**Characterization of the RBM15 protein binding with XCI-escaping long noncoding RNAs**

Caleb Mayes[1], Emma Bose[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

Long noncoding ribonucleic acids (lncRNAs) are transcripts greater than 500 nucleotides in length that do not encode for protein. Through their sequences and structures, lncRNAs interact with other nucleic acids and proteins to regulate a variety of biological processes, including gene expression. The rate at which lncRNAs have been identified exceeds our ability to study their structures and functions. Within the last decade, several tens of thousands have been discovered throughout the human genome, yet fewer than 50 have been structurally characterized. In an attempt to better understand and map the vast molecular networks that underlie biological activity, several genome-wide studies have been carried out to study RNA structures, identify their cognate protein binding partners, and map sites of protein binding. Every class of RNA is represented in these studies, including lncRNAs. In this study, we mined this data to investigate the propensity of X-inactivation (XCI) escaping lncRNAs to interact with the RNA binding protein (RBP) RBM15. We find that XCI-escaping lncRNAs are intricately structured and, depending on their localization (cytoplasm, nucleoplasm, chromatin), adopt different structures. Bioinformatic and experimental analyses reveal that U-containing sequences are bound by RNA recognition motifs (RRMs) 2 and 3 of RBM15. Collectively, the mining of data, carried out by graduate students of the New York University Spring 2023 Macromolecular Chemistry course, and the experimental validation of those results, carried out by high school summer interns from traditionally excluded groups, demonstrates i) an efficient way to make use of the genome-wide studies that have been carried out to better understand how lncRNAs function and ii) a way to engage students from a variety of backgrounds in biological research.

**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].

The rate at which lncRNAs are identified outmatches our ability to structurally and functionally characterize them. Nearly 170,000 lncRNAs have been identified in humans alone [1,8], but fewer than 50 have been structurally investigated to better understand how they interact with RBPs. Examples of studied lncRNAs include XIST, Cyrano, and Braveheart, which are each bound by several RBPs to regulate their functions [9–12]. In the absence of specific studies, there have been several genome-wide chemical probing and crosslinking experiments which encompass lncRNAs, to investigate i) the structures that RNAs adopt, and ii) the proteins that they bind, respectively [13–16]. While this provides valuable insight about how these RNAs function, experimental validation is necessary to understand the molecular mechanisms that underlie these interactions.

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. Through its RRMs and SPOC domain, RBM15 has been identified as a key player in post-transcriptional regulation (e.g., through m6A modification [17] and nuclear export [18,19]) and also serves as a regulatory protein in the process of hematopoietic cell homeostasis [17]. 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) [20].

During X-chromosome inactivation (XCI), approximately 94% of genes of the inactive X-chromosome are transcriptionally silenced by the lncRNA XIST [21]. 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 [22]. The A-repeats, a 450 nucleotides (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 [23]. 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 [17,24–26].

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) [27]. 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 [28–31]. A more recently identified RBP of the lncRNA FIRRE is RBM15 [20]. Binding of this RBP has been localized to a 156-nt LR known as the Repeating RNA Binding Domain (RRD) through crosslinking experiments [32,33].

The structural mechanism underlying how RBM15 interacts with lncRNAs, including XIST and FIRRE, has yet to be established. In this study, we sought to understand how RBM15 interacts with structured lncRNAs. First, we mined chemical probing data available from the RNA Atlas of Structure Probing (RASP) [34] to fold the structures of nine XCI-escaping 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 binds to U-containing structured domains in lncRNAs predominately through two of its three RRMs: RRM2 and RRM3.

## Results

### XCI-escaping RNAs are intricately structured

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 [35–42]. We selected nine XCI-escaping 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**.

