SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Cell culture. HeLa-S3 cells were cultured in DMEM+ 10% FBS + Pen/Strep at 37ºC with 5% CO_2. Flp-In HEK293T-REx cells express doxycycline inducible 3XFLAG-DDX3X at a level comparable to endogenous DDX3X (from YJ Cho lab). The HEK293T cells were cultured with DMEM + 10% Tetracycline-free FBS (Clontech, Tet System Approved FBS, 631106) + Pen/Strep + 2mM Glutamax (Gibco). Cells were induced for 8 hours with 1 µg/ml Doxycycline before AMT crosslinking. J1 mouse ES cells (male) were cultured on gelatin coated plates at 37ºC with mES media: 500ml Knockout DMEM (Gibco), 90ml FBS, 6ml non-essential amino acid (MEM NEAA, 100x, Gibco), 6ml glutamine or glutamax (200mM stock solution), 6ml Pen/Strep, 1ml BME and 60µlLIF (Millipore, ESG1106).

icSHAPE in HEK293 cells. The icSHAPE experiments were performed as previously described (Flynn et al., 2016). The sequencing data were deposited in GEO with the accession number: GSE74353.

PARIS (Step 1. in vivo crosslinking and RNA purification): AMT (Sigma-Aldrich A4330) was dissolved in pure water at a concentration of 1mg/ml. Cells cultured to 70% confluency in 10cm or 15cm plates were treated with the normal culture media plus AMT for 30min in 37ºC incubator (40% v/v media and 0.5mg/ml AMT, adjusting osmolarity to physiological with 10x PBS). After the 30min incubation, the media was replaced with 0.5mg/ml AMT in 1x PBS. The control cells were incubated without AMT in either the incubation phase or the crosslinking phase. The plates in crosslinking solution were placed on ice bed in a Stratalink 2400 UV crosslinker and crosslinked for 30 min under UV365nm bulbs (Calvet and Pederson, 1981; Thompson and Hearst, 1983). The distance between cells and the bulbs is about 15cm. After crosslinking, cells were collected and frozen for further use. In general, the AMT-treated cell pellets have a darker color than the non-treated cells. This color difference can be used as an indicator of successful crosslinking. To purify RNA from the crosslinked cells, cell pellets were dissolved in 3 volumes of urea/SDS buffer (4M urea with 0.1% SDS), and pipetted vigorously. To each tube of dissolved cells (~ 20 million), add S1 nuclease buffer to a final concentration of 1x and add 2µl S1 nuclease (ThermoFisher EN0321), incubate at room temperature for 10min, with frequent pipetting to break viscous material. The S1 nuclease digestion is necessary to extract RNA from the crosslinked cells, since the crosslinking generates insoluble complexes in TRIzol (See Figure S1A-C for details of the digestion results). Both the S1 nuclease and subsequent ShortCut RNase III (NEB M0245) digestion produces 5'-phosphate and 3'-hydroxyl groups that are directly usable for subsequent proximity ligation and library preparation. After S1 digestion, add 10% SDS to a final concentration of 1% and then add Proteinase K (PK) to a final concentration of 10µg/ml. Perform the PK digestion at 50C for 30min to remove all proteins crosslinked to RNA. Even though psoralen crosslinking is selective for nucleic acids, proteins are also crosslinked to a lower degree (Sastry et al., 1997; Sastry et al., 1993). Then RNA is purified by TRIzol (Life Technologies) and resuspended in water. The S1/PK purified RNA was run on a Bioanalyzer to confirm that the crosslinking and purification worked. Successfully crosslinked and extracted RNA has a major broad peak between 1000 and 4000nt not present in the controls (Figure S1B).

PARIS (Step 2. Digestion and 2D purification of crosslinked RNA). The S1/PK purified RNA was further digested with ShortCut RNase III (dsRNase) at 37C for 20min to reduce the size of the RNA fragments. This digestion is critical for proper 2D

