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

Chase A. Weidmann, Anthony M. Mustoe, Parth B. Jariwala, J. Mauro Calabrese and Kevin M. Weeks

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 (RNP) are complexes made up of interacting RNA and protein and govern both messenger RNA (mRNA) regulation and the function of non-coding RNA (ncRNA). Understanding how RNP 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, and high-resolution structural approaches 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. Sites of protein crosslinking to RNA can be mapped transcriptome wide, either without or with identification of binding sites for individual proteins (crosslinking and immunoprecipitation (CLIP) and photo-activatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP)). However, UV-based crosslinking suffers from experimental biases and limited binding site resolution; metabolic labeling probes a single substituted nucleoside-base at a time; 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. 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. When activated with long-wavelength UV, diazirines form carbene or diazo intermediates, 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), 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 (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.

1Department of Chemistry, University of North Carolina, Chapel Hill, NC, USA. 2Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA. 3Department of Pharmacology, University of North Carolina, Chapel Hill, NC, USA. 4E-mail: weeks@unc.edu

NATURE BIOTECHNOLOGY | VOL 39 | MARCH 2021 | 347-356 | www.nature.com/naturebiotechnology
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 complexes3–6, whereas nucleotides distant from bound proteins rarely passed reactivity thresholds (Supplementary Fig. 1).
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 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 and photo-lysine metabolic labeling, 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 (Fig. 2a, and Supplementary Fig. 2), 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) 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 cross-links, multiple crosslink sites can be detected per single RNA molecule. We adapted a G-test framework 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
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 RN endonucleases that cleave distinct substrates\(^\text{20,21}\). 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\(^\text{P}\). 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\(^\text{7}\). 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\(^\text{22}\). 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\(^\text{26}\) (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\(^\text{27-29}\). 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\(^\text{30}\)). 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\(^\text{31}\), obtained in K562 cells\(^\text{32}\)). Although there are differences between HEK293 and K562 cells, both cell line cells maintain silenced X chromosomes and XIST compartments, and eCLIP peaks are shared on genes of similar expression between cell lines\(^\text{33}\). 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\(^\text{34}\) eCLIP\(^\text{35}\) 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\(^\text{27}\). Community members include factors involved in XIST processing, XIST stability and XIST-mediated silencing: UCHL5 (ref. \text{36}), EXOSC5 (ref. \text{37}), HNRNPUL1 (ref. \text{38}) and RBM15 (ref. \text{39}) (Fig. 5). Silencing community members TARDBP and RBM22 are RNA-dependent regulators of transcription\(^\text{40}\). 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\(^\text{41}\). PTBP1, MATR3 and TIA1 each undergo liquid–liquid phase transitions to form RNA granules\(^\text{42-44}\), and PTBP1 and MATR3 interact on other RNAs\(^\text{45}\), features consistent with the formation of an XIST-mediated compartment.

The Splicing community includes proteins that control splicing (U2AF2, SRSF1, TRA2A, AQR and ILF3\(^\text{45-47}\)) and a chromatin modulator (GRWD1)\(^\text{48}\). 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 HNRNP proteins (U and C) that interact with one another\(^\text{49}\) 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.
RNP-MaP reveals interaction sites for Compartmentalization community proteins. The XIST E region, critical for maintaining the silenced X chromosome compartment\textsuperscript{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\textsuperscript{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\textsuperscript{50} and by CLIP methods\textsuperscript{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...
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) 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 eCLIP sites, but only 35% of RNP-MaP sites (of 3,766) fall within eCLIP sites.

PTBP1 and MATR3 interaction with E region controls XIST particle formation. PTBP1 and MATR3, major components of the Compartimentalization 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 CLIP 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 Xist29,30. 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 compartment29,30 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$^7$ for sites and 10$^5$ 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. Our RNP-MaP study now directly reveals that repeat sequences in Xist/XIST are extensively bound...
Many other highly connected protein interaction networks identified by RNP-MaP in Xist/XIST suggest additional biology that warrants future study. 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.$^{10}$ 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 Dro sophila 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 ribonucleoprotein P holozyme. 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. Garzía, 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. & Khavarri, A. P. Methods to study RNA–protein interactions. Nat. Methods 16, 225–234 (2019).
13. Madler, S., Bich, C., Touboul, D. & Zenobi, R. Chemical cross-linking with NO-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. 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).
17. 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).
18. 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).
19. 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).
20. 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).
21. Will, C. In vitro reconstitution of mammalian U1 snRNPs active in splicing: the U1–C1 protein enhances the formation of early (E) spliceosomal complexes. Nucleic Acids Res. 24, 4614–4623 (1996).
22. 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 RNP s. *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 IncRNA 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. Repeat E anchors Xist RNA to the inactive X chromosomal compartment through CDK1A-interacting protein (CIZ1). *Proc. Natl Acad. Sci. USA* **114**, 10654–10659 (2017).