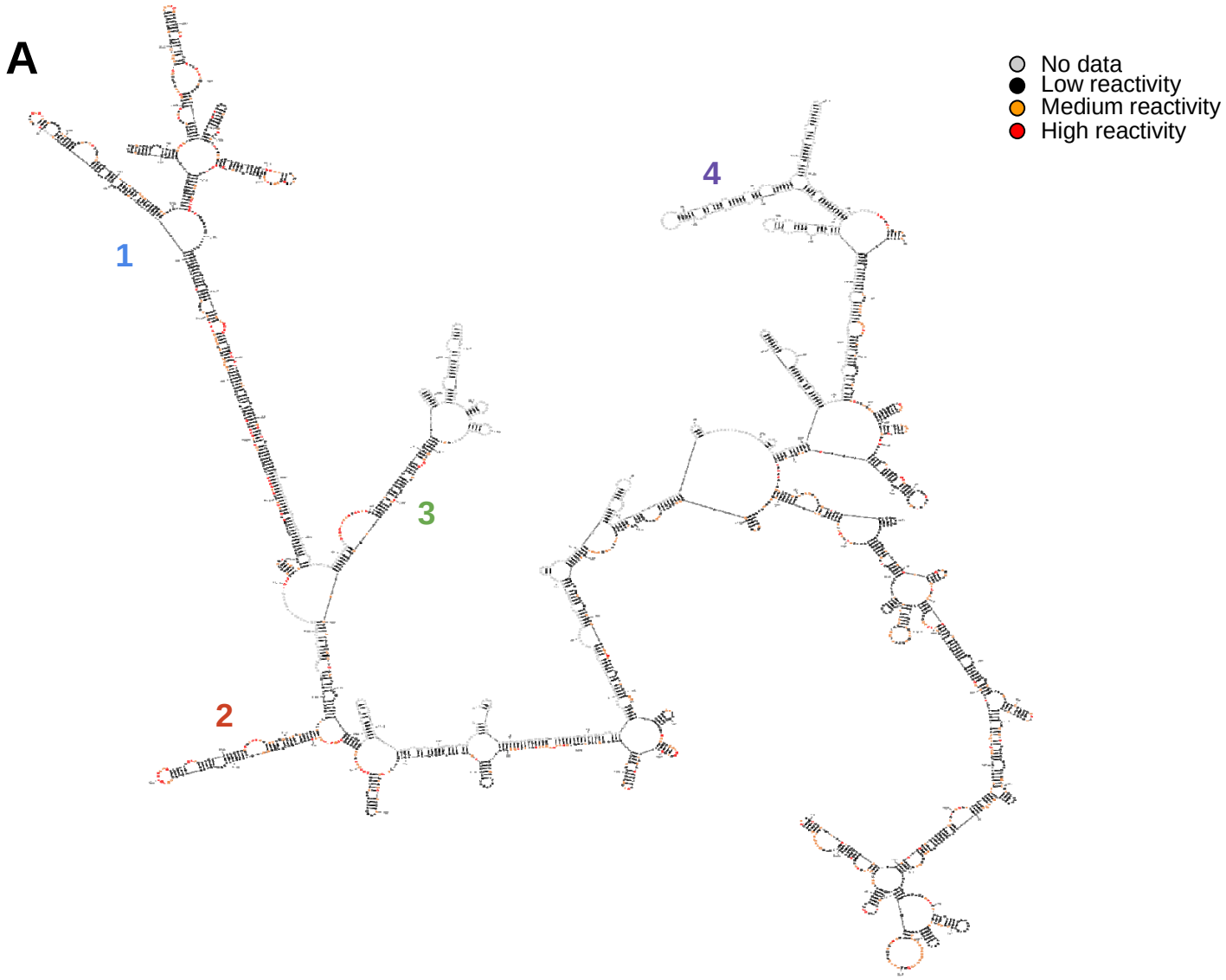
**Table 1. Functional and experimental details regarding the long noncoding RNAs used in this study.**

| LncRNA, RNA length, transcript ID | Function | Probing conditions and coverage |
|---|---|---|
| FTX (-)<br>2597 nucleotides<br>ENST00000429124.5 | Positive regulator of XIST<br>[43] | *NAI-N3 icSHAPE in HEK293* [39]<br>Chromatin *in vitro*: 0.6916<br>Chromatin *in vivo*: 0.6916<br><br>*DMS-MaPseq in HEK293T* [42]<br>*In vivo*: 0.0111 |
| INE1 (+)<br>945 nucleotides<br>ENST00000456273.1 | Continued expression is associated with prostate cancer, likely playing a role in autophagy regulation<br>[44] | *NAI-N3 icSHAPE in HEK293* [39]<br>Chromatin *in vitro*: 0.7704<br>Chromatin *in vivo* 0.7704<br><br>*DMS-MaPseq in HEK293T* [42]<br>*In vivo*: 0.1143 |
| JPX (+)<br>2194 nucleotides<br>ENST00000414209.5 | Activation of the XIST lncRNA<br><br>Aids in chromatin loop organization<br>[45,46,47] | *NAI-N3 icSHAPE in HEK293* [39]<br>Chromatin *in vitro*: 0.6778<br>Chromatin *in vivo*: 0.6778<br>Cytoplasm *in vitro*: 0.1855 |
| MIR503HG (-)<br>646 nucleotides<br>ENST00000457876.5 | Functions as the host gene of miR-503, mediating its expression<br><br>Affects cell proliferation, invasion, metastasis, apoptosis, and angiogenesis<br>[48,49,50] | *NAI-N3 icSHAPE in HEK293* [39]<br>Chromatin *in vitro*; 0.0991<br>Chromatin *in vivo*: 0.0991<br>Nucleoplasm *in vitro*: 0.1084<br><br>*DMS, DIM-2P-seq in HEK293T* [37]<br>*In vivo*: 0.1130 |
| ZNF674-AS1 (+)<br>2078 nucleotides<br>ENST00000421685.2 | Inhibits cell migration and proliferation<br><br>Inhibits proteins in glycolysis<br>[51,52] | *NAI-N3 icSHAPE in HEK293* [39]<br>Cytoplasm *in vitro*: 0.0284<br>Nucleoplasm *in vitro*: 0.2820<br><br>*DMS, DMS-MaPseq in HEK293T* [42]<br>*In vivo*: 0.0823 |
| LINC00630 (+)<br>2101 nucleotides<br>ENST00000440496.5 | Promotes metastasis and the progression of non-small cell lung cancer<br><br>Associated with radioresistance of colorectal cancer<br>[53,54] | *NAI-N3, icSHAPE in HEK293* [39]<br>Chromatin *in vitro*: 0.1142<br><br>*DMS, DMS-MaPseq in HEK293T* [42]<br>*In vivo*: 0.0609 |

