nature biotechnology

ARTICLES

https://doi.org/10.1038/s41587-020-0709-7

Analysis of RNA–protein networks with RNP-MaP defines functional hubs on RNA

Chase A. Weidmann[1,2], Anthony M. Mustoe[1], Parth B. Jariwala[1], J. Mauro Calabrese[2,3] and Kevin M. Weeks[1]✉

RNA–protein interaction networks govern many biological processes but are difficult to examine comprehensively. We devised ribonucleoprotein networks analyzed by mutational profiling (RNP-MaP), a live-cell chemical probing strategy that maps cooperative interactions among multiple proteins bound to single RNA molecules at nucleotide resolution. RNP-MaP uses a hetero-bifunctional crosslinker to freeze interacting proteins in place on RNA and then maps multiple bound proteins on single RNA strands by read-through reverse transcription and DNA sequencing. RNP-MaP revealed that RNase P and RMRP, two sequence-divergent but structurally related non-coding RNAs, share RNP networks and that network hubs define functional sites in these RNAs. RNP-MaP also identified protein interaction networks conserved between mouse and human XIST long non-coding RNAs and defined protein communities whose binding sites colocalize and form networks in functional regions of XIST. RNP-MaP enables discovery and efficient validation of functional protein interaction networks on long RNAs in living cells.

Ribonucleoproteins (RNPs) are complexes made up of interacting RNA and protein and govern both messenger RNA (mRNA) regulation and the function of non-coding RNA (ncRNA)[1,2]. Understanding how RNPs assemble and function, often involving multi-component RNA–protein networks, is critical for characterizing biological mechanisms. Biochemical approaches have defined protein interactions required for several RNP assemblies[1], and high-resolution structural approaches[3–6] have transformed understanding of small and large RNP architectures. Nonetheless, it remains challenging to characterize RNP assemblies and their interacting networks in living cells.

Current methods for characterizing RNPs in live cells suffer from several limitations. Crosslinking with ultraviolet (UV) light, optionally aided by metabolic incorporation of photo-activatable nucleotides, captures RNP information in living cells[7–11]. Sites of protein crosslinking to RNA can be mapped transcriptome wide, either without[10,11] or with identification of binding sites for individual proteins (crosslinking and immunoprecipitation (CLIP) and photo-activatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP))[7,8]. However, UV-based crosslinking suffers from experimental biases and limited binding site resolution[9]; metabolic labeling probes a single substituted nucleobase at a time[8]; and CLIP strategies require a specific antibody or protein tag. Methods, such as mass spectrometry, that focus on cataloging RNA-binding proteins do not readily locate protein-binding sites on RNA and do not easily prioritize proteins in terms of function[12]. Major challenges unaddressed by current approaches are: 1) How do multiple proteins interact with an RNA to form networks, and 2) Which protein interaction networks drive function for an individual RNA?

Here we describe RNP-MaP, an experimentally concise strategy to locate protein interaction sites on RNA in live cells with nucleotide resolution and to reveal multi-protein interaction networks integral to RNP function. After validation on RNPs with known structure, we used RNP-MaP to define functional protein interaction communities within the XIST long non-coding RNA (lncRNA), resulting in the discovery of an RNP network in the XIST E region that controls maintenance of the XIST particle. RNP-MaP should be widely useful for understanding RNP biology, particularly in defining functionally critical domains in large mRNAs and ncRNAs.

Results

RNP-MaP strategy. We identified a cell-permeable reagent, NHS-diazirine (SDA), that rapidly labels RNA nucleotides at sites of protein binding. SDA has two reactive moieties: a succinimidyl ester and a diazirine (Fig. 1a). Succinimidyl esters react to form amide bonds with amines, such that reaction occurs overwhelmingly with lysine side chains[13]. When activated with long-wavelength UV, diazirines form carbene or diazo intermediates[14], which are broadly reactive toward nucleotide ribose and base moieties. Two-step reaction of SDA thus crosslinks protein residues with RNA with a distance governed by SDA linker length (4 Å) and side chain flexibility (~8 Å for lysine). Lysine is the second-most prevalent amino acid in RNA-binding domains (after arginine)[15], and diazirine photo-intermediates are short lived. SDA thus crosslinks short-range RNA–protein interactions relatively independently of local RNA structure or protein properties. Live cells are treated with SDA for 10 min; excess reagent is quenched; and cells are exposed to UV light. SDA-treated cells are then lysed, and crosslinked proteins are digested to short peptide adducts.

We detected SDA-mediated RNA–protein crosslinks using the MaP reverse transcription technology[16] (Fig. 1b). With MaP, a relaxed fidelity reverse transcriptase reads through adduct-containing nucleotides and incorporates non-templated nucleotides into the product DNA at the site of RNA–protein crosslinks. Because reverse transcription reads through the adducts, RNP-MaP detects multiple protein crosslinks that co-occur on single RNA molecules (Fig. 1b). Sequencing the DNA product and locating sites of mutation thus reveals two key features of an RNA–protein complex: RNP-MaP adducts at individual nucleotides report locations of protein binding,

[1]Department of Chemistry, University of North Carolina, Chapel Hill, NC, USA. [2]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA. [3]Department of Pharmacology, University of North Carolina, Chapel Hill, NC, USA. ✉e-mail: weeks@unc.edu

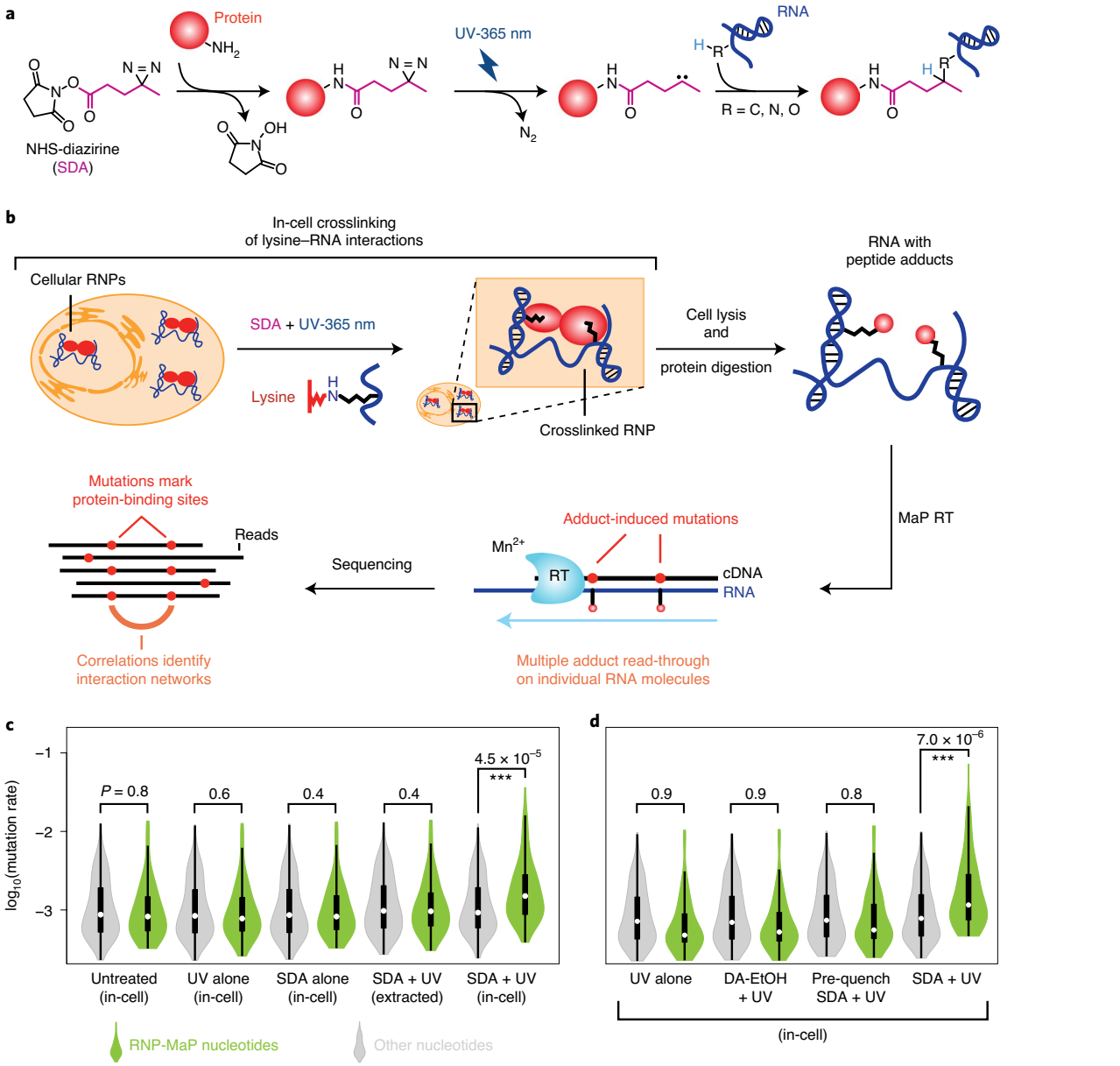


Fig. 1 | RNP-MaP strategy for probing RNA–protein interaction networks in cells. a, Scheme for selective chemical crosslinking of proteins to RNA by SDA. **b**, Workflow of the RNP-MaP experiment. **c**, Nucleotide mutation rates after MaP reverse transcription, separated into RNP-MaP sites and non-sites (green and gray). Combinations of UV (365 nm, 3 J/cm^2) and SDA (10 mM) were applied directly to cells (in-cell) or to protein-free RNA extracted from cells (extracted). The number of nucleotide mutation rates included in each distribution (*n*) are 77 RNP-MaP sites and 162 non-sites. **d**, RNA mutation rates for cells treated with UV and SDA versus non-reactive controls (DA-EtOH or pre-quenched SDA). Representative data for Rmrp RNA from mouse embryonic stem cells (SM33) are shown. The *n* for distributions of RNP-MaP sites and non-sites is 82 and 157, respectively. For violin plots, circles indicate medians; box limits indicate the first and third quartiles; whiskers extend 1.5 times the interquartile range; and smoothed polygons show data density estimates and extend to extreme values. *P* values (Kolmogorov–Smirnov test, one sided) are shown. If *P* values are calculated comparing each condition to their untreated (in **c**) or UV-alone (in **d**) counterparts, only reactivities of RNP-MaP nucleotides in the SDA- and UV-treated in-cell samples are significantly increased ($P < 0.05$).

and correlated crosslinking across multiple nucleotides reveals higher-order protein interaction networks. RNP-MaP thus detects both protein binding location and interaction network information, with no requirement for pre-existing knowledge about the proteins involved. If desired, the involved proteins can subsequently be assigned by comparison to other information, such as CLIP data sets.

RNP-MaP validation. The SDA reactivity at each nucleotide is the ratio of the MaP mutation rate for cells treated with SDA and UV compared to cells treated with UV only (Supplementary Fig. 1). We derived universal normalization factors for each RNA nucleotide (U, A, C and G) based on analysis of RNPs of known structure, enabling identification of protein-bound nucleotides (termed RNP-MaP sites) for arbitrary RNAs of interest (Supplementary Fig. 1). RNP-MaP reactivities were reproducible (Supplementary Fig. 2). Nucleotides with high reactivities are close to lysine amines in human U1, RNase P and ribosome complexes[3–6], whereas nucleotides distant from bound proteins rarely passed reactivity thresholds (Supplementary

