Characterization of RBM15 protein binding with long noncoding RNAs

Emma Bose[$,1], Caleb Mayes[$,1], Lance Ellis[#,1], Corrine Baker[#,1], Sofia Tambalotti[#,1], Shengwei Xiong[#,1], Yaa Pokua Osei Sarpong[#,1], Marwan Shalaby[#,1], Lucas Barry[#,1], Frank Lewis[#,1], Johnson Joseph[#,1], Talaidh Isaacs[#,1], Derik McCarthy[#,1], Dana Katz[#,1], Jingyang Wang[#,1], Victoria Zirimu[#,1], Luis Vargas[#,1], Julian Von Hofe[#,1], Glen Aguilar[#,1], Katherine Buchan[#,1], Lei Zheng[#,1], Gregory Wolfe[1,2], and Alisha N Jones*[,1]

[$,#]These authors contributed equally
*Corresponding author: alisha.jones@nyu.edu

[1]Department of Chemistry, New York University, New York, New York, USA
[2]Department of Physics, New York University, New York, New York, USA

Abstract

The RNA binding motif 15 protein (RBM15) plays a critical role in post-transcriptional regulation. Its role in facilitating m6A modification, specifically through guiding the writer complex (WTAP METTL13 METTL14) to DRACH sequence motifs, has been demonstrated for several RNAs, including long noncoding RNAs (lncRNAs). The structural mechanism that underlies how RBM15 interacts with RNA has yet to be elucidated. In this study, we mined and bioinformatically assessed publicly available genome wide RNA 2D structural probing and RBP cross linking and immunoprecipitation data to investigate the propensity of lncRNAs to interact with RBM15. We then experimentally characterized how this interaction occurs with two lncRNAs, FIRRE and XIST. RBM15, which possesses three RNA recognition motifs (RRMs), primarily interacts with stem loop structures adopted by lncRNAs through its two terminal RRMs, RRMs 2 and 3. Using solution NMR, we find RRMs 2 and 3 are rigidly confined in solution, in the absence of RNA. Altogether, this work provides clarity on the molecular mechanism by which RBM15 interacts with RNAs to govern biological function.

Running title: How the RBM15 protein interacts with structured lncRNAs

Keywords: lncRNA, FIRRE, XIST, RBM15, CatRAPID, iCLIP, protein RNA interaction, chemical probing, RNA structure

Introduction

Long noncoding RNAs, or lncRNAs, are transcripts consisting of more than 500 nucleotides that are not translated into protein [1,2]. Via their sequences and structures, lncRNAs regulate a variety of biological processes, including chromatin remodeling, transcription, post-transcriptional gene expression, messenger RNA stability, translation, and post-translational modification [3–6]. This regulation occurs through lncRNAs interacting with their cognate RNA binding proteins (RBPs), other RNAs, and DNA [7].

Of particular interest to us is the RBP RBM15. RBM15 possesses three N-terminal RNA recognition motifs (RRMs), and a spen orthologue and paralogue C-terminal (SPOC) domain (**Figure 1A**). Through its RRMs and SPOC domain, RBM15 has been identified as a key player in post-transcriptional regulation (e.g., through m6A modification [8] and nuclear export [9,10]) and also serves as a regulatory protein in the process of hematopoietic cell homeostasis [8]. RBM15 has been predicted to bind several lncRNAs, of which include the X-inactive specific transcript (XIST) and the functional intergenic repeating RNA element (FIRRE) [11] (**Figure 1B**).

During X-chromosome inactivation (XCI), approximately 94% of genes of the inactive X-chromosome are transcriptionally silenced by the lncRNA XIST [12]. XIST, which is conserved in placental mammals, is a 19kb long transcript that consists of 6 repeating elements, the A-, B-, C-, D-, E- and F-repeats [13]. The A-repeats, a 450 nucleotide (nt) region located at the 5' end of the transcript, are responsible for transcriptional silencing. In their absence, XIST can coat, but cannot transcriptionally silence the inactive X chromosome [14]. The A-repeats are predicted to adopt a structure that is conformationally dynamic; through this structural ensemble, the A-repeats interact with approximately 15 RBPs, one of which is RBM15 [8,15–17].

Included in the 6% of genes that escape XCI is FIRRE, which is syntenically conserved in mammalian species and is composed of transposable elements known as local repeats (LRs) [18]. Nearly 70 RBPs have been identified as interaction partners of the FIRRE lncRNA, which is implicated in the regulation of various biological processes that include stem cell pluripotency, adipogenesis, nuclear architecture organization, and hematopoiesis [19–22]. A more recently identified RBP of the lncRNA FIRRE is RBM15 [11]. Binding of this RBP has been localized to a 156-nt LR known as the Repeating RNA Binding Domain (RRD) through crosslinking experiments [23,24].

The structural mechanism underlying how RBM15 interacts with lncRNAs has yet to be established. In this study, we sought to understand how RBM15 interacts not only with XIST and FIRRE, but with other structured lncRNAs. First, we mined chemical probing data available from the RNA Atlas of Structure Probing (RASP) [25] to fold the structures of the nine lncRNAs. These lncRNAs are: Just Proximal to XIST (JPX), Five Prime to XIST (FTX), Long Intergenic Non-Protein Coding RNA 00630 (LINC00630), MicroRNA 503 Host Gene (MIR503HG), Long Intergenic Non-Protein Coding RNA 00894 (LINC00894), Functional Intergenic Repeating RNA Element (FIRRE), Inactivation Escape 1 (INE1), RepA X Inactive-Specific Transcript (RepA XIST), and Zinc Finger 674 Antisense RNA 1 (ZNF674-AS1). Following the identification of domains predicted to bind RBM15, we then focused on experimentally validating the predicted interactions with the FIRRE RRD and the A-repeats of XIST. We find that RBM15 primarily binds to stem-loop structured domains in lncRNAs through two of its three RRMs: RRM2 and RRM3. Herein we describe these results.

Results

LncRNAs are folded in solution

The RASP database contains 58 chemical probing and enzymatic cleavage datasets from 8 different studies carried out both in cells and *in vitro*. This includes selective 2' hydroxyl acylation analyzed by primer extension (SHAPE) and dimethyl sulfate (DMS) mutational profiling (MaP) experiments [26–33]. We selected nine lncRNAs from the X-chromosome and predicted the secondary structure of the splice variants that had the most chemical probing data coverage (number of nucleotides showing reactivity versus total number of nucleotides). Coverage varied from 8% to 100% across the lncRNAs. Details regarding the coverage of chemical probing data used to fold the structures are listed in **Table 1**.

The secondary structure predictions reflect intricately folded states for each of the nine lncRNAs (**Supplementary Figures 1-9**). Both *in vitro* and *in vivo* chemical probing data was available for FTX, INE1, JPX, and XIST; a comparison of the 2D structures from the two conditions reveals that each RNA folds differently. Structural differences are also observed in cases where the lncRNA is localized in the nucleus, cytoplasm, or chromatin. This is likely due to protein or RNA binding events occurring in the cell [34]. Indeed, several proteins have been predicted to bind these lncRNAs based on o RNA motifs enrichment in complete transcriptomes (oRNAment) analysis (**Supplementary Figure 10**) [35]. Considering that our focus was on understanding how RBM15 interacts with lncRNAs, we focused on identifying regions of the structured lncRNAs that interact with RBM15.

