum and 100 U/ml penicillin–streptomycin (Life Technologies, 15140) at 37 °C in 5% CO<sub>2</sub>. Approximately 1 × 10<sup>7</sup> HeLa cells were used for RIC-seq library construction. Cells were fixed in 10 ml of PBS containing 1% formaldehyde at room temperature for 10 min and then quenched by 0.125 M glycine. After washing three times with ice-cold PBS, cells were scraped into 15-ml tubes and collected by centrifugation at 2,500 rpm for 10 min at 4 °C. Cells were resuspended in 1 ml of permeabilization buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 0.5% NP-40, 0.3% Triton X-100, 0.1% Tween 20, 1× protease inhibitors (Sigma Aldrich, P8340), 2 U/ml SUPERase In RNase inhibitor (Thermo Fisher, AM2694)), mixed briefly and incubated on ice for 15 min. The cell integrity was examined at this step under a microscope. Cells were spun down at 3,500 rpm for 5 min at 4 °C. After washing 3 times with 1× PNK buffer (50 mM Tris-HCl pH 7.4, 10 mM MgCl<sub>2</sub>, 0.2% NP-40), the cell pellet was briefly incubated with 200 µl of 1× MN mixture (50 mM Tris-HCl pH 8.0, 5 mM CaCl<sub>2</sub>, 0.03 U/µl micrococcal nuclease (Thermo Fisher, EN0181)) for 10 min at 37 °C. The reaction was stopped by washing twice with 1× PNK + EGTA buffer (50 mM Tris-HCl pH 7.4, 20 mM EGTA, 0.5% NP-40) and twice with 1× PNK buffer. Cells were pelleted at 3,500 rpm.

**pCp–biotin labelling and proximity ligation.** Cells were gently resuspended in 100 µl of 1× FastAP buffer plus 10 U of FastAP alkaline phosphatase (Thermo Fisher, EFO651) and incubated at 37 °C for 15 min. The cell pellet was washed twice in 1× PNK + EGTA buffer, twice in 1× high-salt wash buffer (5× PBS, 0.5% NP-40) and three times in 1× PNK buffer (50 mM Tris-HCl pH 7.4, 10 mM MgCl<sub>2</sub>, 0.05% NP-40). To deposit pCp–biotin, cells were resuspended in 100 µl of ligation mixture (10 µl of 10× RNA ligase reaction buffer, 6 µl of 40 U/µl RNase inhibitor, 4 µl of 1 mM pCp–biotin (Thermo Fisher, 20160), 100 U of T4 RNA ligase, 20 µl of nuclease-free water and 50 µl of 30% PEG 20000) and incubated overnight at 16 °C. The pCp-labelled cells were pelleted at 3,500 rpm for 5 min at 4 °C and washed 3 times with 1× PNK buffer. Cells were further treated with 10 U of FastAP alkaline phosphatase in 100 µl of 1× FastAP buffer for 15 min. After thorough washing, the cell pellets were resuspended in a PNK mixture (10 µl of 10× PNK buffer A, 15 µl of 10 mM ATP, 10 µl of T4 polynucleotide kinase (Thermo Fisher, EK0032), 65 µl of nuclease-free water) and incubated at 37 °C for 45 min. The PNK treatment was stopped by washing twice with 1× PNK + EGTA buffer and twice with 1× PNK buffer (0.05% NP-40).

For *in situ* proximity ligation, the ligation mixture (20 µl of 10× RNA ligase reaction buffer, 8 µl of 40 U/µl RNA inhibitor (Thermo Fisher, EO0381), 10 µl of 10 U/µl T4 RNA ligase (Thermo Fisher, EL0021), 20 µl of 1 mg/ml BSA, 142 µl of nuclease-free water) was applied directly to the cell pellets and incubated overnight at 16 °C. Cells were pelleted at 3,500 rpm for 5 min at 4 °C and washed 3 times with 1× PNK buffer.

**RNA purification.** To purify RNA from cell pellets, 200 µl of proteinase K buffer (10 mM Tris-HCl pH 7.5, 10 mM EDTA, 0.5% SDS) and 50 µl of proteinase K (Takara, 9034) were applied and sequentially incubated at 37 °C for 60 min and 56 °C for 15 min. The RNA was extracted with

750 µl of TRIzol LS (Thermo Fisher, 10296028) and 220 µl of chloroform following the manufacturer's instructions. After centrifugation at 13,000 rpm for 15 min at 4 °C, the supernatant was transferred to a 1.5-ml Eppendorf tube and precipitated overnight by mixing with 500 µl of isopropanol and 1 µl of glycoblue (15 mg/ml, Thermo Fisher, AM9515) at -20 °C. RNA was pelleted at 13,000 rpm for 20 min at 4 °C and washed twice with 75% ethanol. The RNA pellet was dissolved in 15 µl of nuclease-free water.

**DNase I treatment.** To remove the potential RNA–DNA ligation by-products mediated by T4 RNA ligase, 20 µg of total RNA was treated with a mixture containing 10 µl of 10× RQ1 DNase I buffer, 3 µl of RNAsin and 5 µl of DNase I (Promega, M6101) at 37 °C for 20 min. RNA was then purified by acid phenol:chloroform extraction and ethanol precipitation. The RNA pellets were resuspended in nuclease-free water and quantified using a 2100 Bioanalyzer.

**RNA fragmentation.** The rRNA can be optionally removed using the Ribo-off rRNA Depletion Kit (Vazyme, N406-01), or one can directly proceed to the next step (to probe the rRNA structure and interactions). For fragmentation, 16 µl of total RNA was mixed with 4 µl of 5× first-strand buffer (250 mM Tris-HCl pH 8.3, 375 mM KCl, 15 mM MgCl<sub>2</sub>) and incubated at 94 °C for 5 min, then quickly put on ice.

**pCp–biotin selection.** To enrich pCp–biotin-marked RNA fragments, which include both the chimeric and singleton (the RNA fragments that have been 3'-end-labelled but not ligated) RNAs, 20 µl of preblocked MyOne Streptavidin C1 beads (Thermo Fisher, 65002), 30 µl of nuclease-free water and 50 µl of 2× TWB buffer were applied to each RNA sample and incubated at room temperature for 30 min by rotating. After washing 4 times with 1× TWB buffer, the beads were resuspended in 100 µl of PK buffer (100 mM NaCl, 10 mM Tris-HCl pH 7.0, 1 mM EDTA, 0.5% SDS) and incubated in Thermomixer at 95 °C for 10 min (1,000 rpm). The tube was quickly placed on a magnet stand for 1 min, and the supernatant was transferred to a new EP tube. This elution process was repeated two more times. Finally, the enriched RNAs were extracted with acid phenol:chloroform and precipitated at a final concentration of 300 mM NaCl, with 3 volumes of -20 °C ethanol and 1 µl of glycoblue. RNA was pelleted at 13,000 rpm and washed twice with 75% ethanol before resuspension in 10 µl of nuclease-free water.

**Strand-specific library preparation.** To synthesize the first-strand cDNA, 0.5 µl of N6 primer (0.1 µg/µl) was added to the streptavidin-enriched RNAs and incubated at 65 °C for 5 min, followed by incubation on ice for 2 min. Reverse transcription was performed by the further addition of 5.5 µl of reverse transcription mix (3 µl of 5× first-strand buffer, 1 µl of 10 mM dNTPs, 0.5 µl of 100 mM DTT, 0.5 µl of RNase inhibitor, 0.5 µl of Superscript II reverse transcriptase (Thermo Fisher, 18064-014)) and incubation at 25 °C for 10 min, 42 °C for 40 min and then 70 °C for 15 min. Notably, the majority of biotin-C at the 3' overhangs of non-ligated RNAs may lose the additional cytosine base at this step.

We next created dUTP second-strand cDNA by adding the reaction mixture (10 µl of 5× second-strand buffer (Thermo Fisher, 10812-014), 0.8 µl of 25 mM dNTPs with 80% of the dTTP replaced by dUTP, 20.5 µl of RNase-free water, 0.2 µl of 5 U/µl RNase H (Thermo Fisher, EN0202) and 2.5 µl of *Escherichia coli* DNA polymerase I (Enzymatics, P705-500)) to the first-strand cDNA and incubating at 16 °C for 2 h. The double-stranded (ds)DNA products were purified using 1.8× AMPure beads (Beckman Coulter, A63881) following the manufacturer's protocol and end-repaired with a mixture containing 10× T4 PNK buffer (5 µl), dsDNA (42 µl), 25 mM dNTPs (0.4 µl), T4 DNA polymerase (1.2 µl, Enzymatics, P7080L), Klenow DNA polymerase (0.2 µl, Enzymatics, P7060L) and T4 polynucleotide kinase (1.2 µl, Enzymatics, Y9040L).

# Article

The repaired dsDNA was purified again by  $1.8\times$  AMPure beads and eluted in  $20\text{ }\mu\text{l}$  EB buffer.

To add adenine to the 3' ends of dsDNA, a mixture containing  $2.3\text{ }\mu\text{l}$  of  $10\times$  blue buffer (Enzymatics, B0110L),  $0.5\text{ }\mu\text{l}$  of Klenow exo<sup>-</sup> (Enzymatics, P7010-LC-L) and  $0.5\text{ }\mu\text{l}$  of dATP ( $5\text{ mM}$ ) was mixed with  $19.7\text{ }\mu\text{l}$  of dsDNA and incubated at  $37^\circ\text{C}$  for 30 min. For paired-end library construction, Illumina Y-shaped adaptors ( $2\text{ pmol}$ ) were ligated to the dsDNA using Quick T4 DNA Ligase (Enzymatics, L6030-HC-L) and incubated at  $20^\circ\text{C}$  for 15 min. The ligated DNA product was purified twice with AMPure beads and eluted in  $16.5\text{ }\mu\text{l}$  of EB buffer.

The adaptor-ligated cDNA was mixed with  $1\text{ }\mu\text{l}$  of  $10\text{ }\mu\text{M}$  Illumina PE1.0,  $1\text{ }\mu\text{l}$  of  $10\text{ }\mu\text{M}$  index primer,  $1\text{ }\mu\text{l}$  of  $50\text{ mM}$  MgSO<sub>4</sub>,  $2.5\text{ }\mu\text{l}$  of  $10\times$  Pfx buffer,  $0.4\text{ }\mu\text{l}$  of  $25\text{ mM}$  dNTPs,  $0.4\text{ }\mu\text{l}$  of  $2.5\text{ U}/\mu\text{l}$  Platinum Pfx DNA polymerase (Thermo Fisher, 11708-013) and  $3\text{ }\mu\text{l}$  of USER enzyme (NEB, M5505S). The mixture was incubated at  $37^\circ\text{C}$  for 15 min to allow the USER enzyme to digest the dUTP strand, then at  $94^\circ\text{C}$  for 2 min. PCR was performed with the following programme: 11 to 13 cycles of  $94^\circ\text{C}$  for 15 s,  $62^\circ\text{C}$  for 30 s and  $72^\circ\text{C}$  for 30 s. The 200–450-bp PCR products were purified from agarose gel by MinElute Gel Extraction Kit (Qiagen, 28606). The DNA concentration was measured by Qubit (Thermo Fisher, Q32854) and sequenced using Illumina HiSeq X Ten. We noticed that the dUTP second-strand marking and Illumina adaptor ligation method performed better than the sequential ligation of RNA adaptors for the library construction because we observed a higher pCp ratio at chimeric junctions and better performance in capturing known RNA structures for the dUTP library construction strategy.

## Reverse transcription qPCR, nuclear and cytoplasmic fractionation, and western blot

Reverse transcription qPCR (RT-qPCR), nuclear and cytoplasmic fractionation, and western blot were performed as previously described<sup>22,26,27</sup>. The primer sequences are listed in Supplementary Table 9. The following antibodies were used: MYC antibody (Proteintech, 10828-1-AP, 1:1,000), GAPDH (ABclonal, AC033, 1:10,000), anti-hnRNPK antibody (Bethyl, A300-676A, 1:5,000), anti-RNA polymerase II (Santa Cruz, sc-899, 1:1,000), lamin A/C (Santa Cruz, sc-7293, 1:200), and rabbit anti-mouse IgG monoclonal antibody (HRP conjugate and light-chain-specific, CST, 58802).

## RAP and ncRNA knockdown

RAP of U1 snRNA was performed as previously described<sup>28</sup>. To block snoRNA, we designed two ASOs targeting different regions of *SNORD22* and synthesized two sense oligonucleotides as negative controls (GenePharma). The ASOs or sense controls were electroporated into  $1 \times 10^6$  of HeLa cells using the 4D-Nucleofector System (Lonza). The final concentrations of ASOs were  $8\text{ }\mu\text{M}$ . Cells were collected in 48 h, and the total RNAs were extracted for RT-qPCR analysis. LNA ASOs were designed and synthesized by Exiqon. For *CCAT1-5L* and eRNA knockdown, 100 pmol of LNA oligonucleotides were transfected into 1 well of a 12-well plate using Lipofectamine RNAiMAX (Thermo Fisher, 13778). Cells were collected after 48 h and examined by RT-qPCR and western blot. The ASO sequences are listed in Supplementary Table 9.

## Plasmid construction

*CCAT1-5L* or *MYC* cDNA was inserted into pLenti-CMV-GFP-Hygro (Addgene, 17446) plasmid between *BamH*I (NEB, R3136S) and *Sall* (NEB, R3138T) restriction sites. As a size-matched negative control, we amplified an approximately 4.8-kb fragment from λDNA and inserted it into the same backbone. The Flag-tagged hnRNPK<sup>29</sup> was obtained from G. Zhou as a gift. The HA-tagged hnRNPK was inserted into pcDNA3 between *BamH*I and *EcoR*I (NEB, R3101S) restriction sites. The hnRNPK synonymous variant and dimerization-dead mutant (T389A/Q391A) containing resistant sites to siRNA no. 2 were generated by following a site-directed mutagenesis protocol<sup>27</sup>.