separation of crosslinked RNA fragments (See Figure S1C for representative results). Each digestion reaction in 50µl included 20µg S1/PK purified RNA and 2µl (low RNase experiment in Figure 1) or 4µl (high RNase in Figure 1 and all other experiments) ShortCut RNase III. $MnCl_2$ and ShortCut buffer were used as described in manufacturer's manual. After digestion, the RNA was usually quantified by Bioanalyzer, and 10-15µg RNA was usually recovered from the digestion of 20µg RNA. Purified RNA was separated by 2D gel electrophoresis as follows. For first dimension electrophoresis, RNA samples were loaded on 12% native polyacrylamide gel (1.5mm x 10cm x 10cm) and run at 150V for 100 min, at room temperature. Higher voltage would heat up the gel and should be avoided. After electrophoresis, the gel was stained with SYBR Gold (ThermoFisher S11494) and imaged under 312nm UV light. The 254nm UV reverses the crosslinking and should be avoided. Gel slices of each lane were embedded and polymerized into the top of the second dimension 20% urea-TBE denatured polyacrylamide gel (1.5mm x 10cm x 10cm). Two to three slices of gels can be multiplexed in one second-dimension gel. The second dimension is typically run at 55W for 50min. The second dimension gel has higher temperature due to the high power, and the high temperature facilitates denaturation of the dsRNA fragments. After the second dimension electrophoresis, the gel is stained with SYBR Gold and imaged. Destaining the gel with TBE increases sensitivity and is recommended since the crosslinked RNA is not abundant. Gel containing crosslinked RNA above the diagonal is cut out and crushed for RNA extraction.

PARIS (Step 3. proximity ligation, reverse crosslinking and library preparation). Purified dsRNA fragments were quantified by Bioanalyzer and the yield is about 50-100 ng for each 2D gel. The proximity ligation was performed under the following conditions. Each ligation reaction contains 10µl RNA, 2µl 10x ligation Buffer, 5µl T4 RNA Ligase 1 (NEB M0437M), 1µl SuperaseIn (Life Technologies), 1µl 0.3mM ATP and 1µl water. The ligation mixture is incubated at room temperature overnight (16-20h). After ligation, the samples were boiled for 2 minutes to terminate the reaction. After heat denaturation, the samples turned murky. These samples were centrifuged to remove the precipitate and then precipitated by ethanol and 1µl glycoblue (Ambion) (no need for Trizol extraction). Samples were resuspended in 15µl water. To reverse crosslinking, samples were transferred to the lid of a multi-well cell culture plate and directly irradiated with UV254nm for 15min on ice bed.

Two different methods were used for library preparation, one for HeLa, and the other for HEK293 and mES cells. For the HeLa PARIS library preparation, the photo-reversed RNA samples were reverse transcribed using a set of barcoded primers to generate cDNA. The primers are designed as follows (ordered from IDTDNA): /5phos/DDDNNXXXXNNNNNNTACCCTTCGCTTCACACACAAG/iSp18/GGATCC/iSp18 /TACTGAACCGCNNNNNN, where the DDD indicates non-cytosine bases, XXXX indicates the barcode, N indicates any of the 4 bases, iSp18 is spacer. The order of the barcodes are the same as described before (Spitale et al., 2015). The first random hexamer region is used to check for PCR duplication, whereas the second random hexamer is the primer region. The subsequent library preparation for the HEK293 and mES cells is essentially the same as described before, starting from step 26 in the Flynn et al. protocol paper (Flynn et al., 2016). Briefly, RNA was first ligated with barcoded adapters and then reverse transcribed using primers annealed to the adapters. In each of the gel purification steps in library preparation, fragments with >40nt inserts are selected. Libraries are multiplexed and sequenced on the Illumina NextSeq. We noticed that the direct ligation of RNA to adapters was a better approach for the library

preparation because of higher reverse transcription efficiency, and therefore used the direct ligation method for all subsequent libraries.

Mapping and filtering of PARIS data. See Figure S1H for an overview of the analysis strategy. The custom scripts for data analysis are deposited in Github: https://github.com/qczhang/ and https://github.com/zhipenglu/. After sequencing of PARIS libraries, adapters from the 3' end were trimmed off using Trimmomatic (Bolger 2014). We then collapsed these reads to remove PCR duplicates with identical sequences including barcode region (using the readCollapse script from the icSHAPE pipeline: https://github.com/qczhang/icSHAPE). The inclusion of random hexamer in the middle of the primer/adapter allows us to remove PCR duplicates efficiently to achieve the goal that each read from the sequencing represent an independent single molecule measurement of an RNA structure. Then the libraries were split based on barcodes (using splitFastqLibrary from icSHAPE pipeline, https://github.com/qczhang/icshape) and 5' adapters were removed using Trimmomatic.