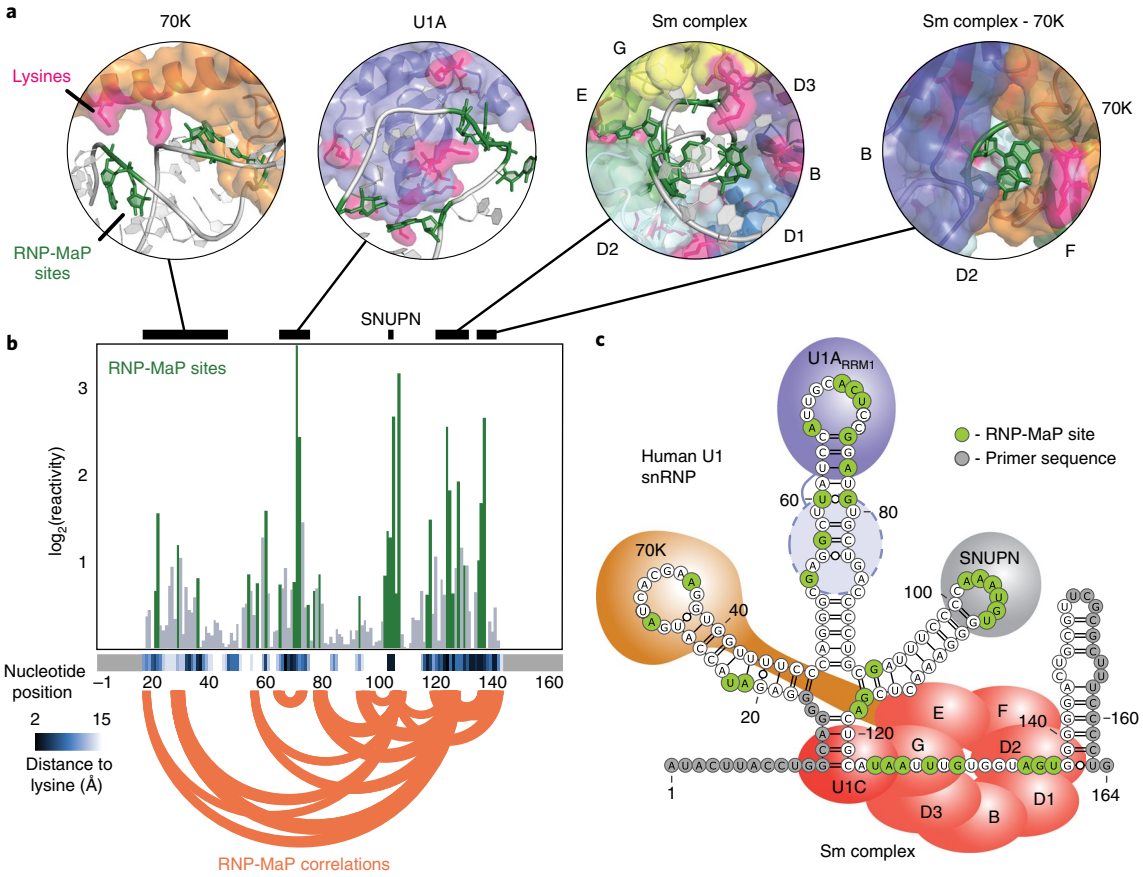


Fig. 2 | RNP-MaP defines protein interaction networks in the U1 snRNP. a, Structures surrounding lysine-RNA crosslinking sites (from 3CW1 (ref. [4]) and 4PKD[5]). **b**, Bar graph of $\log_2$(average reactivity) from replicate experiments performed on U1 snRNA in HEK293 cells. RNP-MaP sites passing thresholds are shown in green. Nucleotide distance (Å) to nearest lysine amine are shown with a heat map (bottom). Black bars (top) indicate locations of structures represented in **a**. RNP-MaP correlations (top 10% in MI strength) are shown as orange arcs. Nucleotides that overlap amplification primers (light gray boxes) are not observable by RNP-MaP. **c**, Secondary structure model of the human U1 RNP showing relative protein positions, RNP-MaP sites and primer regions. Estimated location of U1A RRM2 binding (not visualized in structures) consistent with RNP-MaP signal is shown (dashed line oval).

Fig. 2). RNP-MaP sites occurred in both single-stranded and base-paired regions of RNA. Unpaired RNA regions showed higher reactivities for some RNAs (Supplementary Fig. 2), which likely reflects binding preferences for single-stranded RNA. RNP-MaP sites occurred at all four nucleotides, with higher reactivities at uridine and adenosine (Supplementary Fig. 2). RNP-MaP sites were not detected if UV or SDA was omitted or if RNA was first extracted from cells (removing protein) before treatment (Fig. 1c and Supplementary Fig. 3). RNP-MaP sites were also not detected if SDA was substituted with a diazirine ethanol compound (DA-EtOH, no lysine-reactive group) or if SDA was pre-quenched (Fig. 1d). RNP-MaP signals overlapped those of two orthogonal approaches, ΔSHAPE chemical probing[17] and photo-lysine metabolic labeling[18], which also identify protein-bound sites (Supplementary Fig. 4). In sum, RNP-MaP identifies protein-proximal nucleotides, with good coverage across all four ribonucleotides in diverse structural contexts, in a manner strictly dependent on SDA and UV dosage and the presence of cellular proteins.

RNP-MaP defines protein interaction networks in the U1 small nuclear ribonucleoprotein. In human HEK293 cells, RNP-MaP sites clustered in regions of U1 RNA known to bind proteins[5,19] (Fig. 2a,b), including at all four nucleotides and in both single-stranded and base-paired regions (Fig. 2c). Most RNP-MaP sites were within 9 Å of a lysine residue (Fig. 2b). We also identified RNP-MaP sites in U1 stem loop 2 (positions 48–91) that do not correspond to a known interaction site (Fig. 2c). These RNP-MaP sites presumably reflect binding by the second (currently unvisualized[5]) U1A RRM2 domain or an unidentified protein component of the U1 small nuclear ribonucleoprotein (snRNP).

RNP-MaP uniquely identifies correlated protein-binding events that occur between sites on an RNA (Fig. 1b). Because the MaP reverse transcription process reads through protein–RNA crosslinks, multiple crosslink sites can be detected per single RNA molecule. We adapted a G-test framework[20] to identify pairs of RNA nucleotides that are co-modified in a statistically significant manner, which we term RNP-MaP correlations (Supplementary Figs. 1 and 5). RNP-MaP correlations are distinct from RNP-MaP sites, each providing an independent complementary measure of protein binding to RNA. Correlations require that a single RNA molecule forms at least two crosslinks and arise from three scenarios: 1) a single protein that binds two locations on one RNA; 2) two proteins that interact and bind two locations on one RNA; or 3) two proteins are deposited at two locations on one RNA by a coordinated assembly process.

Networks of RNP-MaP correlations were consistent with the architecture of the U1 snRNP complex (Fig. 2b,c). The highest density and strength of correlations involved nucleotides bound by the Sm protein complex (Fig. 2b and Supplementary Fig. 5), whose initial loading onto U1 is necessary for the maturation of the

snRNP[21]. RNP-MaP correlations also detect long-distance interactions between 70K and the Sm complex important for U1 snRNP assembly. Correlations between the Sm core and the U1A protein were weak, consistent with independent binding by U1A and its expendability for splicing[22]. In sum, RNP-MaP reveals protein interaction networks, pinpoints the central hubs of these networks and identifies interactions important for assembly and function in the U1 snRNP.

RNP-MaP reveals a protein interaction network conserved in RNase P and RMRP RNAs. RNase P and RMRP are divergent, but structurally related, ncRNAs that bind intersecting sets of proteins to form RNP endonucleases that cleave distinct substrates[6,20,23]. Despite substantial differences in sequence, the two RNPs exhibit nearly identical RNP-MaP profiles (Fig. 3a). RNP-MaP sites identify nine out of ten known protein interactions with RNase P[6]. The matching patterns of RNP-MaP sites for RNase P and RMRP core domains suggest that most protein interaction sites are shared. Locations of RNP-MaP sites are also conserved between human and mouse homologs of RNase P and RMRP (Supplementary Fig. 6).

The patterns of through-space RNP-MaP interaction networks for the RNase P and RMRP RNAs are nearly identical after alignment by structural domains (Fig. 3b). The strongest correlations for RNase P define hubs involving the specificity domain, the substrate cleavage site and the RPP25/POP7 dimer (which links the specificity domain and cleavage site). These hubs are conserved in RMRP (Fig. 3c). Omission of proteins that comprise each hub suppresses or eliminates RNase P catalytic function[24]. Our RNase P and RMRP data thus identify shared protein interactions and functional interaction network hubs in sequence divergent RNAs and confirm conservation of RNP architecture in mice and humans.

RNP-MaP identifies conserved protein interaction networks in the XIST lncRNA. The 20-kb X-inactive specific transcript (denoted Xist in mouse and XIST in human) controls X-chromosome dosage compensation in eutherian mammals[25]. The Xist/XIST sequence shows low conservation between mice and humans, despite accomplishing the same functions. We applied RNP-MaP to Xist/XIST protein interaction networks in mouse and human cells by enriching for Xist/XIST RNAs using an RNA antisense pulldown[26] (Supplementary Fig. 7). Data were obtained for 97% of nucleotides in Xist/XIST RNAs (17,410 and 18,708 nts); RNP-MaP sites (2,139 and 3,766, respectively) occurred at all four nucleobases and in both structured and unstructured regions (Supplementary Fig. 8).

High RNP-MaP site density occurred in the Xist/XIST A, B, C, D and E regions (Fig. 4a), which contain repetitive sequences important for Xist localization, assembly and chromatin silencing[27–32]. Conservation of high RNP-MaP density in these regions occurred in both human and mouse RNAs despite changes in copy number (human XIST contains two copies of the B region) and in size, relative position and sequence (C, D and E regions differ extensively between humans and mice)[33]. We also discovered multiple additional regions of Xist/XIST that have not previously been defined as functional but that exhibit strong RNP-MaP signal density clearly conserved between mice and humans (Fig. 4a).

We compared the RNP-MaP signal on human XIST to the most comprehensive set of enhanced CLIP (eCLIP) per-protein binding measurements available (from ENCODE[34], obtained in K562 cells[35]). Although there are differences between HEK293 and K562 cells, both female cell lines maintain silenced X chromosomes and XIST compartments, and eCLIP peaks are shared on genes of similar expression between cell lines[35]. We focused on proteins whose binding sites on XIST were reproducible between eCLIP replicates, yielding 30 proteins (from a total of 120 eCLIP experiments). Regions of XIST with more eCLIP sites had more RNP-MaP sites, especially over the XIST A and E regions (Fig. 4b). eCLIP site density is lower than RNP-MaP density across multiple regions, likely reflecting that only a subset of Xist-binding proteins have been mapped by eCLIP. Together, these data show that RNP-MaP identifies protein-binding sites in lncRNAs that are conserved between species and critical for function, in the absence of pre-existing knowledge about protein-binding sites.

Protein-bound regions in the Xist/XIST RNAs form higher-order interaction networks with distinct levels of interactivity, which are features invisible to alternative strategies that only detect individual RNA–protein interactions. Six highly networked regions occur in XIST, at least five of which are conserved in mouse (Fig. 4a). The E region of Xist/XIST represents an extreme example in which an extended region (spanning 1–1.5 kb) forms a cooperative protein interaction network, as evidenced by high correlation strength densities. In contrast, there also exist highly protein-bound regions, such as the C region of mouse Xist and portions of the Xist/XIST D region, that do not show strong correlations and where proteins, therefore, bind relatively independently of one another (Fig. 4a). Thus, RNP-MaP reveals distinct local patterns of higher-order RNA–protein interaction networks, detected as low and high levels of network interactivity.

Communities of XIST-binding proteins. Using the same ENCODE[34] eCLIP[35] data, we assigned proteins to the interaction networks identified by RNP-MaP correlations. We performed a network analysis of high-confidence eCLIP sites that are linked by highly significant RNP-MaP correlations, revealing communities of proteins whose binding sites on XIST are networked together (Fig. 5 and Supplementary Table 1). We categorized these communities based on the functions of XIST sequences to which the proteins bind, yielding 5′ Silencing, Compartmentalization, Splicing and U/C communities (Fig. 5 and Supplementary Fig. 8). The communities are distinct: correlations between proteins from different communities occur significantly fewer times than expected based on the proximity of their binding sites (Supplementary Table 2).

Proteins in the 5′ Silencing community bind primarily to the 5′ region of XIST, including in the silencing-critical A region[27]. Community members include factors involved in XIST processing, XIST stability and XIST-mediated silencing: UCHL5 (ref. [36]), EXOSC5 (ref. [37]), HNRNPUL1 (ref. [38]) and RBM15 (ref. [39]) (Fig. 5). Silencing community members TARDBP and RBM22 are RNA-dependent regulators of transcription[40]. Binding sites for the 5′ Silencing community members show high interactivity consistent with forming a specific coordinated RNP on XIST.

The strongest inter-protein correlations occur among Compartmentalization community members PTBP1, MATR3 and TIA1, which bind in the XIST E region (Fig. 5 and Supplementary Fig. 8). The XIST E region is critical for maintenance of the silenced X chromosome compartment[30]. PTBP1, MATR3 and TIA1 each undergo liquid–liquid phase transitions to form RNA granules[41–43], and PTBP1 and MATR3 interact on other RNAs[44], features consistent with the formation of an XIST-mediated compartment.

The Splicing community includes proteins that control splicing (U2AF2, SRSF1, TRA2A, AQR and ILF3)[45–47] and a chromatin modulator (GRWD1)[48]. All Splicing community proteins, except for TRA2A, bind to XIST at exon–exon junctions (Supplementary Fig. 8), consistent with a function in splicing of XIST transcripts. The smallest community (U/C) includes two HNRNPs (U and C) that interact with one another[49] but do not strongly interact with other communities and interact sparsely across XIST, suggesting that these proteins play more independent roles in XIST function or that methodological constraints precluded assignment to a single representative community. Together, network analysis reveals how RNP-MaP defines RNP communities with distinct levels of networking (low versus high), each associated with critical lncRNA functions.