## smFISH, DNA and RNA dual FISH, and northern blot

smFISH probes were designed and synthesized by Biosearch Technologies (<https://www.biosearchtech.com/stellaris-designer>). Human *MALAT1* probes with Quasar 670 dye, *NEAT1* 5' with Quasar 570 dye, *PVT1* intron with Quasar 670 dye, *NEAT1* middle with Quasar 570 dye, *NEAT1* 3' non-contact region (about 1.4 kb away from the 3' end) with Quasar 570 dye and *MYC* promoter with Quasar 670 dye or fluorescein dye were purchased from Biosearch Technologies (SMF-2046-1, SMF-2036-1, ISMF-2071-5, SMF-2037-1, SMF-1063-5, SMF-1065-5 and SMF-1025-5). RNA FISH was performed as previously described<sup>30</sup>. Notably, smFISH revealed 2 or 3 foci for *CCAT1-5L*, which probably reflects the ploidy variation in HeLa cells. DNA FISH probes were prepared by PCR amplification of 1-kb genomic fragments, and then fluorescently labelled using a ULYSIS Nucleic Acid Labelling Kit (Thermo Fisher, U21650, U21660). DNA FISH was performed under similar conditions for smFISH. To perform DNA and RNA dual FISH, cells were first fixed with 4% PFA for 15 min and permeabilized with 70% ethanol. The labelled DNA FISH probes were denatured for 10 min at  $95^\circ\text{C}$ , and the cells were denatured in  $2\times$  SSC containing 70% formamide for 10 min at  $75^\circ\text{C}$ . After that, cells were blocked with salmon sperm DNA and hybridized with 1  $\mu\text{g}$  labelled probes overnight at  $37^\circ\text{C}$ . On the next day, slides were washed 3 times with  $2\times$  SSC for 10 min at  $45^\circ\text{C}$ . RNA FISH was then performed. The photographs were taken using an Olympus FV1200 laser scanning microscope. Structured illumination microscopy was performed on a DeltaVisionOMXV4 system (GE Healthcare) equipped with a 603/1.42 NA Plan Apo oil-immersion objective (Olympus) and three laser beams (405, 568 and 647 nm). The northern blot was performed with the DIG Northern Starter Kit (Roche, 12039672910) according to the manufacturer's instructions.

## PiggyBac-Cas9-mediated genome editing and CRISPR interference of *CCAT1-5L*

We used a web tool (<http://crispr.mit.edu/>) to design single-guide (sg)RNAs for deleting the 5' extra extended region in *CCAT1-5L*. The oligonucleotides were synthesized by BGI (Beijing) and annealed to dsDNA, which were inserted into the pXG1 vector digested by *BpI* (Thermo Fisher, FD1014). To combine two sgRNAs in a single vector, we first used *Nhe*I (NEB, R3131S) and *Xho*I (R0146L) restriction enzymes to cut the pXG1-sgRNA1 plasmid and then used *Xba*I (R0145S) and *Xho*I enzymes to cut out U6-sgRNA2 from a pXG1-sgRNA2 plasmid. The purified U6 sgRNA2 fragment was subsequently ligated to the pXG1-sgRNA1 plasmid between the *Nhe*I and *Xho*I sites. The resulting pXG1-sgRNA1-sgRNA2 was further digested with *Sal*I and *Af*II (R0520S) enzymes to get the products of U6-sgRNA1-U6-sgRNA2 fragment, which was inserted into piggyBac-Cas9 plasmid<sup>31</sup> between the *Sal*I and *Af*II restriction sites. We screened two working sgRNAs targeting the upstream region, and three sgRNAs targeting the downstream region, of *CCAT1-5L*. With these working sgRNA sequences, we constructed seven piggyBac-Cas9-2×sgRNA plasmids, which we termed : NC1 + NC2 (two non-targeting sgRNA repeats), Up1 + Down (Dn)2, Up1 + Dn3, Up1 + Dn6, Up2 + Dn2, Up2 + Dn3, and Up2 + Dn6. To establish mutant HeLa cell lines, we cotransfected the PiggyBac-Cas9-2×sgRNA-DsRed and super PiggyBac transposase plasmids into  $4 \times 10^6$  HeLa cells using Lipofectamine 2000 (Invitrogen, 11668019). The transfected cells were sorted with DsRed signals by fluorescence-activated cell sorting (FACS) and selected with  $100\text{ }\mu\text{g}/\text{ml}$  hygromycin B (Invitrogen, 10687010) for 10 days. We performed genotyping with two primers flanking the two deletion sites. All of the successful mutant cell lines were confirmed by Sanger sequencing at BGI.

We used a reported two plasmid system for CRISPR interference (CRISPRi)<sup>32</sup> study: pU6-sgRNA EF1Alpha-puro-T2A-BFP (Addgene, no. 60955) and pHRSFFV-dCas9-BFP-KRAB (Addgene, no. 46911). To block the transcription elongation of Pol II, we designed four sgRNAs targeting the upstream region (up, about 450 bp downstream of the

transcription start site (TSS)) and four sgRNAs targeting the downstream region (Dn, about 2.4 kb from the TSS) of *CCAT1-5L*. The sgRNAs were individually inserted into pU6-sgRNA EF1Alpha-puro-T2A-BFP plasmid between *Bst*XI (NEB, R0113S) and *Xba*I restriction sites. We first made lentiviral particles carrying dCas9-KRAB, as previously described<sup>27</sup>, and infected HeLa cells to generate a stable cell line by FACS with BFP. Next, we prepared lentiviral particles carrying sgRNAs and transduced into the stable cell lines. After selection with puromycin (TaKaRa, 631305) for five days, we collected samples and examined the expression of *CCAT1-5L* and MYC.

### Co-IP, CLIP-seq and ChIP-qPCR

Co-IP, CLIP-seq and ChIP-qPCR were performed as we previously described<sup>27</sup>. The following antibodies were used: anti-hnRNP K (Santa Cruz, sc-28380), anti-hnRNP K (Bethyl, A300-676A), anti-Pol II antibody (Santa Cruz, sc-56767), anti-Flag (Sigma, F1804) and anti-HA (Sigma, H9658). For CLIP-seq, we followed the irCLIP protocol and synthesized a preadenylated 3' DNA linker<sup>33</sup>: 5' rAppATCTCGTATGCCGTCTCTG CTTGTAAAAAAAAAA/iAzideN/AAAAAAAAAAAAA/3Bio/-3'. An infrared dye was further conjugated to the linker by click-chemistry using IRDye-800CW-DBCO (LiCor, cat. no. 929-50000). The final preA-L3-IR800-biotin linker was purified and used for 3' linker ligation. The hnRNP monomer- and dimer-associated RNAs were extracted from the membrane after infrared imaging with a LI-COR Odyssey CLx Imager. Primers used for ChIP-qPCR are listed in Supplementary Table 9.

### Cell proliferation and invasion assays

Cell proliferation was assessed by the MTS Kit (Promega, G3580). After transfection with LNA oligonucleotides or plasmids for 48 h, HeLa cells were seeded into a 96-well plate. MTS reagent was then directly added into the culture medium and incubated for 3 h at 37 °C. The absorbance was measured at an optical density of 490 nm. For the colony formation assay, the transfected cells were seeded into a 6-well plate and maintained in DMEM containing 10% FBS for 2 weeks, during which the medium was changed every 3 days. Colonies were then fixed with cold methanol and stained with 0.1% crystal violet (Sigma, C3886) in PBS at room temperature for 15 min. The stained colonies were counted to determine the colony formation ability. Cell invasion assays were conducted in 24-well fitted inserts (8-μm pore size; Millipore, 3422). After transfection for 48 h, 5 × 10<sup>4</sup> cells were seeded into the upper insert, which was precoated with 200 mg/ml Matrigel (BD Biosciences, 356234), and DMEM containing 10% FBS was added to the lower chamber. After incubation for 36 h, the HeLa cells were fixed and stained with 0.1% crystal violet. The stained cells at the lower surface of the membrane were imaged and counted under a light microscope. The invasion index was calculated as the percentage of invaded cell number to the corresponding control.

### 3C-qPCR

3C-qPCR was performed as previously described<sup>34</sup>. In brief, HeLa cells treated with LNA oligonucleotides were crosslinked with 1% formaldehyde for 10 min and then quenched with 0.125 M glycine. The cells were resuspended in 5 ml of ice-cold lysis buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 5 mM MgCl<sub>2</sub>, 0.1 mM EGTA, 1× complete protease inhibitor) and incubated for 10 min. The nuclei were spun down at 400g for 5 min and digested overnight with 400 U *Dpn*II (NEB, R0543S) at 37 °C. The digested nuclei were transferred to a 50-ml falcon tube, and DNA ligation was performed in a 7-ml ligation mixture containing 100 U T4 DNA ligase (NEB, M0202L). The reaction was kept at 16 °C for 4 h, followed by 30 min at room temperature. The samples were decrosslinked with proteinase K at 65 °C overnight. After treatment with RNase A, the genomic DNA was extracted with phenol-chloroform and precipitated in the presence of a 1/10 volume of 3 M sodium acetate and 2.5 volumes of 100% ethanol. The DNA was pelleted down at 2,200g for 45 min at

4 °C and washed twice with 75% ethanol. The pellet was briefly dried at room temperature and dissolved in 150 μl of 10 mM Tris-HCl pH 7.5. After quantification by NanoDrop, the genomic DNA was diluted to 200 ng/μl for qPCR analysis. The data were normalized to the GAPDH loading control. The primers and TaqMan probes used were listed in Supplementary Table 9.

### ChIP-MS

ChIP-MS was performed as previously described<sup>35</sup>. In brief, 8 × 10<sup>7</sup> of HeLa cells were crosslinked and resuspended in 4 ml of lysis buffer (50 mM Tris-Cl pH 7.0, 10 mM EDTA, 1% SDS, 1× protease inhibitor cocktail, 2 U/ml SUPERase In RNase inhibitor). The lysate was sonicated at the maximum setting (BRANSON SLPe, output setting 4, 10 s per cycle) until the mixture turned clear. The sonicated samples were centrifuged at 13,000 rpm for 10 min at 4 °C, and the soluble chromatin was transferred to a new tube. The 100 pmol of biotinylated ODD, EVEN and LacZ probes were individually applied to 1 ml of soluble chromatin and incubated at 37 °C for 16 h with rotations at a hybridization oven. After the hybridization was completed, 100 μl of streptavidin magnetic C1 beads (Life Technology, 65001) was added to each tube and incubated at 37 °C for an additional 30 min. After thoroughly washing, proteins were eluted into LDS sample buffer (Invitrogen, NP0008) and boiled at 95 °C for 30 min with occasional mixing for reverse crosslinking. The eluted proteins were analysed by mass spectrometry or western blot.

### RIC-seq data processing and analysis

**Mapping and filtering of RIC-seq data.** RIC-seq libraries were paired-end-sequenced at Novogene. The analysis strategy is illustrated in Extended Data Fig. 1d. Adapters were trimmed off with the Trimmomatic program (v.0.36)<sup>36</sup>, and PCR duplicates were removed by home-made scripts. The poly(N) tails at the 3' end were further clipped with the Cutadapt program (v.1.15)<sup>37</sup>. After filtering, the paired reads were first aligned to 45S pre-rRNA, and the remaining reads were mapped to the human reference genome (assembly version: hg19) using the STAR software (v.020201)<sup>38</sup>. The parameters used were as follows: STAR--runMode alignReads--genomeDir index--readFilesIn read.fq--outFileNamePrefix outprefix --outFilterMultimapNmax 100 --outSAMattributes All --alignIntronMin 1 --scoreGapNoncan -4 --scoreGapATAC -4 --chimSegmentMin 15 --chimJunctionOverhangMin 15. Three additional parameters were further applied (--alignSJoverhangMin 15 --alignSJDBoverhangMin 10 --alignSjstitchMismatchNmax 5-15 5) to improve the mapping quality. The STAR mapping produced normally mapped reads (Aligned\_out.sam) and chiasmatically mapped reads (Chimeric\_out.sam). The low-quality and secondary mapping results were filtered using the SAMtools package (v.0.1.19)<sup>39</sup>. We next extracted the gapped reads from the normally mapped reads and further discarded the reads resulting from RNA splicing. The chiasmatically mapped reads and remaining gapped reads constituted the set of chimeric reads representing useful data for determining RNA structures and RNA–RNA interactions. To identify high-confidence intramolecular interactions, we followed the DG algorithm<sup>5</sup>, which is widely used for PARIS data analysis. The basic assumption is that high-confidence interactions should contain at least two unique chimeric reads with different termini. Then, a connection score is computed by dividing the total number of reads connecting two different RNA fragments to the coverage of chimeric reads at these two RNA fragments (coverage A\_B/((coverage A × coverage B))). We used the connection score of 0.01 as a cut-off to remove the low-scoring RNA–RNA interactions. For intermolecular interactions, we adopted a previously described Monte Carlo simulation to detect significant contacts by comparing observed pairwise interactions to the counts of random ligations<sup>40</sup>, in which 100,000 simulations were performed. RNA–RNA interactions with local-background-corrected *P* values lower than 0.05 were used for downstream analysis. Although pCp–biotin labelling leads to enrichment of both the chimeric and singleton reads (that is, those derived

# Article

from RNAs fragments that have been 3'-end-labelled but not ligated), it enables us to enrich pairwise-interacting RNAs efficiently. Notably, compared with RNA-seq reads (about 45.3% in introns), we found that around 73.2% of the singleton reads were mapped to the intronic region, indicating an enrichment for nascent transcripts.