RBM15 interacts with stem-loop structured motifs through RRM2 and RRM3

RBM15 has been proposed to interact with FIRRE and XIST [8,11]. We were interested in determining whether the other lncRNAs in this study interact with RBM15. We queried the RBPs of these lncRNAs in oRNAment and searched through iCLIP data available from POST-traAnscriptional Regulation (POSTAR3) [36] and find that of the 9 lncRNAs investigated in this study, all but INE2 interact with RBM15 [36,37]. In agreement with the RNA binding sequence motif for RBM15 (**Figure 2A**), the crosslinking experiments revealed U and A-containing sequences in each of the 8 lncRNAs (**Figure 2B, Supplementary Table 1**) [38–41]. Notably, the RBM15 crosslinking sites are recurring in stem-loop structured motifs. This binding preference occurs even when the RNA localizes in different parts of the cell (e.g., cytoplasm versus chromatin).

We used CatRAPID Signature to determine which region of RBM15 is predicted to interact with RNA [42]. In agreement with characteristics of typical RNA recognition motifs, regions corresponding to the RRM domains of RBM15 (RRM1: 170-252, RRM2: 374-451, and RRM3: 455-529) are predicted to bind with RNA (**Supplementary Figure 11**). CatRapid Graphic (for RNAs < 1,200 nt) and CatRAPID Fragments (for RNAs > 1,200 nt) [43,44] were then used to determine which amino acid residues of RBM15 interact with the lncRNAs. Interestingly, we observed RRMs 2 and 3 (amino acids 374 - 529) bind in the best capacity for the majority of lncRNAs (**Figure 3**). The nucleotides involved in these interactions overall gave analogous results (relative to the POSTAR3 crosslinking sites) for the majority of the RNAs investigated. For example, the RBM15 crosslinking sites for lncRNAs FIRRE and XIST predominantly occur at the 5' end of the transcripts; CatRAPID also showed a high propensity for these interactions at the 5' end of the transcripts. The same is true for the lncRNAs LINC00894 and ZNF674-AS1. MIR503HG has four RBM15 crosslinking sites that are interspersed through the transcript, mostly occurring in the half of the transcript closer to the 3' end, which agrees with the CatRAPID prediction. Three of the lncRNAs, JPX, FTX, and LINC00630, have RBM15 crosslinking sites that did not completely agree with the CatRAPID prediction. For JPX, there are 15 crosslinking sites that are majorly near the 5' end of the transcript, however CatRAPID predicted the interaction occurring more towards the 3' end of JPX. Similarly, there are four crosslinking sites for FTX, yet CatRAPID predicted the interaction between FTX and RBM15 being sparse, localized at the 3' end of the transcript. And for LINC00630, the nine crosslinking sites occur throughout the transcript, however CatRAPID predicts that the interaction between

RBM15 and LINC00630 occurs only at the 3' end of the transcript. These results suggest that CatRAPID can provide useful information regarding protein-RNA interactions.

Encouraged by these findings, we carried out electrophoretic mobility shift assays (EMSA) for various RRM domains of RBM15 (RRM1, RRM2, RRM3, RRM12, RRM23; we were unable to obtain a soluble RRM123 construct of adequate concentration for EMSA experiments) with the FIRRE RRD and with the XIST A-repeats (**Figure 4A, B**).

RBM15 RRM12 shows weak binding with the FIRRE RRD RNA (**Figure 4A**). This is indicated by the decrease of the intensity of the RNA band at 20 μM. On the other hand, the intensity of the FIRRE RRD band is diminished between 0.6 and 1.2 μM with RBM15 RRM23. This suggests that RRM23 binds with better affinity than RRM12. We quantified the band intensities for the interaction between RRM23 and FIRRE and calculated a K_d of 1.56 ± 0.46 μM. Individually, RRMs 1 and 3 do not show appreciable binding with the FIRRE RRD. Addition of RRM2, however, results in modest binding; the band intensity gradually decreases as the protein concentration increases. Altogether, this data supports the notion that tandem RRM domains are required for binding, and that the interaction may be driven by RRM2.

Since the A-repeats fold independently of the rest of the XIST lncRNA [45], we focused on binding shift assays with the A-repeats. Similar to the FIRRE RRD, RBM15 RRM23 binds to the XIST A-repeats with higher affinity than RRM12 (**Figure 4B**). A decrease in RNA band intensity is observed when the concentration of RRM23 is between 1.2 and 2.4 μM, whereas no change is observed in EMSAs with RRM12. RRM23 binds the XIST A-repeats with a dissociation constant of 0.94 ± 0.21 μM. The individual RRMs, RRM1, RRM2, and RRM3, do not bind as well as the tandem RRM constructs; the intensity of the free RNA only modestly decreases as the concentration of protein increases. Similar to FIRRE, these results suggest that tandem RRMs are required for binding.

Considering the CLIP and chemical probing data supported the notion that stem loop-structured RNAs are bound by RBM15, we carried out binding shift assays with hairpin structured motifs of the FIRRE RRD and XIST A Repeat lncRNAs. We find that RRM23 binding occurs with the FIREE RRD stem loop RNA and the XIST A-repeat duplex with a binding affinity within the range of full length RNA binding (**Figure 5A,B**). A decrease in the free RNA band intensity is observed as early as 2 micromolar RBM15 RRM23 for both RNAs.

NMR and AlphaFold structural modeling reveal RBM15 RRMs 2 and 3 interact

The poor binding observed between the individual RRMs (in particular RRM2 and RRM3) with the RNAs suggest that RRMs 2 and 3 may interact, adopting a structural orientation that is required to form a stably structured RNA complex. Interestingly, RRMs 2 and 3 are separated by just four amino acids, an indication that they may share a structural interface. We used NMR to determine whether RRM 2 and 3 interact (**Figure 6A**). Overlaying ^{1}H ^{15}N HSQC spectra corresponding to RRM2 and RRM3 with an RRM23 construct show strong chemical shift perturbations for several of the amide resonances. These residues may correspond to the binding interface that exists between the two RRMs. Because no structure exists for RBM15, we used AlphaFold [46,47] to model the 3D structure. This revealed that a structural interface is predicted to form between RRM2 and RRM3 (**Figure 6B**), further supporting our conclusions from NMR analysis.

Discussion

It is widely accepted that lncRNAs function through their sequences and structures to facilitate the biological processes they regulate. While several genome-wide crosslinking studies have been carried out to identify

sites of protein binding, knowledge of the molecular mechanisms by which these interactions occur is lacking. In this study, we mined chemical probing data available from the RASP database to investigate the structure of nine lncRNAs, and identified RBM15 sites of binding throughout these transcripts based on iCLIP data from the POSTAR3 database. We then validated these results computationally, and experimentally validated our finds for two lncRNAs, XIST and FIRRE.