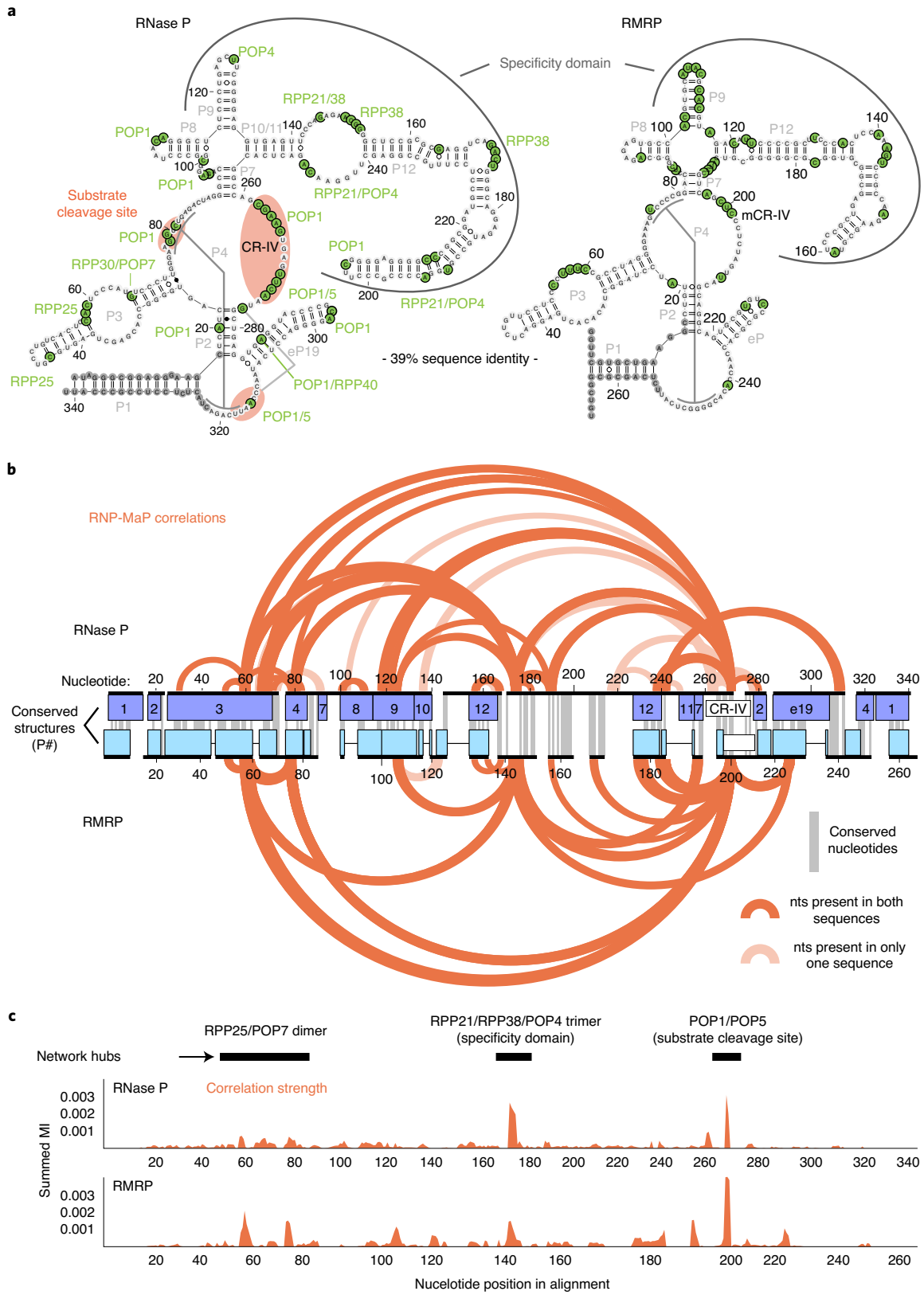


Fig. 3 | RNP-MaP reveals conserved protein interaction networks in RNase P and RMRP RNAs. a, Secondary structures of human RNase P and RMRP RNAs annotated with RNP-MaP sites (green). Experiments performed with HEK293 cells. Proteins proximal to each site (based on nearest lysine) are labeled. Functional domains are indicated, and conserved base-paired structural regions (P#) are labeled (gray). **b**, RNP-MaP correlations for human RNase P and RMRP, plotted on a structure-based sequence alignment. Correlations shown correspond to the top 10% of MI strength. Correlations that reflect linkages between nucleotides present in only one of the RNAs are shown in light shading. **c**, Total strength of correlations (by MI) at each nucleotide in the human RNase P and RMRP RNAs, plotted on the shared structure-based alignment. Protein interaction network hubs are labeled.

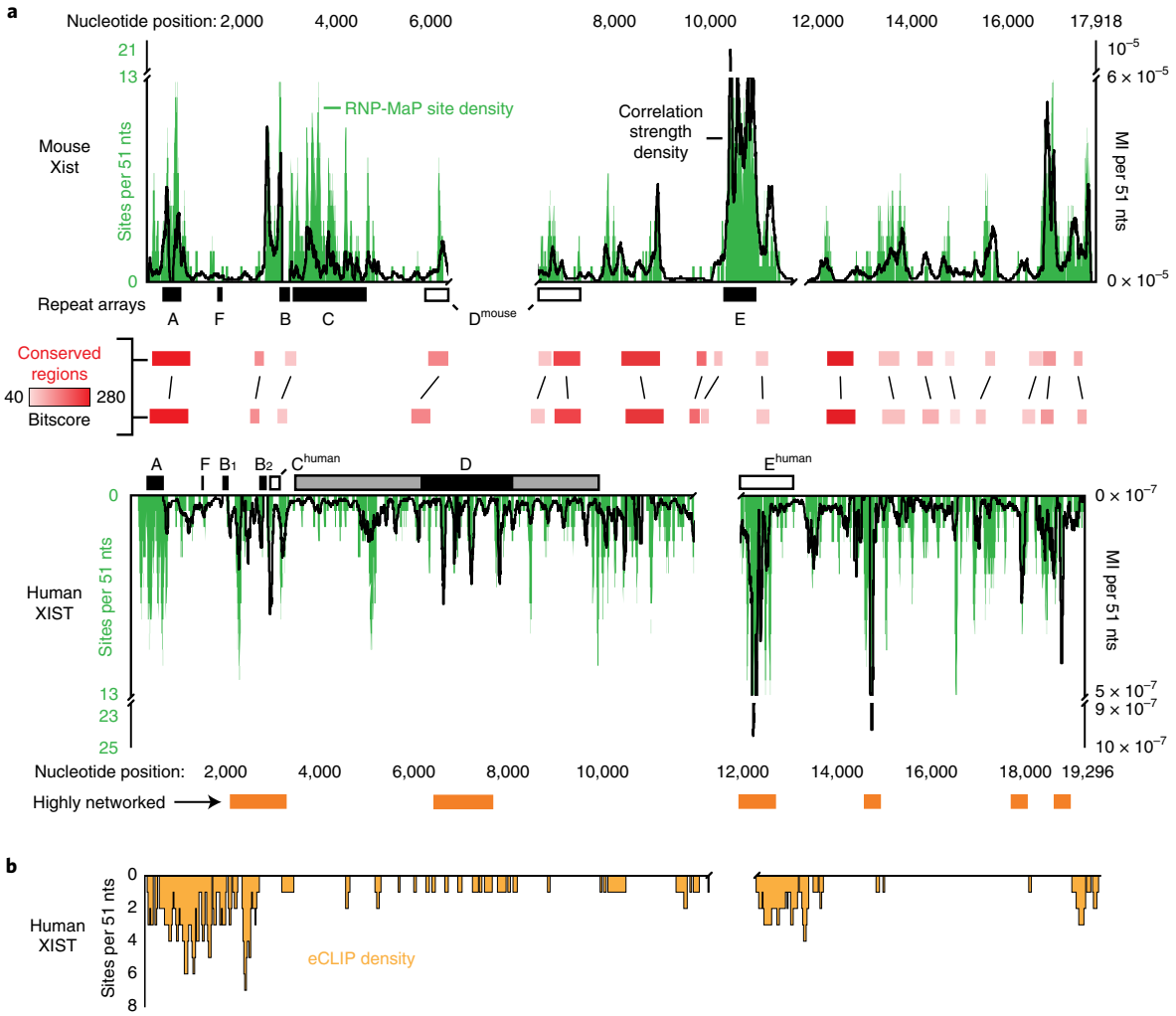


Fig. 4 | RNP-MaP identifies conserved protein interaction networks in the XIST lncRNA. a, Density of RNP-MaP sites (green, the total number of RNP-MaP sites per 51-nt window, left axis) across mouse Xist (top) and human XIST (bottom). Black lines indicate correlation strength densities (depth-normalized MI, per 51-nt window, right axis). Human XIST regions that are highly networked (orange boxes, bottom) or have conserved local sequence alignments with mouse Xist (red boxes, middle) are emphasized. Rectangles labeled A–F show locations of tandem repeat arrays within each RNA, including well-defined repeat sequences (solid rectangles) and regions identified as homologous based on RNP-MaP similarity (open rectangles). The D region in human XIST contains a distinct repeat core (black) flanked by more degenerate repetitive elements (gray). Gaps were introduced to align RNAs by conservation and RNP-MaP similarity. Low sequencing depth prevented identification of correlations in the human XIST A region; high repeat content prevented read alignment to Xist and XIST B regions. Experiments were performed in SM33 and HEK293 cells. **b**, Density of eCLIP sites along the human XIST RNA, shown as the number of eCLIP sites within 51-nt windows. Based on 151 total sites from 30 proteins mapped reproducibly to XIST in K562 cells[57].

RNP-MaP reveals interaction sites for Compartmentalization community proteins. The XIST E region, critical for maintaining the silenced X chromosome compartment[29,30], is distinguished by strong inter-protein network connectivity (Fig. 4a) and includes binding sites for the proteins PTBP1, MATR3 and TIA1 in the Compartmentalization community (Fig. 5 and Supplementary Fig. 8). PTBP1, MATR3 and TIA1 proteins are implicated in formation of RNA foci[41–43], likely through multivalent RNA–protein interactions, consistent with the highly interactive protein network we observed in the XIST E region (Fig. 4a). We, therefore, investigated the role that these protein interaction networks play in XIST particle formation.

We examined binding by PTBP1, MATR3 and TIA1 in a simplified system using recombinant proteins and a synthetic RNA spanning the human XIST E region. PTBP1, MATR3 and TIA1 each bound the XIST E region RNA with similar RNP-MaP patterns (Fig. 6a and Supplementary Fig. 9). Binding occurs at pyrimidine-rich sequences, similarly to motifs previously defined in vitro[50] and by CLIP methods[51–53]. However, binding to pyrimidine-rich motifs (termed class 1 here) was not significantly enriched (Fig. 6a) relative to the abundance of these motifs in the XIST E region. Instead, a purine-rich motif (termed class 2 here) was significantly enriched in RNP-MaP sites for each protein, both in the reconstituted system and in cells (Fig. 6a). Class 1 and 2 motifs each align to reveal a 4–6-nt core sequence motif. Under our simplified conditions, PTBP1, MATR3 and TIA1 show higher RNP-MaP reactivity with class 2 motifs than with class 1 motifs (Fig. 6b). In cells, owing to the differing conditions and proteins present, strong RNP-MaP signals occur at both class 1 and 2 motifs and throughout the E region (Fig. 6b).