**Visualization of RIC-seq data in Integrative Genomics Viewer.** RNA–RNA interactions were displayed using the Integrative Genomics Viewer (IGV) visualization tool (v.2.3.92)<sup>40</sup>. The chimeric tag information was first stored and sorted in a BED format file, in which each row contains three columns: the chromosome name and the coordinates of two junction sites for each of the chimeric reads. For visualizing RNA structure, each row in the BED file recorded the coordinates of nucleotides that constitute a base pair. After adding a track line ‘track graphType=arc’ to the head of the BED files, RNA structures or RNA interactions were presented as arcs across a reference transcript in IGV.

**Visualization of RIC-seq data in Juicebox.** The Juicebox tool (v.1.6.2) was applied to construct RNA–RNA interaction maps (<http://aidenlab.org/juicebox>). The chimeric read information was stored in a .medium format file, in which each row recorded the genomic coordinates of two tags for a chimeric read. After sorting by chromosome names, this file was converted into .hic format using the pre command in Juicer tools<sup>13</sup> and further visualized as a heat map in Juicebox. In the two-dimensional heat maps, the colour intensities represented junction frequency.

**Structural analysis of pre-miRNA.** The genomic coordinates of pre-miRNA were extracted from the human GENCODE gene set (annotation release version: v19; genome assembly version: hg19). The count of RIC-seq chimeric reads along each pre-miRNA was normalized as a Z-score, and the pre-miRNAs with at least one chimeric read were displayed ( $n = 280$ ). As previously described<sup>41</sup>, we used normalized complexity to evaluate the enrichment of pCp along pre-miRNA. The VARNA tool<sup>42</sup> was used to visualize the RNA structure of pre-miR-296.

**Structural analysis of representative RNAs.** RIC-seq reads were individually mapped to the reference sequence of rRNAs, snoRNAs, *TERC*, 7SK, 7SL, *RPPH1* and snRNAs. The RefSeq accession numbers are: NR\_046235.3 for rRNA, NR\_006880.1 for U3 snoRNA, NR\_001566.1 for *TERC*, NR\_002312.1 for *RPPH1*, NR\_001445.2 for 7SK, NR\_002715.1 for 7SL, NR\_003925.1 for U4 snRNA, and NR\_004394.1 for U6 snRNA. The chimeric reads with overlapping gaps were assembled into a cluster, and a unique identifier (ID) was assigned to each cluster. Accordingly, the unique read group tag ‘RG: Z: Cluster\_id’ was also appended to the alignments of chimeric reads in SAM format. After being converted into BAM format and sorted by SAMtools package (v.0.1.19), the alignments were further visualized in IGV directly, and chimeric reads belonging to the same cluster could be grouped with the same colour according to the assigned cluster ID.

Proximal interacting RNA fragments from the same or multiple RNAs can be randomly cut by MNase at the protein-free region and subsequently labelled with pCp–biotin (Extended Data Fig. 1k). Therefore, the alignment of the cytosine-containing chimeric junction sequence to its parent reference transcript enables us to deduce the RNA duplex directly or peripherally interacting with proteins, and to identify long-range RNA proximal interactions from single or multiple RNAs. We placed a black line at the last base of the left arm in chimeric read to improve the visualization of ‘loops’.

**Analysis of U1–*MALAT1* and snoRNA–rRNA interactions.** *MALAT1* (NR\_002819.4) and U1 (NR\_004430.2) sequences were assembled into a mini-genome. RIC-seq and PARIS reads<sup>5</sup> were mapped to this mini-genome using the STAR program. An identical pipeline was

applied to identify and collect chimeric reads that support the interaction between U1 and *MALAT1*. The U1 motif was first compiled from all annotated constitutive 5' splice sites in the human genome, as previously described<sup>43</sup>, and then was used for searching significant matches ( $P < 0.01$ ) in *MALAT1* RNA. We defined RIC-seq-specific U1 targeting sites as the chimeric read coverage at a specific region that was four times greater than the PARIS signal, and vice versa for determining PARIS-specific locations. The remaining U1 contact sites containing at least two chimeric reads by these two methods were defined as common sites. Sequences for snoRNA were downloaded from snoRNABase (<https://www-snorna.bioutl.fr>)<sup>44</sup>. A total of 192 nonredundant snoRNAs and rRNA was assembled as a super-genome and used in the analysis.

**RNA enriched in different subcellular fractions.** RNA-seq data of different subcellular fractions in HeLa cells were extracted from the Gene Expression Omnibus (GEO) database (accession code GSE110028)<sup>45</sup>. The clean reads were mapped to the human reference genome (hg19) using HISAT2 (v.2.1.0)<sup>46</sup>. The Cuffnorm program in Cufflinks package (v.2.2.1)<sup>47</sup> was used to calculate RNA abundance with a ‘geometric’ library normalization method. An RNA was defined as chromatin-enriched if the chromatin/nucleoplasmic and the chromatin/cytoplasmic ratios were both higher than 1.5. The nucleoplasmic- or cytoplasmic-enriched RNAs were identified in a similar way. The RNAs enriched in different fractions are listed in Supplementary Table 2.

**Comparison with PARIS, proximity RNA-seq and APEX-seq.** PARIS duplex groups defined in HEK293T cells<sup>48</sup> were used for comparison with RIC-seq-detected RNA–RNA interactions in HeLa cells. Because the PARIS duplex group was derived from the chimeric reads aligned to the transcriptome, we chose RIC-seq interactions that happened in the exonic region of RNAs that have fragments per kilobase of transcript per million mapped reads (FPKM) values  $\geq 1$  in both cell types for the comparison. RIC-seq-identified intra- or intermolecular interactions were classified as an overlapping group if both arms in the chimeric reads overlapped with a duplex group from PARIS, or were classified as a RIC-seq-specific group otherwise. We used the RNAhybrid<sup>49</sup> program to calculate the minimum free energy of the RNA duplexes in the overlapping group and RIC-seq-specific group. The colocalized RNAs previously detected by proximity RNA-seq<sup>10</sup> in SH-SY5Y cells were compared with the significant intermolecular contacts revealed by RIC-seq in HeLa cells. Only co-expressed RNAs with an FPKM  $\geq 1$  in both cell types were used for the analysis. The colocalized RNAs previously detected by APEX-seq<sup>11</sup> in four subcellular locations—including cytosol ( $n = 46$ ), mitochondrial ( $n = 15$ ), nucleolus ( $n = 324$ ), and nucleus ( $n = 217$ )—were selected for comparison. Because APEX-seq libraries were generated from polyA<sup>+</sup> RNAs, in which intron sequences were largely removed, we chose RIC-seq detected significant intermolecular interactions ( $P \leq 0.05$ ) in the exonic region for analysis. The in silico random RNA–RNA interaction (denoted as expected) between 602 transcripts was compared with the observed intermolecular contacts revealed by RIC-seq. Among the 602 transcripts, RIC-seq revealed 71 significant ( $P \leq 0.05$ ) intracompartment and 33 intercompartment interactions.

***MALAT1*- and *NEAT1*-interacting motifs.** We first divided the genome into 500-bp bins and sorted these bins according to the number of chimeric reads that had one end linked to the *MALAT1* or *NEAT1* transcripts. The genome sequences in the top 1,000 bins were chosen for motif analysis using the MEME package (v.4.11.4)<sup>50</sup>. The parameters were as follows: -nmotifs 10 -evt 1e-10 -minw 10 -maxw 40 -p 20 -maxsize 1000000 -revcomp -dna.

**Random ligation rate estimation.** The human (hg19) and fly reference genomes (dm6) were first assembled into a pan-genome. The RIC-seq

reads generated from the mixed HeLa and *Drosophila* S2 cells at a 1:5 ratio were then aligned to the pan-genome. The chimeric reads with one arm mapped to the human genome, and the other arm mapped to the fly genome, were classified as cross-species RNA–RNA misligations. The random ligation rate was calculated by dividing the number of misligations to the number of total chimeric reads. To generate the heat map shown in Extended Data Fig. 1m, we calculated the intra- and interchromosomal RNA interactions for each human and fly chromosome. The relative enrichment score  $E$  was defined as  $L_{ij}/T_i \times T_j$ , in which  $L_{ij}$  indicates the number of pairwise RNA interactions between chromosome  $i$  and  $j$ ,  $T_i$  and  $T_j$  stand for the total numbers of chimeric reads for chromosome  $i$  and chromosome  $j$ , respectively.

**Structural analysis of 28S rRNA.** The RIC-seq reads aligned to 45S pre-rRNA (NR\_046235.3) were collected and used to construct the interaction matrix shown in Fig. 1h. A Knight–Ruiz normalization algorithm, widely used in the normalization of Hi-C contact matrices<sup>51</sup>, was applied to eliminate sequencing bias along rRNA. For building the physical interaction map of 28S rRNA, the cryo-EM model of human 80S ribosome (RCSB Protein Data Bank (PDB) ID 4V6X) was downloaded, and the spatial distances between every 5-nt bin in 28S rRNA were calculated using the mean spatial coordinates of carbon atoms in each 5-nt bin. Watson–Crick and non-Watson–Crick base pairs were identified using the DSSR software<sup>52</sup>. The 3D structure of the ribosome was visualized by the PyMOL system (Educational version, <https://pymol.org/2/>). For the missing structures in 28S rRNA, we combined intramolecular RNA–RNA interactions detected by RIC-seq with the RNAstructure algorithm<sup>53</sup> to deduce their 2D structures.

**ROC analysis.** The ROC approach was adopted to evaluate the performance of RIC-seq in detecting proximal RNA interactions. The physical proximity information derived from the cryo-EM model of the 28S rRNA was used as a gold standard. Notably, the missing part in the cryo-EM model was not used for generating the ROC curve. For each pairwise 5-nt bin, the true-positive datasets were defined by the average Euclidean distance of the pairs being within 25 Å, and the true-negative datasets were otherwise. To evaluate the RIC-seq data, we designed a proximal index to classify each pairwise 5-nt bin into a proximal or distal group. Specifically, after generating a normalized interaction matrix, we calculated the proximal index for every pairwise 5-nt bin as the percentage of pairwise bins with normalized RIC-seq interaction signal higher than 50 among the 121 surrounding pairwise bins in the interaction matrix. For each pairwise 5-nt bin, we classified it as a proximal or distal interaction on the basis of whether the proximal index was higher or lower than a threshold, respectively. The ROC curve was obtained by varying the threshold of the proximal index from 0 to 1 and counting the rates of true positives and false positives.

**RNA topological domain analysis.** We used an iterative algorithm to split full RNA into separate domains on the basis of the RIC-seq signal. To eliminate the effect of splicing, we focused on those genes in which more than 80% of chimeric reads originated from precursor transcripts. First, the full RNA was considered as a single domain, and we tried to find candidate sites that could split it into separate domains on the basis of RIC-seq data. Next, we iteratively chose the optimal site that maximized the ratio of intradomain and interdomain RIC-seq density as a new candidate domain boundary. The iteration was stopped when the ratio began to decrease. Finally, the RNA was split into domains with the maximum ratio of intradomain and interdomain RIC-seq density. We also randomly separated each RNA into the same number of domains 100 times to construct a background domain set (denoted as random control).

**Inferring RIC-seq-detected active promoter–enhancer interaction network.** The ChIP-seq data of H3K4me3, H3K4me1 and H3K27ac

generated in HeLa cells were extracted from ENCODE (<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHistone>), and peaks were identified by MACS2 software (v.2.2.4)<sup>54</sup> in narrow-peak mode with default parameters. As previously described<sup>55</sup>, promoters were defined as H3K4me3-enriched peaks in regions within  $\pm$  5 kb of TSSs. For enhancers, the H3K27ac peaks within  $\pm$  2.5 kb of known gene promoters were first removed, and the remaining H3K27ac peaks clustered within a 12.5-kb region were then stitched together and classified as enhancers. As some lncRNAs show enhancer-like activity<sup>56</sup>, we also included H3K27ac-enriched peaks in regions  $\pm$  5 kb from the TSS of the noncoding gene for the definition of enhancers. Super-enhancers were identified by using the Rank Ordering of Super-Enhancers (ROSE) algorithm based on the coverage of H3K27ac ChIP-seq signals<sup>16</sup>. Using these defined promoter and enhancer regions, we next adopted a Monte Carlo simulation strategy<sup>10</sup> to detect significant interactions between enhancer- and promoter-derived RNAs in HeLa cells. The whole-genome networks were built based on the significant intrachromosomal RNA–RNA interactions ( $P \leq 0.05$ ) and sufficient trans-chromosomal interactions needed to depict chromosome–chromosome relationships. The interaction networks were displayed by Cytoscape (version 3.7.2)<sup>20</sup> using a self-organized layout algorithm (edge-repulsive spring-electric layout), which was supported by the third-party app of AllegroLayout.

**t-distributed stochastic neighbour embedding analysis for hub RNAs.** t-distributed stochastic neighbour embedding (t-SNE) has been widely used for displaying cells on the basis of gene-expression profiles in two dimensions. Similarly, to assess the specificity of hub RNAs, we also used t-SNE to give each hub RNA a position in a two-dimensional map on the basis of their target profiles. In brief, the t-SNE analysis relied on a matrix in which each entry denotes the significance ( $-\log_{10}(P\text{ values})$ ) of an interaction between a hub RNA (rows in the matrix,  $n = 642$ ) and one of their interacting transcripts (columns in the matrix,  $n = 4,140$ ). Low-confidence interactions with  $P$  value  $>0.1$  or supported by less than 3 chimeric reads were removed, and transcripts interacted with less than ten hub RNAs were also filtered. The entries with  $-\log_{10}(P\text{ values})$  greater than five were restricted as five. The matrix was normalized by subtracting the mean of the matrix from each entry as previously described<sup>10</sup>. Finally, the t-SNE analysis was performed on the basis of the revised matrix using the Rtsne package (<https://github.com/jkrijthe/Rtsne>) in R with the following parameters: seed 4, perplexity = 50, theta = 0.2.

