SUPPLEMENTAL INFORMATION

EXTENDED EXPERIMENTAL PROCEDURES

Cell Culture and Flavopiridol Treatment
For most RAP experiments, we used V6.5 male mouse ES cells grown in 2i + LIF on plates precoated with 0.2% gelatin and 3.5 μg/mL laminin. For Xist RAP-DNA, we used pSM33 male mouse ES cells induced with doxycycline for three hours to activate Xist expression as previously described (Engreitz et al., 2013). For Xist RAP-RNA, we used pSM33s induced with doxycycline for six hours to increase total Xist abundance and thus provide a control with expression better matched to that of Malat1. For transcription inhibition, we treated pSM33s (in the absence of doxycycline) for one hour with 1 μM flavopiridol or vehicle only (DMSO). We first made a 1 mM flavopiridol stock solution by resuspending flavopiridol hydrochloride hydrate (Sigma) in DMSO, and then added flavopiridol to 1 μM in cell culture media. We harvested cells after one hour to minimize effects due to flavopiridol toxicity. We note that replicate RAP experiments across the two male mouse ES cell lines (V6.5 and pSM33) produced similar data.

Probe Design and Generation
To generate tiled probe sets to capture target RNAs, we synthesized barcoded oligo pools as described previously for Malat1, Xist, and Hdac2 (Engreitz et al., 2013) (Table S4). To capture the 165-nt U1 snRNA, we designed three 50-nt ssDNA oligos that did not overlap the U1 sequence that hybridizes to pre-mRNAs (Table S4). We find that these oligos capture all of the mouse U1 homologs (Rnu1b1, Rnu1b2, Rnu1b6, Rnu1a1) despite small sequence differences. We ordered the U1 oligos with a 5’ biotin from Integrated DNA Technologies (IDT) and mixed them at equimolar concentrations for a total of 1 μM probe stock solution in H2O. We used a similar strategy to design 50-nt probes for U12 (Rnu12), U3 (Rnu3b1), and U17a (Snora73a) (Table S4).

RAP with FA-DSG Crosslinking
We performed RAP-RNA[FA-DSG] and RAP-DNA similar to Engreitz et al. (2013), with modifications to improve the sensitivity and specificity. First, we increased the amount of probe used in the hybridization reaction to improve the efficiency of capture for low abundance transcripts. Second, we used DNA probes instead of RNA probes to reduce background resulting from direct, artificial interactions between the RNA probes and proteins. Finally, we used RNase H to specifically elute captured complexes off the streptavidin-coated beads, similar to previous approaches (Simon et al., 2011).

We crosslinked mouse ES cells with 2 mM disuccinimidyl glutarate (DSG) for 45 min at room temperature followed by 3% formaldehyde (FA) for 10 min at 37°C. We lysed crosslinked or noncrosslinked cells and solubilized chromatin using a combination of sonication and DNase treatment as previously described (Engreitz et al., 2013), except that we increased sonication time from one minute to two minutes to improve chromatin solubilization. For each RAP experiment, we used lystate from 5 million cells, 50 pmol of ssDNA probe, and 625 μl streptavidin-coated beads. We performed the hybridization at 37°C and washed six times in 1× original bead volume RNase H Elution Buffer. Following these washes, we resuspended the streptavidin-coated beads once in 1× original bead volume RNase H Elution Buffer and 7.5 μl sample volume RNase H Binding Buffer and 1.9× sample volume 100% ethanol during the binding step to remove residual RT primer. We resuspended ssDNA probes in H2O at 100 ng/μL.

For DNA sequencing, we incubated the Proteinase K reaction at 60°C overnight to completely reverse crosslinks and isolated DNA using silane-coated magnetic beads as previously described (Engreitz et al., 2013), eluting in 12.5 μl H2O. We used the NEBNext Ultra DNA Library Prep Kit (New England Biolabs), scaled to 1/4 volume in the manufacturer’s protocol, to generate DNA sequencing libraries.