| | | |
|---|---|---|
| LINC00894 (+)<br>284 nucleotides<br>ENST00000455777.1 | Promotes the progression of breast cancer through cell proliferation and invasion and is upregulated in kidney cancer<br><br>[55,56] | *NAI-N3, icSHAPE in HEK293* [39]<br>Chromatin *in vitro*: 0.2042 |
| FIRRE (-)<br>2928 nucleotides<br>ENST00000427391.1 | Chromosomal organization<br><br>Positioning of $X_i$ to the perinucleolar region via CCCTC binding factor (CTCF)<br><br>Pluripotency and adipogenesis<br><br>[27,28,31] | *DMS, DMS-MaPseq in HEK293T* [42]<br>*In vivo*: 0.0888 |
| XIST RepA (-)<br>1260 nucleotides<br>ENST00000429829.5 | Transcriptional silencing of the inactive X-chromosome<br><br>[57,58] | *NAI-N3, icSHAPE in HEK293* [39]<br>Chromatin *in vitro*: 1.0<br>Chromatin *in vivo*: 1.0<br>Cytoplasm *in vitro*: 1.0<br>Cytoplasm *in vivo*: 1.0<br>Nucleoplasm *in vitro*: 1.0<br>Nucleoplasm *in vivo*: 1.0<br><br>*NAI-N3, icSHAPE in HEK293* [38]<br>*In vitro*: 1.0<br>*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

***FTX*:** The lncRNA five prime to XIST was folded based on NAI-N3 SHAPE experiments. Data was collected under two conditions: chromatin *in vitro* and chromatin *in vivo* (**Figure 1A, B**). Overall, FTX is intricately structured, possessing several elongated helical stems. While there are four structured domains predicted to fold both *in vitro* and *in vivo*, nearly half of the RNA folds differently between the two conditions. This is likely due to protein or RNA binding events occurring in the cell [59]. Indeed, several proteins have been predicted to bind FTX based on o RNA motifs enrichment in complete transcriptomes (oRNAment) analysis [60] (**Supplementary Figure 1**). In cells, these interactions can block the chemical probe from modifying the RNA, resulting in differential reactivity profiles when compared to *in vitro* chemical probing data.