The structures predicted for the nine lncRNAs reveal that when the RNA localizes to different regions of the cell, the RNA can adopt a different structure. This is not unexpected given that most lncRNAs interact with proteins to facilitate function [3,48,49]. These proteins, which also localize to different regions of the cell, can modulate the structure adopted by the lncRNA. In support of this, comparing *in vitro* and *in vivo* folded RNAs reveals differential reactivity profiles, which is an indication of protein binding events [50].

We find that RBM15 binding sites occur in stem-loop structured regions across each of the lncRNAs investigated. These regions typically contain uracil residues, which is in agreement with what has been previously reported [38–41]. CatRAPID analysis was performed to identify amino acid and nucleotide residues involved in the binding interface. RBM15 interacts with RNA through its two terminal RRMs, RRM2 and RRM3. This was validated by EMSA experiments with the FIRRE and XIST lncRNAs. We find that, while the CatRAPID amino acid predictions are accurate (based on our binding shift assays), RNA nucleotides predicted to serve as binding sites for RBM15 are not wholly consistent with experimental data.

The functional role of the interactions between RBM15 and the majority of these lncRNAs has yet to be extensively explored; however, RBM15 is known to play a key role in the recruitment of Wilms' tumor 1-associating protein (WTAP) and N6-adenosine-methyltransferase catalytic subunit (METTL3) to RNA to facilitate the N6 methylation of adenosine [8]. These adenosines are nested in DRACH sequences (where D = G, A, or U, R = G or A, and H = A, C, or U). XIST is known to undergo m6A methylation, and possesses DRACH sequences throughout its primary sequence. However, not all adenosines nested within DRACH sequences are modified. RBM15, which has specific binding properties with RNA, is believed to regulate which DRACH sequences harbor adenosines that will be modified. This is accomplished through its affinity for U-rich sequences that may be proximal to m6A sites. Indeed, RBM15 crosslinks with U-containing sequences in each of the RNAs investigated in this study. Furthermore, each of the RNAs show m6A enrichment in their RNA sequences (**Figure 7**).

While further investigation is needed to fully comprehend how RBM15 is implicated in the regulatory roles of the lncRNAs that we focused on here, our results nevertheless provide structural insight about how RBM15 recognizes and binds RNA. This information is useful when considering, for example, how single nucleotide polymorphisms occurring in RBM15 binding sites may influence the role RBM15 has in biological function.

Experimental Procedures

Identifying lncRNAs and obtaining chemical probing data

The LNCipedia [51] was used to curate a list of all X-chromosome lncRNAs known to date. Genome-wide chemical probing datasets were downloaded from the RASP database [25]. An in-house generated script was used to identify lncRNAs possessing chemical probing data in the RASP database and is available at: https://github.com/gpwolfe/match_lncRNA_chromosome.

The Ensembl project (release 109) was used to identify the Homo sapiens Ensembl Canonical transcript for each of the nine RNAs. Chemical probing data for the canonical transcript was mined from the RASP

database. Using RNAStructure [52] the secondary structure of each RNA was predicted (using chemical probing data as soft constraints) and the lowest energy structure was visualized using RNAcanvas [53].

Identification of RNA-Protein Binding Partners

Potential protein binding partners for each lncRNA were identified with oRNAment [35]. To identify the binding sites of these potential protein binders, the computational tool CatRAPID was used. Putative RBPs were further validated against enhanced/individual-nucleotide crosslinking immunoprecipitation (eCLIP, iCLIP) data available from the POSTAR3 database [36].

***In vitro* transcription and purification of RNAs**

The XIST A-repeat and the FIRRE RRD RNAs were *in vitro* transcribed using in-house prepared T7 polymerase. Briefly, 50 µL transcriptions containing 600 ng dsDNA, 20-80 µM $MgCl_2$, 8 mM of each rNTP, 50% polyethylene glycol 8000, 1X transcription buffer (5 mM Tris pH 8, 5 mM spermidine, 10 mM dithiothreitol (DTT), and 0.6 mg of T7 polymerase were incubated at 37°C for 1 hour. Following transcription, the RNAs were purified on (6.5% XIST A-repeats, 8% FIRRE RRD, 20% for stem loop constructs) urea denaturing polyacrylamide gels, followed by extraction of the RNA from the gel using electro-elution. Electro-eluted RNAs were equilibrated against water and stored at -20°C until further use.

Protein expression and purification

DNA plasmid encoding each RBM15 construct was transformed into BL21 DE3 *E. coli* competent cells. Colonies were grown up in 1 L of lysogeny broth (or minimal media supplemented with $^{15}NH_4Cl$ for uniform labeling) at 37°C to an optical density (OD_{600}) of 0.9, and induced with 1 M isopropyl beta-D-1-thiogalactopyranoside (IPTG) at 18°C. Cells were lysed by sonication in a buffer containing 300 mM NaCl, 10 mM imidazole, 50 mM Tris, pH 8, and 5 mM beta-mercaptoethanol. Protein (possessing an N-terminal His_6-tag followed by a tobacco enterovirus (TEV) protease cleavage site) was purified by immobilized metal affinity chromatography (IMAC) against an increasing imidazole concentration gradient. The N-terminal His_6-tag was cleaved using in house-prepared TEV protease, followed by reverse IMAC. Protein was purified by size exclusion chromatography on a Superdex S75 purification column and eluted in a buffer containing 150 mM NaCl, 50 mM sodium phosphate pH 6.5, and 5 mM DTT. Fractions containing purified protein were evaluated by sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS PAGE) and stored at -80°C in buffer containing 150 mM NaCl, 25 mM sodium phosphate, pH 6.4, and 5 mM DTT until further use.

Electrophoretic mobility shift assays

RNAs were diluted to 250 nM in a buffer containing 125 mM NaCl and 25 mM sodium phosphate (pH 6.4) and incubated with increasing concentration of protein (as indicated in the figures). Reactions were supplemented with 30% glycerol prior to loading on a 1% TBE (0.1 M Tris-base, 0.1 M Boric Acid, 2 mM Ethylenediaminetetraacetic acid (EDTA)) agarose or 12% native PAGE gel stained with SYBR Safe or SYBR gold, respectively (Invitrogen). Gels were run at 60V for 45 minutes (agarose) or 110 V for 50 minutes (PAGE), followed by visualization on a GelDoc imager. Band quantification was carried out in ImageLab and K_ds were determined in Prism using the Hill equation: $y = \frac{B_{max} \times X^h}{{K_{half}}^h + X^h}$. Three replicates were collected for each binding assay.

Nuclear magnetic resonance spectroscopy

^{15}N-labeled RBM15 protein samples were concentrated to 0.3 mM in a buffer containing 150 mM NaCl, 25 mM sodium phosphate (pH 6.4), 5 mM DTT and 10% D_2O. 2D ^1H ^{15}N HSQC experiments were collected on each construct at 298 K on a cryo-equipped 800 MHz spectrometer. Spectra were processed in TopSpin.

Data Availability

All data is included in this manuscript.

Acknowledgments

The authors would like to thank members of the Jones lab for useful discussions and the Sattler lab for RBM15 plasmids.