Despite their highly significant RNP-MaP signal, class 2 motifs have not been detected by CLIP for PTBP1, MATR3 or

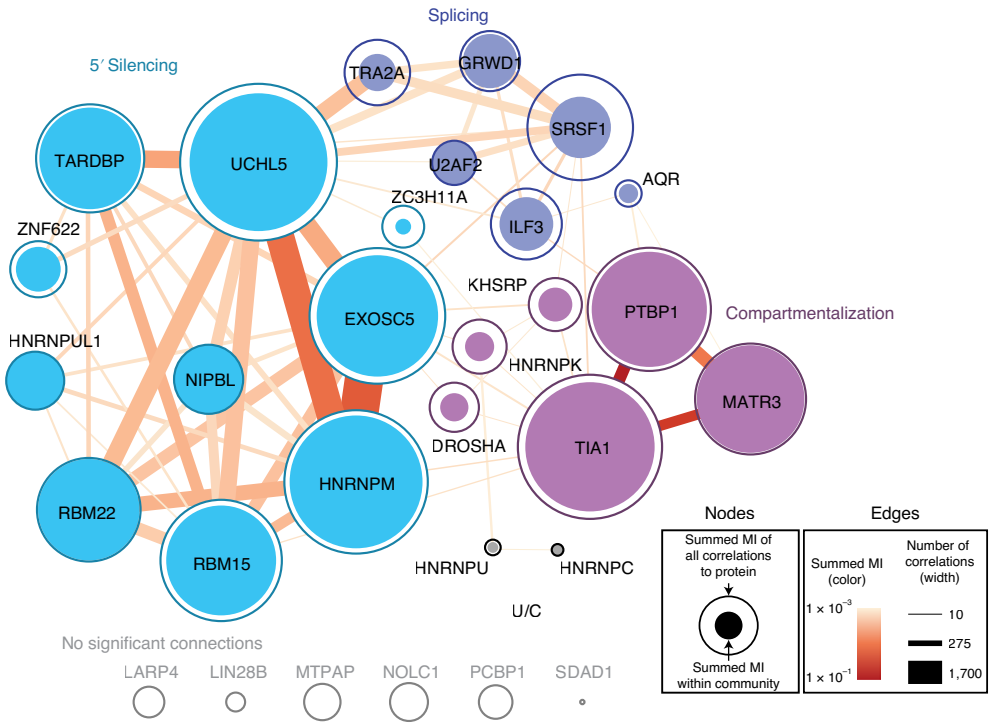


Fig. 5 | Communities of XIST-binding proteins. Network graph of protein-bound sites (nodes) from eCLIP data on human XIST RNA linked by RNP-MaP correlations (edges). Communities were identified by maximizing the modularity of the network[58,59] while weighting by correlation strength (MI). Summed MI for all correlations linking each protein node to other proteins in its community (filled circles) or to all proteins (open circles) is represented as circles with proportional areas. Summed MI of all correlations in an edge (color) and total number of correlations in each edge (width) is shown. Low read depth in A and B repeat regions prevented correlation analysis of a portion of RBM15, SRSF1, UCHL5 and HNRNPK eCLIP sites at these locations, potentially underestimating their contributions to each network. Additionally, the known XIST-binding protein SPEN is not represented in eCLIP data.

TIA1 (Supplementary Fig. 9). This discrepancy is consistent with two models: 1) conditions and proteins in cells alter the intrinsic binding preferences of PTBP1, MATR3 and TIA1, or 2) nucleotide biases of CLIP methods (for uridines[9]) mask binding to pyrimidine-poor class 2 motifs (Supplementary Fig. 10). Overall, RNP-MaP identifies a larger set of protein-binding sites than CLIP: 93% of eCLIP sites in XIST (of 151 analyzed) contain three or more RNP-MaP sites (enrichment $P=0.019$, compared to randomized eCLIP sites), but only 35% of RNP-MaP sites (of 3,766) fall within an eCLIP site.

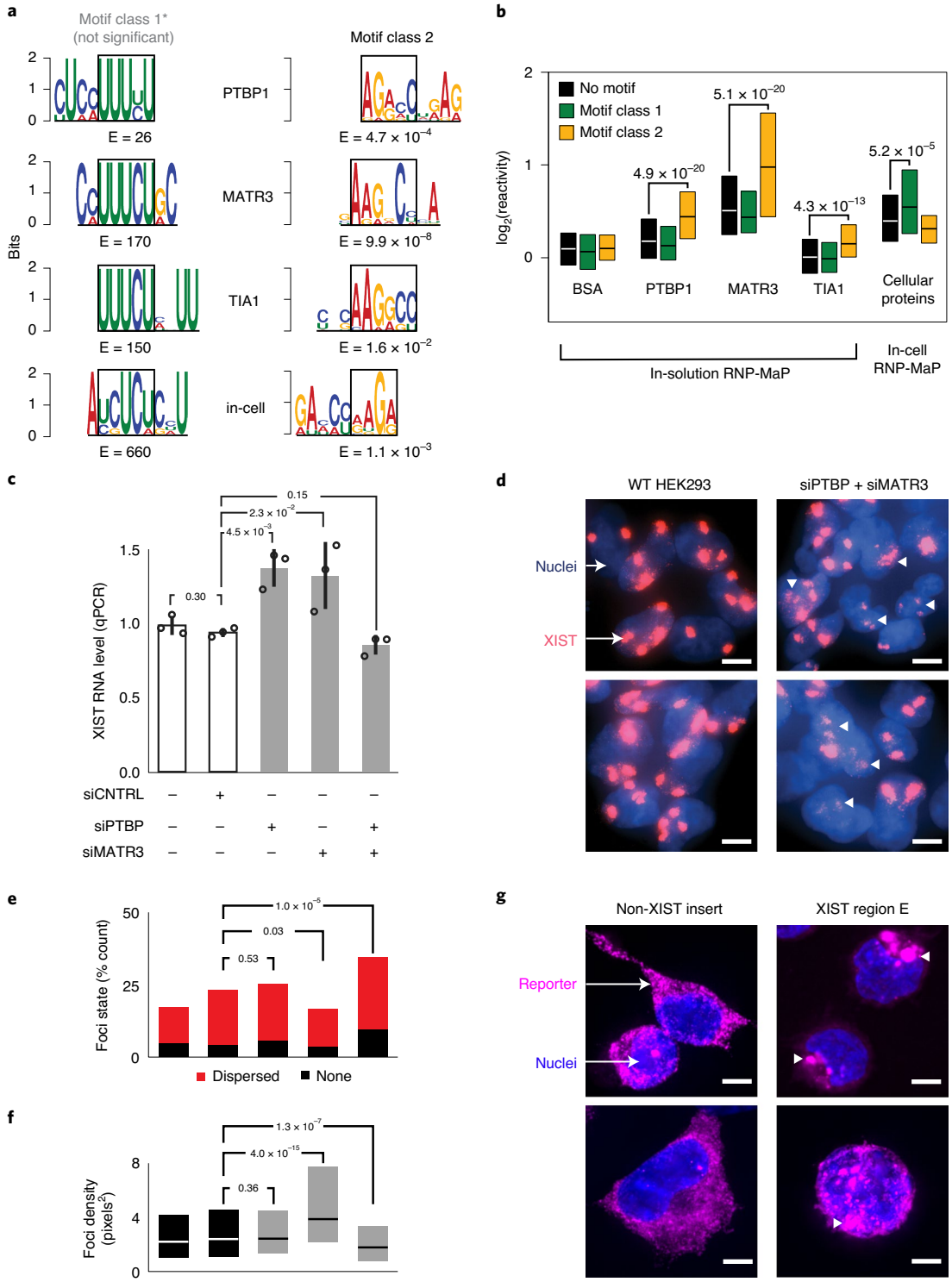
PTBP1 and MATR3 interaction with E region controls XIST particle formation. PTBP1 and MATR3, major components of the Compartmentalization community, bound to the E region RNA at lower concentrations and showed higher overall RNP-MaP reactivities than TIA1 (Fig. 6b and Supplementary Figs. 9 and 10), and PTBP1 and MATR3 have more high-frequency eCLIP signals in the XIST E region (nts in the top 0.001%) than TIA1 (324 and 539 versus 73, respectively). We, therefore, focused further functional analysis on PTBP1 and MATR3 and depleted each protein in HEK293 cells by RNA interference, achieving 75% and 85% knockdown, respectively. Knockdown of either PTBP1 or MATR3 individually resulted in a ~40% increase in XIST RNA levels, whereas co-depletion of both PTBP1 and MATR3 returned XIST to normal levels (Fig. 6c). Compared to normal HEK293 cells (which have 2–5 silenced X chromosomes per cell), XIST foci in cells depleted of both PTBP1 and MATR3 were more dispersed or frequently absent (Fig. 6d,e), and remaining foci were significantly less dense (Fig. 6f). The dispersion observed upon PTBP1 and MATR3 depletion resembles the effects of CIZ1 depletion and E region deletion in mouse Xist[29,30]. These data suggest that PTBP1 and MATR3, whose binding in the E region forms an exceptionally interactive network, function to maintain the human XIST particle.

We inserted the highly protein-interactive E region into an RNA reporter and compared its expression and localization to reporters containing other highly protein-interactive XIST regions or a non-XIST sequence. The E region-containing reporter, but not other reporters, formed large foci in cells (Fig. 6g). These E region foci, although formed in the cytoplasm, are similar in size to native XIST particles observed in HEK293 cells (Fig. 6d). E region foci appear to trigger cellular deformations, and the E region-containing reporter is less stable than other tested sequences (Supplementary Fig. 10). Highly interactive RNP networks, partially deleterious out of context, thus appear to intrinsically assemble on XIST E region RNA. E foci likely include granule-associated proteins such as PTBP1, MATR3 and TIA1, and further work is necessary to identify components of these non-canonical granules. Still, these data support a role for the E region in organization of the XIST compartment[29,30] and highlight the ability of RNP-MaP to discover and characterize novel motifs in ncRNAs, whose functions reflect interconnected RNA–protein networks.

Discussion

RNP-MaP enables rapid and concise characterization of functionally important RNA–protein interaction networks. Protein-binding sites are identified across an RNA with low sequence and structure biases; interaction networks are distinguishable by their correlation patterns; and functionally important hubs are revealed by their binding site density and interconnectivity.

RNP-MaP currently requires read depths of 10^3 for sites and 10^4 for correlations, and maximum correlation distance is governed by the length of reverse transcriptase products (currently ~500 nts).



RNP-MaP detects RNA–protein interactions conserved between species and critical for function, without pre-existing knowledge of the interacting proteins, and can be integrated with other information to reveal protein identity. Coupling RNP-MaP with complementary CLIP and mass spectrometry approaches will enable definition of cellular RNP networks in unprecedented detail.

RNP-MaP revealed insights into the assembly of small RNPs U1, RNase P and RMRP. Each RNP has multiple interaction network hubs, and the strongest interaction hubs in each RNA correspond to regions central to RNP assembly and activity: the Sm complex assembly site in U1 and the substrate cleavage sites in RNase P and RMRP. The unique ability of RNP-MaP to distinguish interaction networks by their correlation strength and density will aid in discovery and prioritization of functional elements in large non-coding, messenger and viral RNAs.

Previous analyses of the mouse Xist RNA revealed that repeat-containing regions are structurally dynamic and accessible for protein binding, and these regions were proposed to function as 'landing pads' for proteins[17]. Our RNP-MaP study now directly reveals that repeat sequences in Xist/XIST are extensively bound

Fig. 6 | PTBP1 and MATR3 interactions with E region and functional control of XIST particle formation. a, RNP-MaP enriched 9-mer sequence motifs (by MEME)[60] using purified components for each indicated protein and comparison with motifs observed in cell. Motifs are shown as position-weighted matrices and E-values, aligned to a shared core sequence (black boxes) for class 1 and class 2 motifs. **b**, Box plots of RNP-MaP $\log_2$(reactivity) at class 1 and 2 motifs or other nucleotides (no motif) in the XIST E region, either for simplified conditions with synthetic RNA and indicated recombinant protein (in-solution) or with native XIST and cellular proteins in HEK293 cells (in-cell). The number of nucleotide reactivities included in each distribution (*n*), from left to right, were 631, 230, 293, 651, 235, 293, 651, 235, 293, 651, 235, 292, 668, 229 and 293. Box limits indicate first and third quartiles; central lines indicate medians. Significant increases in reactivity compared to no-motif nucleotides are indicated, and, with the exception of TIA1 no-motif and class 1 motif sites, reactivities in-cell and in-solution are all significantly increased compared to BSA ($P < 0.05$, Kolmogorov–Smirnov test, one sided). **c**, Relative expression levels of XIST RNA in HEK293 cells, as a function of PTBP and MATR3 knockdown by siRNAs. Standard deviation (error bars), means (bar heights) of replicate measurements ($n = 3$, open circles) and *P* values (Student's *t*-test, two sided) are shown. **d**, Visualization of XIST foci for wild-type (WT) HEK293 cells and cells depleted of PTBP and MATR3 by siRNA treatment. Dispersed and non-punctate XIST foci are emphasized with white triangles. XIST in cells imaged by FISH (red) using labeled antisense oligonucleotides; nuclei labeled with DAPI (blue). **e**, Quantification of dispersed and absent XIST foci as a function of PTBP and MATR3 depletion (HEK293 cells). Conditions are ordered as per **c**. *P* values (chi-square goodness-of-fit test) are shown. The total number of cells counted (*n*) were 107, 396, 200, 286 and 304, respectively. **f**, Effect of PTBP and MATR3 depletion on XIST foci density. Conditions are ordered as per **c**. *P* values (Kolmogorov–Smirnov test, two sided) are shown. The total number of XIST foci observed (*n*) were 311, 1,000, 746, 533 and 863, respectively. Box limits indicate first and third quartiles; central lines indicate medians. **g**, Localization of non-XIST control and XIST E region-containing RNA reporters (by FISH, magenta); nuclei labeled by DAPI (blue). Condensed foci formed by XIST E region reporters are emphasized by white triangles. The percentage of cells with granule phenotypes was 9% (1 of 11) and 89% (8 of 9), respectively. *P* values (binomial test and chi-squared goodness of fit) were significant (7.8×10^{-8} and 1.13×10^{-13}, respectively). Data from **e** and **f** were combined from two biological replicate experiments (performed and imaged on different days); **d** and **g** are representative images. Scale bars in **d** and **g** are 5 μm.

by highly networked proteins. Depleting the highly networked Compartmentalization community members PTBP1 and MATR3 induces dispersal of native XIST RNA foci in cells, and E region RNA is sufficient to promote foci formation in a heterologous reporter RNA in cells. These data suggest that protein interaction networks in the E region support formation of a phase-separated particle in vivo, which is consistent with recent orthogonal analyses of Xist particle shape and composition[54] and a conserved role in this process for MATR3, PTBP1 and the Xist E region in mice[55]. The many other highly connected protein interaction networks identified by RNP-MaP in Xist/XIST suggests additional biology that warrants future study.