After primary preprocessing, reads were mapped to hg38 or mm10 genome indices using the STAR program (Dobin 2013). The parameters used are as follows. STAR --runMode alignReads --genomeDir STAR_index --readFilesIn fastq_file --outFileNamePrefix name_prefix --outReadsUnmapped Fastq --outFilterMultimapNmax 100 --outSAMattributes All --alignIntronMin 1 --scoreGapNoncan -4 --scoreGapATAC -4 --chimSegmentMin 15 --chimJunctionOverhangMin 15. The parameters chosen here reduce the penalty for gapped reads and allow mapping of chiastic reads. For the detailed analysis of a select subset of repetitive or intron-containing RNA genes, such as 45S rRNAs, mitochondrial rRNAs, snRNAs and XIST, STAR indices were made for each one individually. The human snRNA index contains the complete set of 9 snRNAs (U1, U2, U4, U6, U5, U11, U12, U4atac and U6atac). The human XIST index contains the XIST RNA without the introns. To make these mini-genome STAR indices, it is important to note that the STAR index generation step requires a custom value for the --genomeSAindexNbases option (see STAR manual for details). To map reads to the mini-genomes, the STAR parameters were adjusted so that --outFilterMultimapNmax is 1 and --alignSJoverhangMin is 15.

The STAR mapping produces normally mapped reads (xxxAligned_out.sam) and chiastically mapped reads (xxxChimeric_out.junction and xxxChimeric_out.sam). The gapped reads were extracted from the normally mapped reads and combined with chiastically mapped reads for making duplex groups. These gapped reads (including normal and chiastic reads) are the useful data for determining RNA structures and RNA:RNA interactions. We obtained 2.5%-6% gapped reads in all mapped reads (Table S1). The higher percentage of gapped reads in PARIS [compared to 0.28% in RPL (Ramani et al., 2015) and 2% in hiCLIP (Sugimoto et al., 2015)], demonstrates the efficiency of the experimental strategy. The ligation could happen on both ends of the crosslinked helices, and these two ligation events would generate reads in normal "splicing-like" configuration or in a chiastic manner (e.g. normal: LLLLLL-RRRRRR or chiastic: RRRRRR-LLLLLL, L for bases from the left arm, while R for the right arm). The presence of reads in a DG with both configurations suggest that these reads not only come from distinct molecules but also ligated in different ways, thus lending stronger support that these gapped reads come from structures.

Processing of gapped and chiastic reads to read groups. We implemented a pipeline that automatically processes the mapped and filtered reads into duplex groups

(DG tag in the SAM file). The processing first removed gapped reads that are gapped as a result of splicing and further removed PCR duplicates. This step of analysis was implemented in the samPairingCalling.pl script in the paris pipeline (https://github.com/qczhang/paris). Furthermore, we implemented an additional tag NG (non-overlaping group) in the SAM file to maximally pack DGs. The XG (chiastic group, XG:i:0 for normal gapped, XG:i:1 for chiastic on the same strand of the same chromosome, and XG:i:2 for all others) tag was implemented to differentiate normal gapped reads and chiastic reads, where the two arms are swapped in relative position. Chiastic reads also include additional reads that are mapped to different strands or different chromosomes.

The DG assembly uses a two-step greedy algorithm to maximize speed while maintaining the reasonable clustering of reads into distinct groups that support individual RNA helices. Before DG assembly, gapped reads are first sorted by coordinates and examined in one round. In the first step of DG assembly, we generate intermediate DGs by grouping gapped together. Each read is either added to an existing DG or used to establish a new DG based on this criterion: all reads in a DG must share at least 5nt in both arms. Because every intermediate DG (and also the final DG) is represented using the coordinates of the core regions where all reads share (this is to ensure that all reads in one DG overlap with each other), some intermediate DGs that are not overlapped by core regions actually represent the same duplex structure. So in the second step, we assembly final DGs by merging those intermediate DGs as long as the maximum gap of both arms between two intermediate DGs is less than 10nt and the maximum length of both arms of the final DG less than 30nt.