Author Contributions

CM, EB, SX, YPOS, MS, LB, FL, JJ, TI, DM, DK, JW, VZ, LV, JVH, GA, KB, LZ, and ANJ carried out data mining, folding of RNA structures, and POSTAR3 and CatRAPID analysis. ANJ, EB, LE, CB, and ST prepared samples and carried out binding shift assays. ANJ carried out NMR experiments. GW wrote scripts for assessing data. ANJ wrote the manuscript and ANJ, CM, and EB revised the manuscript.

Funding and Additional Information

This work was funded by the National Science Foundation, 2243667 to A.N.J.

Conflicts of interest

A.N.J. is a consultant for NextRNA.

Figure Legends

Figure 1. RBM15, XIST, and FIRRE. A) RBM15 possesses three RRMs and a SPOC domain. RBM15 forms a complex with WTAP, METTL3, and METTL14 to facilitate m6A modification of RNA. B) The lncRNAs FIRRE and XIST escape transcriptional silencing and are proposed to interact with RBM15.

Figure 2. RBM15 CLIP binding sites with lncRNAs. A) Binding sequence motif logo for RBM15 based on iCLIP data [41]. B) iCLIP RBM15 crosslinking RNA binding sites for RBM15 with FIRRE, FTX, JPX, LINC00630, LINC00894, MIR503HG, XIST RepA, and ZNF674-AS1. Crosslinking sites are highlighted in green.

Figure 3. CatRAPID-predicted interaction domains for RBM15 withlncRNAs. CatRAPID Graphic was used for RNAs shorter than 1,200 nts, including FIRRE RRD, XIST A-repeats, INE1, MIR503HG, and LINC00894. Red indicates a higher propensity for an interaction. CatRAPID Fragments was used for RNAs longer than 1,200 nts, including LINC00630, FTX, JPX, and ZNF674-AS1. The RRM domains of RBM15 are indicated with green vertical lines.

Figure 4. The interaction of RBM15 RRMs with the FIRRE RRD and XIST A-repeat lncRNAs. A) EMSA of RBM15 RRM12, 23, 1, 2, and 3 with the FIRRE RRD domain. B) EMSA of RBM15 RRM12, 23, 1, 2, and 3 with the XIST A-repeats. The asterisk indicates a second conformation sampled by the XIST A-repeats.

Figure 5. The interaction of RBM15 RRM23 with stem loop structured elements from the A) FIRRE RRD and B) XIST A-repeat lncRNAs.

Figure 6. Structure of RBM15. A) ^{1}H ^{15}N HSQC overlay of RRM2 and RRM3 with RRM23. B) Alphafold structural model of RBM15 RRMs 123.

Figure 7. m6A enrichment mapped to studied lncRNAs. Data obtained from UCSC genome browser.

References

1. Ulitsky, I. & Bartel, D. P. lincRNAs: genomics, evolution, and mechanisms. *Cell* **154**, 26–46 (2013).
2. Mattick, J. S. *et al.* Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat. Rev. Mol. Cell Biol.* (2023) doi:10.1038/s41580-022-00566-8.
3. Statello, L., Guo, C.-J., Chen, L.-L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**, 96–118 (2021).
4. Han, P. & Chang, C.-P. Long non-coding RNA and chromatin remodeling. *RNA Biol.* **12**, 1094–1098 (2015).
5. Sebastian-delaCruz, M., Gonzalez-Moro, I., Olazagoitia-Garmendia, A., Castellanos-Rubio, A. & Santin, I. The Role of lncRNAs in Gene Expression Regulation through mRNA Stabilization. *Noncoding RNA* **7**, (2021).
6. Briata, P. & Gherzi, R. Long Non-Coding RNA-Ribonucleoprotein Networks in the Post-Transcriptional Control of Gene Expression. *Non-Coding RNA* **6**, 40 (2020).
7. Chillón, I. & Marcia, M. The molecular structure of long non-coding RNAs: emerging patterns and functional implications. *Crit. Rev. Biochem. Mol. Biol.* **55**, 662–690 (2020).
8. Patil, D. P. *et al.* m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537**, 369–373 (2016).
9. Lindtner, S. *et al.* RNA-binding motif protein 15 binds to the RNA transport element RTE and provides a direct link to the NXF1 export pathway. *J. Biol. Chem.* **281**, 36915–36928 (2006).
10. Zolotukhin, A. S. *et al.* Nuclear export factor RBM15 facilitates the access of DBP5 to mRNA. *Nucleic Acids Res.* **37**, 7151–7162 (2009).
11. Graindorge, A. *et al.* In-cell identification and measurement of RNA-protein interactions. *Nat. Commun.* **10**, 5317 (2019).
12. Berletch, J. B., Yang, F., Xu, J., Carrel, L. & Disteche, C. M. Genes that escape from X inactivation. *Hum. Genet.* **130**, 237–245 (2011).

13. Brockdorff, N. *et al.* The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **71**, 515–526 (1992).

14. 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).

15. Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404–416 (2015).

16. Jones, A. N. & Sattler, M. Challenges and perspectives for structural biology of lncRNAs-the example of the Xist lncRNA A-repeats. *J. Mol. Cell Biol.* **11**, 845–859 (2019).

17. Jones, A. N., Gabel, F., Bohn, S., Wolfe, G. & Sattler, M. Cotranscriptional folding of the lncRNA Xist A-repeats indicates a modular structure. *bioRxiv* (2022) doi:10.1101/2022.07.26.501616.

18. Hacisuleyman, E., Shukla, C. J., Weiner, C. L. & Rinn, J. L. Function and evolution of local repeats in the Firre locus. *Nat. Commun.* **7**, 11021 (2016).

19. Hacisuleyman, E. *et al.* Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat. Struct. Mol. Biol.* **21**, 198–206 (2014).

20. Lewandowski, J. P. *et al.* The Firre locus produces a trans-acting RNA molecule that functions in hematopoiesis. *Nat. Commun.* **10**, 5137 (2019).

21. Sun, L. *et al.* Long noncoding RNAs regulate adipogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 3387–3392 (2013).

22. Yang, F. *et al.* The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biol.* **16**, 52 (2015).

23. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

24. Luo, Y. *et al.* New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.* **48**, D882–D889 (2020).

25. Li, P., Zhou, X., Xu, K. & Zhang, Q. C. RASP: an atlas of transcriptome-wide RNA secondary structure probing data. *Nucleic Acids Res.* **49**, D183–D191 (2021).

26. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505**, 701–705 (2014).

27. Wan, Y. *et al.* Landscape and variation of RNA secondary structure across the human transcriptome.

Nature **505**, 706–709 (2014).