**Hub RNA enriched motifs.** Known RBP motifs were downloaded from the CISBP-RNA database (<http://cisbp-rna.ccb.utoronto.ca/>, *Homo sapiens*, version 0.6) and presented as position weight matrices. After repeat-masking, the position weight matrix matches in the hub RNA chimeric reads covered regions were counted using the FIMO package<sup>57</sup> with the following parameters: --norc --thresh 0.001 --motif-pseudo 0.1.

**Datasets analysed.** The RNA-seq data on HeLa cells were generated previously (NCBI Sequence Read Archive database accession code SRR628455)<sup>26</sup>. The CLIP-seq data generated in HeLa cells were downloaded from NCBI GEO database (GSM2078375 for HuR, GSM2078374 for hnRNP, GSM1048186 for PTBP1, GSM1509288 for U2AF65, GSM2409898 for NONO, GSM2409900 for PSF, GSM1215139 for FUS, GSM1167169 for RBM6, GSM1001331 for eIF4AIII, GSM850171 for hnRNP, GSM1167172 for RBM10, GSM1048187 for AGO2, GSM2107199 for ALKBH1, GSM2298848 for UPF1, GSM2409884 for EWS, GSM2712721 for YTHDC2 and GSM1135012 for METTL3). Topologically associating domains (TADs) were downloaded from a previous publication<sup>18</sup>.

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

**Data availability**

RIC-seq and hnRNP CLIP-seq data that support the findings of this study have been deposited in the GEO under accession number GSE127188.

**Code availability**

Custom codes used for data analysis in this paper can be found at <https://github.com/caochch/RIC-seq>.

26. Xue, Y. et al. Direct conversion of fibroblasts to neurons by reprogramming PTB-regulated microRNA circuits. *Cell* **152**, 82–96 (2013).
27. Chen, J. et al. The RNA-binding protein ROD1/PTBP3 cotranscriptionally defines AID-loading sites to mediate antibody class switch in mammalian genomes. *Cell Res.* **28**, 981–995 (2018).
28. Engeitz, J. M. et al. RNA–RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell* **159**, 188–199 (2014).
29. Qin, G. et al. Long noncoding RNA p53-stabilizing and activating RNA promotes p53 signaling by inhibiting heterogeneous nuclear ribonucleoprotein K deSUMOylation and suppresses hepatocellular carcinoma. *Hepatology* **71**, 112–129 (2020).
30. Orjalo, A. V., Jr & Johansson, H. E. Stellaris® RNA fluorescence in situ hybridization for the simultaneous detection of immature and mature long noncoding RNAs in adherent cells. *Methods Mol. Biol.* **1402**, 119–134 (2016).
31. Wang, X. W. et al. A microRNA-inducible CRISPR–Cas9 platform serves as a microRNA sensor and cell-type-specific genome regulation tool. *Nat. Cell Biol.* **21**, 522–530 (2019).
32. Gilbert, L. A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
33. Zarnegar, B. J. et al. irCLIP platform for efficient characterization of protein–RNA interactions. *Nat. Methods* **13**, 489–492 (2016).
34. Hagège, H. et al. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat. Protocols* **2**, 1722–1733 (2007).
35. Chu, C. et al. Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404–416 (2015).
36. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
37. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetJ.* **17**, 10–12 (2011).
38. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
39. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
40. Robinson, J. T. et al. Integrative Genomics Viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
41. Xue, Y. et al. Genome-wide analysis of PTB–RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol. Cell* **36**, 996–1006 (2009).
42. Darty, K., Denise, A. & Ponty, Y. VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**, 1974–1975 (2009).
43. Almada, A. E., Wu, X., Kriz, A. J., Burge, C. B. & Sharp, P. A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**, 360–363 (2013).
44. Lestrade, L. & Weber, M. J. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.* **34**, D158–D162 (2006).
45. Nojima, T. et al. Derepressed expression of mammalian lncRNA through loss of SPT6 induces R-loop formation, replication stress, and cellular senescence. *Mol. Cell* **72**, 970–984 (2018).
46. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
47. Trapnell, C. et al. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
48. Gong, J. et al. RISE: a database of RNA interactome from sequencing experiments. *Nucleic Acids Res.* **46**, D194–D201 (2018).
49. Krüger, J. & Rehmsmeier, M. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.* **34**, W451–W454 (2006).
50. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
51. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
52. Lu, X. J., Bussemaker, H. J. & Olson, W. K. DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res.* **43**, e142 (2015).
53. Reuter, J. S. & Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 129 (2010).
54. Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
55. Li, X. et al. GRID-seq reveals the global RNA–chromatin interactome. *Nat. Biotechnol.* **35**, 940–950 (2017).
56. Ørom, U. A. et al. Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46–58 (2010).
57. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).

**Acknowledgements** We thank A. M. Pyle, B. Zhou and J. Hu for critical reading of this manuscript; Y. Wang for sharing the PiggyBac-Cas9 system; G. Shan for assistance with the DNA FISH; L.-L. Chen for HT29 cells; G. Zhou for sharing the hnRNP expression plasmid; and Y. Wang, Y. Feng and S. Li for SIM imaging and analysis. This work was supported by the National Natural Science Foundation of China (91740201, 91940306, 31522015 and 81921003), the Ministry of Science and Technology of China (2017YFA0504400 and 2016YFC0900400) and the Strategic Priority Program of CAS (XDB19000000) to Y.X., by the Beijing Municipal Natural Science Foundation (5182024) to C.C., and by the National Natural Science Foundation of China (31771438, 31970610) to L.J.

**Author contributions** Y.X. initiated and planned the study; Z.C. cultured cells, constructed clones, performed smFISH and created the RIC-seq library; L.J. performed Cas9 deletion, CRISPRi, DNA FISH, ChIP-MS and functional validation of CCAT1-5L with the help of C.X. and S.W.; C.C. performed the bioinformatics analysis, led the project and prepared the figures with the help of Z.D. and N.H.; R.Y. tested RIC-seq conditions, validated enhancer–promoter interactions and constructed hnRNP CLIP-seq libraries; D.W. performed 3C-qPCR analysis; J.C. and X. Yu performed co-IP and ChIP-qPCR experiments; S.H., L.W. and X. Yang advised on bioinformatics analysis, Y.X., Z.C. and C.C. wrote the manuscript.

**Competing interests** Z.C., C.C. and Y.X. have filed a joint application for the patent of RIC-seq technology.

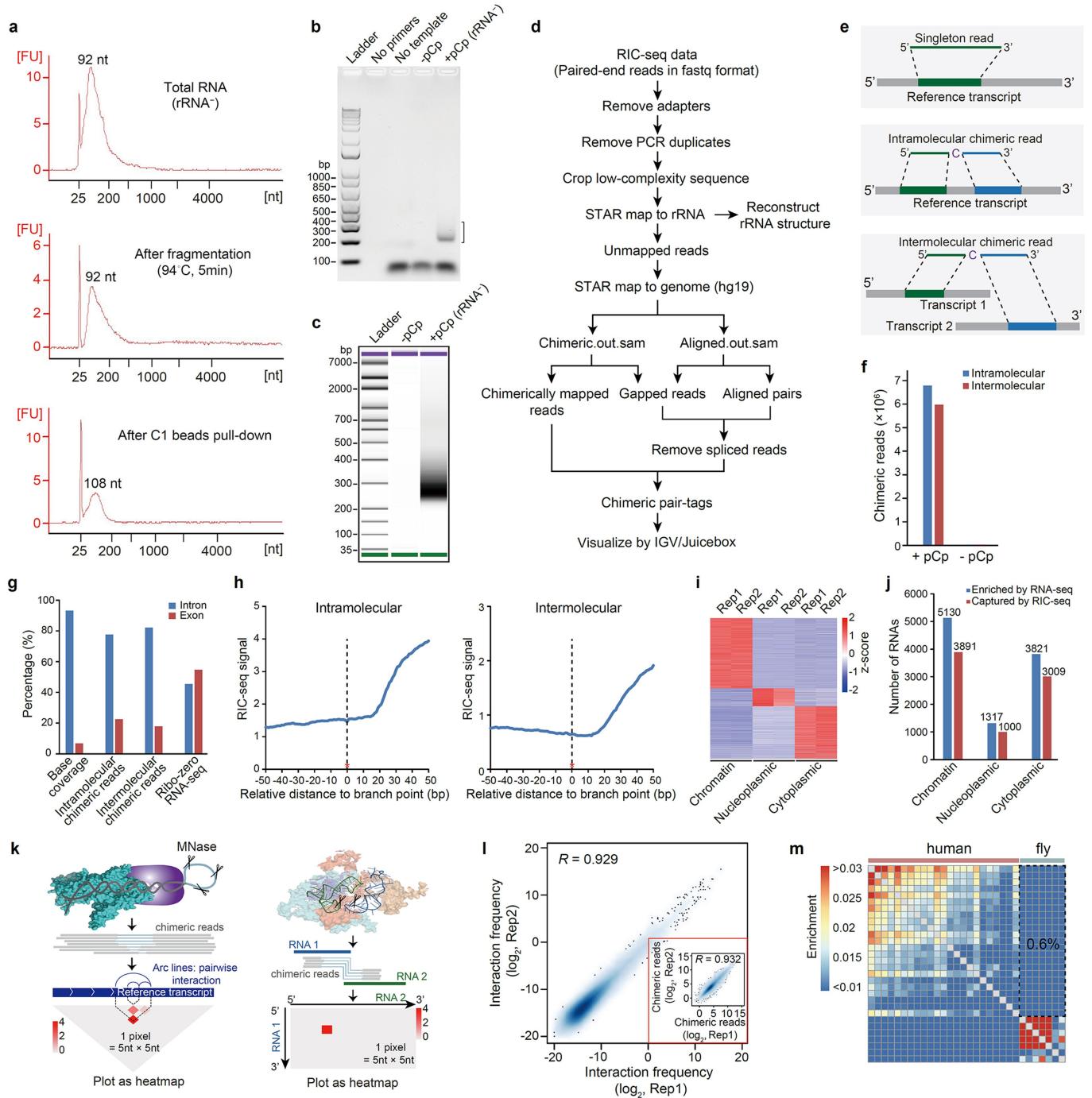
**Additional information**

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2249-1>.

**Correspondence and requests for materials** should be addressed to Y.X.

**Peer review information** *Nature* thanks Scott Tenenbaum, Jernej Ule and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

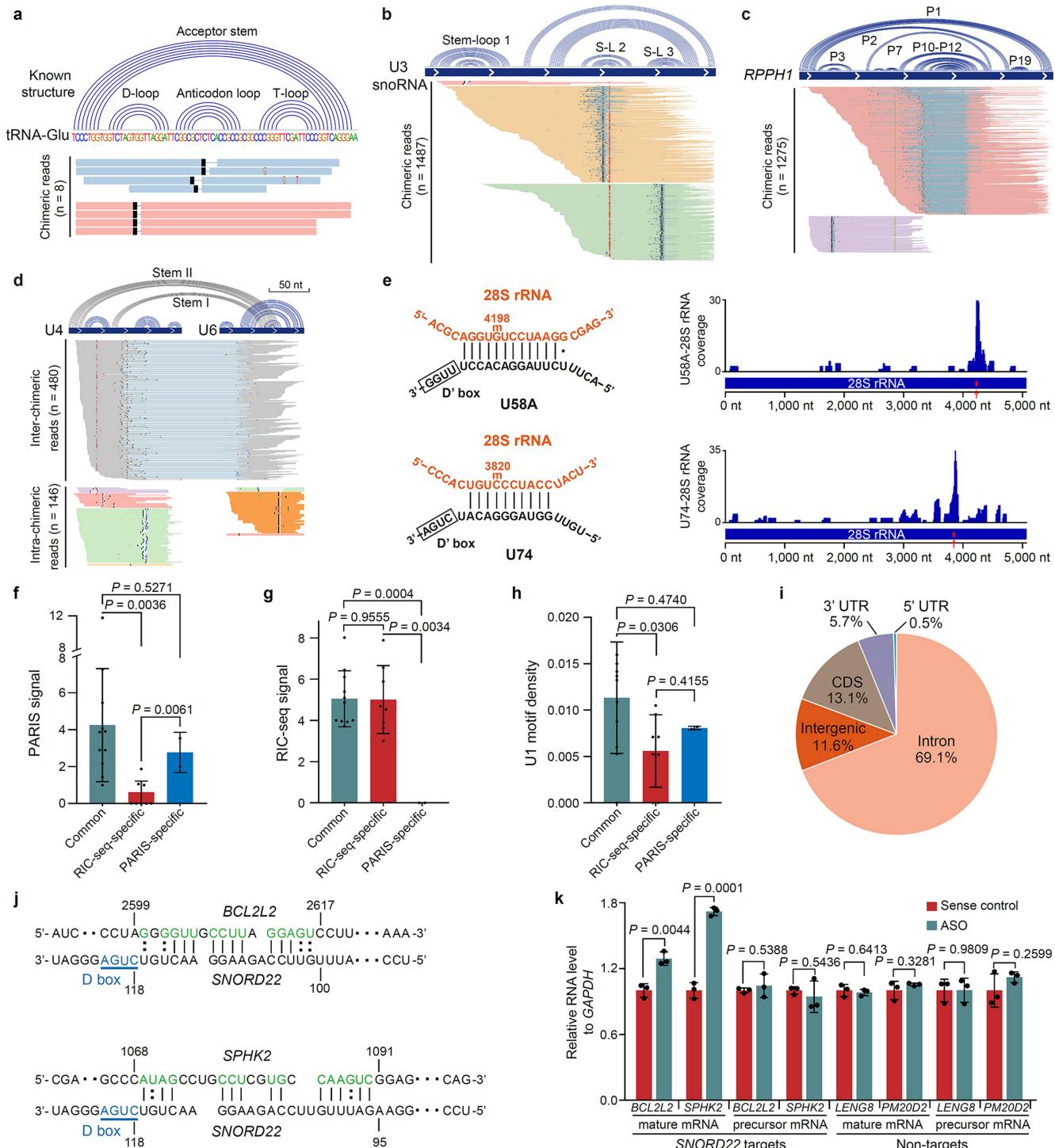
**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