To identify DGs with high confidence, we used two criteria. First each DG must have at least two unique gapped reads that have different termini, which should come from structures in two individual RNA molecules. The ligation in solution ensures that similar gapped reads are unlikely to form multiple times in solution by chance alone. Second, we computed a connection score for the two arms of a DG, as follows: number of reads connecting the two arms divided by the coverage of gapped reads at the two arms (connection_A_B/sqrt(coverage_A * coverage_B), A and B representing the two arms). The coverage of the two arms are different from the gapped reads connecting them because each region could be covered by multiple DGs, and some of them are likely to be alternative structures, which are pervasive in the transcriptome (Figure 4). We used a cutoff of 0.01 to remove low score DGs. This second criterion in effect normalizes for transcript abundance and ensures that our RNA duplex map is not dominated by low frequency duplexes in abundant transcripts.

The NG assembly algorithm is summarized as follows (https://github.com/zhipenglu/). All DGs are first ranked by read numbers (more reads, higher rank). Then going through all the DGs, each DG is either considered a new NG if it overlaps with all existing NGs, or added to the first existing non-overlapping NG. The end result is that for any NG, none of the DGs overlap with each other.

PARIS determines RNA helices with two arms of ~20-30nt each. At this size, the inferred base pairs are nearly always unique, which justifies the term “near base pair resolution”. This is in contrast to 1D flexibility measurements of RNA structures (e.g. DMS-seq, SHAPE etc.), where the data are nucleotide resolution but the precise base pairs are not determined. Although PARIS is performed on cells rather than single molecules, the

information we obtain is for single molecules (one-arm to one-arm Watson-Crick basepairs for each structured RNA fragment).

Visualization of RNA structure models and PARIS data on IGV genome browser (Robinson et al., 2011). As described above, gapped reads were organized into DGs based on newly implemented tags, each providing strong support for a potential RNA duplex At the same time, DGs can be visualized as arcs connecting the two arms in zoomed-out view of transcripts, highlighting the architecture of entire transcripts. PARIS-constrained predicted base pairs are also displayed as arcs in zoomed-in views. This arrangement allows easy integration with flexibility measurements of RNA structures (like icSHAPE and DMS-seq), protein-binding sites (as determined by various CLIP methods), together with phylogenetic information, DNA sequence variants, copy number variation, and additional annotations of functional motifs in the transcriptome. This integration provides a holistic view of the structural basis of RNA functions.

Specifically, RNA structure models were prepared in bed format where each pair of second and third column coordinates constitutes a base pair. This format is similar to the connect format described by the mfold program (Zuker, 2003). This method allows simultaneous presentation of alternative and complex structures, and easy comparison with other types of data on genome browser tracks. The PARIS gapped reads alignment data were visualized with the following options in IGV. Gapped reads can be grouped by tag XG (chiastic group), DG (duplex group) or NG (non-overlapping group), colored by tag XG or DG as needed (see IGV website for detailed updates). Visualization of entire bam files is not recommended since long-range and especially inter-molecular interactions would extremely compress the DGs in arcs or read alignments. Instead, individual RNAs should be extracted from the bam files (aligned gapped reads) and bed files (arcs representing DGs) using the SAMtools and BEDtools programs. Examples and instructions can be downloaded from the following link https://www.dropbox.com/s/1oqkcfzlfafdahq/PARIS_visdata.tgz?dl=0

Analysis of RNA:RNA interactions. PARIS determines RNA:RNA interactions in a “all-to-all” fashion. This is the strength of the method, yet it makes the analysis very complicated. The majority of the human genome is duplicated sequences such as repetitive DNA, genes with multiple copies and/or pseudogenes. This property makes unambiguous identification of RNA:RNA interactions very difficult on a genomic scale. To identify true RNA-RNA interactions from PARIS data, the reads were mapped to selected subsets of RNAs, each one as a small “chromosome”. The non-redundant sets of RNA families from Rfam, most of which are noncoding, were used to make the human and mouse reference “genomes”. One sequence from each family was randomly chosen for each species. The 12S, 16S, 18S and 28S rRNAs were added since they are not extracted well from the Rfam. STAR mapping was performed with these specific parameters: --outFilterMultimapNmax 10 --alignSJoverhangMin 15 --outSAMattributes All --alignIntronMin 1 --scoreGapNoncan -4 --scoreGapATAC -4 --chimSegmentMin 15 --chimJunctionOverhangMin 15. Mapped reads were assembled into duplex groups and filtered to remove interactions within RNAs and ones with identical break points. To identify the interactions between ncRNAs and the 45S pre-rRNA, we mapped the reads to the Rfam non-redundant ncRNAs plus one copy of the 45S rRNA. To identify the interactions between ncRNAs and mRNA/lncRNAs, we constructed a reference containing the Rfam ncRNAs described above, all the miRNAs from miRBase and all RNAs longer than 300nt from UCSC annotations