A
No data
Low reactivity
Medium reactivity
High reactivity
1
2
3
4

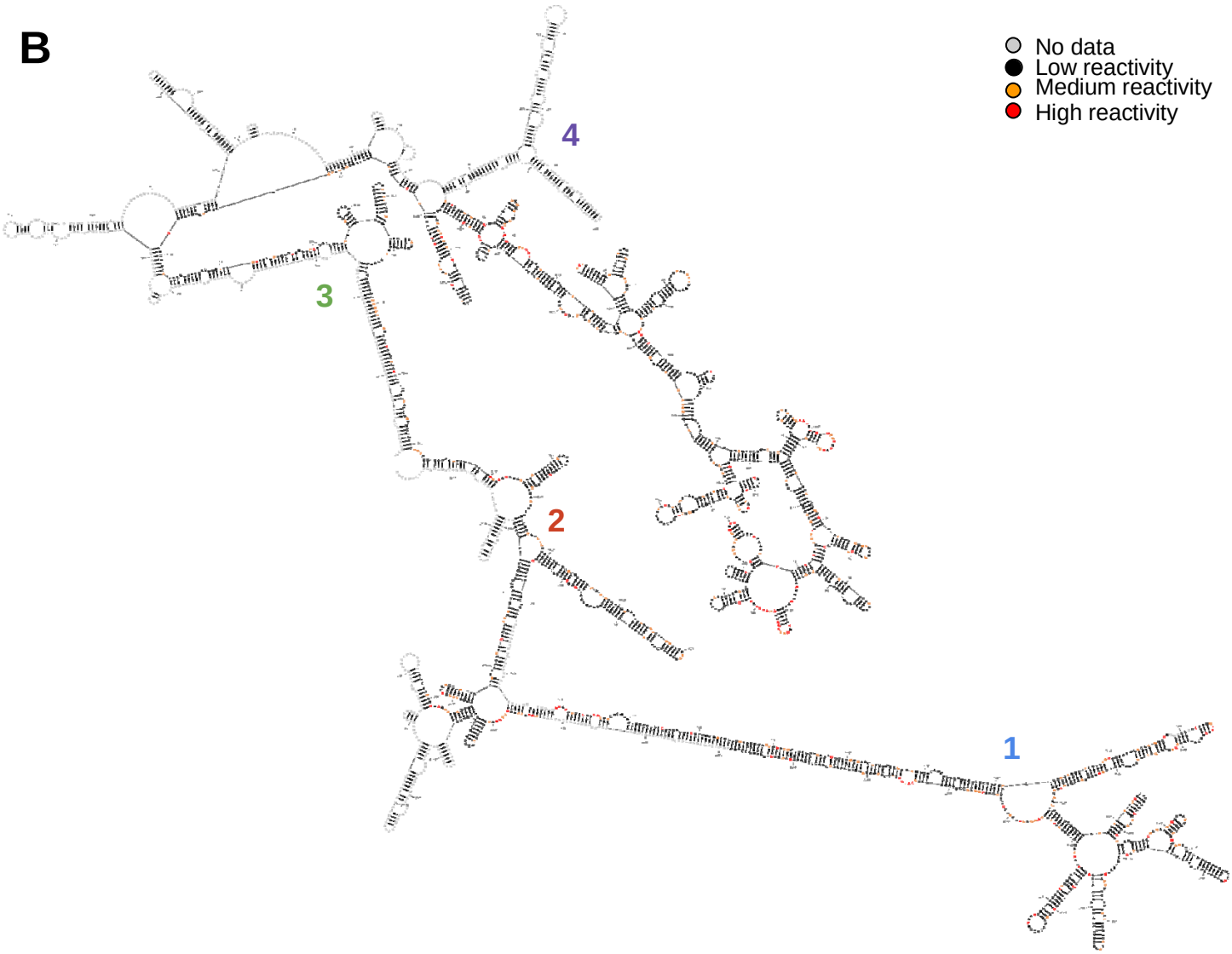



**Figure 1.** Secondary structure of the lncRNA FTX based on NAI-N3 SHAPE chemical probing data. Four domains (1-4) are similar between the A) FTX chromatin *in vitro* and B) FTX chromatin *in vivo* predicted structures. Chemical probing data is mapped to each structure (gray = no data, black = low reactivity, orange = medium reactivity, and red = high reactivity).

***INE1:*** The lowest energy structures of the INE1 lncRNA in chromatin *in vitro* and *in vivo* in HEK293 cells reveal that this lncRNA is intricately structured. DMS probing reveals that the lncRNA adopts several shared structured motifs in cells and when probed following extraction of the RNA from its cellular environment (**Figure 2A, B**). Similar to FTX, differences between the two structures may be attributed to cellular conditions, such as interactions with proteins like embryonic lethal, abnormal vision-like 1 (ELAVL1), and heterogeneous nuclear ribonucleoprotein H2 (HNRNPH2) (**Supplementary Figure 1**).

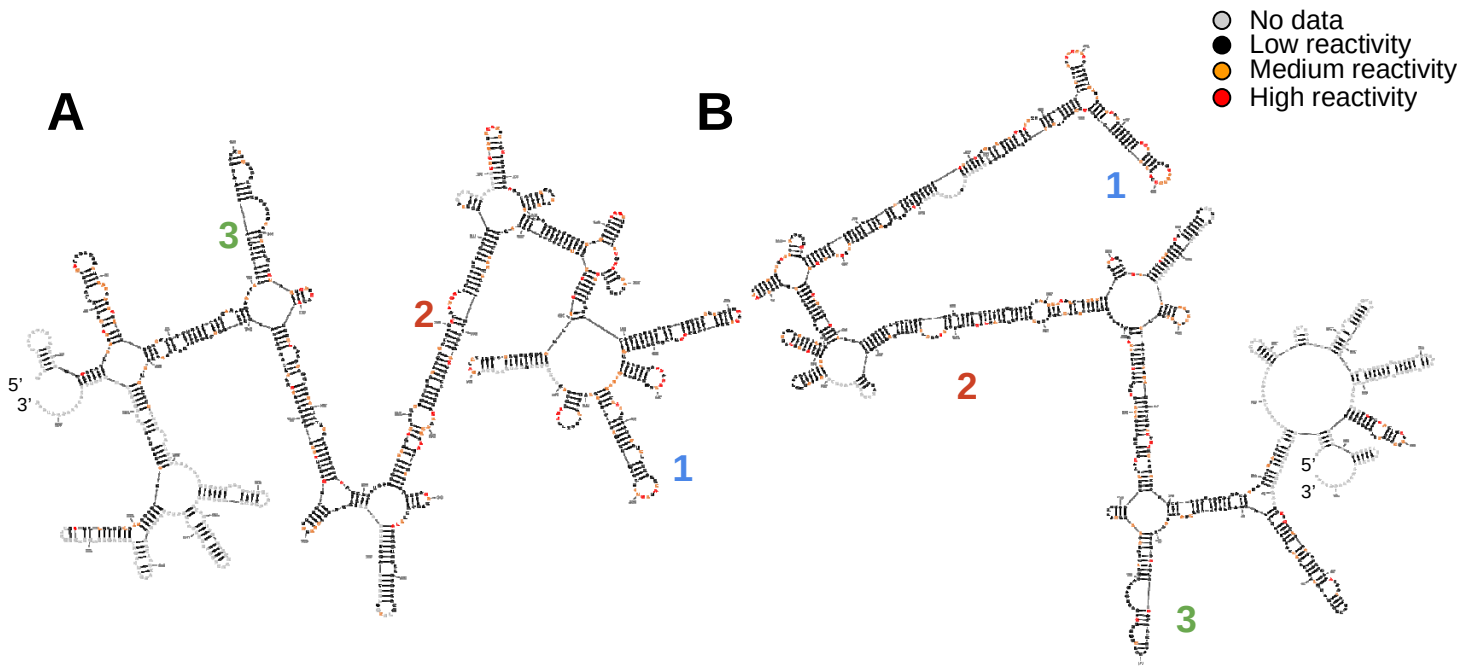



**Figure 2.** Secondary structure of the lncRNA INE1 based on NAI-N3 SHAPE probing data. Three domains (1-3) are similar between the A) INE1 chromatin *in vitro* and B) INE1 chromatin *in vivo* predicted structures. Chemical probing data is mapped to each structure (gray = no data, black = low reactivity, orange = medium reactivity, and red = high reactivity).