29. 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).

30. Nesterova, T. B. et al. Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation. *Nat. Commun.* **10**, 3129 (2019).

31. Brockdorff, N. Local tandem repeat expansion in Xist RNA as a model for the functionalisation of ncRNA. *Nucleic Acids Res.* **43**, 3898–3911 (2015).

32. Uemura, Y. et al. Matrin3 binds directly to intronic pyrimidine-rich sequences and controls alternative splicing. *Genes Dev.* **167–174** (2002).

33. Sahakyan, A., Yang, Y. & Plath, K. The role of Xist in X-chromosome dosage silencing. *Nature* **503**, 711–719 (2020).

34. Dey, I. et al. The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. *Genes Dev.* **31**, 876–888 (2017).

35. Van Nostrand, E. L. et al. A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 508–514 (2020).

36. Ong, S. L. et al. Systematic analysis of eukaryotic RNase P RNPs. *Nucleic Acids Res.* **57**, 369–373 (2019).

37. Ciaudo, C. et al. Nuclear mRNA degradation pathway(s) are implicated in Xist regulation and X chromosome inactivation. *Proc. Natl Acad. Sci. USA* **785**, 785–798 (1981).

38. Nesterova, T. B. et al. Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation. *Nat. Commun.* **10**, 3129 (2019).

39. Brockdorff, N. Local tandem repeat expansion in Xist RNA as a model for the functionalisation of ncRNA. *Nucleic Acids Res.* **43**, 3898–3911 (2015).

40. Brea, 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 Biology* (1994).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020
Methods

Cell culture. Adherent mammalian cells used in chemical probing experiments, either SM33 (ref. 5) or HEK293 cells, were grown to 80–90% confluence in either six-well plates (for refolding 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 mg/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 replicas, chemical probing and sequencing library preparation were performed on distinct populations of cells on different days.

In-cell crosslinking with SDA. SDA (NHS-diazirine, 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 and were pelleted, SDA was added 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 3/3 cm of 365-nm wavelength UV light (about 9 min in a UVP CL-1000 equipped with five 8 W FST5 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 MgCl2, 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. Nuclear samples were collected and then pelleted at 1,500 g 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 MgCl2, 1 mM CaCl2, 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 1.1x folding buffer (111 mM HEPES (pH 8.0), 111 mM NaCl and 5.55 mM MgCl2), followed by two extractions with 1 volume of chloroform. RNA was buffer exchanged into 1.1x folding buffer over a desalting column (PD-10, GE Healthcare). RNA was pelleted at 13,000 g at 37 °C. RNA solution was split into two equal portions: one was added to a 1/9 volume of 250 mM SNA 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% confluence 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/cm² 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 above, were crosslinked by SDA. Cells treated by SDA, then with photo-lysine 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 1x DNase buffer and incubated with 2 units of DNase (TURBO, Thermo Scientific) for 10 min at 37 °C. After the first incubation, 5 µl of 1x DNase buffer was 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 nuclelease-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 1.5x 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 1x hybridization buffer. After magnetic separation, the pre-cleared supernatant was retained. Biotinylated antisense RNA capture probes26 (Guttmann 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 1x 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 hybridization, 30 µl of streptavidin magnetic beads were added, and resuspended in 1x 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 1x 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 kit, Qiagen) to reduce non-target RNA. RNA eluates 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 mM in 200 µl of KNP crosslinking buffer (1× PBS (pH 7.4), 1 mM MgCl2, 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 4µl
MaP reverse transcription. MaP reverse transcription was performed using a
revised protocol as described22-24. 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
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, 5µl of freshly made 2.22×
MaP buffer (111 mM Tris-HCl (pH 8.0), 167 mM KCl, 13.3 mM MgCl2, 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 1× 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 described22-24. Briefly, 3µl of complementary DNA (cDNA) from
the reverse transcription reaction was used as template for step 1 PCR, using 20
cycles of gene-specific PCR (QS Hot Start polymerase, New England Biolabs): 30
µl 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 1 primer contained the same
added handles to prime step 2 PCR (Supplementary Table 3), in which Illumina
adapter and multiplex indexing sequences were appended to the libraries. For
1 PCR products were purified (SPIR beads, Mag-Bind TotalPure NGS, Omega
Bio-tek, at a 1x ratio), and 2 µg of product was used as template for step 2 PCR.
Step 2 PCR involved 30 s at 98 °C, 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 SPIR 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 8µl 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 2.5h. The double-stranded DNA (dsDNA) products
were purified with SPIR 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, 5µl of 0.2 µg µl−1 dsDNA was
combined with 10µl of Tagment DNA Buffer and 5µl of Amplicom Tagment Mix
(Nextera XT DNA Library Prep Kit, Illumina). Libraries 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
5µl 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 SPIR beads at a 0.6×x ratio and eluted into 15 µl of nuclease-free water. For
low-concentration Xist and XIST capture libraries, 8µl of capture product was
fragmented with only 2 µl of Amplicom Tagment Mix; the concentration of index
primers was halved during PCR and PCR cycles were decreased to 20 cycles.