**Extended Data Fig. 1 | Characterization of RIC-seq technology.** **a**, The RNA fragment sizes were quantified using a 2100 Bioanalyzer. Top, total RNA after rRNA depletion. Middle, RNA after fragmentation. Bottom, RNA after C1 beads pull-down. **b, c**, The RIC-seq libraries were purified from the gel (the bracketed region) and quantified using a 2100 Bioanalyzer. –pCp denotes samples without pCp-biotin labelling. **d**, The mapping pipeline for RIC-seq data. **e**, Read types and their relationship to reference transcripts. **f**, The numbers of intra- and intermolecular chimeric reads. **g**, The intra- and intermolecular chimeric reads show a 4.2- and 5.5-fold increase, respectively, in the intron-to-exon ratio, as compared with RNA-seq data. **h**, The intra- and intermolecular RIC-seq signals do not decrease after the branch point (red star). **i**, Transcripts enriched in different subcellular fractions revealed by RNA-seq<sup>45</sup> in HeLa cells. Two replicates for each fraction (rep 1 and rep 2). **j**, The number of RNAs enriched (blue) in each subcellular fraction and the numbers of RNAs that could be

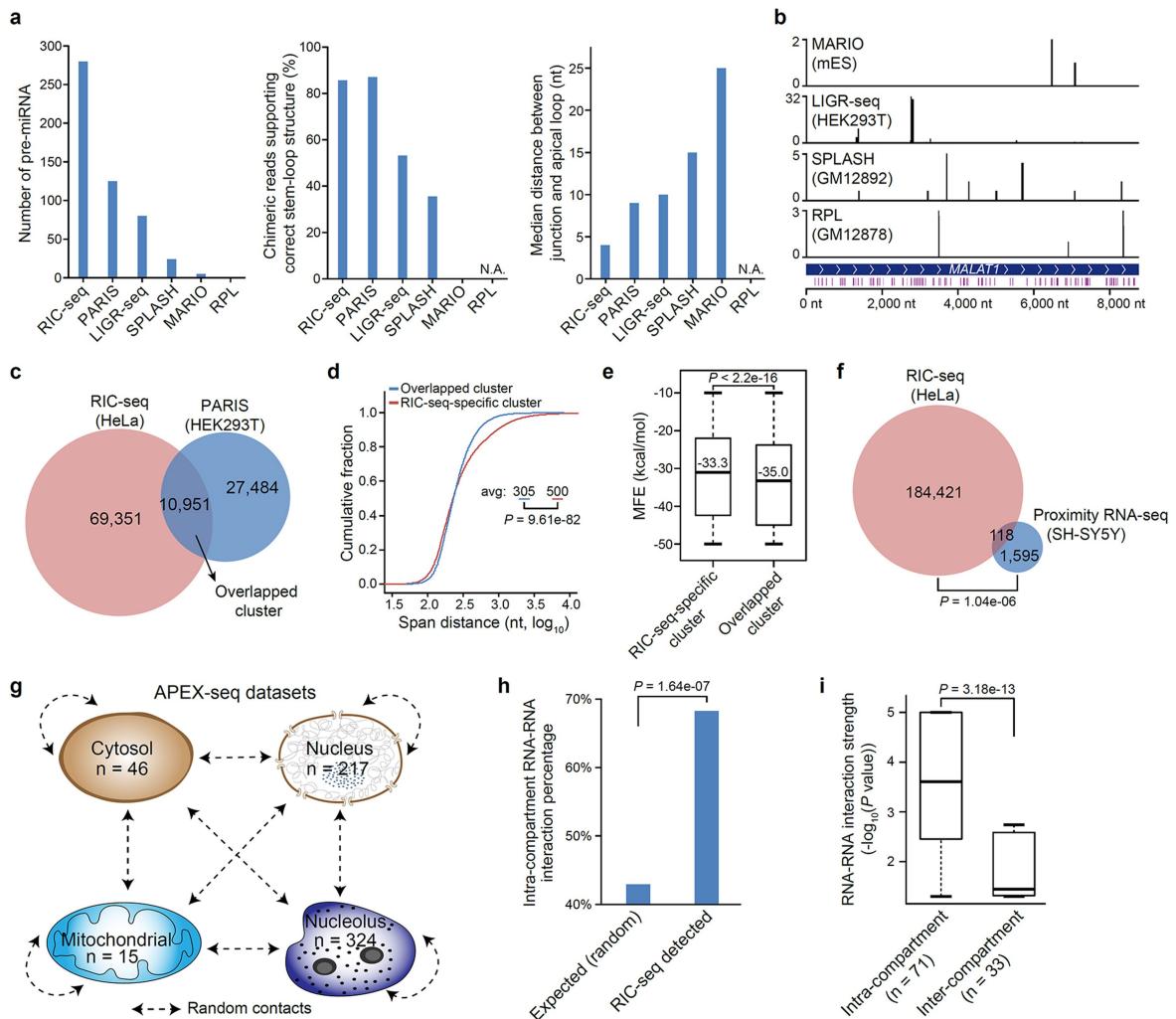
captured by RIC-seq (red). **k**, Cartoon of RBP-mediated RNA proximal interactions and data presentation. The structural models of the RBPs were generated in PyMOL on the basis of PDB accessions 5B16 and 4V6X. Light blue lines between chimeric reads represent gaps. Pairwise junctions are visualized as arc lines or plotted as a heat map. Colour intensity indicates the number of junctions within each 5 × 5-nt pixels. **l**, Scatter plots showing the reproducibility of RIC-seq-detected pairwise interactions in two biological replicates. The reproducibility for chimeric reads per transcript ( $R = 0.932$ ) is displayed in a red box. RNA abundance is normalized for both scatter plots. **m**, Heat map showing the cross-species RNA–RNA interactions observed by the cell mixing strategy. The boxed region in dashed lines represents random ligations. The experiments in **b, c** were independently repeated three times with similar results.

# Article



**Extended Data Fig. 2 | RIC-seq recapitulates known and newly identified RNA–RNA interactions.** **a–c**, RIC-seq recapitulates known structures of tRNA, U3 snoRNA and RPPH1 lncRNA. Known structural elements are marked above specific arc lines. For U3 snoRNA, only three clusters of chimeric reads that support stem-loops 1 to 3 are shown. Mismatches based on variation from the reference transcriptome sequence are marked as red dots (C>T), and green dots (G>A). **d**, RIC-seq recapitulates known intra- and intermolecular interactions of U4 and U6 snRNAs. Red dots, C>T mismatches. **e**, RIC-seq recapitulates snoRNA-interacting sites in 28S rRNA. The red arrow indicates known modification sites. The boxed region represents the D' box. **f–h**, PARIS

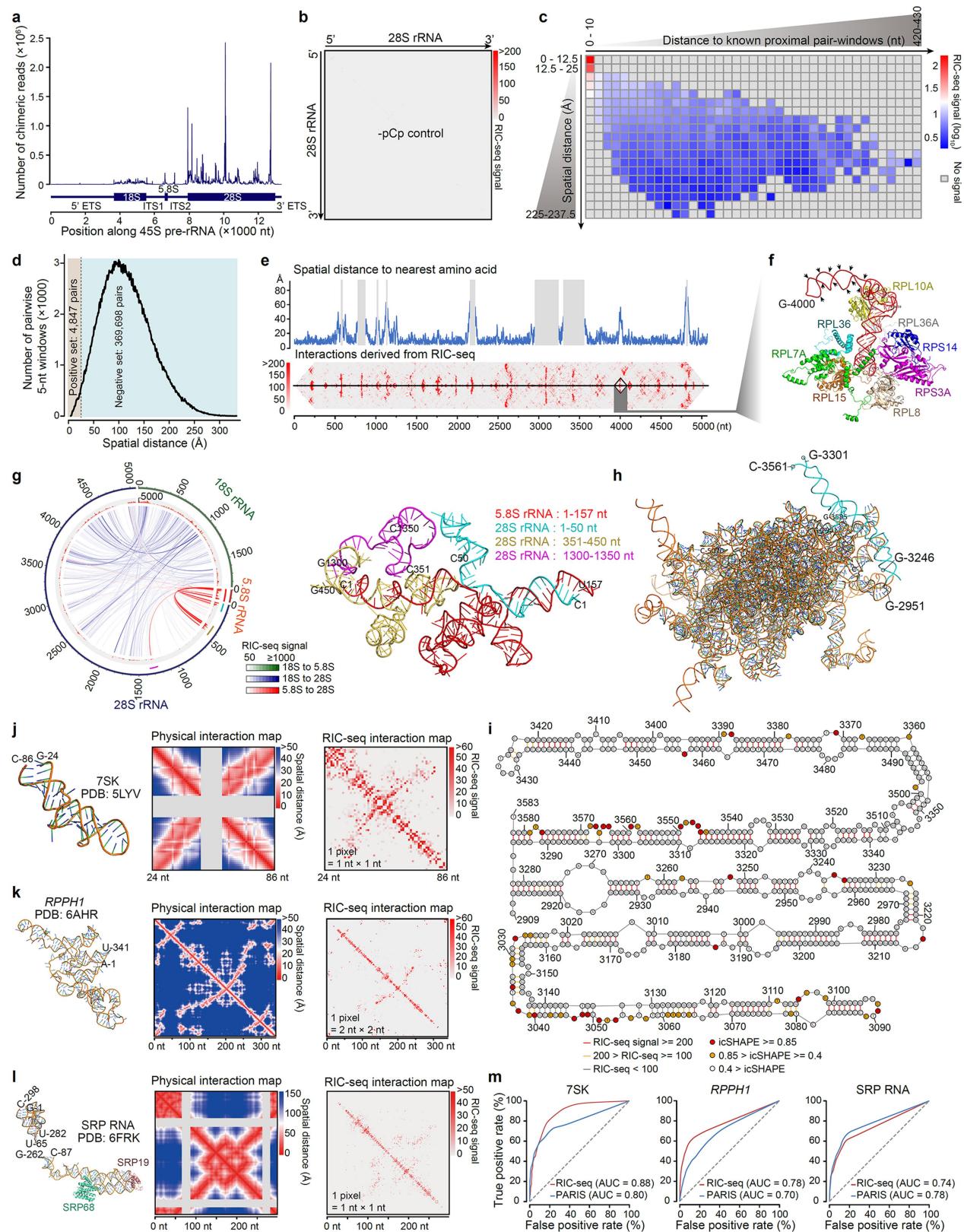
signal, RIC-seq signal and U1-motif density in three different groups of U1-MALAT1 contact sites. Common regions,  $n=11$ ; RIC-seq-specific regions,  $n=8$ ; PARIS-specific regions,  $n=2$ . **i**, The genomic distribution of snoRNA-interacting sites detected by RIC-seq. **j**, The SNORD22 interacting sites in SPHK2 and BCL2L2 RNA. The D box is shown in blue. **k**, qPCR showing reduced mRNA levels of BCL2L2 and SPHK2 upon knockdown of SNORD22 with ASOs. LENG8 and PM20D2 served as negative controls. Data are mean  $\pm$  s.d.,  $n=3$  biological replicates. Two-tailed unpaired  $t$ -test was used to calculate the  $P$  values in **f–h**, **k**.



**Extended Data Fig. 3 | Comparison with existing methods.** **a**, The sensitivity, accuracy and resolution of existing methods for detecting the structure of pre-miRNA. Sensitivity, the number of detected pre-miRNAs; accuracy, the percentage of chimeric reads that support correct stem-loop structures; resolution, the mean distance of chimeric junctions from the apical loop. Notably, RNA proximity ligation (RPL) did not detect pre-miRNA. N.A., not available. **b**, U1 binding sites in *MALAT1* detected by mapping RNA interactome in vivo (MARIO), ligation of interacting RNA followed by high-throughput sequencing (LIGR-seq), sequencing of psoralen crosslinked, ligated, and selected hybrids (SPLASH) and RPL methods. U1 motifs, purple lines; y axis, chimeric read number. **c**, Venn diagram showing overlapping duplexes detected by RIC-seq and PARIS methods in different cell lines. RNAs with fragments per kilobase of transcript per million mapped reads (FPKM) values  $\geq 1$  in both cell types are used for the analysis. **d**, The transcriptomic span distance of RIC-seq-specific (red,  $n = 69,351$ ) and overlapped clusters (blue,  $n = 10,951$ ). The transcriptomic span excludes introns. Two-sided Kolmogorov–Smirnov test was used to calculate the  $P$  value. **e**, Box plot showing the minimum free energy (MFE) of duplexes among RIC-seq-specific and overlapped groups. **f**, Venn diagram showing the proximal RNAs detected by RIC-seq and proximity RNA-seq in two different cell lines. Two-sided Fisher's exact test was used to calculate the  $P$  value. **g**, Cartoon depicting the number of colocalized RNAs in four subcellular compartments revealed by APEX-seq. Dashed lines, in silico random contacts. **h**, The percentage of intracompartiment RNA–RNA interactions observed by RIC-seq ( $n = 71$ ) is higher than that of random contacts. The one-sided binomial test was used to calculate the  $P$  value. **i**, Box plot showing that the RIC-seq signal strength of intracompartiment RNA–RNA interactions is stronger than for intercompartment interactions. Two-tailed unpaired  $t$ -test was used to calculate the  $P$  values in **e** and **i**. For the box plots in **e** and **i**, the centre line represents the median, the box borders represent the first (Q1) and third (Q3) quartiles, and the whiskers are the most extreme data points within  $1.5 \times$  the interquartile range (from Q1 to Q3).