***JPX:*** The lncRNA JPX localizes to both chromatin and the cytoplasm. NAI-N3 SHAPE chemical probing data from chromatin *in vitro* and *in vivo*, and from the cytoplasm *in vitro* reveal structurally identical folds for JPX (**Figure 3**). The secondary structure of the first 343 nucleotides of JPX was recently reported; in this study, the *in vivo* and *in vitro* structures (probed in HEK293 cells with the NAI SHAPE probe) are 70% similar [61]. JPX interacts with the CCCTC-binding factor protein (CTCF) [47]; this interaction is attributed, in part, to the structural differences observed between the two states. We compared the secondary structure predicted for JPX 1 - 343 with the structure predicted here and found several structural similarities. The Y-shaped domain at the 5' end, along with several terminal loops, are in agreement (**Figure 3**, blue asterisks). Folding the full-length transcript reveals that the first 343 nucleotides of JPX can base pair with downstream nucleotides.

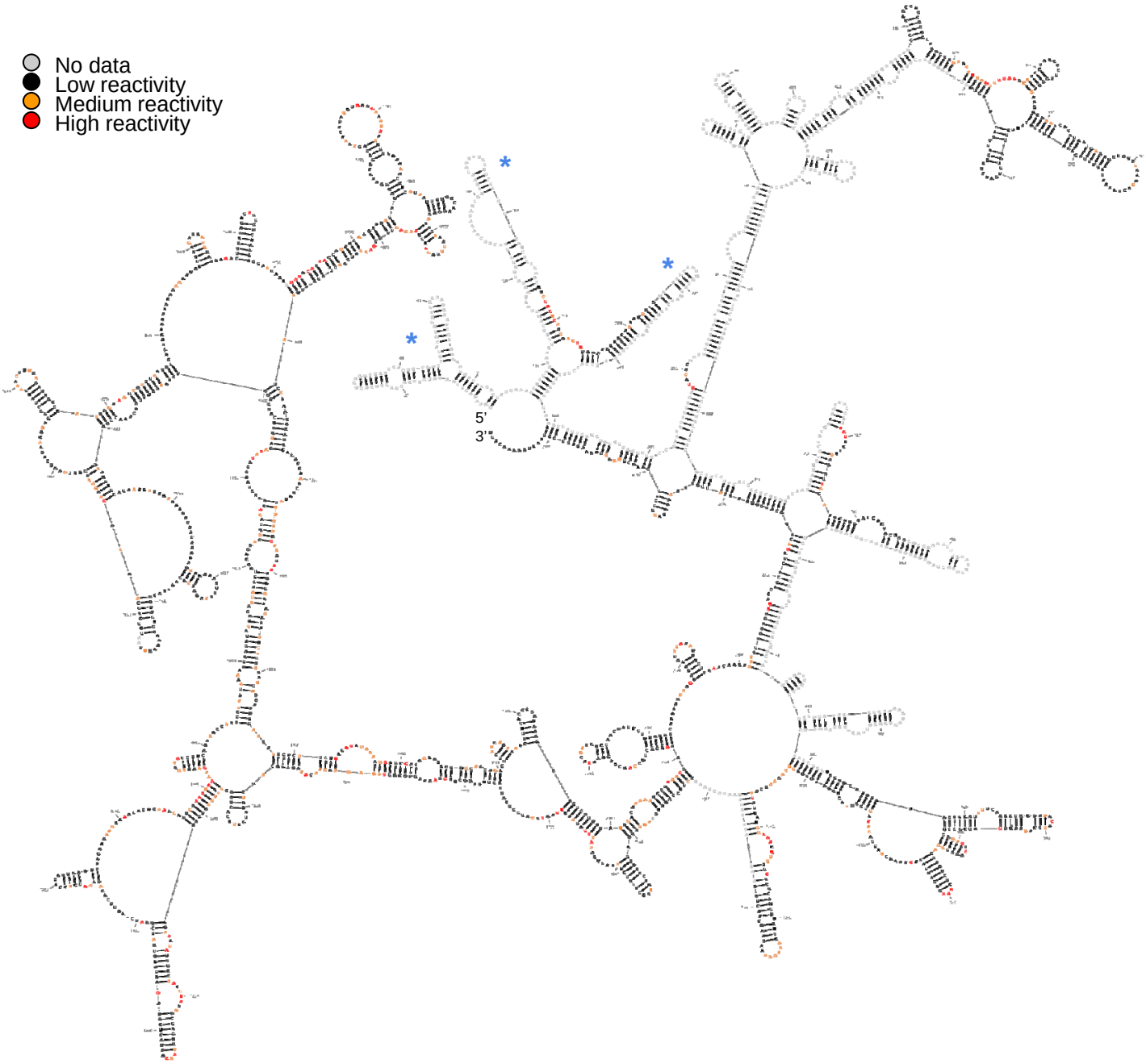



**Figure 3.** Secondary structure of the lncRNA JPX (chromatin *in vitro*) based on NAI-N3 SHAPE chemical probing data. Blue asterisks indicate structured domains with a similar fold to a previously reported structure corresponding to nucleotides 1 - 343 of JPX [61]. Chemical probing data is mapped to each structure (gray = no data, black = low reactivity, orange = medium reactivity, and red = high reactivity).

***XIST RepA*:** The RepA domain of XIST, which is composed of both the A and F repeats, localizes to the cytoplasm, nucleoplasm, and chromatin in HEK293 cells based on NAI-N3 chemical probing experiments. Both *in vivo* and *in vitro* data is available for each location. The secondary structures between the *in vivo* and *in vitro* pairs differ; similar to other lncRNAs previously discussed, this may be due to protein binding interactions (**Supplementary Figure 1**). Surprisingly, the lowest energy structure predicted for RepA in the cytoplasm and nucleoplasm are nearly identical (**Figure 4A, B**), with two structured domains adopting a similar fold in the chromatin-based secondary structure (**Figure 4C**). Only the mouse RepA XIST structure has been previously reported [62]; however, due to differences in primary sequence, the two structures cannot be compared.