Although we cannot give an accurate estimate of the proportion of reads from RNA:RNA interactions in all gapped reads for the reasons described above, we counted the number of gapped reads for several abundant RNA:RNA interactions, such as snRNAs (U4:U6, U2:U6, U4atac:U6atac etc.) and rRNAs (5.8S:28S) in the HEK293_1 dataset, to provide a glimpse of the RNA:RNA interaction abundance. In the 45S rRNA mini-genome, 5.8S:28S interaction is supported by 34862 reads for the 3 helices. U8:28S, 87 reads (shown in Figure 6), 376 reads in the mES dataset. U4:U6, 1734 reads. U2:U6, 1105 reads. U4atac:U6atac, 365 reads. U12:U6atac, 4 reads (filtered out due to the high stringency of the analysis).

MicroRNAs are detected in our PARIS data sets but generally toward the lower end of read coverage. For example, the Mir-295 cluster of microRNAs, which make up more than 50% of total miRNA in mouse ES cells, generated less than 20 gapped reads of in our mES dataset. We can detect the precursor structure of these miRNAs but not interactions with the known targets such as Casp2 or Ei24 mRNAs. Some possible explanation for the low coverage may be that the miRNA-mRNA duplex is typically short (~5 nt of the miRNA seed), and the Argonaute proteins entirely envelop miRNAs, which may blocks the psoralen crosslinking.

To examine the evolutionary conservation and covariation of the RNA:RNA interactions, the relevant sequences were downloaded from Rfam, and the interacting regions were aligned structurally using LocARNA (Will et al., 2007).

Comparison of PARIS data with STAU1 hiCLIP data. The STAU1 hiCLIP raw data from Sugimoto et al. were mapped to hg38 using STAR with the same parameters described above (Sugimoto et al., 2015). This mapping strategy yields even better results than the original pipeline since STAR deals with gapped mapping natively and chiastic reads are also mapped properly at the same time. The adapters from the hiCLIP were simply ignored by the STAR mapper. The mapped reads were then processed in the same way as the PARIS data. To compare the PARIS and STAU1 hiCLIP data, the duplex groups (DG) in bed12 format were intersected using the bedtools package using the following parameters.

Analysis of dsRBP CLIP data (DICER1 and DGCR8). The HEK293 cell DICER1 PAR-CLIP raw data rep1 and rep2 (Rybak-Wolf et al., 2014) were mapped to hg38 genome using STAR (Dobin et al., 2013) and the following parameters: --outFilterMultimapNmax 10 --outSAMattributes All. Significant target sites were called using the PARalyzer program with default parameters (Corcoran et al., 2011). Numbers of DICER1 binding sites identified are 8899 and 3496, respectively. The HEK293 cell DGCR8 HITS-CLIP mapped data D8_2 and T7_2 in bedgraph format were lifted to hg38 coordinates. Significant DGCR8 binding sites were called using the Piranha program with the following parameters: -s -a 0.7 -b 40 -u 0. Numbers of DGCR8 binding sites identified are 57110 and 7290, respectively. To compare CLIP data with PARIS defined structures, binding sites that are common in both replicates were extracted. A total of 2383 DICER1 and 6415 DGCR8 target sites were used in the comparison.

Phylogenetic analysis of RNA structure. The two arm intervals of each DG were used to extract multiple alignments from whole-genome alignments of 23 amniote vertebrate species (Ensembl, hg38 version) with the python script maf_extract_ranges_indexed.py (bxpython package, https://github.com/bxlab/bx-python). RNAalifold was used to predict

a consensus structure from the alignments for each DG with or without inter-arm base-pairing constraints (Lorenz et al., 2011). The significance of each conserved structure was assessed using SISSIz shuffling with the RIBOSUM matrix (Gesell and von Haeseler, 2006).