28. Wu, X. & Bartel, D. P. Widespread Influence of 3'-End Structures on Mammalian mRNA Processing and Stability. *Cell* **169**, 905–917.e11 (2017).
29. Lu, Z. *et al.* RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* **165**, 1267–1279 (2016).
30. Sun, L. *et al.* RNA structure maps across mammalian cellular compartments. *Nat. Struct. Mol. Biol.* **26**, 322–330 (2019).
31. Guo, J. U. & Bartel, D. P. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **353**, (2016).
32. Weng, X. *et al.* Keth-seq for transcriptome-wide RNA structure mapping. *Nat. Chem. Biol.* **16**, 489–492 (2020).
33. Zubradt, M. *et al.* DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods* **14**, 75–82 (2017).
34. Liu, F. *et al.* Long noncoding RNA FTX inhibits hepatocellular carcinoma proliferation and metastasis by binding MCM2 and miR-374a. *Oncogene* **35**, 5422–5434 (2016).
35. Benoit Bouvrette, L. P., Bovaird, S., Blanchette, M. & Lécuyer, E. oRNAment: a database of putative RNA binding protein target sites in the transcriptomes of model species. *Nucleic Acids Res.* **48**, D166–D173 (2020).
36. Zhao, W. *et al.* POSTAR3: an updated platform for exploring post-transcriptional regulation coordinated by RNA-binding proteins. *Nucleic Acids Res.* **50**, D287–D294 (2022).
37. Huppertz, I. *et al.* iCLIP: protein-RNA interactions at nucleotide resolution. *Methods* **65**, 274–287 (2014).
38. Ma, H., Wen, H., Xue, Z., Li, G. & Zhang, Z. RNANetMotif: Identifying sequence-structure RNA network motifs in RNA-protein binding sites. *PLoS Comput. Biol.* **18**, e1010293 (2022).
39. Van Nostrand, E. L. *et al.* A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 (2020).
40. Dominguez, D. *et al.* Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Mol. Cell* **70**, 854–867.e9 (2018).
41. Feng, H. *et al.* Modeling RNA-Binding Protein Specificity In Vivo by Precisely Registering Protein-RNA

Crosslink Sites. *Mol. Cell* **74**, 1189–1204.e6 (2019).

42. Bellucci, M., Agostini, F., Masin, M. & Tartaglia, G. G. Predicting protein associations with long noncoding RNAs. *Nat. Methods* **8**, 444–445 (2011).
43. Livi, C. M., Klus, P., Delli Ponti, R. & Tartaglia, G. G. catRAPID signature: identification of ribonucleoproteins and RNA-binding regions. *Bioinformatics* **32**, 773–775 (2016).
44. Armaos, A., Colantoni, A., Proietti, G., Rupert, J. & Tartaglia, G. G. catRAPID omics v2.0: going deeper and wider in the prediction of protein-RNA interactions. *Nucleic Acids Res.* **49**, W72–W79 (2021).
45. Fang, R., Moss, W. N., Rutenberg-Schoenberg, M. & Simon, M. D. Probing Xist RNA Structure in Cells Using Targeted Structure-Seq. *PLoS Genet.* **11**, e1005668 (2015).
46. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
47. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
48. Bridges, M. C., Daulagala, A. C. & Kourtidis, A. LNCcation: lncRNA localization and function. *J. Cell Biol.* **220**, (2021).
49. Guo, C.-J. *et al.* Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. *Cell* **181**, 621–636.e22 (2020).
50. 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. U. S. A.* **113**, 10322–10327 (2016).
51. Volders, P.-J. *et al.* LNCipedia 5: towards a reference set of human long non-coding RNAs. *Nucleic Acids Res.* **47**, D135–D139 (2019).
52. Reuter, J. S. & Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 1–9 (2010).
53. Johnson, P. Z. & Simon, A. E. RNAcanvas: interactive drawing and exploration of nucleic acid structures. *Nucleic Acids Res.* **51**, W501–W508 (2023).
54. Chureau, C. *et al.* Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. *Hum. Mol. Genet.* **20**, 705–718 (2011).
55. Li, J. *et al.* Identification of potential autophagy-associated lncRNA in prostate cancer. *Aging* **13**, 13153–13165 (2021).

56. Sun, X., Zhang, J. & Cao, C. CTCF and Its Partners: Shaper of 3D Genome during Development. *Genes* **13**, (2022).

57. Tian, D., Sun, S. & Lee, J. T. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell* **143**, 390–403 (2010).

58. Sun, S. *et al.* Jpx RNA activates Xist by evicting CTCF. *Cell* **153**, 1537–1551 (2013).

59. Long, J., Ou, C., Xia, H., Zhu, Y. & Liu, D. MiR-503 inhibited cell proliferation of human breast cancer cells by suppressing CCND1 expression. *Tumour Biol.* **36**, 8697–8702 (2015).

60. Han, X., Li, B. & Zhang, S. MIR503HG: A potential diagnostic and therapeutic target in human diseases. *Biomed. Pharmacother.* **160**, 114314 (2023).

61. Yang, Y. *et al.* MiR-503 targets PI3K p85 and IKK-β and suppresses progression of non-small cell lung cancer. *Int. J. Cancer* **135**, 1531–1542 (2014).

62. Li, D. *et al.* LncRNA ZNF674-AS1 regulates granulosa cell glycolysis and proliferation by interacting with ALDOA. *Cell Death Discov* **7**, 107 (2021).

63. Li, D., Xie, Y., Sun, J., Zhang, L. & Jiang, W. LncRNA ZNF674-AS1 Hinders Proliferation and Invasion of Hepatic Carcinoma Cells through the Glycolysis Pathway. *J. Oncol.* **2022**, 8063382 (2022).

64. Mao, G., Jin, H. & Wu, L. DDX23-Linc00630-HDAC1 axis activates the Notch pathway to promote metastasis. *Oncotarget* **8**, 38937–38949 (2017).

65. Liu, F. *et al.* Long noncoding RNA LINC00630 promotes radio-resistance by regulating BEX1 gene methylation in colorectal cancer cells. *IUBMB Life* **72**, 1404–1414 (2020).

66. Zhou, Q. *et al.* A novel lncRNA-miRNA-mRNA competing endogenous RNA regulatory network in lung adenocarcinoma and kidney renal papillary cell carcinoma. *Thorac Cancer* **12**, 2526–2536 (2021).

67. Chen, B., Liu, D., Chen, R., Guo, L. & Ran, J. Elevated LINC00894 relieves the oncogenic properties of thyroid cancer cell by sponging let-7e-5p to promote TIA-1 expression. *Discov Oncol* **13**, 56 (2022).

68. Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131–137 (1996).

69. Galupa, R. & Heard, E. X-Chromosome Inactivation: A Crossroads Between Chromosome Architecture and Gene Regulation. *Annu. Rev. Genet.* **52**, 535–566 (2018).