For RNA sequencing, we incubated the Proteinase K reaction at 60°C for two hours to reverse crosslinks as completely as possible without excessive fragmentation of captured RNAs. We purified nucleic acids using silane-coated magnetic beads. We eluted in 31 μl H2O, then added 10 μl 5X FNK Buffer (50 mM Tris-HCl pH 7.5, 5 mM MgCl2, 0.6 mM CaCl2, 50 mM KCl, 10 mM DTT, 0.01% Triton X-100), 1 μl Murine RNase Inhibitor (New England Biolabs), 3 μl FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific), 3 μl T4 Polynucleotide Kinase (New England Biolabs), 1 μl TURBO DNase (Life Technologies), and 1 μl Exonuclease I (New England Biolabs). This reaction digests residual ssDNA probe and genomic dsDNA and repairs the 5’ and 3’ RNA ends for adaptor ligation. We
incubated this reaction at 37 °C for 30 min, then cleaned the reaction with silane-coated magnetic beads and eluted in 6 μl H2O. From this point, we used the strand-specific RNA sequencing protocol described previously (Engreitz et al., 2013), starting with 3’ adaptor ligation.

**RAP-RNA with Formaldehyde Crosslinking**

For RAP-RNA[FA], we made slight modifications to the crosslinking and lysis protocols. We crosslinked V6.5 mouse ES cells with 2% formaldehyde for 10 min at 37 °C (no DSG). We enriched nuclei as above, but prior to sonication we resuspended nuclei from 20 million cells directly in 1 ml of RAP Hybridization Buffer. We solubilized chromatin and fragmented RNA by sonicating for 7 min at 10 W, omitting the DNase treatment step used above. This resulted in RNA fragments that averaged ~150 nucleotides. Following sonication, we increased the total volume of RAP Hybridization Buffer to match the lysate concentration used in RAP-RNA[FA-DSG]. Further steps were identical to RAP-RNA[FA-DSG].

**RAP-RNA with AMT Crosslinking**

We modified the RAP protocol further for RAP-RNA[AMT]. First, we fixed in vivo RNA-RNA interactions using 4’-aminomethyltrioxalen (AMT), a psoralen crosslinker (Sigma). We note that because psoralens require opposing uridines to generate interstrand crosslinks, direct RNA-RNA interactions that do not contain opposing uridines are likely inaccessible when using a psoralen crosslinking reagent (including ~5% of annotated 5’ splice sites that deviate from the consensus motif and do not contain AU or UA dinucleotides). We washed adherent ES cells (approximately 25 million in a 15-cm tissue culture dish) in room temperature PBS. We trypsinized and pelleted the cells, then washed the cells once more with PBS. We made a 0.5 mg/ml AMT solution in PBS by first resuspending AMT in water at a concentration of 1 mg/ml and then adding an equal volume of 2 × PBS. We chilled this solution in the dark on ice, then resuspended the cells in 4 ml of ice-cold AMT solution (or ice-cold PBS alone for –AMT control). We incubated the cells on ice for 15 min in the dark to allow the AMT to permeate the cells. Next, we transferred samples to a prechilled 10-cm tissue culture dish and placed the cells on ice under a long-wave UV bulb (350 nm) in a UV Stratalinker 2400 (Stratagene). The cells were positioned so they were approximately 3-4 cm underneath the light bulb. We exposed the cells to UV light at maximum power for 7 min with mixing every 2 min, thus activating the AMT crosslinking. We transferred irradiated cells to cold tubes and spun at 330 × g for 4 min to pellet cells. We isolated crosslinked RNA using TRIzol reagent (Life Technologies) following by two chloroform extractions and ethanol precipitation. To provide high resolution of U1 binding sites, we fragmented RNA to a median size of ~100 nt by incubating at 70 °C for 3 min in 1 × Fragmentation Buffer (Ambion). Finally, we digested residual DNA with TURBO DNase and re-cleaned the samples before beginning the RAP protocol. The removal of both protein and DNA from the hybridization eliminated potential sources of background as well as RNA-RNA interactions mediated by protein or chromatin intermediates.