For the direct comparison between human and mouse PARIS determined structures, the mouse DGs were lifted from mm10 to hg38 coordinates using the liftOver utility and the mm10ToHg38.over.chain file (UCSC). The liftOver program was run with the following parameters: liftOver -minMatch=0.2 -minBlocks=0.2 -fudgeThick. The minMatch was reduced from the default so that most regions can be properly aligned between species. In order to visualize the mouse PARIS reads on the human genome in IGV, the mouse PARIS reads were first converted to bed format using bedtools, lifted to hg38 coordinates, and then converted back to bam format using bedtools. It is noted that this strategy is limited by the quality of the available genome alignments, and improvement of these alignments is beyond the scope of the current study. To visualize the multiple alignments of RNA structures, the aligned sequences were plotted using the R-chie program using the default settings, except the choice of color codes (Lai et al., 2012).

Analysis of alternative structures. We defined alternative structures as helices that overlap on one arm by more than 50% (see the scripts in https://github.com/zhipenglu/duplex). These alternative structures are critical for the dynamic remodeling of the spliceosome during splicing. DGs were intersected with each other to identify pairs of DGs that have one pair of overlapped arms (left-left, left-right or right-right), but not two pairs at the same time. Inter-arm structures were predicted and significant overlapping of base pairs were used as another filter for alternative structures (at least 50% overlap). For alternative structures in the HEK293T and mES cells, we also confirmed them using the icSHAPE data from matched cell types (Figure 4G) and conservation/covariation data (Figure S7C). We note that interlocked RNA structures are also detected, likely representing alternative structures or pseudoknots. Although we can not definitively show that the interlocked structures are indeed pseudoknots, the results represent a stepforward in the identification of pseudoknots, which were previously only identified by crystallography or pure prediction.

Determination of XIST structure. Among the three cell types we performed PARIS on, only HEK293 cells express XIST and undergo X inactivation even though their karyotype is hypotriploid. HEK293 cells have been used to study human X inactivation. In 293 cells, human Xist silences linked genes in an A-repeat-dependent manner; 293 cells are thus a reasonable model for our Xist structural analysis (Chow et al., 2007). HeLa cells are female, but do not undergo X inactivation, while the J1 mES cells are male. To facilitate the analysis and visualization of the intron-containing XIST RNA, PARIS reads were mapped to a single mature XIST RNA reference without the introns. Since the XIST RNA contains multiple repetitive regions, only uniquely mapped reads were used. Gapped reads with mismatches were removed from both the Aligned.out.sam and Chimeric.out.sam files before duplex group assembly. For the repA region, although we cannot definitively determine whether the inter-repeat structures are intramolecular or between several identical XIST RNA molecules, the tendency for proximal interactions suggests that the inter-repeat contacts are within the same molecule, as would arise during co-transcriptional folding. *Trans* interactions would not follow such distance constraints (Figure 6).

In vitro SPEN iCLIP

In iCLIP experiments, we use UV radiation to crosslink direct RNA-protein interactions, and then use RNase to digest the RNA, followed by end-labeling the RNA with ^{32}P to visualize the RNA-protein complex. A brief summary of the protocol is as follows. Cloning and Purification of Recombinant SPEN RRMs as previously described (Arieti et al., 2014). RepA and control RNA were in vitro transcribed with T7 polymerase (MEGAscript T7 Transcription Kit), DNase Treated (TURBO DNase), and then purified with TRIzol. Purified RNA was resuspended and folded as described (Maenner et al., 2010). Until completion of UV crosslinking, samples were kept at room temperature. In a 1:20 RNA:protein molar ratio, 400ng of control or RepA RNA was incubated with the recombinant RRMs in the Maenner buffer for 20 minutes. The mixture was crosslinked twice with 2500 x 100 uJ using a UV Stratalinker 2400 (254nm bulbs), with mixing by pipette in between. SPEN-RNA crosslinked samples were treated with RNaseA, and stopped with RNase inhibitor. RNaseA was removed by spin filtration using a 30kDa Amicon column (Millipore) three times at 6000xg for 5 minutes, using RNase-free water for each dilution. Samples were then radiolabeled by 5' kinasing with gamma-32P ATP (Optikinase). The completed kinase reactions were heated in denaturing, reducing sample buffer. The denatured samples were run on SDS Bis-Tris PAGE, transferred to nitrocellulose, and visualized by radioblot. The membrane in the region >8 kDa above the apparent protein size was cut out for RNA isolation. The RNA was isolated by Proteinase K treatment at 55C followed by acid phenol chloroform extraction and alcohol precipitation.