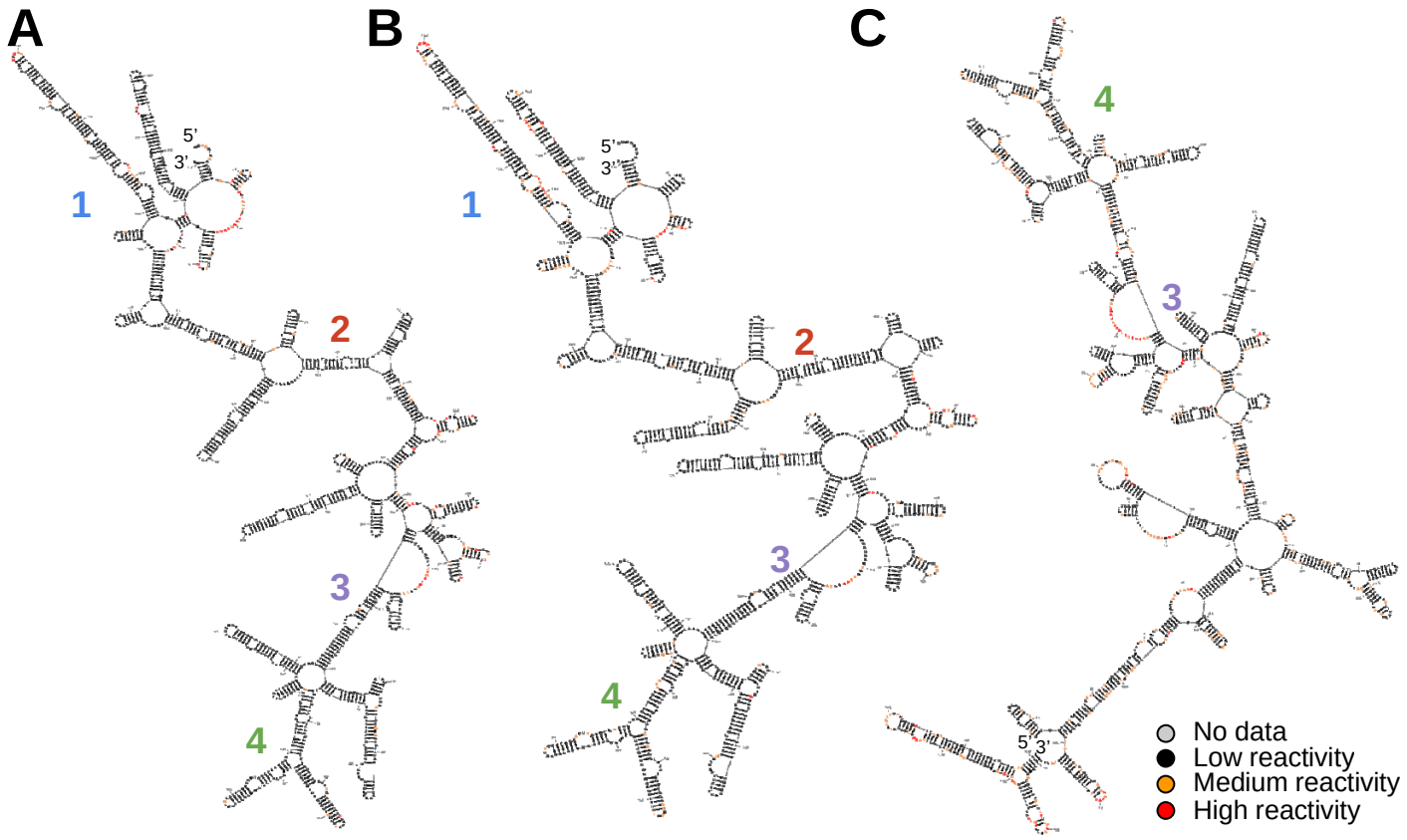



**Figure 4.** Secondary structures of XIST RepA (*in vivo*) based on NAI-N3 SHAPE chemical probing data. Structure of XIST RepA in A) the cytoplasm B) the nucleoplasm, and C) the chromatin. Numbers correspond to structured regions that are maintained despite the localization of the transcript. Chemical probing data is mapped to each structure (gray = no data, black = low reactivity, orange = medium reactivity, and red = high reactivity).

***LINC00894:*** The ENST00000455777.1 splice variant of the lncRNA LINC00894 localizes to the chromatin in HEK293 cells. *In vitro* chemical probing data (NAI-N3) reveals the RNA folds to form an extended branched structure, with 5 stem loop domains (**Figure 5**). Nucleotides located on the 5' half of the transcript lack chemical probing data, which may influence how the RNA is predicted to fold.