Smirnov test was used to calculate the  $P$  value. **e**, Box plot showing the minimum free energy (MFE) of duplexes among RIC-seq-specific and overlapped groups. **f**, Venn diagram showing the proximal RNAs detected by RIC-seq and proximity RNA-seq in two different cell lines. Two-sided Fisher's exact test was used to calculate the  $P$  value. **g**, Cartoon depicting the number of colocalized RNAs in four subcellular compartments revealed by APEX-seq. Dashed lines, in silico random contacts. **h**, The percentage of intracompartiment RNA–RNA interactions observed by RIC-seq ( $n = 71$ ) is higher than that of random contacts. The one-sided binomial test was used to calculate the  $P$  value. **i**, Box plot showing that the RIC-seq signal strength of intracompartiment RNA–RNA interactions is stronger than for intercompartment interactions. Two-tailed unpaired  $t$ -test was used to calculate the  $P$  values in **e** and **i**. For the box plots in **e** and **i**, the centre line represents the median, the box borders represent the first (Q1) and third (Q3) quartiles, and the whiskers are the most extreme data points within  $1.5 \times$  the interquartile range (from Q1 to Q3).

# Article

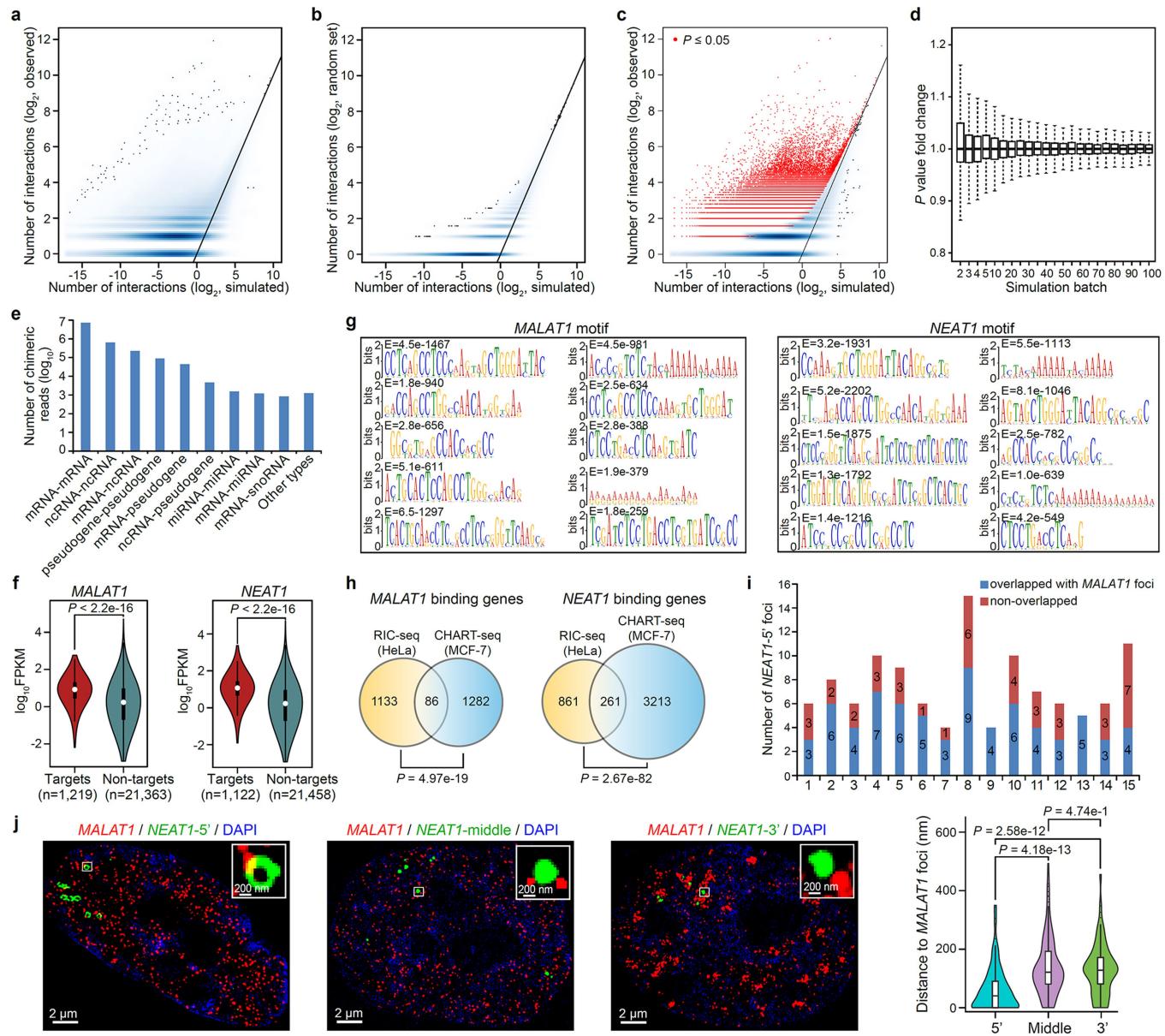


Extended Data Fig. 4 | See next page for caption.

**Extended Data Fig. 4 | RIC-seq precisely recaptures rRNA and lncRNA 3D interactions.** **a**, The chimeric reads mapped to rRNA. **b**, The RNA–RNA interactions of 28S rRNA in pCp<sup>-</sup> samples. **c**, The RIC-seq and cryo-EM data of 28S rRNA are highly correlated. The RIC-seq signal markedly decreases with increased distance to known pairwise-interacting sites in 28S rRNA (x axis) and with the increased spatial distance (y axis). **d**, The true-positive and true-negative datasets are generated from the cryo-EM model of 28S rRNA. **e**, The boundaries of RNA duplexes in 28S rRNA tend to be occupied by RBPs. The spatial distance to the nearest amino acid is calculated from the cryo-EM model. Regions absent from the structure are shown in light grey. **f**, RIC-seq captures the G4000 duplex protruding from the RBP (in different colours) complex. The arrowhead represents MNase random cut and pCp–biotin labelling position. **g**, The intermolecular interactions between 5.8S, 18S and

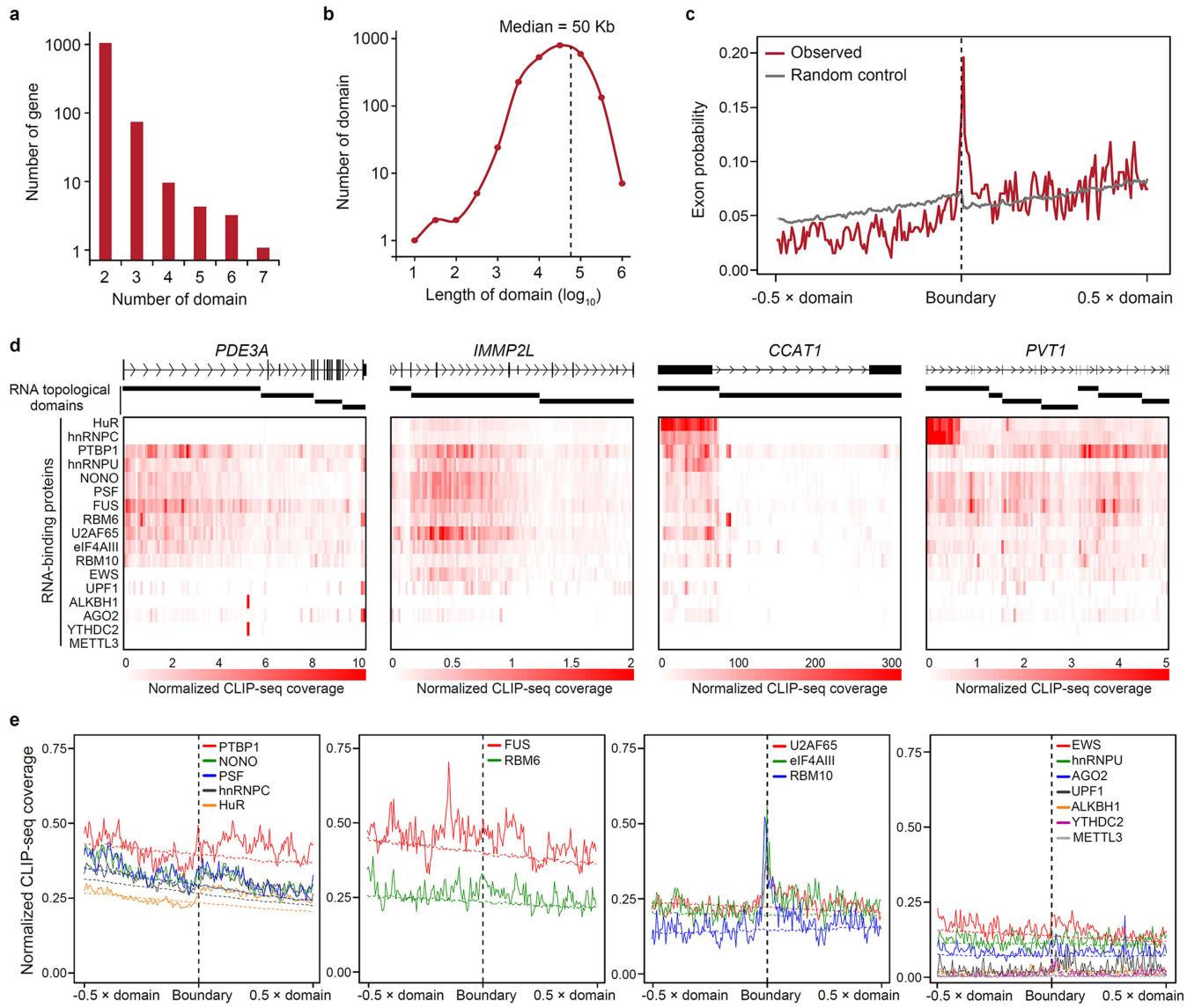
28S rRNA. The histogram on the inner circle represents the number of chimeric reads at the given positions. The red arc lines marked interactions between 5.8S and 28S rRNA are shown on the right. **h**, The modelled structure of 28S rRNA (PDB ID 4V6X). **i**, The secondary structure deduced from RIC-seq data for two missing 28S rRNA regions (nt 2951–3246 and nt 3301–3561). The bases are marked by different colours based on *in vivo* click selective 2-hydroxyl acylation and profiling experiment (icSHAPE) scores, and the base pairs are denoted by different lines on the basis of the strength of the RIC-seq signal. **j–l**, The structural model, physical interaction map and RIC-seq interaction map of 7SK 5'-hairpin, SRP (7SL RNA) and *RPPH1*. Grey regions in the physical maps indicate no structural data available. **m**, RIC-seq showed comparable performance to PARIS in detecting the structure of 7SK, SRP RNA and *RPPH1*. Dashed line, random classifier.

# Article



**Extended Data Fig. 5 | Monte Carlo simulation to identify significant RNA–RNA interactions and lncRNA targets.** **a**, Intermolecular RNA–RNA interactions ( $n = 2,088,874$ ) were plotted against the average pairwise interaction counts from 100,000 simulations. **b**, A single random dataset ( $n = 3,536,556$ ) was taken for further 100,000 simulations and plotted as in **a**. **c**, The observed pairwise interactions were compared with simulated random counts to identify high-confidence interactions ( $P \leq 0.05$ , red dots). **d**, Box plot showing how the  $P$  value distribution changed as the Monte Carlo simulation progressed using  $n = 2,088,874$  intermolecular RNA–RNA interaction events. The 100,000 simulations were divided into 100 batches. **e**, RNA–RNA interaction types and their numbers of RIC-seq chimeric reads. **f**, The violin plot shows the expression levels of *MALAT1*- and *NEAT1*-targeting genes. **g**, The enriched motifs among *MALAT1* or *NEAT1* chimeric targets. **h**, Overlaps of