SPEN iCLIP library preparation and analysis

Preadenylated 3' biotinylated adapters were ligated to the isolated RNA through 3' end repair by a T4 PNK (NEB) dephosphorylation reaction, followed by RNA Ligase 1 (NEB) ligation reaction without ATP. Free adapter was removed by digestion with RecJ exonuclease. Subsequent library preparation (cDNA synthesis, circularization, library amplification and purification) were all completed as described by the FAST-iCLIP method (Flynn et al., 2015). Libraries were sequenced on an Illumina MiSeq using a 150-cycle v3 kit for single-end 75 base reads. Reads were mapped to the XIST or GFP RNA reference and RT stop sites were extracted.

Supplemental References

Anger, A.M., Armache, J.P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D.N., and Beckmann, R. (2013). Structures of the human and Drosophila 80S ribosome. Nature *497*, 80-85.

Arieti, F., Gabus, C., Tambalo, M., Huet, T., Round, A., and Thore, S. (2014). The crystal structure of the Split End protein SHARP adds a new layer of complexity to proteins containing RNA recognition motifs. Nucleic acids research *42*, 6742-6752.

Calvet, J.P., and Pederson, T. (1981). Base-pairing interactions between small nuclear RNAs and nuclear RNA precursors as revealed by psoralen cross-linking in vivo. Cell *26*, 363-370.

Chow, J.C., Hall, L.L., Baldry, S.E., Thorogood, N.P., Lawrence, J.B., and Brown, C.J. (2007). Inducible XIST-dependent X-chromosome inactivation in human somatic cells is reversible. Proceedings of the National Academy of Sciences of the United States of America *104*, 10104-10109.

Corcoran, D.L., Georgiev, S., Mukherjee, N., Gottwein, E., Skalsky, R.L., Keene, J.D., and Ohler, U. (2011). PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. Genome biology *12*, R79.

Ding, Y., Tang, Y., Kwok, C.K., Zhang, Y., Bevilacqua, P.C., and Assmann, S.M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature *505*, 696-700.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Duga, S., Asselta, R., Malcovati, M., Tenchini, M.L., Ronchi, S., and Simonic, T. (2000). The intron-containing L3 ribosomal protein gene (RPL3): sequence analysis and identification of U43 and of two novel intronic small nucleolar RNAs. Biochimica et biophysica acta *1490*, 225-236.

Flynn, R.A., Martin, L., Spitale, R.C., Do, B.T., Sagan, S.M., Zarnegar, B., Qu, K., Khavari, P.A., Quake, S.R., Sarnow, P.*, et al.* (2015). Dissecting noncoding and pathogen RNA-protein interactomes. Rna *21*, 135-143.

Flynn, R.A., Zhang, Q.C., Spitale, R.C., Lee, B., Mumbach, M.R., and Chang, H.Y. (2016). Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. Nature protocols *11*, 273-290.

Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H.*, et al.* (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. Nature *462*, 58-64.

Gesell, T., and von Haeseler, A. (2006). In silico sequence evolution with site-specific interactions along phylogenetic trees. Bioinformatics *22*, 716-722.

Gong, C., and Maquat, L.E. (2011). lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature *470*, 284-288.

Grow, E.J., Flynn, R.A., Chavez, S.L., Bayless, N.L., Wossidlo, M., Wesche, D.J., Martin, L., Ware, C.B., Blish, C.A., Chang, H.Y.*, et al.* (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature *522*, 221-225.

Hausner, T.P., Giglio, L.M., and Weiner, A.M. (1990). Evidence for base-pairing between mammalian U2 and U6 small nuclear ribonucleoprotein particles. Genes & development *4*, 2146-2156.

Hung, T., Pratt, G.A., Sundararaman, B., Townsend, M.J., Chaivorapol, C., Bhangale, T., Graham, R.R., Ortmann, W., Criswell, L.A., Yeo, G.W.*, et al.* (2015). The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. Science *350*, 455-459.