Figures and Tables
Characterization of RBM15 protein binding with long noncoding RNAs

Emma Bose[$,1], Caleb Mayes[$,1], Lance Ellis[#,1], Corrine Baker[#,1], Sofia Tambalotti[#,1], Shengwei Xiong[#,1], Yaa Pokua Osei Sarpong[#,1], Marwan Shalaby[#,1], Lucas Barry[#,1], Frank Lewis[#,1], Johnson Joseph[#,1], Talaidh Isaacs[#,1], Derik McCarthy[#,1], Dana Katz[#,1], Jingyang Wang[#,1], Victoria Zirimu[#,1], Luis Vargas[#,1], Julian Von Hofe[#,1], Glen Aguilar[#,1], Katherine Buchan[#,1], Lei Zheng[#,1], Gregory Wolfe[1,2], and Alisha N Jones*[,1]

[$,#]These authors contributed equally
*Corresponding author: alisha.jones@nyu.edu

[1]Department of Chemistry, New York University, New York, New York, USA
[2]Department of Physics, New York University, New York, New York, USA

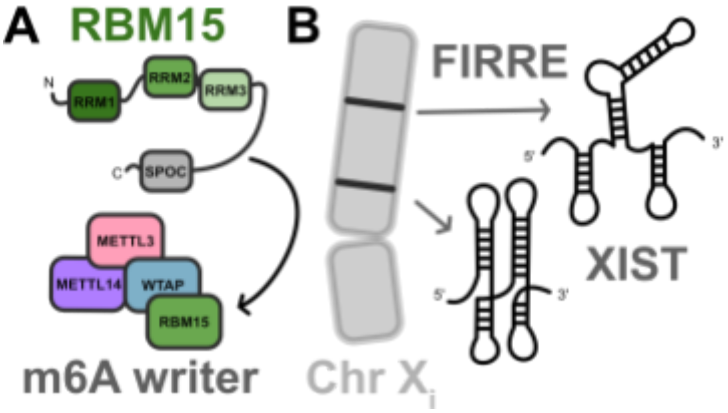

Figure 1. RBM15, XIST, and FIRRE. A) RBM15 possesses three RRMs and a SPOC domain. RBM15 forms a complex with WTAP, METTL3, and METTL14 to facilitate m6A modification of RNA. B) The lncRNAs FIRRE and XIST escape transcriptional silencing and are proposed to interact with RBM15.

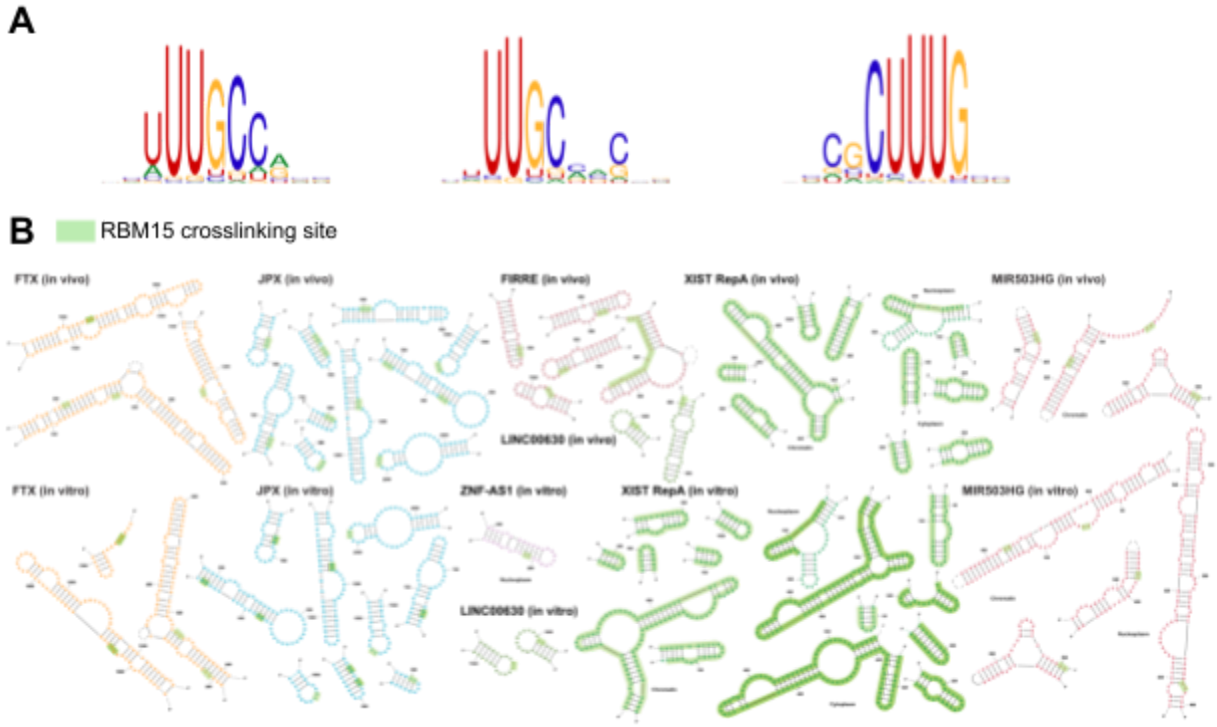

Figure 2. RBM15 CLIP binding sites with lncRNAs. A) Binding sequence motif logo for RBM15 based on iCLIP data [41]. B) iCLIP RBM15 crosslinking RNA binding sites for RBM15 with FIRRE, FTX, JPX, LINC00630, LINC00894, MIR503HG, XIST RepA, and ZNF674-AS1. Crosslinking sites are highlighted in green.