Most ncRNAs and untranslated regions of mRNAs function by recruiting proteins and forming higher-order RNA–protein complexes. The ability of RNP-MaP to identify function-critical RNA regions and their interconnected protein networks will enable focused exploration of the thousands of ncRNAs and mRNA untranslated regions whose overall functions and specific internal functional elements are unexplored[56]. RNP-MaP can be further applied to reveal how protein interaction networks form and dissociate in both coding RNAs and ncRNAs, how networks differ between cell types and how networks change in response to stimuli.

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/s41587-020-0709-7.

Received: 4 January 2020; Accepted: 16 September 2020;
Published online: 19 October 2020

References

1. Gehring, N. H., Wahle, E. & Fischer, U. Deciphering the mRNP code: RNA-bound determinants of post-transcriptional gene regulation. *Trends Biochem. Sci.* **42**, 369–382 (2017).
2. Guttman, M. & Rinn, J. L. Modular regulatory principles of large non-coding RNAs. *Nature* **482**, 339–346 (2012).
3. Anger, A. M. et al. Structures of the human and *Drosophila* 80S ribosome. *Nature* **497**, 80–85 (2013).
4. Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K. W., Li, J. & Nagai, K. Crystal structure of human spliceosomal U1 snRNP at 5.5 resolution. *Nature* **458**, 475–480 (2009).
5. Kondo, Y., Oubridge, C., van Roon, A. M. M. & Nagai, K. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5′ splice site recognition. *eLife* **4**, 1–19 (2015).
6. Wu, J. et al. Cryo-EM structure of the human ribonuclease P holoenzyme. *Cell* **175**, 1393–1404.e11 (2018).
7. Ule, J., Hwang, H. W. & Darnell, R. B. The future of cross-linking and immunoprecipitation (CLIP). *Cold Spring Harb. Perspect. Biol.* **10**, a032243 (2018).
8. Garzia, A., Meyer, C., Morozov, P., Sajek, M. & Tuschl, T. Optimization of PAR-CLIP for transcriptome-wide identification of binding sites of RNA-binding proteins. *Methods* **118–119**, 24–40 (2017).
9. Wheeler, E. C., Van Nostrand, E. L. & Yeo, G. W. Advances and challenges in the detection of transcriptome-wide protein–RNA interactions. *Wiley Interdiscip. Rev. RNA* **9**, e1436 (2018).
10. Freeberg, M. A. et al. Pervasive and dynamic protein binding sites of the mRNA transcriptome in *Saccharomyces cerevisiae*. *Genome Biol.* **14**, R13 (2013).
11. Schueler, M. et al. Differential protein occupancy profiling of the mRNA transcriptome. *Genome Biol.* **15**, R15 (2014).
12. Ramanathan, M., Porter, D. F. & Khavari, P. A. Methods to study RNA–protein interactions. *Nat. Methods* **16**, 225–234 (2019).
13. Mädler, S., Bich, C., Touboul, D. & Zenobi, R. Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities. *J. Mass Spectrom.* **44**, 694–706 (2009).
14. Das, J. Aliphatic diazirines as photoaffinity probes for proteins: recent developments. *Chem. Rev.* **111**, 4405–4417 (2011).
15. Krüger, D. M., Neubacher, S. & Grossmann, T. N. Protein–RNA interactions: structural characteristics and hotspot amino acids. *RNA* **24**, 1457–1465 (2018).
16. Smola, M. J., Rice, G. M., Busan, S., Siegfried, N. A. & Weeks, K. M. Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.* **10**, 1643–1669 (2015).
17. Smola, M. J. et al. SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. *Proc. Natl Acad. Sci. USA* **113**, 10322–10327 (2016).
18. Yang, T., Li, X. M., Bao, X., Fung, Y. M. E. & Li, X. D. Photo-lysine captures proteins that bind lysine post-translational modifications. *Nat. Chem. Biol.* **12**, 70–72 (2016).
19. Kühn-Hölsken, E. et al. Mapping the binding site of snurportin 1 on native u1 snRNP by cross-linking and mass spectrometry. *Nucleic Acids Res.* **38**, 5581–5593 (2010).
20. Mustoe, A. M., Lama, N. N., Irving, P. S., Olson, S. W. & Weeks, K. M. RNA base-pairing complexity in living cells visualized by correlated chemical probing. *Proc. Natl Acad. Sci. USA* **116**, 24574–24582 (2019).
21. So, B. R. et al. A U1 snRNP-specific assembly pathway reveals the SMN complex as a versatile hub for RNP exchange. *Nat. Struct. Mol. Biol.* **23**, 225–230 (2016).
22. Will, C. In vitro reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein enhances the formation of early (E) spliceosomal complexes. *Nucleic Acids Res.* **24**, 4614–4623 (1996).
23. Esakova, O. & Krasilnikov, A. S. Of proteins and RNA: the RNase P/MRP family. *RNA* **16**, 1725–1747 (2010).

24. Perederina, A., Berezin, I. & Krasilnikov, A. S. In vitro reconstitution and analysis of eukaryotic RNase P RNPs. *Nucleic Acids Res.* **46**, 6857–6868 (2018).
25. Sahakyan, A., Yang, Y. & Plath, K. The role of Xist in X-chromosome dosage compensation. *Trends Cell Biol.* **28**, 999–1013 (2018).
26. Engreitz, J. M. et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
27. Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat. Genet.* **30**, 167–174 (2002).
28. Colognori, D., Sunwoo, H., Kriz, A. J., Wang, C. Y. & Lee, J. T. Xist deletional analysis reveals an interdependency between Xist RNA and polycomb complexes for spreading along the inactive X. *Mol. Cell* **74**, 101–117 (2019).
29. Ridings-Figueroa, R. et al. The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. *Genes Dev.* **31**, 876–888 (2017).
30. Sunwoo, H., Colognori, D., Froberg, J. E., Jeon, Y. & Lee, J. T. Repeat E anchors Xist RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1). *Proc. Natl Acad. Sci. USA* **114**, 10654–10659 (2017).
31. Lee, H. J. et al. En bloc and segmental deletions of human XIST reveal X chromosome inactivation-involving RNA elements. *Nucleic Acids Res.* **47**, 3875–3887 (2019).
32. Nesterova, T. B. et al. Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation. *Nat. Commun.* **10**, 3129 (2019).
33. Brockdorff, N. Local tandem repeat expansion in Xist RNA as a model for the functionalisation of ncRNA. *Noncoding RNA* **4**, 28 (2018).
34. Davis, C. A. et al. The encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* **46**, D794–D801 (2018).
35. Van Nostrand, E. L. et al. A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 (2020).
36. Moindrot, B. et al. A pooled shRNA screen identifies Rbm15, Spen, and Wtap as factors required for Xist RNA-mediated silencing. *Cell Rep.* **12**, 562–572 (2015).
37. Ciaudo, C. et al. Nuclear mRNA degradation pathway(s) are implicated in Xist regulation and X chromosome inactivation. *PLoS Genet.* **2**, e94 (2006).
38. Sakaguchi, T. et al. Control of chromosomal localization of Xist by hnRNP U family molecules. *Dev. Cell* **39**, 11–12 (2016).
39. Patil, D. P. et al. M6 A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537**, 369–373 (2016).
40. Xiao, R. et al. Pervasive chromatin-RNA binding protein interactions enable RNA-based regulation of transcription. *Cell* **178**, 107–121 (2019).
41. Yap, K. et al. A short tandem repeat-enriched RNA assembles a nuclear compartment to control alternative splicing and promote cell survival. *Mol. Cell* **72**, 525–540 (2018).
42. Rayman, J. B., Karl, K. A. & Kandel, E. R. TIA-1 self-multimerization, phase separation, and recruitment into stress granules are dynamically regulated by Zn 2. *Cell Rep.* **22**, 59–71 (2018).
43. Gallego-Iradi, M. C. et al. N-terminal sequences in matrin 3 mediate phase separation into droplet-like structures that recruit TDP43 variants lacking RNA binding elements. *Lab. Investig.* **99**, 1030–1040 (2019).
44. Attig, J. et al. Heteromeric RNP assembly at LINEs controls lineage-specific RNA processing. *Cell* **174**, 1067–1081 (2018).
45. Long, J. C. & Caceres, J. F. The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.* **417**, 15–27 (2009).
46. De, I. et al. The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes. *Nat. Struct. Mol. Biol.* **22**, 138–144 (2015).
47. Rigo, F. et al. Synthetic oligonucleotides recruit ILF2/3 to RNA transcripts to modulate splicing. *Nat. Chem. Biol.* **8**, 555–561 (2012).
48. Sugimoto, N. et al. Cdt1-binding protein GRWD1 is a novel histone-binding protein that facilitates MCM loading through its influence on chromatin architecture. *Nucleic Acids Res.* **43**, 5898–5911 (2015).
49. Hein, M. Y. et al. A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* **163**, 712–723 (2015).
50. Dominguez, D. et al. Sequence, structure, and context preferences of human RNA binding proteins. *Mol. Cell* **70**, 854–867 (2018).
51. 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).
52. Uemura, Y. et al. Matrin3 binds directly to intronic pyrimidine-rich sequences and controls alternative splicing. *Genes Cells* **22**, 785–798 (2017).
53. Meyer, C. et al. The TIA1 RNA-binding protein family regulates EIF2AK2-mediated stress response and cell cycle progression. *Mol. Cell* **69**, 622–635 (2018).
54. Cerase, A. et al. Phase separation drives X-chromosome inactivation: a hypothesis. *Nat. Struct. Mol. Biol.* **26**, 331–334 (2019).
55. Pandya-Jones, A. et al. A protein assembly mediates Xist localization and silencing. *Nature* https://doi.org/10.1038/s41586-020-2703-0 (2020).
56. Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R. & Johnson, R. Towards a complete map of the human long non-coding RNA transcriptome. *Nat. Rev. Genet.* **19**, 535–548 (2018).
57. Van Nostrand, E. L. et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514 (2016).
58. Bastian, M., Heymann, S. & Jacomy, M. Gephi: an open source software for exploring and manipulating networks visualization and exploration of large graphs. *Int. AAAI Conf. Weblogs Soc. Media* https://doi.org/10.13140/2.1.1341.1520 (2009).
59. Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* P10008 (2008).
60. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In *Proc. Second International Conference on Intelligent Systems for Molecular Biololgy* **2**, 28–36 (1994).

Methods

Cell culture. Adherent mammalian cells used in chemical probing experiments, either SM33 (ref. [26]) or HEK293 cells, were grown to 80–90% confluency in either six-well plates (for targeted priming) or 10-cm dishes (for RNA antisense pulldown). HEK293 cells were cultured in DMEM with 10% FBS. SM33 cells were cultured in embryonic stem cell media (DMEM high glucose with sodium pyruvate, 15% FBS, 0.1 mM non-essential amino acids (Gibco), 2 mM L-glutamine, 0.1 mM β-mercaptoethanol and 1,000 U ml^{-1} leukemia inhibitory factor (ESGRO, Millipore Sigma)). Cultures were grown with 100 U ml^{-1} penicillin and 100 µg ml^{-1} streptomycin. To induce expression of the Xist RNA, SM33 cells were supplemented with 2 µg ml^{-1} doxycycline 16 h before treatment. For all experiments when performing biological replicates, chemical probing and sequencing library preparation were performed on distinct populations of cells on different days.

In-cell crosslinking with SDA. SDA (NHS-diazirine, succinimidyl 4,4′-azipentanoate, Thermo Fisher Scientific) was selected from a small screen of commercially available hetero-bifunctional reagents capable of crosslinking RNA and protein. For six-well plates, cells were washed once in 1 ml of PBS and then covered with 900 µl of PBS. To these cells, 100 µl of 100 mM SDA in DMSO was added with concurrent manual mixing. For controls, 100 µl of neat DMSO was added. Cells were treated with SDA for 10 min in the dark at 37 °C, and then excess SDA was quenched with a 1/9 volume of 1 M Tris-HCl, pH 8.0 (111 µl). For SM33 cells, which remained adherent during treatment, quenching was performed for 5 min in the dark at 37 °C. For HEK293 cells, which detached upon treatment, cells were pelleted at 1,000*g* for 3 min immediately after the addition of quencher. Cells were washed once with PBS (and pelleted again if not adherent) and then resuspended in 400 µl of PBS in a well of a six-well plate. To crosslink labeled proteins to RNAs, SDA-treated and untreated cells were placed on ice and exposed to 3 J/cm^2 of 365-nm-wavelength UV light (about 9 min in a UVP CL-1000 equipped with five 8-W F8T5 black lights) at a distance of 4 inches from the light source. When the amount of SDA used for treatment, the amount of UV light exposure or the compound used for crosslinking were varied, no other changes were made to the procedure. When performing the crosslinking procedure on cells grown in 10-cm dishes, reagent volumes used were multiplied by a factor of 5 relative to the six-well procedure.