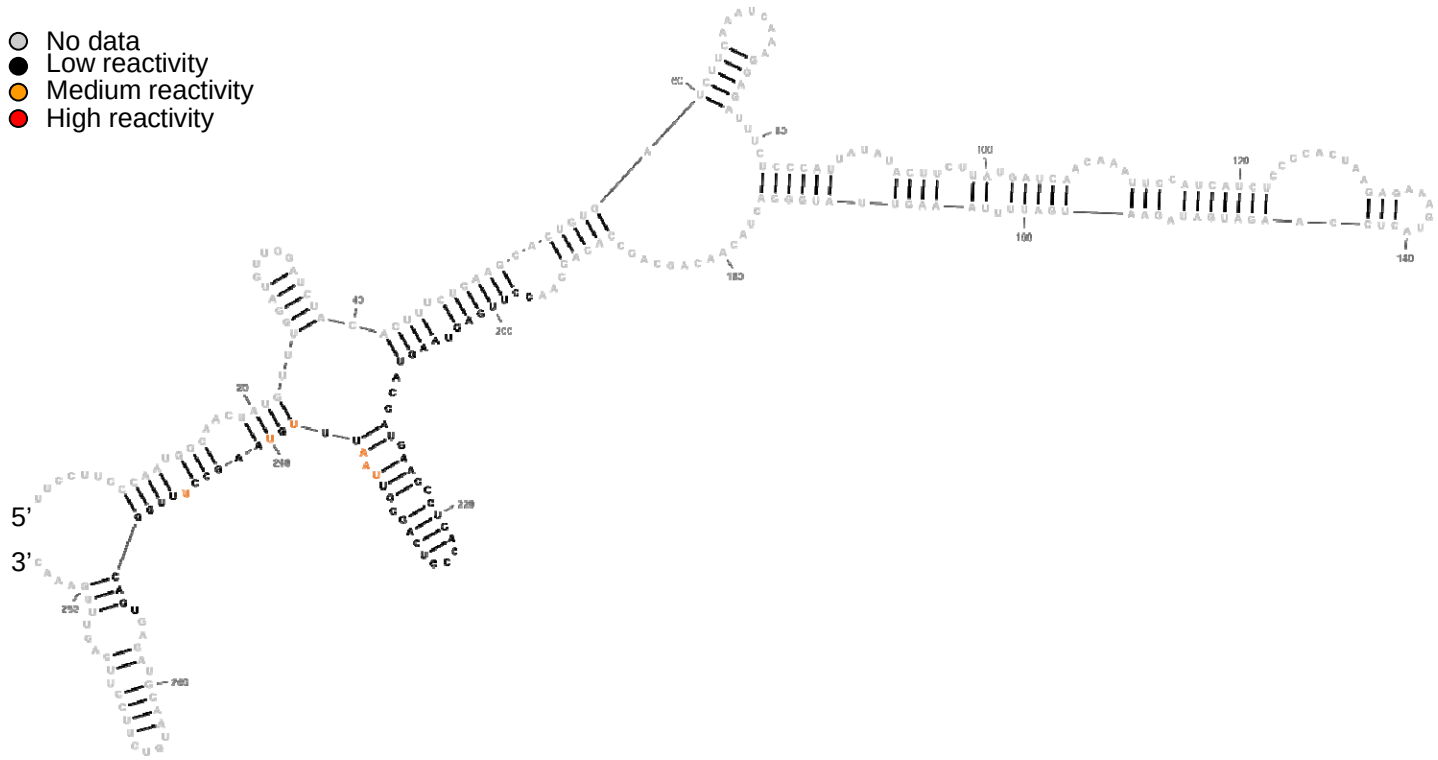



**Figure 5**. Secondary Structure of the lncRNA LINC00894 (*in vitro*) based on NAI-N3 SHAPE chemical probing data. Chemical probing data is mapped to each structure (gray = no data, black = low reactivity, orange = medium reactivity, and red = high reactivity).

***ZNF674-AS1:*** ZNF674-AS1 is present in the nucleoplasm and cytoplasm. The nucleoplasm *in vitro* dataset had the highest coverage, revealing a structure that forms several multiway junctions, extended hairpins, and short stem loop motifs (**Figure 6**). We did not fold the structure of ZNF674-AS1 based on cytoplasm data due to the low coverage. Similar to the other lncRNAs discussed, ZNF674-AS1 is also predicted to bind several RBPs (**Supplementary Figure 1**).