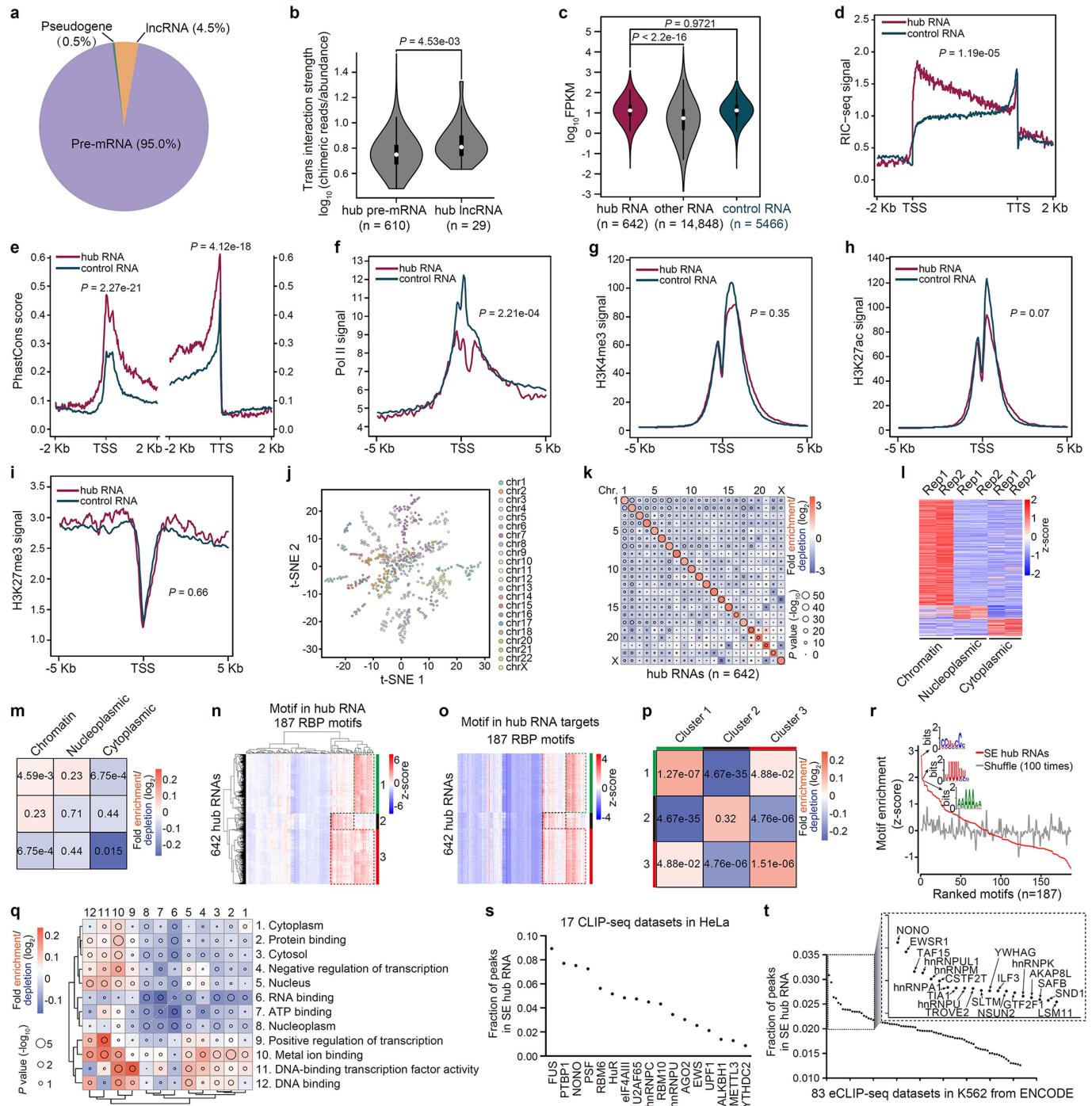
*MALAT1* and *NEAT1* binding sites identified by RIC-seq and CHART-seq. Two-sided Fisher's exact test was used to calculate the  $P$  value. **i**, Summary of *NEAT1* foci and their overlap with *MALAT1* in 15 cells. **j**, Structured illumination microscopy analysis showing the localization of *MALAT1* with *NEAT1* 5' region, *NEAT1* middle region or *NEAT1* 3' non-contact region. The regions marked by white boxes are magnified at the top right. The violin plot illustrates the distance from *NEAT1* foci to the nearest *MALAT1* puncta in 20 cells. 5' end,  $n = 131$ ; middle,  $n = 169$ ; 3' non-contact region,  $n = 166$ . Two-tailed unpaired  $t$ -test was used to calculate the  $P$  values in **f** and **j**. For the violin or box plots in **d**, **f** and **j**, the white centre point represents the median, the box limits represent the Q1 and Q3, the whiskers are the most extreme data points within  $1.5 \times$  the interquartile range (from Q1 to Q3), and the upper–lower limits represent the maximum–minimum values.



**Extended Data Fig. 6 | The features of RNA topological domain.** **a**, The number and distribution of topological domains in genes. **b**, The length distribution of the topological domains. Dashed line, median length. **c**, The domain boundaries frequently overlap with exons. Grey line, random control.

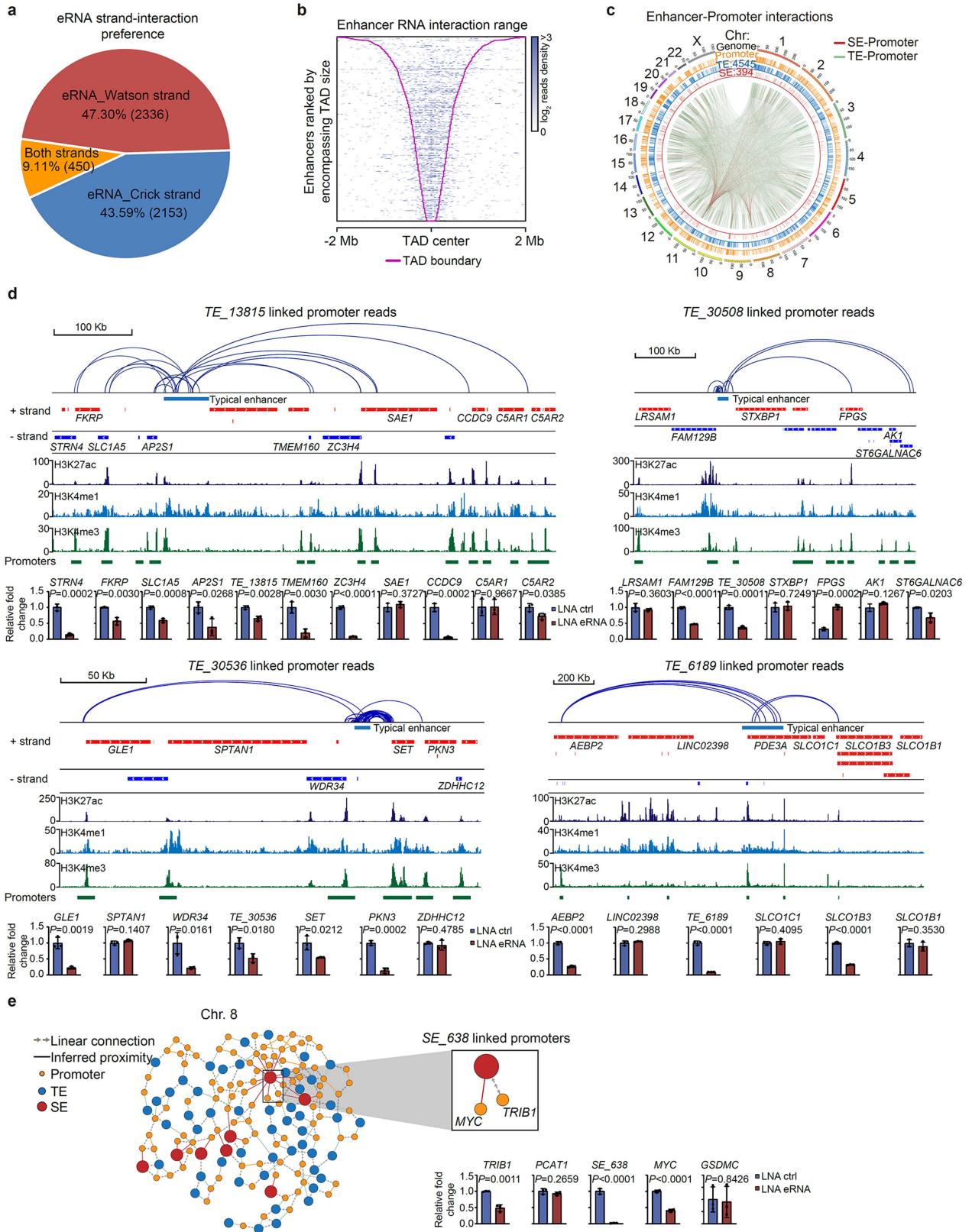
**d**, The RBP binding pattern around topological domains (black box) in pre-mRNA and pre-lncRNA. CLIP-seq coverage of each RBP is shown at the bottom. **e**, RBP binding profiles are plotted around the domain boundary. The dashed lines in colours represent random controls.

# Article



**Extended Data Fig. 7 | The characterization of hub RNAs.** **a**, The classification of hub RNAs. **b**, Hub IncRNA ( $n = 29$ ) show stronger *trans* interaction strength than hub pre-mRNA ( $n = 610$ ). **c–i**, Comparing the expression levels, RIC-seq signals, conservation, Pol II signals, H3K4me3 signals, H3K27ac signals and H3K27me3 signals of hub RNA ( $n = 642$ ) to control RNA ( $n = 5,466$ ). Two-tailed unpaired *t*-test was used to calculate the *P* values in **b** and **c**. The two-sided Kolmogorov–Smirnov test was used to calculate the *P* values in **d–i**. **j**, *t*-SNE visualization of 642 hub RNAs on the basis of the strength of their interaction with target RNAs. **k**, Hub RNAs from the same chromosome tend to interact with each other. **l**, Expression pattern of hub RNAs in different subcellular fractions. **m**, Hub RNAs prefer to interact with each other in the same subcellular fraction (chromatin,  $n = 493$ ; nucleoplasmic,  $n = 61$ ; cytoplasmic,  $n = 80$ ). **n, o**, Hub RNAs and their targets are classified into

three distinct clusters based on 187 RBP motifs. **p**, Hub RNAs in the same cluster tends to interact with each other. Cluster 1 to 3,  $n = 292, 79, 271$ . **q**, Hub RNAs in the same and different Gene Ontology terms are clustered on the basis of their relative interaction strength. From 1 to 12,  $n = 186, 379, 253, 51, 213, 65, 79, 206, 61, 108, 54$  and 59. **r**, Over-represented motifs in super-enhancer-related hub RNAs. **s**, RBPs enriched in super-enhancer-related hub RNAs. **t**, The eCLIP peaks enriched in super-enhancer-related hub RNAs. For the violin plots in **b** and **c**, the white centre point represents the median, the box limits represent Q1 and Q3, the whiskers are the most extreme data points within  $1.5 \times$  the interquartile range (from Q1 to Q3), and the upper–lower limits represent the maximum–minimum values. The two-sided binomial test was used to calculate the *P* values shown in **k, m, p** and **q**.

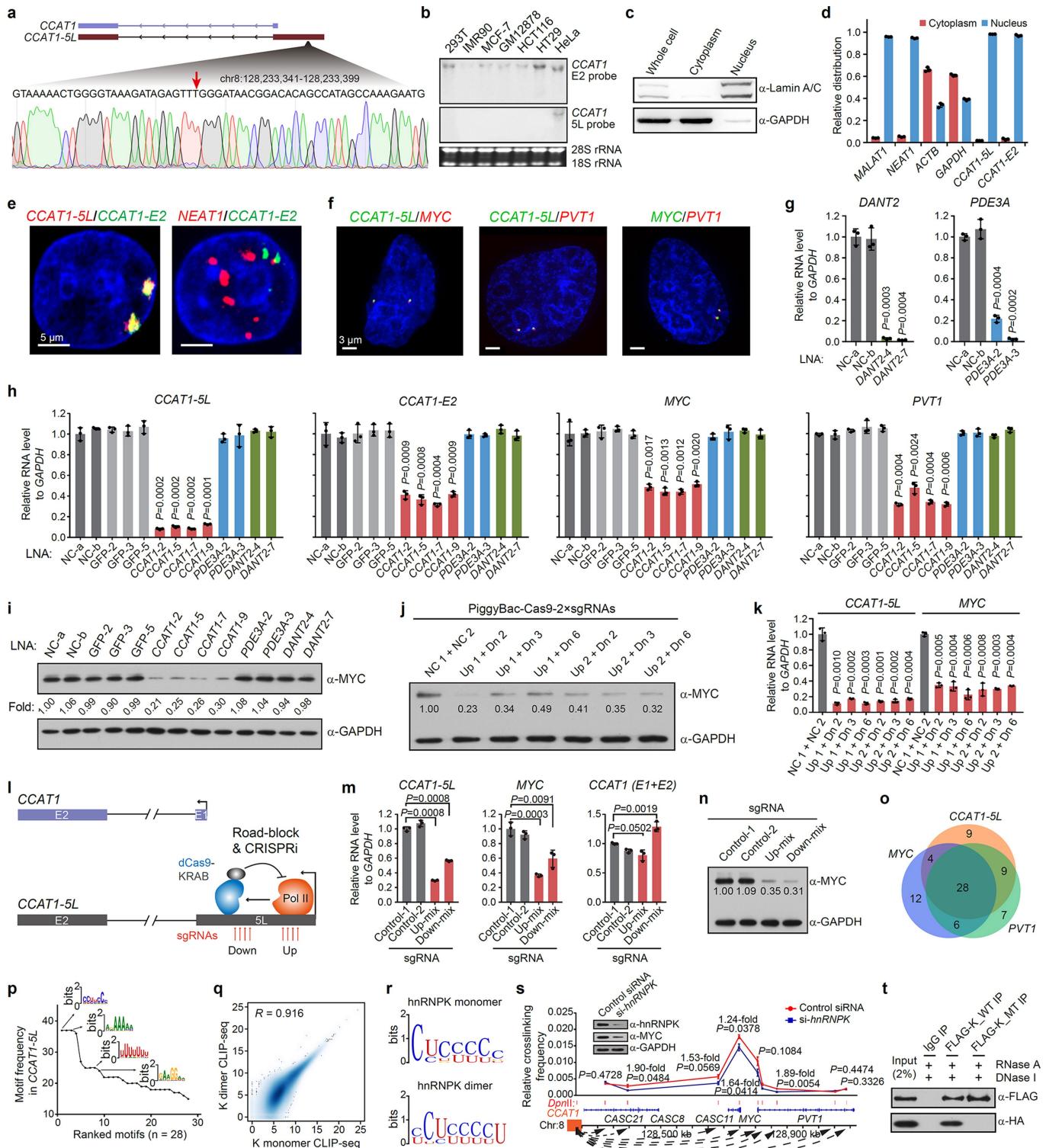


**Extended Data Fig. 8** | See next page for caption.

## Article

**Extended Data Fig. 8 | Features and validation of enhancer–promoter interaction.** **a**, Pie chart shows the strand preference of eRNAs interacting with promoter RNAs. **b**, eRNAs tend to interact with other RNAs within TADs. Colour intensity indicates read density in 40-kb windows. The purple line denotes the boundary of TADs. **c**, Circos plot showing whole-genome enhancer–promoter contacts on the basis of their pairwise-interacting RNAs. Red circle, super-enhancers; blue circle, typical enhancers; yellow circle, promoters. Red and light green arc lines illustrate inferred super-enhancer–promoter and typical-enhancer–promoter interactions, respectively. **d**, qPCR validation of enhancer–promoter interactions at four typical enhancers upon

depletion of eRNAs with LNA ASOs. Specific-enhancer-linked promoter reads are shown as blue arc lines above the genes at positive and negative strands. *CSAR1*, *LRSAM1*, *AK1*, *SPTAN1*, *ZDHHC12*, *LINCO2398*, *SLCO1C1* and *SLCO1B1* served as locus-specific controls. The relative fold change is normalized to the LNA control. **e**, Inferred enhancer–promoter and promoter–promoter interaction networks on chromosome 8. qPCR showing the expression level of super-enhancer-638-linked genes (box) upon depletion of super-enhancer RNAs with LNA ASOs. *PCAT1* and *GSDMC* served as locus-specific negative controls. Data in **d**, **e** are mean  $\pm$  s.d.;  $n=3$  biological replicates, two-tailed, unpaired *t*-test.