For each RAP-RNA[AMT] experiment (+/ – AMT), we used 2 μg of input RNA, 15 pmol ssDNA probe, and 200 μl streptavidin-coated magnetic beads (Life Technologies). We denatured input RNA and probes separately in 5 μl water at 85 °C for 3 min, then mixed the probe and input RNA in 300 μl of LiCl Hybridization Buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 500 mM LiCl, 1% Triton X-100, 0.2% SDS, 0.1% sodium deoxycholate, 4 μm ligation) preheated at 55 °C. We incubated the hybridization reaction for two hours at 55 °C, shaking at 1200 r.p.m. We captured the magnetic beads and associated complexes, and washed the beads by resuspending completely 3 × in 250 μl of Low Stringency Wash Buffer (0.1 × SSPE, 0.1% SDS, 1% NP-40, 4 M urea) and 3 × with 250 μl of High Stringency Wash Buffer (1 × SSPE, 0.1% SDS, 1% NP-40, 4 M urea) for a total of six washes, each for 6 min at 60 °C. We eluted using RNase H, purified RNA using silane-coated magnetic beads, and proceeded to RNA sequencing as described above.

**Negative Controls for RAP**

To ensure that enrichment in RAP-RNA and RAP-DNA did not result due to technical artifacts, we performed numerous negative controls to eliminate potential sources of background as well as RNA-RNA interactions mediated by protein or chromatin intermediates. For RAP-RNA[AMT], these include off-target hybridization between the probes and other RNAs, and interactions between other RNAs and the streptavidin beads. To control for these potential issues, we performed the experiment in parallel with a –AMT control, which was treated identically to the +AMT sample except that AMT was not added prior to UV irradiation. When comparing the two experiments, 5’ splice sites and U1 motifs were enriched in U1 RAP-RNA +AMT but not –AMT.

For RAP-DNA and RAP-RNA[FA-DSG], the potential sources of background additionally include off-target hybridization with DNA and nonspecific interactions between the probes and proteins that might be crosslinked to RNA. To account for these potential issues, we performed two controls. First, we performed the RAP-RNA and RAP-DNA experiments in noncrosslinked lysate to control for off-target capture through direct probe-RNA hybridization. Second, we purified Xist (an abundant nuclear RNA) and Hdac2 (a randomly selected mRNA expressed in ES cells) from crosslinked lysate to control for nonspecific capture mediated by protein interactions with the ssDNA probes or streptavidin-coated beads. In RAP-DNA, noncrosslinked purifications did not generate DNA libraries with sufficient yield to quantify, and sequencing of this tiny amount of DNA did not yield peaks in any regions of the genome except for the endogenous locus (data not shown).

In RAP-RNA[FA-DSG], the control purifications strongly enriched for the target RNA (which typically comprised > 90% of the total sequencing reads). These controls did not significantly enrich for any regions of the transcriptome with the exception of certain GC-rich simple or tandem repeats and the surrounding RNA sequences. We note that these artificial signals are clearly recognizable.
To identify motifs enriched in RAP-RNA[AMT], we counted all 8-mer sequences contained within 30 bases upstream of the second testing. We considered either all reads (Figure 1C) or all reads mapping to introns that are greater than 200 nucleotides away and input. Significant 8-mers were defined using Fisher's exact test (p < 0.001 after Bonferroni correction for multiple hypothesis due to intramolecular crosslinks or UV damage. We compared the fraction of reads containing each 8-mer between RAP-RNA[AMT] and RAP-RNA[FA] versus input. An equal number of random sites matching the same characteristics (intronic, > 200 nucleotides from the nearest exon and counted the number of reads mapping within 30 nucleotides of the motif in RAP-RNA and input. We next filtered reads aligning to highly abundant RNA transcripts, including ribosomal RNAs, snRNAs, and repetitive elements, as defined by RefSeq and RepeatMasker. A FASTA file containing these sequences is available at the Gene Expression Omnibus (GSE55914). We aligned all remaining reads to the mouse transcriptome (RefSeq) and genome (mm9) using Tophat (version 2.0.8) and discarded reads with MAPQ < 30.

RNA Sequencing Alignment

To analyze RNA sequencing data, we first used Bowtie2 (version 2.1.0) to filter out reads mapping to the PCR tags on the capture probes, where applicable (Malat1, Xist, and Hdadc2). All reads passing this initial probe filter were included in the total read count for each experiment. We next filtered reads aligning to highly abundant RNA transcripts, including ribosomal RNAs, snRNAs, and repetitive elements, as defined by RefSeq and RepeatMasker. A FASTA file containing these sequences is available at the Gene Expression Omnibus (GSE55914). We aligned all remaining reads to the mouse transcriptome (RefSeq) and genome (mm9) using Tophat (version 2.0.8) and discarded reads with MAPQ < 30.