Sequencing of MaP Libraries. Size distributions and purities of ampiclon 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.

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 software38 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 events24. 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 ampiclon 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
different profiles. ShapeMapper 2 was run with min-depth 3,000 and output-classified
flags, with all other values set to defaults. In an RNP-MaP experiment, the SDA
+ UV-treated samples were passed as the ‘modified’ samples and UV-only treated
samples as ‘unmodified’ samples. The outputs ‘profile.txt’, ‘parsed.mesh’ 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 ShapeMapper. ShapeMapper analysis software38 was used with in-cell
and cell-extracted 5tnA experimental SHAPE data from mouse Xist and human XIST
to inform RNA structure modeling by RNAstructure41. Default parameters were
used to generate base-pairing probabilities for all nucleotides (with a max pairing
distance of 600nt), Shannon entropies for each nucleotide and minimum free
energy structure models.

SHAPE of mouse RNase P and Rmrp. Normalized SHAPE reactivities for
5tnA-treated mouse RNase P and Rmrp RNAs were compared between in-cell
treated samples and those treated after cell extraction using the ΔSHAPE
program7. Default parameters were used, and the 5’ primer sequence, the 3’
polyadenylation nucleotides transcribed in reverse transcription step were set as
to mask all excluded 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
into uncrosslinked mutation frequency (SDA + UV / SDA + 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.

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 each RNA replicate,
reactions were further grouped by nucleotide identity (U, A, C and G). The
nearest amino acids were assumed to be less than 4 Å. For each RNA replicate,
deviations (SDall X) of reactivities for all nucleotides included in both the
10-Å groups to create relative threshold factors (TX):

\[
T_X = \frac{[\text{SGD}_{\text{UV}} - \text{SGD}_{\text{UV}}]}{\text{SGD}_{\text{UV}}} - \frac{\text{SGD}_{\text{UV}}}{\sqrt{\text{reads}_{\text{UV}}}}
\]

where \( \text{reads}_{\text{UV}} \) is the number of reads counted in the UV-treated
condition.

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
to the SNUNP protein has been mapped by crosslinking and mass spectrometry41,
and distances between the nucleotides surrounding the crosslink site and
nearest amino acids were assumed to be less than 4 Å. For each RNA replicate,
reactivity libraries 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 (BGX>10) and compared to the median (MEDx) and standard
deviations (SDx) of reactivities for all nucleotides included in both the <10-Å and
>10-Å groups to create relative threshold factors (TX):
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 databases; 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®. RNP-MaP crosslinking sites were superimposed onto atomic resolution structure models using PyMol®. Secondary structure projection images were generated using the VARNA visualization appli for RNA®.

**RNP-MaP correlation analysis.** Correlations between RNP-MaP sites were computed over 3-nucleotide windows using a previously described G-test framework (RingMapper)®. 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 < 10⁻⁴⁰) in the SDA + UV-treated sample and G < 0.001 (P > 0.05) 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.

**RNP-MaP correlation analysis.** Correlations between RNP-MaP sites were computed over 3-nucleotide windows using a previously described G-test framework (RingMapper)®. 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 < 10⁻⁴⁰) in the SDA + UV-treated sample and G < 0.001 (P > 0.05) 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.

**RNP-MaP correlation analysis.** Correlations between RNP-MaP sites were computed over 3-nucleotide windows using a previously described G-test framework (RingMapper)®. 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 < 10⁻⁴⁰) in the SDA + UV-treated sample and G < 0.001 (P > 0.05) 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.

**RNP-MaP correlation analysis.** Correlations between RNP-MaP sites were computed over 3-nucleotide windows using a previously described G-test framework (RingMapper)®. 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 < 10⁻⁴⁰) in the SDA + UV-treated sample and G < 0.001 (P > 0.05) 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.
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 Hybirdization 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 µl 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 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. Guideline 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. Britter, R. G. & Matthews, 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. Darly, 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.J. conducted experiments. C.A.W., A.M.M. and K.M.W. analyzed data. C.A.W. and P.J. conducted experiments. 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.
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

Ourweb 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 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 | Methods |
|-----------------------------------|---------|
| n/a                               | n/a     |
| ☒ Antibodies                      | ☒ Involved in the study |
| ☒ Eukaryotic cell lines           | ☒ ChiP-seq |
| ☒ Palaeontology and archaeology   | ☒ Flow cytometry |
| ☒ Animals and other organisms     | ☒ MRI-based neuroimaging |
| ☒ Human research participants     |         |
| ☒ Clinical data                   |         |
| ☒ Dual use research of concern    |         |

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 ICCLABS, register): No commonly misidentified cell lines were used.