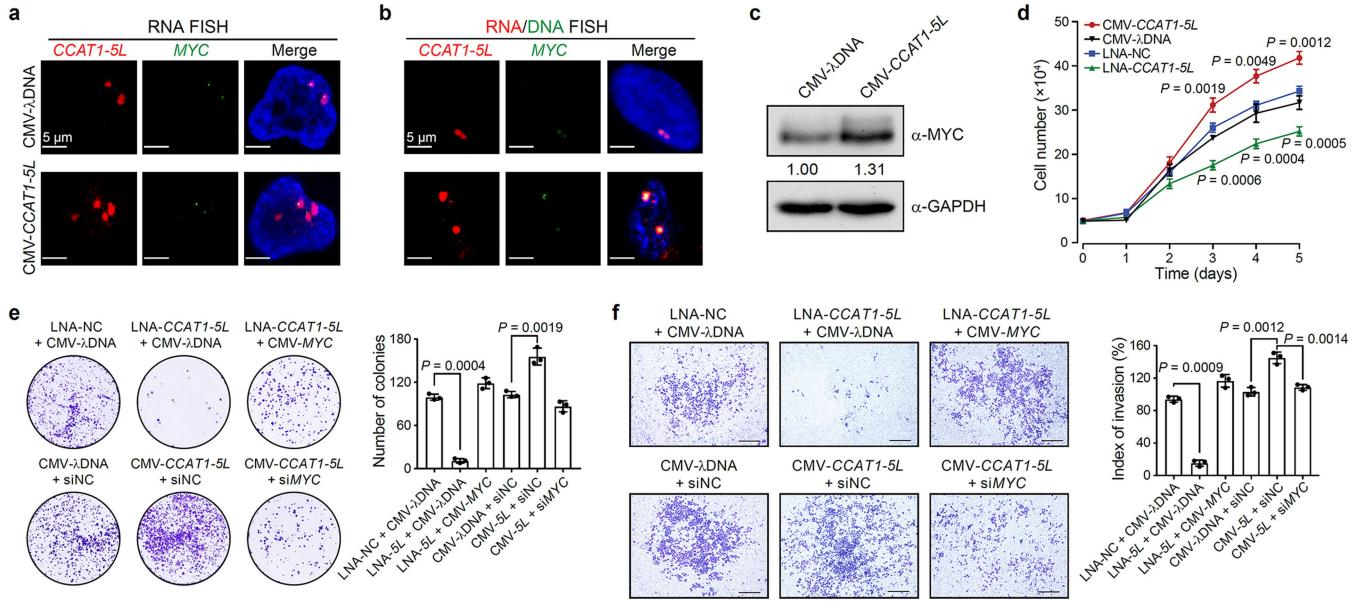
**Extended Data Fig. 9 |** See next page for caption.

# Article

## Extended Data Fig. 9 | Characterization of *CCAT1-SL* and its interaction partner.

**a**, Genotyping results showing no integration of human papillomavirus at *CCAT1-SL* locus. **b**, Northern blot analysis of *CCAT1-SL* expression across diverse cell lines. **c**, Western blot showing the purity of nuclear and cytoplasmic fractions. **d, e**, *CCAT1-SL* is localized in the nucleus revealed by qPCR and smFISH. *CCAT1-E2*, *CCAT1* exon 2. **f**, DNA FISH showing *CCAT1-SL*, *MYC* and *PVT1* loci are colocalized. **g**, qPCR showing efficient knockdown of hub RNAs *DANT2* and *PDE3A* with LNA ASOs. Non-targeting LNAs, NC-a and NC-b. **h**, qPCR showing reduced expression of *CCAT1-SL*, *CCAT1-E2*, *MYC* and *PVT1* upon *CCAT1-SL* knockdown with LNA ASOs. **i**, Western blot showing reduced MYC levels upon *CCAT1-SL* knockdown. **j**, Western blot showing reduced MYC levels upon the deletion of the extra extended region of *CCAT1-SL* by CRISPR–Cas9. **k**, qPCR showing reduced levels of *CCAT1-SL* and

*MYC* in mutant cell lines. **l**, A cartoon depicts the transcriptional blocking of the extra extended region of *CCAT1-SL* by CRISPRi. **m**, qPCR showing the expression levels of *CCAT1-SL*, *MYC* and *CCAT1* exon 1 and 2 (*CCAT1*(*E1+E2*)) upon CRISPRi. **n**, Western blot showing reduced MYC levels upon the CRISPRi of *CCAT1-SL*. **o**, Motifs enriched in *CCAT1-SL*, *MYC* and *PVT1*. **p**, Enriched motifs in *CCAT1-SL* RNA. **q**, HnRNPK monomer and dimer CLIP-seq are highly correlated in HeLa cells ( $n = 313,762$  windows). **r**, Consensus motifs identified by hnRNP K CLIP-seq. **s**, 3C-qPCR analysis of the long-distance interactions at the *CCAT1-SL*, *MYC* and *PVT1* loci upon *hnRNPK* knockdown with siRNA. **t**, Co-IP showing hnRNP K (T389A/Q391A) mutant (Flag\_KMT) did not form a dimer in HeLa cells. Data in **d, g, h, k, m** are mean  $\pm$  s.d.; Data in **s** are mean  $\pm$  s.e.m.;  $n = 3$  biological replicates, two-tailed, unpaired *t*-test. The experiments in **b, c, e, f, i, j, n, t** were independently repeated three times with similar results.



**Extended Data Fig. 10 | CCAT1-5L promotes cell proliferation and metastasis via MYC.** **a**, smFISH showing the ectopically expressed *CCAT1-5L* lncRNA and *MYC* promoter RNA are colocalized. *CCAT1-5L*, red; *MYC*, green; DAPI, blue. **b**, Dual RNA–DNA FISH showing the ectopically expressed *CCAT1-5L* RNA is colocalized with *MYC* locus. *CCAT1-5L* lncRNA, red; *MYC* locus, green; DAPI, blue. **c**, Western blot showing the ectopically expressed *CCAT1-5L* boosts *MYC* levels in HeLa cells. GAPDH served as a loading control. The *MYC* level is

quantified in the middle. **d**, The depletion or ectopic expression of *CCAT1-5L* (*SL*) in HeLa cells influences the proliferation rate. **e**, Knockdown or ectopic expression of *CCAT1-5L* affects colony formation. **f**, *CCAT1-5L* is critical for cell metastasis in a transwell assay. Scale bar, 50  $\mu$ m. The experiments in **a–c** were independently repeated three times with similar results. Data in **d–f** are mean  $\pm$  s.d.;  $n=3$  biological replicates, two-tailed, unpaired *t*-test.

# 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.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- 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

RIC-seq read sequences were acquired using Illumina instrumentation and software (HiSeq X Ten, bcl2fastq2 Conversion Software v2.16). For smFISH, probes were designed by Stellaris® Probe Designer (<https://www.biosearchtech.com/stellaris-designer>), SIM distances were measured using the Imaris software (v9.2, Bitplane AG).

Data analysis

All software and code used in this study has been described in published literature (Trimmomatic v0.36, Cutadapt v1.15, STAR v020201, SAMtools v0.1.19, IGV v2.3.92, Juicebox v1.6.2, VARNA v3-93, HISAT2 v2.1.0, Cufflinks v2.2.1, RNAhybrid v2.1.2, MEME v4.11.4, DSSR v1.7.7, PyMOL Educational version, MACS2 v2.2.4, Cytoscape v3.7.2) or are custom scripts available on GitHub (<https://github.com/caochch/RIC-seq>).

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

RIC-seq and hnRNPK CLIP-seq data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE127188.

# 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](https://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  No statistical methods were used to predetermine sample sizes. The sample sizes were determined based on previous experience.

Data exclusions  No data was excluded.

Replication  Each experiment was performed independently at least two times. All experiments were highly reproducible.

Randomization  No randomization was required for this study since no comparisons were made between samples/experimental groups.

Blinding  No blinding was required for this study since no comparisons were made between samples/experimental groups.

## 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                                     |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Antibodies            |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology                    |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms      |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants      |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                    |

### Methods

- |                                     |   |
|-------------------------------------|---|
| n/a                                 | Involved in the study                           |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

## Antibodies

### Antibodies used

Antibody (clone), Biological source, Supplier name, Catalog number, Lot number, Dilution, Application  
 GAPDH, Mouse, ABclonal, AC033, 910033001, 1:10000, WB  
 MYC, Rabbit, Proteintech, 10828-1-AP, 00081188, 1:1000, WB  
 LaminA/C (346), Mouse, Santa Cruz Biotechnology, sc-7293, F2113, 1:400, WB  
 Pol II (N-20), Rabbit, Santa Cruz Biotechnology, sc-899, D2715, 1:1000, WB  
 Goat anti-mouse IgG (H+L) HRP conjugate, Invitrogen, 31430, TC262979, 1:10000, WB  
 Goat anti-rabbit IgG (H+L) HRP conjugate, Invitrogen, 31460, TC260069, 1:10000, WB  
 Rabbit Anti-Mouse IgG (Light Chain Specific) (D3V2A) mAb (HRP Conjugate), CST, #58802, 5, 1:5000, WB  
 hnRNP K, Bethyl, A300-676A, A300-676A-1, 1:5000, WB/CLIP-seq/IP  
 HA, Mouse, Sigma, H9658, 127M4869V, IP/WB  
 FLAG, Mouse, Sigma, F1804, SLBT7654, IP/WB  
 Pol II (8WG16), Mouse, Santa Cruz Biotechnology, sc-56767, G0219, ChIP-qPCR  
 hnRNP K (D-6), Mouse, Santa Cruz Biotechnology, sc-28380, B0615, ChIP-qPCR

### Validation

All antibodies have been validated by the manufacturer and checked in the lab by western blot with human cell lysate. The reference and manufacturer validations are listed as follows:  
 GAPDH: <https://abclonal.com.cn/catalog/AC033>  
 MYC: <http://www.ptgcn.com/products/MYC-Antibody-10828-1-AP.htm>  
 Lamin A/C: <https://www.scbt.com/zh/p/lamin-a-c-antibody-346?requestFrom=search>  
 Pol II (N-20): see reference at <https://www.sciencedirect.com/science/article/pii/S0092867413005035?via%3Dihub>  
 Goat anti-mouse IgG (H+L) HRP conjugate: <https://www.thermofisher.com/cn/zh/antibody/product/Goat-anti-Mouse-IgG-H-L-Secondary-Antibody-Polyclonal/31430>  
 Goat anti-rabbit IgG (H+L) HRP conjugate: <https://www.thermofisher.com/cn/zh/antibody/product/Goat-anti-Rabbit-IgG-H-L-Secondary-Antibody-Polyclonal/31460>  
 Rabbit Anti-Mouse IgG (Light Chain Specific) (D3V2A) mAb (HRP Conjugate): <https://www.cellsignal.com/products/secondary-antibodies/rabbit-anti-mouse-igg-light-chain-specific-d3v2a-mab-hrp->

conjugate/58802?\_=1580714122786&Ntt=58802&tahead=true  
hnRNPk:  
<https://www.biomol.com/products/antibodies/primary-antibodies/general/anti-hnrnp-k-a300-676a-t?number=A300-676A>  
HA: <https://www.sigmaldrich.com/catalog/product/sigma/h9658?lang=zh&region=CN>  
FLAG:  
<https://www.sigmaldrich.com/catalog/product/sigma/f1804?lang=zh&region=CN>  
Pol II (8WG16):  
<https://www.scbt.com/zh/p/pol-ii-antibody-8wg16?requestFrom=search>  
hnRNPK (D-6):  
<https://www.scbt.com/p/hnrnp-k-antibody-d-6?requestFrom=search>

## Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

HeLa (CCL-2), HCT116 (CCL-247), MCF-7 (HTB-22), 293T (CRL-3216), IMR-90 (CCL-186), and Schneider's Drosophila Line 2 (CRL-1963) are from ATCC. HT29 are from Dr. Ling-ling Chen. GM12878 are from Coriell Institute.

Authentication

Identity of HeLa cells was frequently checked by the morphological features and the RNA expression profile, but not authenticated. The other cell lines were not authenticated but verified by morphological characteristics.

Mycoplasma contamination

No mycoplasma contamination was detected in these cells.

Commonly misidentified lines  
(See [ICLAC](#) register)

No commonly misidentified cell line was used.