RAP-RNA Enrichment Calculations

To calculate the enrichment for a target RNA (e.g., U1), we divided the fractions of all reads mapping to the target RNA in RAP-RNA and input (e.g., (reads mapping to U1 in RAP-RNA / all RAP-RNA reads) / (reads mapping to U1 in input / all input reads)). To calculate the enrichment for all other nontarget RNAs or sets of RNAs (e.g., U2, introns), we performed the same calculation, except that we subtracted the number of reads mapping to the target RNA from the total read counts (e.g., (reads mapping to introns in U1 RAP-RNA / all RAP-RNA reads that do not map to U1) / (reads mapping to introns in input / all input reads that do not map to U1)). This adjustment was performed because a substantial fraction of reads in RAP-RNA experiments map to the target RNA, skewing the enrichment calculations. We note that in Figure 2A, the enrichment for U1 (11-fold versus input) was lower than the enrichment for Hdadc2 (6,500-fold) not because the U1 purification was less efficient but because U1 is much more abundant, representing 1/60 of the sequencing reads from nuclear-enriched input RNA, and thus cannot be enriched more than 60-fold versus input.

RAP-RNA[AMT] and RAP-RNA[FA] Read Count Analysis

To accurately map the positions of crosslink sites in RAP-RNA[AMT], we examined the 5’ base of the second read in each pair; in our RNA-sequencing protocol, this corresponds to the last nucleotide added by RT. Theoretically, RT could halt due to many reasons, including reaching the end of an RNA fragment, reaching the site of UV-induced RNA damage, or reaching the site of an AMT cross-link (either intramolecular or intermolecular). To focus on sites of AMT crosslinks, we normalized to the –AMT RAP-RNA control, which has RT stops due to all of the same reasons except for AMT crosslinks. For each read-end, we calculated the distance to the nearest 5’ splice site (defined by RefSeq). In Figure S2, read-pairs for snRNA RAP-RNA[FA] experiments are counted and plotted in the same manner to allow direct visual comparison of AMT and FA results; for RAP-RNA[FA], however, the fragment ends do not necessarily correspond with crosslink sites because formaldehyde crosslinks are reversed prior to reverse transcription. U12-intron locations were downloaded from U12DB: The U12 Intron Database (Alioto, 2007).

RAP-RNA[AMT] N-mer Analysis

To identify motifs enriched in RAP-RNA[AMT], we counted all 8-mer sequences contained within 30 bases upstream of the second read of each read-pair (i.e., upstream of the original RNA fragment). We extended 30 bases upstream to account for RT stops due to intramolecular crosslinks or UV damage. We compared the fraction of reads containing each 8-mer between RAP-RNA[AMT] and input. Significant 8-mers were defined using Fisher’s exact test (p < 0.001 after Bonferroni correction for multiple hypothesis testing). We considered either all reads (Figure 1C) or all reads mapping to introns that are greater than 200 nucleotides away from the nearest exon (Figure 1D). We chose 200 nucleotides because it represented the distance from the 5’ splice site at which U1 RAP-RNA[AMT] no longer showed enrichment over input (Figure 1B).

5’ Splice Site Motif Analysis

We enumerated and scored all 5’ss motifs as strong, medium, or weak matches using a maximum-entropy model trained on all known 5’ splice sites as described previously (Almada et al., 2013). For Figure 1E, we considered all 5’ss motif matches that occur in introns greater than 200 nucleotides from the nearest exon and counted the number of reads mapping within 30 nucleotides of the motif in RAP-RNA and input. An equal number of random sites matching the same characteristics (intronic, > 200 nucleotides from exons) were chosen as negative controls. We note that the same proportion (>95%) of strong, medium, and weak 5’ss motif matches contained AU or UA dinucleotides and thus were potentially discoverable with psoralen crosslinking.
RAP-RNA Peak Calling
To call significant peaks in the Malat1 transcript for U1 RAP-RNA[AMT], we compared read-end counts in 10-nucleotide windows in U1 RAP-RNA +AMT to two controls, U1 RAP-RNA –AMT and Input +AMT. Each window was scored using a binomial test parameterized by \( n \) (the number of read-ends occurring in the window in RAP-DNA +AMT), \( k \) (the number of read-ends mapping to the window in the control), and \( p \) (the total number of reads in RAP-DNA +AMT divided by the sum of the totals for RAP-DNA +AMT and control). U1 RAP-RNA +AMT was compared separately to each of the two controls, and the least significant \( p \) value of the two was chosen to represent the \( p \) value for a given window. Windows with a Bonferroni-corrected \( p \) value < 0.001 were called significant and adjacent significant windows were merged. We used a similar approach to identify significant peaks in FA-crosslinked and noncrosslinked RAP-RNA experiments.