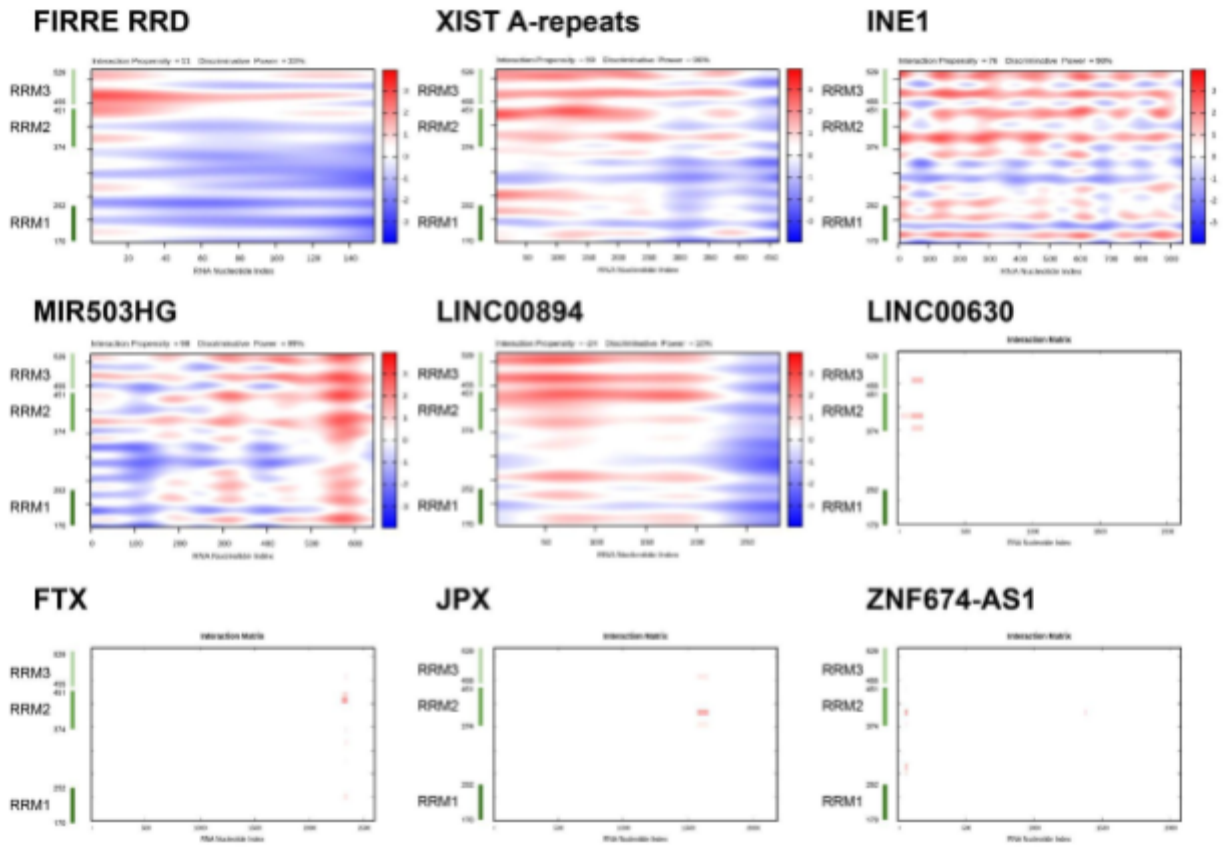

Figure 3. CatRAPID-predicted interaction domains for RBM15 withlncRNAs. CatRAPID Graphic was used for RNAs shorter than 1,200 nts, including FIRRE RRD, XIST A-repeats, INE1, MIR503HG, and LINC00894. Red indicates a higher propensity for an interaction. CatRAPID Fragments was used for RNAs longer than 1,200 nts, including LINC00630, FTX, JPX, and ZNF674-AS1. The RRM domains of RBM15 are indicated with green vertical lines.

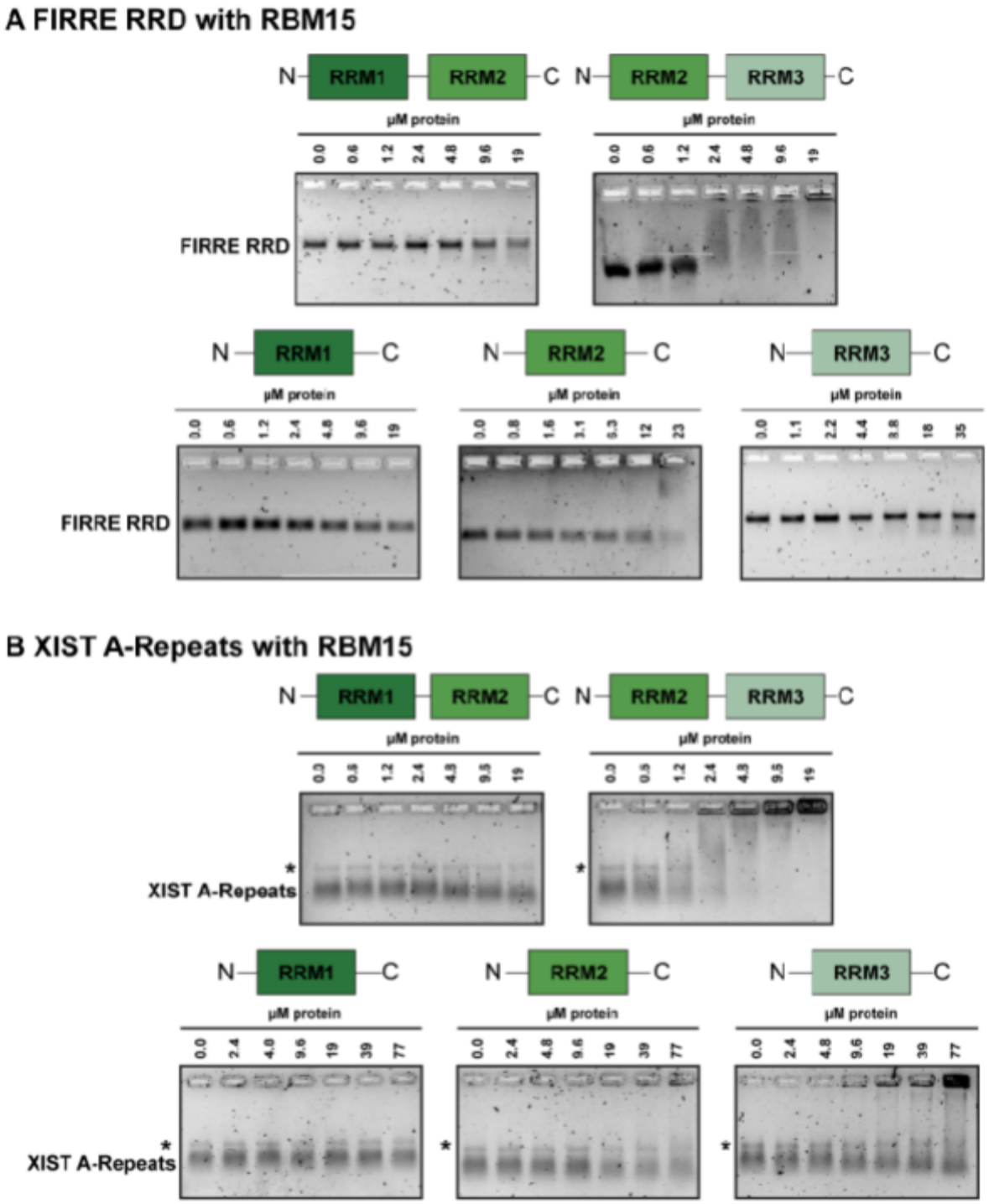

Figure 4. The interaction of RBM15 RRMs with the FIRRE RRD and XIST A-repeat lncRNAs. A) EMSA of RBM15 RRM12, 23, 1, 2, and 3 with the FIRRE RRD domain. B) EMSA of RBM15 RRM12, 23, 1, 2, and 3 with the XIST A-repeats. The asterisk indicates a second conformation sampled by the XIST A-repeats.

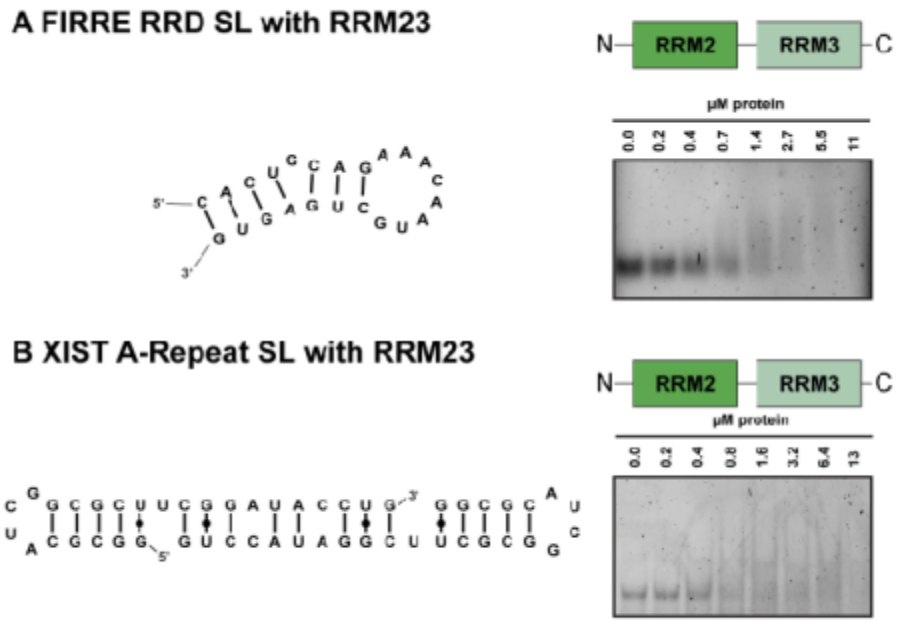

Figure 5. The interaction of RBM15 RRM23 with stem loop structured elements from the A) FIRRE RRD and B) XIST A-repeat lncRNAs.