Cellular fractionation and proteinase K lysis of SDA-treated cells. Crosslinked cells were pelleted at 1,500*g* for 5 min at 4 °C, washed once in cold PBS and pelleted again and resuspended in cytoplasmic lysis buffer (10 mM KCl, 1.5 mM $MgCl_2$, 20 mM Tris-HCl (pH 8.0), 1 mM DTT and 0.1% Triton X-100). Cells were lysed for 10 min at 4 °C with agitation. Nuclei were pelleted at 1,500*g* for 5 min at 4 °C, and cytoplasmic lysates were separated into new tubes. Nuclei were washed once in low-salt solution (10 mM KCl, 1.5 mM $MgCl_2$, 20 mM Tris-HCl (pH 8.0) and 1 mM DTT), incubated with agitation at 4 °C for 2 min, pelleted again and then resuspended in proteinase K lysis buffer (40 mM Tris-HCl (pH 8.0), 200 mM NaCl, 20 mM EDTA, 1.5% SDS and 0.5 mg ml^{-1} of proteinase K). Components were added to cytoplasmic lysates to adjust to proteinase K lysis buffer concentrations. For samples from six-well plates, 500 µl of cytoplasmic lysis buffer and 500 µl of proteinase K lysis buffer were used; 2.5 ml of each were used for 10-cm dish samples. Nuclear and cytoplasmic fractions were incubated for 2 h at 37 °C with intermittent mixing. Nucleic acid was recovered through two extractions with 1 volume of 25:24:1 phenol:chloroform:isoamyl alcohol (PCA) and two extractions with 1 volume of chloroform.

Control SDA treatment of protein-free RNA extracted from cells. SM33 cells in 10-cm dishes were washed once in 5 ml of PBS and then lysed in 2.5 ml of proteinase K lysis buffer at 23 °C for 45 min. Nucleic acid was recovered through two extractions with 1 volume of PCA and two extractions with 1 volume of chloroform, and the resulting solution was buffer exchanged into PBS (PD-10 columns, GE Healthcare). The nucleic acid solution was incubated at 37 °C for 20 min before splitting into two equal volume portions (1.75 ml each). To one portion, a 1/9 volume of 100 mM SDA in DMSO was added; a 1/9 volume of neat DMSO was added to the other. Each sample was incubated at 37 °C for 10 min in the dark. Each sample was spread evenly over a new 10-cm dish, placed on ice and exposed to 3 J/cm^2 of 365-nm-wavelength UV light.

In-cell treatment with 5-nitroisatoic anhydride SHAPE reagent. SM33 mouse embryonic stem cells were grown in six-well plates. In-cell 5NIA treatment proceeded as described[61]. Cells were washed once in PBS and then covered with 900 µl of serum-free embryonic stem cell media. To these cells, 100 µl of 250 mM 5-nitroisatoic anhydride (5NIA, AstaTech) in anhydrous DMSO was added with concurrent manual mixing. For controls, 100 µl of neat DMSO was added instead. Cells were treated with 5NIA for 10 min at 37 °C; cells were washed once with 1 ml of PBS; and then RNA was harvested from cells with TRIzol (Invitrogen) according to the manufacturer's specifications.

5NIA treatment of cell-extracted RNA. SM33 cells on 10-cm dishes were washed once in ice-cold PBS and resuspended in 2.5 ml of ice-cold lysis buffer (40 mM Tris-HCl (pH 8.0), 25 mM NaCl, 6 mM $MgCl_2$, 1 mM $CaCl_2$, 256 mM sucrose, 0.5% Triton X-100, 1,000 U ml^{-1} of RNasin (Promega) and 450 U ml^{-1} of DNase I (Roche)). Cells were lysed for 5 min at 4 °C with agitation. Nuclei were pelleted at 1,500*g* for 5 min at 4 °C, resuspended in 2.5 ml of proteinase K digestion buffer and incubated for 45 min at 23 °C with agitation. RNA was extracted twice with 1 volume of PCA that had been pre-equilibrated with 1.1× folding buffer (111 mM HEPES (pH 8.0), 111 mM NaCl and 5.55 mM $MgCl_2$), followed by two extractions with 1 volume of chloroform. RNA was buffer exchanged into 1.1× folding buffer over a de-salting column (PD-10, GE Healthcare). After incubating for 20 min at 37 °C, RNA solution was split into two equal portions: one was added to a 1/9 volume of 250 mM 5NIA in DMSO; the other was added to a 1/9 volume of neat DMSO. Both portions were incubated for 10 min at 37 °C.

In-cell crosslinking with photo-lysine. HEK293 cells in six-well plates at ~60% confluency were washed once with PBS and then cultured for 16 additional hours in media with either 2 mM natural lysine or 2 mM photo-lysine (MedChemExpress). After 16 h, cells were washed once with 1 ml of PBS and then coated with a thin layer of 400 µl of PBS. Cells were then crosslinked on ice with 10 J/cm^2 of 365-nm-wavelength UV light. Cells were washed once in PBS, pelleted at 1,500*g* and resuspended in proteinase K lysis buffer. Proteins were digested for 2 h at 37 °C. Nucleic acid was recovered through two extractions with 1 volume of PCA and two extractions with 1 volume of chloroform.

RNA precipitation and DNase treatment. Nucleic acids, including those treated after extraction from cells and those collected by TRIzol or PCA after in-cell treatments, were precipitated by the addition of a 1/25 volume of 5 M NaCl and 1 volume of isopropanol, incubation for 10 min at 23 °C and centrifugation at 10,000*g* for 10 min. The precipitate was washed once in 75% ethanol and pelleted by centrifugation at 7,500*g* for 5 min. Pellets from six-well plates were resuspended in 50 µl of 1× DNase buffer and incubated with 2 units of DNase (TURBO, Thermo Fisher Scientific) at 37 °C for 1 h. After the first incubation, 2 more units of TURBO DNase were added, and samples were incubated at 37 °C for 1 h. Volumes were doubled for samples derived from 10-cm dishes. RNA was purified with Mag-Bind TotalPure NGS SPRI beads (Omega Bio-tek): a 1.8 volume of beads was added to DNase reactions and incubated at 23 °C for 5 min, followed by magnetic separation for 2 min. The solution was discarded, and beads were washed three times with 70% ethanol. RNA was eluted into 30 µl of nuclease-free water.

Antisense-mediated purification of Xist and XIST. In 50 µl of nuclease-free water, 10 µg of total nuclear RNA (from SM33 or HEK293 cells) was heated at 70 °C for 5 min and then immediately placed on ice for 2 min. To the RNA, 100 µl of 1.5× hybridization buffer (15 mM Tris-HCl (pH 7.0), 7.5 mM EDTA, 750 mM LiCl, 0.15% Triton X-100 and 6 M urea), pre-warmed to 55 °C, was added. RNA was pre-cleared for 15 min at 55 °C with 15 µl of streptavidin-conjugated magnetic beads (Dynabeads MyOne Streptavidin C1, Thermo Fisher Scientific) that were pre-washed and resuspended in 1× hybridization buffer. After magnetic separation, the pre-cleared supernatant was retained. Biotinylated antisense RNA capture probes[26] (Guttman laboratory, CalTech), specific to either mouse Xist or human XIST, were heated at 70 °C for 5 min, cooled on ice for 2 min and then diluted in 1× hybridization buffer. Each pre-cleared RNA sample was mixed with 72 ng of capture probes, and mixtures were incubated at 55 °C for 80 min with shaking. After probe hybridization, 30 µl of streptavidin magnetic beads, pre-washed and resuspended in 1× hybridization buffer, was added to RNA probe mixtures, and incubation was continued at 55 °C with shaking for 20 min. Beads were captured by magnetic separation and washed twice with 200 µl of 1× hybridization buffer for 5 min each at 55 °C. Beads were resuspended in 60 µl of NLS elution buffer (20 mM Tris-HCl (pH 8.0), 10 mM EDTA, 2% N-lauroylsarcosine and 10 mM TCEP). RNA was eluted from beads with three heating/cooling cycles where the temperature was ramped down from 95 °C to 4 °C and up to 95 °C in 1.5-min cycles. Beads were captured, and RNA eluates were saved. The same beads were then resuspended in 40 µl of NLS elution buffer, and the elution procedure was repeated; the 40-µl eluate was added to the original 60-µl eluate. Captured RNA was purified (RNeasy MinElute Cleanup Kit, Qiagen). To reduce nontarget RNA in the sample, RNAs were enriched again via a second capture: the procedure was identical to the first capture except the pre-clear step was omitted.

In vitro SDA crosslinking of T7-transcribed XIST E region with recombinant proteins. The E region of human XIST RNA (nucleotides 11,900–13,100 of NCBI NR_001564.2) was transcribed from a DNA template using T7 RNA polymerase (MEGAscript, Thermo Fisher Scientific), treated with DNase I (TURBO, Thermo Fisher Scientific) and purified via denaturing polyacrylamide gel electrophoresis. Product RNA was eluted from gels in nuclease-free water for 2 h at 23 °C and concentrated with centrifugal filters (Amicon Ultra 10K, Millipore Sigma). Before SDA crosslinking, RNA was heat denatured at 98 °C for 2 min and then cooled on ice for 2 min before being diluted to 10 nM in 200 µl of RNP crosslinking buffer (1× PBS (pH 7.4), 1 mM $MgCl_2$ and 1 mM DTT) containing varying concentrations of recombinant XIST-binding proteins PTBP1, MATR3 or TIA1 (HEK293 recombinant, OriGene) or BSA control protein (Millipore Sigma). RNPs were allowed to assemble for 30 min at 23 °C. Then, 196 µl of mixtures was added to 4 µl

of 100 mM SDA (in DMSO) in wells of a six-well plate and incubated in the dark for 15 min at 23 °C. RNPs were crosslinked with 3 J/cm^2 of 365-nm-wavelength UV light. To digest unbound and crosslinked proteins, reactions were adjusted to 1.5% SDS, 20 mM EDTA and 0.5 mg ml^{-1} of proteinase K and incubated for 2 h at 37 °C. RNA was purified once with 1.8× SPRI magnetic beads, purified again over an RNeasy MinElute column (Qiagen) and eluted into 14 µl of nuclease-free water.

MaP reverse transcription. MaP reverse transcription was performed using a revised protocol as described[61,62]. For smaller RNA targets (U1, RNase P and RMRP), 2 pmol of gene-specific primers (Supplementary Table 3) was mixed with 500 ng of total nuclear RNA (or unfractionated total RNA when indicated). For MaP reverse transcription of enriched Xist/XIST RNAs or in vitro crosslinked XIST E region, 7 µl of final RNA product was mixed with 200 ng of random nonamer DNA oligonucleotides. When performing MaP reverse transcription on ribosomal RNA, 3 µg of total cytoplasmic RNA was mixed with 200 ng of random nonamers. To RNA primer mixes, 20 nmol of dNTPs (5 nmol each base) was added (10 µl of total volume of RNA, primers and dNTPs), heated to 70 °C for 5 min and then immediately placed at 4 °C for 2 min. To this template solution, 9 µl of freshly made 2.22× MaP buffer (111 mM Tris-HCl (pH 8.0), 167 mM KCl, 13.3 mM $MnCl_2$, 22 mM DTT and 2.22 M betaine) was added, and the mixture was incubated at 25 °C for 2 min. After adding 200 units of SuperScript II reverse transcriptase (Thermo Fisher Scientific), reaction mixtures were incubated for 10 min at 25 °C and for 90 min at 42 °C, cycled ten times between 42 °C and 50 °C with each temperature incubation 2 min long and then heated to 70 °C for 10 min to inactivate enzyme. Reverse transcription reactions were buffer exchanged into TE buffer (10 mM Tris-HCl (pH 8.0) and 1 mM EDTA) (Illustra G-50 Microspin columns, GE Healthcare).