RAP-RNA[FA-DSG] Enrichment in Introns and Exons
To calculate the enrichment in introns and exons, we used gene definitions from RefSeq. To focus on annotations with sufficient read coverage to confidently calculate enrichment ratios, we considered introns or exons with a read count \( \geq 10 \) in the input library. For gene-level enrichment calculations, we calculated the ratio of all reads mapping to introns or exons in a given gene in RAP-DNA versus input. Because both Malat1 and U1 significantly enriched for introns across most active genes, we defined highly enriched genes as those with enrichment in the top 5% of expressed genes. To compare introns with high density of U1 motifs, we considered only strong 5’ss motif matches (Almada et al., 2013) and normalized the number of strong motif matches to the length of the intron.

Gene Set Enrichment
We searched for significantly enriched Gene Ontology (GO) categories using DAVID 6.7 (Huang et al., 2009). We used the list of all genes expressed in ES cells as background. We reported all GO categories with a false discovery rate (FDR) < 0.001.

DNA Sequencing Alignment
We first filtered out reads mapping to the PCR tags on the capture probes, where possible (Malat1, Hdac2, Xist). We aligned remaining reads to the genome (mm9) using Bowtie2, removed duplicate read pairs with Picard (http://picard.sourceforge.net), and discarded reads with MAPQ < 30. For subsequent data analysis, we connected read-pairs into fragments defined by the span of the two reads.

RAP-DNA Data Analysis
To identify regions of the genome significantly enriched in RAP-DNA versus input, we used a sliding window approach at 2-kb resolution using 10-kb windows. In each window, we counted the number of fragments mapping to that window in RAP-DNA and in input and calculated a \( p \) value using a binomial test parameterized by \( n \) (the number of fragments mapping to the window in RAP-DNA), \( k \) (the number of fragments mapping to the window in input), and \( p \) (the total number of fragments in RAP-DNA divided by the sum of the totals for RAP-DNA and input). We ignored windows containing significant repeat content as previously described (Engreitz et al., 2013). We also ignored windows where the number of fragments in the input library was more than 100 \( \times \) the average across the genome (these windows represented subtelomeric or other repetitive regions that are not annotated in RepeatMasker). We counted a window as significant if the RAP-DNA experiment contained \( \geq 10 \) fragments and if the Bonferroni-corrected \( p \) value was less than 0.001. For Figure 4A, windows that did not pass these filters were not plotted. For RAP-DNA enrichment figures for specific genomic sites, signal was interpolated over repeat-masked windows as previously described (Engreitz et al., 2013).

Transcription Abundance Defined by Chromatin-Associated RNA
To estimate the level of transcription at each gene in the genome, we isolated chromatin-associated RNA as previously described (Bhatt et al., 2012; Pandya-Jones and Black, 2009). We sequenced chromatin-associated, soluble nuclear, and cytoplasmic RNA fractions without poly(A) selection using the strand-specific RNA-sequencing protocol described above. Known chromatin-associated RNAs like Tsix were strongly enriched in the chromatin fraction (data not shown). To estimate transcription levels, we calculated the number of reads per kilobase per million (RPKM) in the chromatin-associated fraction at each RefSeq gene, including both introns and exons. To correlate RAP-DNA enrichment with chromatin-associated RNA abundance, we calculated the enrichment of reads within the gene locus, including both introns and exons, in RAP-DNA versus input.