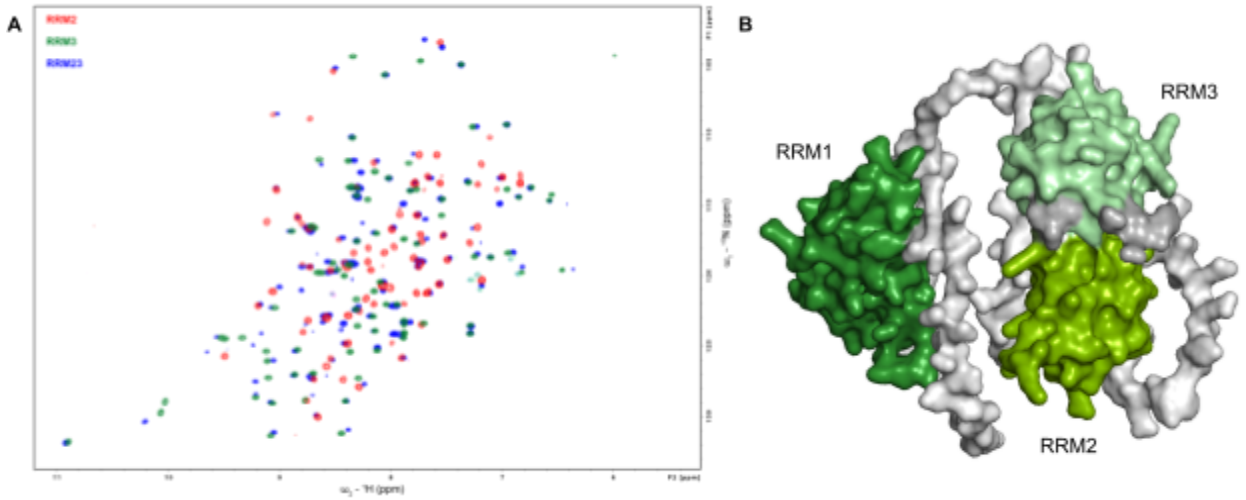

Figure 6. Structure of RBM15. A) ^{1}H ^{15}N HSQC overlay of RRM2 and RRM3 with RRM23. B) Alphafold structural model of RBM15 RRMs 123.

Figure 7. m6A enrichment mapped to studied lncRNAs. Data obtained from UCSC genome browser.

Tables

Table 1. Functional and experimental details regarding the long noncoding RNAs used in this study.

LncRNA, RNA length, transcript ID, locus	Function	Probing conditions and coverage
FTX (-) 2597 nucleotides ENST00000429124.5	Positive regulator of XIST 54	*NAI-N3 icSHAPE in HEK293* [30] Chromatin *in vitro*: 0.6916 Chromatin *in vivo*: 0.6916 *DMS-MaPseq in HEK293T* [33] *In vivo*: 0.0111
INE1 (+) 945 nucleotides ENST00000456273.1	Continued expression is associated with prostate cancer, likely playing a role in autophagy regulation 55	*NAI-N3 icSHAPE in HEK293* [30] Chromatin *in vitro*: 0.7704 Chromatin *in vivo* 0.7704 *DMS-MaPseq in HEK293T* [33] *In vivo*: 0.1143
JPX (+) 2194 nucleotides ENST00000414209.5	Activation of the XIST lncRNA Aids in chromatin loop organization 56,57,58	*NAI-N3 icSHAPE in HEK293* [30] Chromatin *in vitro*: 0.6778 Chromatin *in vivo*: 0.6778 Cytoplasm *in vitro*: 0.1855
MIR503HG (-) 646 nucleotides ENST00000457876.5	Functions as the host gene of miR-503, mediating its expression Affects cell proliferation, invasion, metastasis, apoptosis, and angiogenesis 59,60,61	*NAI-N3 icSHAPE in HEK293* [30] Chromatin *in vitro*; 0.0991 Chromatin *in vivo*: 0.0991 Nucleoplasm *in vitro*: 0.1084 *DMS, DIM-2P-seq in HEK293T* [28] *In vivo*: 0.1130
ZNF674-AS1 (+) 2078 nucleotides ENST00000421685.2	Inhibits cell migration and proliferation Inhibits proteins in glycolysis 62,63	*NAI-N3 icSHAPE in HEK293* [30] Cytoplasm *in vitro*: 0.0284 Nucleoplasm *in vitro*: 0.2820 *DMS, DMS-MaPseq in HEK293T* [33] *In vivo*: 0.0823
LINC00630 (+) 2101 nucleotides	Promotes metastasis and the progression of non-small cell lung cancer	*NAI-N3, icSHAPE in HEK293* [30] Chromatin *in vitro*: 0.1142

ENST00000440496.5	Associated with radioresistance of colorectal cancer [64,65]	*DMS, DMS-MaPseq in HEK293T* [33] *In vivo*: 0.0609
LINC00894 (+) 284 nucleotides ENST00000455777.1	Promotes the progression of breast cancer through cell proliferation and invasion and is upregulated in kidney cancer [66,67]	*NAI-N3, icSHAPE in HEK293* [30] Chromatin *in vitro*: 0.2042
FIRRE (-) 2928 nucleotides ENST00000427391.1	Chromosomal organization Positioning of X_i to the perinucleolar region via CCCTC binding factor (CTCF) Pluripotency and adipogenesis [18,19,22]	*DMS, DMS-MaPseq in HEK293T* [33] *In vivo*: 0.0888
XIST RepA (-) 1260 nucleotides ENST00000429829.5	Transcriptional silencing of the inactive X-chromosome [68,69]	*NAI-N3, icSHAPE in HEK293* [30] Chromatin *in vitro*: 1.0 Chromatin *in vivo*: 1.0 Cytoplasm *in vitro*: 1.0 Cytoplasm *in vivo*: 1.0 Nucleoplasm *in vitro*: 1.0 Nucleoplasm *in vivo*: 1.0 *NAI-N3, icSHAPE in HEK293* [29] *In vitro*: 1.0 *In vivo*: 1.0

Abbreviations: Human Embryonic Kidney 293 (*HEK293*), *In vivo* click selective 2-hydroxyl acylation and profiling experiment (*icSHAPE*), 2-(Azidomethyl)nicotinic acid imidazolide (*NAI-N3*), Dimethyl Sulfate Mutational Profiling Sequencing (*DMS-MaPseq*), DMS-induced mutations mapped by 2P-seq (*DIM-2P-seq)*

(+) corresponds to transcripts from the positive strand

(-) corresponds to transcripts from the negative strand