Kim, D.D., Kim, T.T., Walsh, T., Kobayashi, Y., Matise, T.C., Buyske, S., and Gabriel, A. (2004). Widespread RNA editing of embedded alu elements in the human transcriptome. Genome research *14*, 1719-1725.
Kim, H., Abeysirigunawarden, S.C., Chen, K., Mayerle, M., Ragunathan, K., Luthey-Schulten, Z., Ha, T., and Woodson, S.A. (2014). Protein-guided RNA dynamics during early ribosome assembly. Nature *506*, 334-338.
Kim, Y.K., Furic, L., Desgroseillers, L., and Maquat, L.E. (2005). Mammalian Staufen1 recruits Upf1 to specific mRNA 3'UTRs so as to elicit mRNA decay. Cell *120*, 195-208.
Lai, D., Proctor, J.R., Zhu, J.Y., and Meyer, I.M. (2012). R-CHIE: a web server and R package for visualizing RNA secondary structures. Nucleic acids research *40*, e95.
Lorenz, R., Bernhart, S.H., Honer Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms for molecular biology : AMB *6*, 26.
Macias, S., Plass, M., Stajuda, A., Michlewski, G., Eyras, E., and Caceres, J.F. (2012). DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. Nature structural & molecular biology *19*, 760-766.
Maden, B.E. (1986). Identification of the locations of the methyl groups in 18 S ribosomal RNA from Xenopus laevis and man. Journal of molecular biology *189*, 681-699.
Maenner, S., Blaud, M., Fouillen, L., Savoye, A., Marchand, V., Dubois, A., Sanglier-Cianferani, S., Van Dorsselaer, A., Clerc, P., Avner, P.*, et al.* (2010). 2-D structure of the A region of Xist RNA and its implication for PRC2 association. PLoS biology *8*, e1000276.
McGinnis, J.L., and Weeks, K.M. (2014). Ribosome RNA assembly intermediates visualized in living cells. Biochemistry *53*, 3237-3247.
Nawrocki, E.P., Kolbe, D.L., and Eddy, S.R. (2009). Infernal 1.0: inference of RNA alignments. Bioinformatics *25*, 1335-1337.
Rinke, J., Appel, B., Digweed, M., and Luhrmann, R. (1985). Localization of a base-paired interaction between small nuclear RNAs U4 and U6 in intact U4/U6 ribonucleoprotein particles by psoralen cross-linking. Journal of molecular biology *185*, 721-731.
Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nature biotechnology *29*, 24-26.
Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J.S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. Nature *505*, 701-705.
Rybak-Wolf, A., Jens, M., Murakawa, Y., Herzog, M., Landthaler, M., and Rajewsky, N. (2014). A variety of dicer substrates in human and C. elegans. Cell *159*, 1153-1167.
Sastry, S.S., Ross, B.M., and P'Arraga, A. (1997). Cross-linking of DNA-binding proteins to DNA with psoralen and psoralen furan-side monoadducts. Comparison of action spectra with DNA-DNA cross-linking. The Journal of biological chemistry *272*, 3715-3723.
Sastry, S.S., Spielmann, H.P., Hoang, Q.S., Phillips, A.M., Sancar, A., and Hearst, J.E. (1993). Laser-induced protein-DNA cross-links via psoralen furanside monoadducts. Biochemistry *32*, 5526-5538.
Spitale, R.C., Flynn, R.A., Zhang, Q.C., Crisalli, P., Lee, B., Jung, J.W., Kuchelmeister, H.Y., Batista, P.J., Torre, E.A., Kool, E.T.*, et al.* (2015). Structural imprints in vivo decode RNA regulatory mechanisms. Nature *519*, 486-490.
Sugimoto, Y., Vigilante, A., Darbo, E., Zirra, A., Militti, C., D'Ambrogio, A., Luscombe, N.M., and Ule, J. (2015). hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. Nature *519*, 491-494.
Tarn, W.Y., and Steitz, J.A. (1996). Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. Science *273*, 1824-1832.
Thompson, J.F., and Hearst, J.E. (1983). Structure of E. coli 16S RNA elucidated by psoralen crosslinking. Cell *32*, 1355-1365.

Will, S., Reiche, K., Hofacker, I.L., Stadler, P.F., and Backofen, R. (2007). Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. PLoS computational biology *3*, e65.
Xue, S., Tian, S., Fujii, K., Kladwang, W., Das, R., and Barna, M. (2015). RNA regulons in Hox 5' UTRs confer ribosome specificity to gene regulation. Nature *517*, 33-38.
Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic acids research *31*, 3406-3415.