Controlling for Transcript Abundance in Gene Set Analyses
For Malat1 RAP-DNA, we wished to determine whether specific gene sets were preferentially enriched. We calculated the DNA enrichment across the entire gene body, and used a Mann-Whitney test to first compare enrichments in genes in a given gene set with enrichments at all other active genes. Because much of the variance in DNA enrichment at genes is determined by expression level, however, we performed a second analysis to normalize for this factor. We used a linear regression to predict log2 RPKM of chromatin-associated RNA and a binary variable indicating whether each gene is a member of the gene set of interest. Both RNA binding and alternative spliced genes scored in this linear model as positively associated with RAP-DNA enrichment level with an F-test \( p \) value < 0.05.
We used a similar approach to test whether U1 RAP-DNA enrichment at the 5’ ends of genes depended on the distance to the first 5’ss motif: the linear regression included a binary variable that indicated whether the first 5’ss motif occurred greater or fewer than 200 base-pairs away from the transcription start site.

**Relationship between U1 Localization and 5' Splice Sites near TSS**

We considered two possibilities to explain how U1 localizes to DNA at the 5’ ends of genes. First, because transcription initiation produces a short ~50 nucleotide transcript at the 5’ ends of genes before proximal polymerase pausing (Core et al., 2008), U1 localization to chromatin might occur through this short nascent RNA, which would exist even after inhibiting transcription elongation. In this case, U1 localization to chromatin at 5’ ends of genes might depend on the presence of 5’ss motifs in this short nascent RNA. Alternatively, U1 might interact with chromatin independent of this short nascent transcript, perhaps through associations with protein complexes at the 5’ ends of genes. To inform these possibilities, we examined active genes where the first 5’ss motif occurs > 200 bases downstream of the transcription start site (n = 1,001) and thus should not be present in the short nascent RNA. These genes showed the same pattern of U1 RAP-DNA enrichment as genes with 5’ss motifs closer to the transcription start site, both with and without flavopiridol treatment (Figures S4J and S4K). This analysis suggested that U1 localization to the 5’ ends of genes may not depend on interactions with U1 motifs in the short nascent RNA.

**SUPPLEMENTAL REFERENCES**


A

Direct RNA-RNA interaction (via hybridization)
RNA-1
RNA-2
RNA-3

Indirect RNA-RNA interactions (via protein intermediates)
RNA-1
RNA-2
RNA-3

B

RAP-RNA

Crosslinking: AMT
Fragmentation: Strong
Digest protein and DNA: Yes

Capture target RNA:

Sequence co-purified RNAs:

C

U1 (Rnu1b1)

Direction of reverse transcription

Read ends (relative counts)

Position (nt)

D

7SK (Rn7sk)

Read ends (relative counts)

Position (nt)
Figure S1. Comparison of RAP-RNA Protocols and Validation of AMT Crosslinking, Related to Figure 1

(A) Intermolecular RNA-RNA interactions can be categorized as direct through hybridization (purple) or indirect through protein intermediates (blue). Known examples (gray text) involve small nuclear RNAs (snRNAs), messenger RNAs (mRNAs), transfer RNAs (tRNAs), small nucleolar RNAs (snoRNAs), and ribosomal RNAs (rRNAs).

(B) We developed three RAP-RNA protocols that vary with respect to crosslinking, RNA fragmentation, and timing of protein digestion, each of which is best suited to specific applications. Crosslinking: AMT specifically crosslinks direct RNA-RNA interactions, provided that the resulting hybrid contains opposing uridine bases (RNA 1). FA crosslinks RNA-protein and protein-protein interactions, thus capturing indirect RNA-RNA interactions that occur through protein intermediates as well as direct interactions that are flanked or caged by proteins (RNA 2). The addition of DSG provides stronger protein-protein crosslinking than FA alone, allowing recovery of RNA-RNA interactions that occur through multiple protein intermediates (RNA 3). Fragmentation: The AMT and FA protocols use strong RNA fragmentation prior to capture, allowing high resolution mapping of RNA-RNA interactions. The FA-DSG protocol, in contrast, preserves RNA integrity prior to capture, thus potentially improving capture of intact IncRNAs. As a consequence, the FA-DSG method provides the most inclusive protocol to identify RNAs that interact either directly or indirectly with a target RNA, but does not provide high resolution for specific regions within interacting transcripts. Digest protein: The AMT protocol digests protein and DNA prior to capture to eliminate any possibility of capturing indirect RNA-RNA interactions. In contrast, the FA and FA-DSG protocols digest protein after capture in order to capture indirect as well as direct interactions.