Two-step polymerase chain reaction of small RNA MaP libraries. Small RNA sequencing libraries were generated using a two-step polymerase chain reaction (PCR) strategy as described[61,62]. Briefly, 3 µl of complementary DNA (cDNA) from the reverse transcription reaction was used as template for step 1 PCR, using 20 cycles of gene-specific PCR (Q5 Hot Start polymerase, New England Biolabs): 30 s at 98 °C; 20 cycles of 10 s at 98 °C, 30 s at gene-specific annealing temperature and 20 s at 72 °C; and 2 min at 72 °C. Each set of step 1 primers contained the same added handles to prime step 2 PCR (Supplementary Table 3), in which Illumina adapters and multiplex indexing sequences were appended to the libraries. Step 1 PCR products were purified (SPRI beads, Mag-Bind TotalPure NGS, Omega Bio-tek, at a 1× ratio), and 2 ng of product was used as template for step 2 PCR. Step 2 PCR involved 30 s at 98 °C, 10× (10 s at 98 °C, 30 s at 66 °C and 20 s at 72 °C) and 2 min at 72 °C. Step 2 PCR products were purified with SPRI beads at a 0.8× ratio and eluted into 15 µl of nuclease-free water.

Second-strand synthesis, fragmentation and amplification of long RNA MaP libraries. For products of randomly primed MaP reverse transcription, buffer-exchanged cDNA was diluted to 68 µl with nuclease-free water. Each diluted cDNA was mixed with 8 µl of 10× Second Strand Synthesis Reaction Buffer and 4 µl of Second Strand Synthesis Enzyme Mix (NEBNext, New England Biolabs), and reactions were incubated at 16 °C for 2.5 h. The double-stranded DNA (dsDNA) products were purified with SPRI beads at a 0.8× ratio to favor longer products and exclude probe-templated products. Products were eluted into 15 µl of nuclease-free water. The dsDNA libraries were fragmented, multiplex indexed and PCR amplified. To fragment libraries from total cytoplasmic RNA, 5 µl of 0.2 ng µl^{-1} dsDNA was combined with 10 µl of Tagment DNA Buffer and 5 µl of Amplicon Tagment Mix (Nextera XT DNA Library Prep Kit, Illumina). Mixtures were incubated at 55 °C for 5 min and then cooled to 10 °C. As soon as the temperature reached 10 °C, 5 µl of NT Buffer (Nextera XT DNA Library Prep Kit, Illumina) was added to neutralize the reaction, which was then incubated at 23 °C for 5 min. The entire reaction volume was used as a template for PCR with 15 µl of Nextera PCR Master Mix and 5 µl each of forward and reverse indexing primers (Nextera XT DNA Library Prep Kit, Illumina): 72 °C for 3 min, 95 °C for 30 s; 12 cycles of 95 °C for 10 s, 55 °C for 30 s and 72 °C for 30 s; and 72 °C for 5 min. The final PCR products were purified with SPRI beads at a 0.65× ratio and eluted into 15 µl of nuclease-free water. For low-concentration Xist and XIST capture libraries, 8 µl of capture product was fragmented with only 2 µl of Amplicon Tagment Mix; the concentration of index primers was halved during PCR; and PCR cycles were increased to 20.

Sequencing of MaP libraries. Size distributions and purities of amplicon and randomly primed libraries were verified (2100 Bioanalyzer, Agilent). Step 2 amplicon libraries (about 120 amol of each) were sequenced on a MiSeq instrument (Illumina) with 2 × 150 or 2 × 250 paired-end sequencing, depending on the length of the RNA target. Libraries derived from total cytoplasmic RNA were sequenced with 2 × 300 paired-end sequencing on a MiSeq instrument, combining reads from multiple runs until desired ribosomal RNA sequencing depth was achieved. Xist and XIST capture libraries were sequenced to desired depth via a combination of 2 × 300 paired-end runs on a MiSeq and 2 × 150 paired-end runs on a NextSeq 500 instrument.

Mutation counting and SHAPE profile generation with ShapeMapper 2 software. FASTQ files from sequencing runs, with the exception of capture libraries, were directly input into ShapeMapper 2 software[63] for read alignment and mutation counting. Crosslink-induced termination events are specifically omitted in this analysis because such stops do not contribute information beyond that measured in read-through events[61]. To ensure that mutation rates were not affected by reduced fidelity at reverse transcription initiation sites, reads from capture libraries were trimmed by 14 nucleotides (primer length + 5 nts) after adapter sequences on each end. To accomplish this step for amplicon libraries, target FASTA files input to ShapeMapper 2 had primer-overlapping sequences and the first 5 nucleotides transcribed in reverse transcription set to lowercase, which eliminates these positions from analysis. To expedite analysis of long RNAs such as Xist/XIST, corresponding FASTQs were split into ~10 subsets and run in multiple parallel ShapeMapper 2 instances before having their outputs recombined into single profiles. ShapeMapper 2 was run with min-depth 5,000 and output-classified flags, with all other values set to defaults. In an RNP-MaP experiment, the SDA + UV-treated samples are passed as the 'modified' samples and UV-only treated samples as 'unmodified' samples. The outputs 'profile.txt', 'parsed.mut', and '.map' files are required for RNP-MaP site, RNP-MaP correlation and SHAPE analyses.

Identification of low SHAPE, low Shannon entropy regions of Xist and XIST using SuperFold. SuperFold analysis software[16] was used with in-cell and cell-extracted 5NIA experimental SHAPE data from mouse Xist and human XIST to inform RNA structure modeling by RNAstructure[64]. Default parameters were used to generate base-pairing probabilities for all nucleotides (with a max pairing distance of 600 nt), Shannon entropies for each nucleotide and minimum free energy structure models.

ΔSHAPE of mouse RNase P and Rmrp. Normalized SHAPE reactivities for 5NIA-treated mouse RNase P and Rmrp RNAs were compared between in-cell treated samples and those treated after cell extraction using the ΔSHAPE program[65]. Default parameters were used, and the 5′ primer sequence, the 3′ primer sequence and the first 5 nucleotides transcribed during the reverse transcription step were all masked to exclude them from analysis. Only nucleotides that passed the included Z′ factor and standard score significance testing were mapped as ΔSHAPE sites.

Post-processing of mutation frequencies into RNP-MaP reactivities. Per-nucleotide mutation frequencies (number of mutation events/effective read depth) for both crosslinked (SDA + UV-treated) and uncrosslinked (UV-treated) samples were calculated from output ShapeMapper 2 profiles. RNP-MaP 'Reactivity' was computed as the ratio of nucleotide crosslinked mutation frequency to un-crosslinked mutation frequency (SDA + UV rate/UV-only rate). Exceptions are in Fig. 1c and Supplementary Fig. 1, where 'Reactivity' refers to the ratio with a no-treatment control as the denominator (treatment rate/no-treatment rate). To be designated as RNP-MaP sites, nucleotide positions had to pass three quality filters: 1) sites were required to have at least 50 more mutation events in the SDA + UV-treated sample than the UV-treated sample; 2) site reactivities had to exceed the nucleotide-dependent empirical thresholds described in the next section; and 3) nucleotide reactivities were required to achieve a Z′ factor greater than 0.

$$\text{Zfactor} = 1 - \frac{2.575\left(\boldsymbol{\sigma}_{\text{SDA+UV}} + \boldsymbol{\sigma}_{\text{UVonly}}\right)}{\left|\text{mutation rate}_{\text{SDA+UV}} - \text{mutation rate}_{\text{UVonly}}\right|}$$

$$\text{where } \boldsymbol{\sigma}_{nt} = \sqrt{\text{mutation rate}_{nt}}/\sqrt{\text{reads}_{\text{nt}}}$$

Empirical derivation of RNP-MaP site nucleotide reactivity thresholds. Two biological replicates of RNP-MaP were performed on human U1 small nuclear RNA (snRNA), RNase P RNA and 18S and 28S rRNAs, each a part of RNA–protein complexes where atomic resolution structural data are available, enabling separation of nucleotides into two groups: those within 10 Å of protein (<10 Å) and those father than 10 Å from protein (>10 Å). For U1 snRNA, the binding site of the SNUPN protein has been mapped by crosslinking and mass spectrometry[19], and distances between the three nucleotides surrounding the crosslink site and nearest amino acids were assumed to be less than 4 Å. For each RNA replicate, reactivities were further grouped by nucleotide identity (U, A, C and G). The 90% reactivity values of nucleotides in a >10-Å group were set as background thresholds ($BG_{X>10}$) and compared to the median ($MED_{\text{all X}}$) and standard deviations ($SD_{\text{all X}}$) of reactivities for all nucleotides included in both the <10-Å and >10-Å groups to create relative threshold factors (T_X):

$$T_X = \frac{\left(BG_{X>10} - MED_{\text{all X}}\right)}{SD_{\text{all X}}}, \text{ where X is U, A, C, or G}$$

Relative threshold factors for each nucleotide from all eight replicates of the four RNAs were then averaged together, weighted by the number of nucleotides measured in each RNA, to obtain final empirically derived nucleotide relative threshold factors: 0.59 for U, 0.29 for A, 0.93 for C and 0.78 for G. These factors represent the number of standard deviations from the median of nucleotide

reactivity that must be achieved to be considered an RNP-MaP site. Factors can be applied to any RNA: to get exact nucleotide thresholds for an RNP-MaP experiment, the median reactivities for each nucleotide group (U, A, C or G) are multiplied by their corresponding threshold factors. Factors were calculated from existing comprehensive datasets; however, including more data from other RNPs or improving upon existing atomic resolution RNP structures could increase the precision of these threshold factors in the future.

Graphical display of RNP-MaP reactivities and crosslinking sites. Violin plots representing distributions of RNP-MaP reactivities were generated using the vioplot package in R through the web tool BoxPlotR[66]. RNP-MaP crosslinking sites were superimposed onto atomic resolution structure models using PyMol[67]. Secondary structure projection images were generated using the VARNA visualization applet for RNA[68].

RNP-MaP correlation analysis. Correlations between RNP-MaP sites were computed over 3-nucleotide windows using a previously described G-test framework (RingMapper)[20]. Windows were required to be separated by >4 nucleotides, jointly covered by more than 10,000 sequencing reads, jointly co-mutated >50 times and have background mutation rates below 6% (Supplementary Fig. 2). Pairs of windows exhibiting G-test statistics >20 ($P < 10^{-5}$) in the SDA + UV-treated sample and G <10.83 ($P > 0.001$) in the UV-only sample were determined to be significantly correlated. Current technical limitations of MaP reverse transcription processivity (500–600 nucleotides) and sequencing instrument clustering (<1,000 nucleotides) limit distances of readily measured correlations to <500 nucleotides.

RNase P and RMRP structural alignment. Corresponding helices, loops and intervening regions in RNase P and RMRP RNAs were separated into structural domains (Fig. 3) and separately aligned with MUSCLE[69]. Region alignments were recombined to create the final alignment (Supplemental Data 1).

Calculation of Xist/XIST RNP-MaP site, correlation strength and eCLIP site densities. To calculate the RNP-MaP site density for Xist and XIST (Fig. 4a and Supplementary Fig. 8), nucleotides whose reactivities were in the top 5% of all reactivities (U, A, C and G nucleotides evaluated separately) were identified. Site density was defined as the number of nucleotides in a centered 51-nucleotide window that were top 5% sites. Correlation strength density was defined as the sum of the mutual information (MI) of all nucleotides within the window normalized to (that is, divided by) the read depth of the central nucleotide. We selected high-confidence K562 cell eCLIP sites[57] (Supplemental Data 2) within XIST that passed irreproducible discovery rate thresholds (www.encodeproject.org/eclip), meaning that sites were defined by signal peaks of similar amplitude in both eCLIP replicates. Although other eCLIP experiments have been performed in more directly XIST-relevant cell lines (including human embryonic stem cells), the limited number of proteins tested and the lack of a shared systematic approach did not support inclusion in this analysis. The eCLIP site density (Fig. 4b) was defined as the total number of observed eCLIP sites within the 51-nucleotide window. For densities in Fig. 6, Supplementary Fig. 9 and Supplementary Fig. 10, all RNP-MaP sites were included (because signal only comes from a single protein), and centered nucleotide windows were shortened to 25 long (because the E region is only 1,200 nts).

Identification of conserved sequence regions between mouse Xist and human XIST. To identify and rank areas of significant conservation between mouse Xist (NCBI NR_001463.3) and human XIST (NCBI NR_001564.2) RNA sequences, we performed a local alignment (BLASTn[70]) and retained all segments with E-values above 0 and with lengths >100 nucleotides and ranked these segments by alignment bitscore.

Network analysis of XIST RNP-MaP correlation-linked eCLIP sites. To create a list of protein-linking correlations, we first counted the number of times nucleotides within our high-confidence eCLIP sites were correlated in RNP-MaP data (links) and measured the summed total of MI within links. To ensure that links were not simply a product of eCLIP site number and proximity or the average length and density of RNP-MaP correlations, we randomly shuffled the location of RNP-MaP correlations and counted links achieved between protein pairs iteratively 2,000 times, generating *P* values for each protein pairing based on the number of links between them and the strength of those links (Supplementary Table 1). Resulting links between protein pairs with *P* values less than 0.05 were then used as edges connecting nodes (proteins) on a network map (Fig. 5). A maximum modularity of the network (0.419), weighted by the strengths of included links (MI) and without changing resolution, was calculated using Gephi[58], and node sizes were adjusted manually to convey indicated relationships.