(C and D) Nucleotide-resolution mapping of intramolecular AMT crosslinks for two RNAs with known structures: (C) U1 and (D) 7SK. Counts (left) represent second-strand reads mapping to each RNA for AMT-crosslinked (+AMT) and mock-crosslinked (-AMT) input RNA. Each read contributes one count in the position of the 5' end of the read, corresponding to the 3' end of the reverse-transcribed cDNA. Peaks represent common positions where reads end and largely correspond to the sites of crosslinks predicted based on the secondary structure (right). Orange arrows represent positions predicted to be crosslinked by AMT (uridines on opposing strands that do not occur in the middle of a stretch of complementary base-pairs). For 7SK, only the subsection of the RNA shown has predicted AMT crosslinks. RNA structures are adapted from Guthrie and Patterson (1988) (for U1) and Marz et al. (2009) (for 7SK). We note that some of the peaks in either +AMT or −AMT that do not correspond to predicted AMT crosslinks occur near U-U dinucleotides and perhaps result from UV-induced intrastrand crosslinks (e.g., U1 positions 72-74, 7SK positions 246-248).
Figure S2. RAP-RNA Is a General Method for Mapping RNA-RNA Interactions at High Resolution, Related to Figure 2
(A) Relative expression of small RNAs examined in this study (mean of three replicates ± s.d.) spans several orders of magnitude. Abundance is measured in reads per million and is not normalized by kilobase because all RNAs are approximately the same size (range: 150 to 215 nt) and many sequenced fragments represent the entire, intact transcript.
(B) Comparison of RAP-RNA[AMT] versus RAP-RNA[FA] for U3 and (C) Snora73a, which directly hybridize to specific sites on the pre-ribosomal 45S transcript to guide pre-rRNA processing. Graph shows the significance of enrichment (−log10 binomial p value) across the entire 45S pre-ribosomal RNA transcript (13.4 Kb) at 20-nucleotide resolution. Each read-pair contributes a count at the position corresponding to the 5’ end of the original RNA fragment. Known sites of direct RNA-RNA hybrids for U3 (Borovjagin and Gerbi, 2004; Hughes, 1996) and Snora73a (Fayet-Lebaron et al., 2009) are highlighted in gray. For U3, the AMT approach captures three known direct interaction sites with 45S but misses the fourth; in comparison, the FA approach identifies all four known sites and highlights several additional enriched sites at approximately positions 1,250 and 1,850 in the 18S transcript. Because the four previously known sites were found using psoralen crosslinking (Beltrame and Tollervey, 1992; Maser and Calvet, 1989; Stroke and Weiner, 1989) or genetic analyses (Borovjagin and Gerbi, 2004), these new contacts found by RAP-RNA[FA] may represent direct or indirect contacts inaccessible by previous methods.
Figure S3. Malat1 Indirectly Interacts with Pre-mRNAs Encoding RNA Binding Proteins, Related to Figure 3

(A) Enrichment (Malat1 RAP-RNA[AMT]) for every possible 8-mer sequence for all reads mapping to the genome (not to repetitive RNAs, see Experimental Procedures). 8-mer counts include the second strand read and the 30 bases upstream of the read. No 8-mers are significantly enriched (p < 0.001 after Bonferroni correction and fold enrichment ≥ 4) in Malat1 RAP-RNA versus input.

(B) Relationship between GC content and enrichment in RAP-RNA[FA-DSG] for U1 (left) and Malat1 (right). GC content and enrichment were calculated in 100-nt windows across introns and exons; we focused the 200 genes with the highest expression in nuclear-enriched input RNA to ensure that the denominator of the enrichment ratio (input) was well-represented. Boxplots compare the enrichment distributions of 100-nt windows in the indicated ranges of GC content (light gray), which were chosen based on the quartiles of GC content in exons.

(C) All enriched GO categories (FDR < 0.001) for genes highly enriched in Malat1 RAP-RNA[FA-DSG] versus input. Fold enrichment represents the over-representation of genes in each GO category compared to the universe of genes expressed in ES cells.

(D) Each point represents the average enrichment versus input of introns in a given gene for Malat1 and U1 RAP-RNA[FA-DSG]. Points above the gray dashed line represent the 5% of expressed genes with the highest enrichment in Malat1 versus U1 RAP-RNA.

(E) Enriched GO categories (FDR < 0.001) for genes enriched in Malat1 versus U1 RAP-RNA[FA-DSG]. Genes encoding RNA processing proteins are over-represented in genes preferentially enriched in Malat1 versus U1 RAP-RNA.

(F) Enrichment for individual transcripts or classes of RNAs for RAP-RNA[FA-DSG] replicate experiments comparing the RNA-RNA interactions of Malat1 and U1 to those of two other abundant nuclear ncRNAs (Xist and U3). The enrichment for introns (nascent transcripts) is unique to Malat1 and U1, indicating that the enrichments observed for Malat1 and U1 represent unique interactions compared to those of other abundant nuclear RNAs. RBP pre-mRNA indicates the enrichment for introns in genes encoding RNA-binding proteins.
Figure S4. Malat1 and U1 Localize to Active Genes through Two Mechanisms, Related to Figure 4

(A) Malat1 and U1 RAP-DNA enrichment versus input in 1-Kb windows at chr19:5,650,000-6,002,954. Malat1 enrichment extends beyond the y axis limit to a maximum of 800 at the Malat1 locus (red). Malat1 localizes strongly at the neighboring locus encoding Neat1. Genes within 1 megabase of the Malat1 locus on chromosome 19 have relatively strong Malat1 RAP-DNA enrichment as well as intron enrichment in Malat1 RAP-RNA (Table S3). Gene track merges exons and introns.

(B) Malat1 and U1 enrichment at chr3:95,863,882-96,148,881. U1 but not Malat1 localizes strongly to histone genes (purple). This region includes one genomic locus that encodes a U1 paralog (Rnu1b6, red). Gene track distinguishes exons from introns.

(C) Flavopiridol treatment reduces levels of introns by an average of 40%. Each point represents for one gene the average fold change in input nuclear-enriched RNA (flavopiridol versus DMSO) for its introns (black) or exons (gray).

(D) Cumulative distribution of the fold change of introns (black) or exons (gray) in input nuclear-enriched RNA (flavopiridol versus DMSO).

(E and F) RAP-DNA enrichment versus input in 1-Kb windows for Malat1 and U1 in cells treated with flavopiridol (+Flav) or DMSO (-Flav, control) at chr18:34,679,634-35,163,272 (same as Figure 4B). Gray box highlights a region shown at higher resolution in (F): chr18:34,730,000-34,830,000.

(G) Same as (E) at chr17:29,148,115-29,226,237.

(H) Malat1 RAP-DNA enrichment averaged over all active genes (black) and inactive genes (blue). Shaded regions represent 95% confidence intervals for the average enrichment. TSS: transcription start site. PAS: polyadenylation site. See also Figure 4D.

(I) U1 RAP-DNA enrichment averaged over active genes that span more than 20 kb in cells treated with DMSO (solid line, -Flav) or flavopiridol (dashed line, +Flav). See also Figure 4F.

(J) U1 RAP-DNA localization at 5' ends of genes in cells treated with DMSO or (K) flavopiridol. Each row represents one gene, ordered by increasing distance from the transcription start site (TSS) to the first strong or medium 5'ss motif (middle, Extended Experimental Procedures). Heatmap represents the fragment coverage in U1 RAP-DNA (left) and input (right) across a 3-kb region surrounding the TSS (white = low, red = high). Each row is normalized to the average coverage across the 3-kb region. The depletion of reads just upstream of the TSS is due to DNase digestion of hypersensitive sites at promoters (Engreitz et al., 2013). Dashed lines separate genes with a distance to the first 5'ss motif of more than 200 bp. Note that U1 strongly localizes to the 5' ends of genes regardless of the distance to the first 5'ss motif, suggesting that U1 localization to the 5' ends of genes may not depend on interactions with U1 motifs in the short nascent RNA produced by transcription initiation.