Identification of RNP-MaP enriched motifs in the XIST E region. MEME[60] was used to identify motifs enriched by RNP-MaP in vitro. We first expanded each RNP-MaP nucleotide into 9-mer sites (extending by 4 on either side). Overlapping sites were combined iteratively until no overlapping sequences remained. Combined sites were used as MEME input in classic mode using a zero-order model of sequences, allowing for any number of motif repetitions in each sequence, and explicitly looking for 9-mer motifs. The top two motifs were retained for each in vitro experiment, and locations of all matching motifs in the XIST E region were found using FIMO[71] with a *P* value threshold of 10^{-3}. For the in-cell experiment, only sites from the E region were considered in MEME, and the first and fourth most significant motif (class 2 and class 1, respectively) were included in Fig. 6 and Supplementary Figs. 9 and 10. Use of background sets created from the XIST E region (nucleotides 11,900–13,100) or increasing the order of the background model further strengthened the significance of class 2 motifs and weakened the significance of class 1 motifs.

XIST RNA reporter plasmid design. To create XIST region-containing reporters, we used inverse PCR and re-ligation to insert a multiple cloning site into the 3′ end of the pNL 3.2.CMV vector (Promega) between XbaI and FseI sites and to add the second intron from human *HBA1* at nucleotide position 196 in the nanoluciferase coding region (native Xist/XIST is spliced in its central region). Plasmids with varying regions of XIST were subcloned into the XhoI and KpnI restriction sites, and a control plasmid was generated through inverse PCR (Supplementary Table 3).

Plasmid transfection and purification of XIST reporter RNA for qPCR. HEK293 cells were plated at 100,000 cells per ml in 2-ml volumes per well of six-well plates and then cultured for 24 h at 37 °C. Each well was then transfected with a mixture of 0.6 µg of reporter plasmid and 1.8 µl of FuGENE 6 transfection reagent (Promega) in 200 µl of serum-free DMEM, and cells were cultured for an additional 48 h at 37 °C. Transfected cells were pelleted at 1,500*g* for 5 min at 4 °C, washed once in cold PBS and pelleted again and resuspended in 500 µl of cytoplasmic lysis buffer. Cells were lysed for 10 min at 4 °C with agitation. Nuclei were pelleted at 1,500*g* for 5 min at 4 °C, and cytoplasmic lysates were separated into new tubes. Nuclei were washed once in low-salt solution, incubated with agitation at 4 °C for 2 min, pelleted again and then resuspended in 500 µl of proteinase K lysis buffer. NaCl, EDTA, SDS and proteinase K were added to cytoplasmic lysates up to proteinase K lysis buffer concentrations. Nuclear and cytoplasmic fractions were incubated for 2 h at 37 °C with intermittent mixing. Nucleic acid was recovered through two extractions with 1 volume of 25:24:1 PCA, two extractions with 1 volume of chloroform and precipitation with 1/25 volume of NaCl and 1 volume of isopropanol.

Small interfering RNA transfection and purification of endogenous RNA for quantitative PCR. HEK293 cells were plated at 20% confluence into wells of either six-well (2 ml, for qPCR) or 12-well (1 ml, with each well containing a poly-L-lysine-coated #1.5 coverslip (Neuvitro) plates for microscopy) and then cultured for 24 h at 37 °C. Media were replaced with DMEM + 10% FBS lacking antibiotics, and each well was then transfected with a mixture of Opti-MEM (Gibco), small interfering RNA (siRNA) and Lipofectamine RNAiMAX (Thermo Fisher Scientific) according to the manufacturer's protocol. Final siRNA concentrations were 20 nM in transfections, and siRNAs included the MISSION siRNA Universal Negative Control #1 (Millipore Sigma), PTBP1 siRNA SASI_Hs01_00216644 (Millipore Sigma), PTBP2 siRNA SASI_Hs01_00201967 (Millipore Sigma) and MATR3 siRNA HSS114732 (Thermo Fisher Scientific). For siPTBP samples, siRNAs for both PTBP1 and PTBP2 were included. Twenty-four hours after transfection, media were exchanged for new media including antibiotics. Seventy-two hours after transfection, slides (using cells from 12-well plates) were prepared for microscopy, or cellular RNA was harvested (from six-well plates; TRIzol, Invitrogen) for quantitative PCR (qPCR).

Plasmid transfection for XIST reporter microscopy. HEK293 cells (1 ml) were plated at a concentration of 10,000 cells per ml in wells of a 12-well plate, with each well containing a poly-L-lysine-coated #1.5 coverslip (Neuvitro). After 24 h of growth, 50 µl of transfection mix (20 µl of 15 ng µl^{-1} reporter plasmid, 0.9 µl of FuGENE 6 (Promega) and 79 µl of DMEM) was added to cells.

qPCR of endogenous and reporter RNAs. Equal quantities of RNA (30 ng) from nuclear, cytoplasmic or total TRIzol RNA fractions were used to generate first-strand cDNA in random hexamer-primed reverse transcription reactions (SuperScript II, Thermo Fisher Scientific). Triplicate mixtures of 2.5 µl of template cDNA, 2.5 µl of 2 µM primers and 12.5 µl of Maxima SYBR Green qPCR Master Mix (Thermo Fisher Scientific) in 25-µl total reactions were prepared for each sample. Reaction mixtures from matched no-reverse-transcriptase controls were also prepared. Primer sets (Supplementary Table 3) included those specific to the reporter gene, an endogenous RNA and 18S ribosomal RNA (as a normalization control). qPCR was performed on a QuantStudio 6 Flex Real-Time PCR System (Thermo Fisher Scientific) with steps of 5 min at 95 °C and 40 cycles of 15 s at 95 °C, 30 s at 65 °C and 40 s at 72 °C, using a melting curve to confirm single major products. Fluorescence readings were taken at elongation steps (72 °C). All specific signals were observed at least eight cycle thresholds earlier (256-fold more signal) than no-reverse transcriptase controls. Signals were averaged across triplicate qPCRs, normalized to 18S rRNA signal and then normalized to either control reporter or control siRNA signals, depending on the experiment. XIST reporters were expressed at 50% the level of endogenous XIST, as assessed by reverse transcription qPCR.

RNA fluorescence in situ hybridization of XIST RNA and XIST RNA reporters. Fluorescence in situ hybridization (FISH) probes labeled with Quasar 570 Dye and antisense to human XIST (SMF-2038-1) and the nanoluciferase mRNA (custom) were ordered from LGC Biosearch. Custom design parameters were masking level 5, oligo length 19 and minimum spacing length 2. At 72 h (for siRNA) or 48 h (for reporter expression) after transfection, each well was washed once with 1 ml of PBS, fixed with 1 ml of 3.7% formaldehyde in PBS for 10 min at room temperature and then washed twice with PBS. Cells were permeabilized with 1 ml of 70% ethanol overnight at 4 °C. After removing ethanol, cells were incubated in wash buffer 1 (20% Wash Buffer A (LGC Biosearch) and 10% formamide) for 5 min at room temperature, and then coverslips were transferred to a humidified chamber with cells facing down onto 100 µl of hybridization buffer (90% Stellaris RNA FISH Hybridization Buffer (LGC Biosearch), 10% formamide and 125 nM antisense probes). After overnight incubation in the dark at 37 °C, coverslips were transferred into 12-well dishes and incubated in 1 ml of wash buffer 1 for 30 min at 37 °C in the dark and then counterstained for 30 min at 37 °C with 5 ng µl^{-1} of DAPI in 1 ml of wash buffer 1. Coverslips were washed a final time in 1 ml of Wash Buffer B (LGC Biosearch) before being mounted onto a microscope slide with 12 µl of VECTASHIELD Mounting Medium (Vector Laboratories) and sealed. RNA FISH z-stack images of XIST reporters were captured using a ×100/1.3 oil objective on an Olympus IX81 microscope and were deconvoluted using the AutoQuant X software. z-stack images of native XIST particles were captured using a ×100/1.3 oil objective on a Nikon Eclipse Ti microscope. z stacks were collapsed into maximum intensity projections and quantified using the Fiji software. XIST foci density was measured as the pixel area within XIST foci passing a fluorescence intensity background threshold (3,680).

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

Data availability

Raw and processed sequencing data sets analyzed in this report will be made available upon reasonable request and have been deposited in the Gene Expression Omnibus database (GSE152483).

Code availability

ShapeMapper2, ΔSHAPE, SuperFold and RingMapper software used for analysis are available at http://weeks.chem.unc.edu/software.html and https://github.com/Weeks-UNC. MEME, VARNA, PyMol and Gephi are all third-party, open-source software.

References

61. Busan, S., Weidmann, C. A., Sengupta, A. & Weeks, K. M. Guidelines for SHAPE reagent choice and detection strategy for RNA structure probing studies. *Biochemistry* **58**, 2655–2664 (2019).
62. Sengupta, A., Rice, G. M. & Weeks, K. M. Single-molecule correlated chemical probing reveals large-scale structural communication in the ribosome and the mechanism of the antibiotic spectinomycin in living cells. *PLoS Biol.* **17**, e3000393 (2019).
63. Busan, S. & Weeks, K. M. Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA* **24**, 143–148 (2018).
64. Reuter, J. S. & Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 129 (2010).
65. Smola, M. J., Calabrese, J. M. & Weeks, K. M. Detection of RNA–protein interactions in living cells with SHAPE. *Biochemistry* **54**, 6867–6875 (2015).
66. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2011).
67. DeLano, W. Pymol: an open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography* **40**, 82–92 (2002).
68. Darty, K., Denise, A. & Ponty, Y. VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**, 1974–1975 (2009).
69. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
70. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
71. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).

Acknowledgements

The work was supported by grants from the National Science Foundation (MCB-1121024) and the National Institutes of Health (R35 GM122532) to K.M.W. C.A.W. is a postdoctoral fellow at the American Cancer Society (ACS 130845-RSG-17-114-01-RMC). J.M.C. was supported by National Institutes of Health grant R01 GM121806. Xist and XIST antisense probes were provided by the M. Guttman laboratory (CalTech), and we thank M. Blanco (CalTech) for his initial support in their application. XIST eCLIP data from published works were provided by the G.W. Yeo laboratory (UCSD), and we thank G.W. Yeo (UCSD), M. Corley (UCSD) and D. Sprague (UNC) for support in formatting these data for integration into this work and for helpful comments on the project.

Author contributions

C.A.W. and P.B.J. conducted experiments. C.A.W., A.M.M. and K.M.W. analyzed data. C.A.W., J.M.C. and K.M.W. designed and interpreted experiments. The manuscript was written by C.A.W. and K.M.W. with input from all authors.

Competing interests

A.M.M. is an advisor to and K.M.W. is an advisor to and holds equity in Ribometrix, to which mutational profiling technologies have been licensed.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41587-020-0709-7.

Correspondence and requests for materials should be addressed to K.M.W.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s): *Double-blind peer review submissions: write DBPR and your manuscript number here instead of author names.*

Last updated by author(s): Aug 30, 2020

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 our Editorial Policies 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	No Software was used for data collection.
Data analysis	Shapemapper2, deltaSHAPE, SuperFold, RingMapper software used for analysis are available here (http://weeks.chem.unc.edu/software.html) and here (https://github.com/Weeks-UNC). AutoQuant X 3.0.4, software is commercially available. Fiji (ImageJ 1.52g), MEME 5.1.1, VARNA 3.93, PyMol 2.2.2, Gephi 0.9.2, vioplot in R 0.2, and MUSCLE 3.8.31 are open source software available on the web.

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 and 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

Raw and processed sequencing datasets analyzed in this report are available in the Gene Expression Omnibus database (GSE152483)

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	Sample Sizes were not predetermined. Original Sample sizes were retained as effect sizes were large enough to discriminate real effects.
Data exclusions	No data exclusions to declare.
Replication	When performing biological replicates for all experiments, chemical probing and sequencing library preparation or microscopy imaging were performed on distinct populations of cells on different days.
Randomization	Randomization is not typical for analysis of common cell line experiments.
Blinding	Phenotypic assignments of microscope images were made after image processing, so the selection of cells analyzed was blinded. Frames captured were randomly chosen and no phenotype could be assigned until after image processing (z-stack maximum projection and signal quantification). Assignment and analysis were not blinded, however no data were excluded from analysis in an attempt to eliminate bias.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293 (ATCC), SM33 (provided by Guttman Laboratory, Caltech)
Authentication	no additional authentication was performed
Mycoplasma contamination	Lines were not tested for mycoplasma after receipt.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.