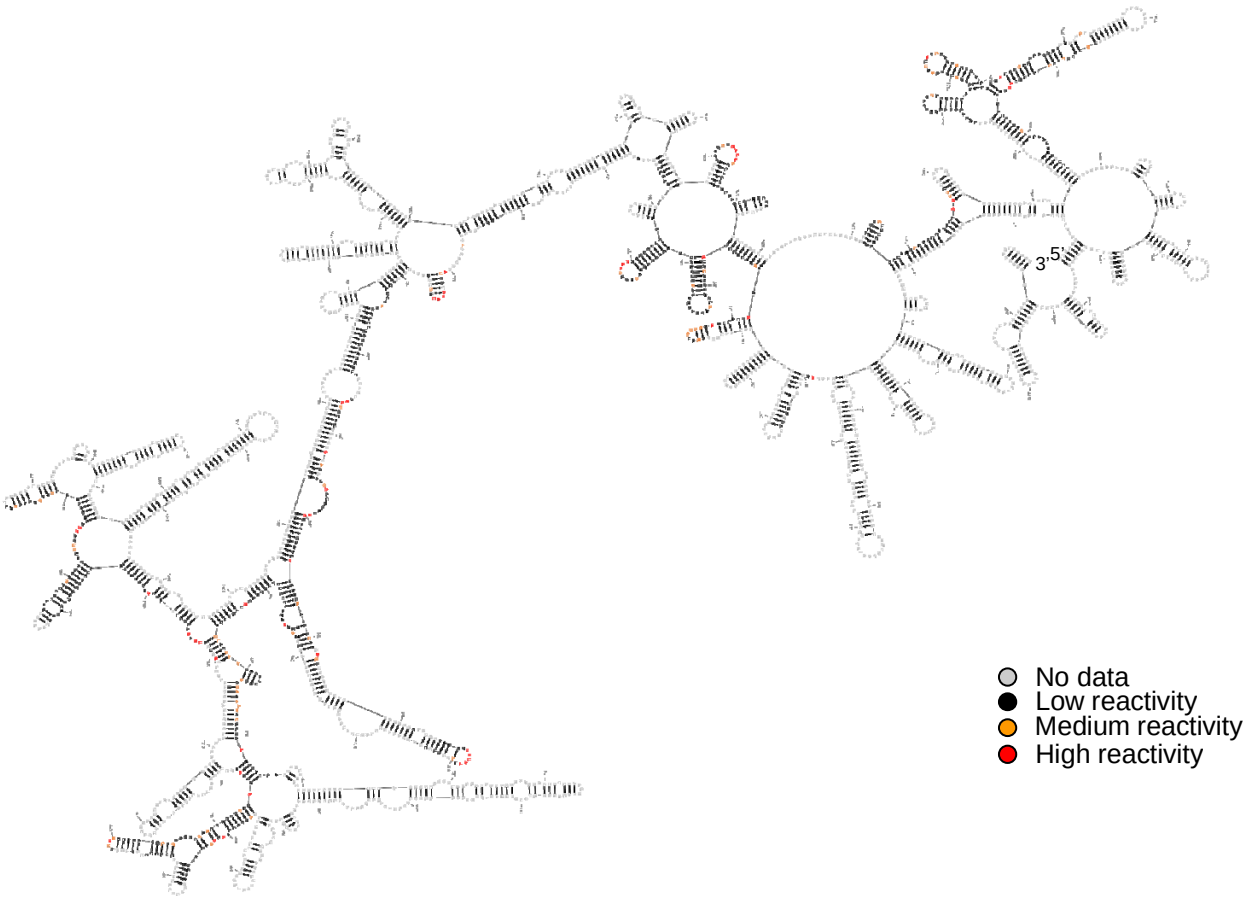



**Figure 6.** Secondary structure of the lncRNA ZNF674-AS1 (*in vitro*) based on DMS-MaPseq chemical probing data from the nucleoplasm. Chemical probing data is mapped to each structure (gray = no data, black = low reactivity, orange = medium reactivity, and red = high reactivity).

***LncRNAS with low coverage: MIR503HG, LINC00630, and FIRRE***

Three lncRNAs, MIR503HG, LINC00630, and FIRRE, have chemical probing coverage less than 20% (**Table 1**). This may be attributed to several factors, including the abundance levels of the lncRNA, compactness of the RNA (decreased accessibility of flexible nucleotides), and experimental biases (e.g., how fragmentation occurs for RNAs during library preparation for MaP experiments). A lack of reactivity can directly impact the accuracy of RNA secondary structure prediction. We used the available chemical probing data to fold the lncRNAs (**Supplementary Figures 2-4)**; however, we acknowledge that more chemical probing data is required to validate the predicted secondary structures.

***MIR503HG*:** The lncRNA MIR503HG localizes to both chromatin and the nucleoplasm. A comparison of nucleotides possessing chemical reactivity from the chromatin-based probing data (*in vitro* and *in vivo*) reveals a structurally similar 5-stem multiway junction (**Supplementary Figure 2A, B**). One of the stem loops that comprises this junction is also predicted to fold in the nucleoplasm (**Supplementary Figure 2C**). MIR503HG interacts with several proteins, including polypyrimidine tract binding protein 1 (PTBP1), heterogeneous nuclear ribonucleoproteins HNRNP, A2B1, and specific protein 1 (SP1) to mediate endothelial to mesenchymal transition, regulate hnRNP A2B1 expression, and inhibit activation of TMEFF1, respectively [34,49]. These protein interactions may impact how MIR503HG folds in chromatin and in the nucleoplasm.

***LINC00630*:** The lncRNA LINC00630 localizes to chromatin; NAI-N3 chemical probing for this lncRNA covers only ~11% of the RNA. This reactivity is more abundant in the middle of the transcript and supports the formation of multiple stem-loop structured motifs (**Supplementary Figure 3**).

***FIRRE:*** Structural data for the lncRNA FIRRE is limited (coverage ~8%, **Table 1**). The structural probing data that is available supports the notion that the lncRNA FIRRE is broadly split into two structured domains (**Supplementary Figure 4**). Each half is composed of a network of multiway junctions and stem-loop motifs. More chemical probing data is required to ensure higher confidence in the RNA structure.

## RBM15 interacts with XCI-escaping lncRNAs through RRM2 and RRM3

RBM15 has been proposed to interact with FIRRE and XIST [17,20]. 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) [63] (**Supplementary Figure 1**) and find that of the 9 lncRNAs investigated in this study, all but INE2 interact with RBM15 [63,64]. In agreement with the RNA binding sequence motif for RBM15 (**Figure 7A**), the crosslinking experiments revealed U and A-containing sequences in each of the 8 lncRNAs (**Figure 7B, Supplementary Table 1**) [13–16].

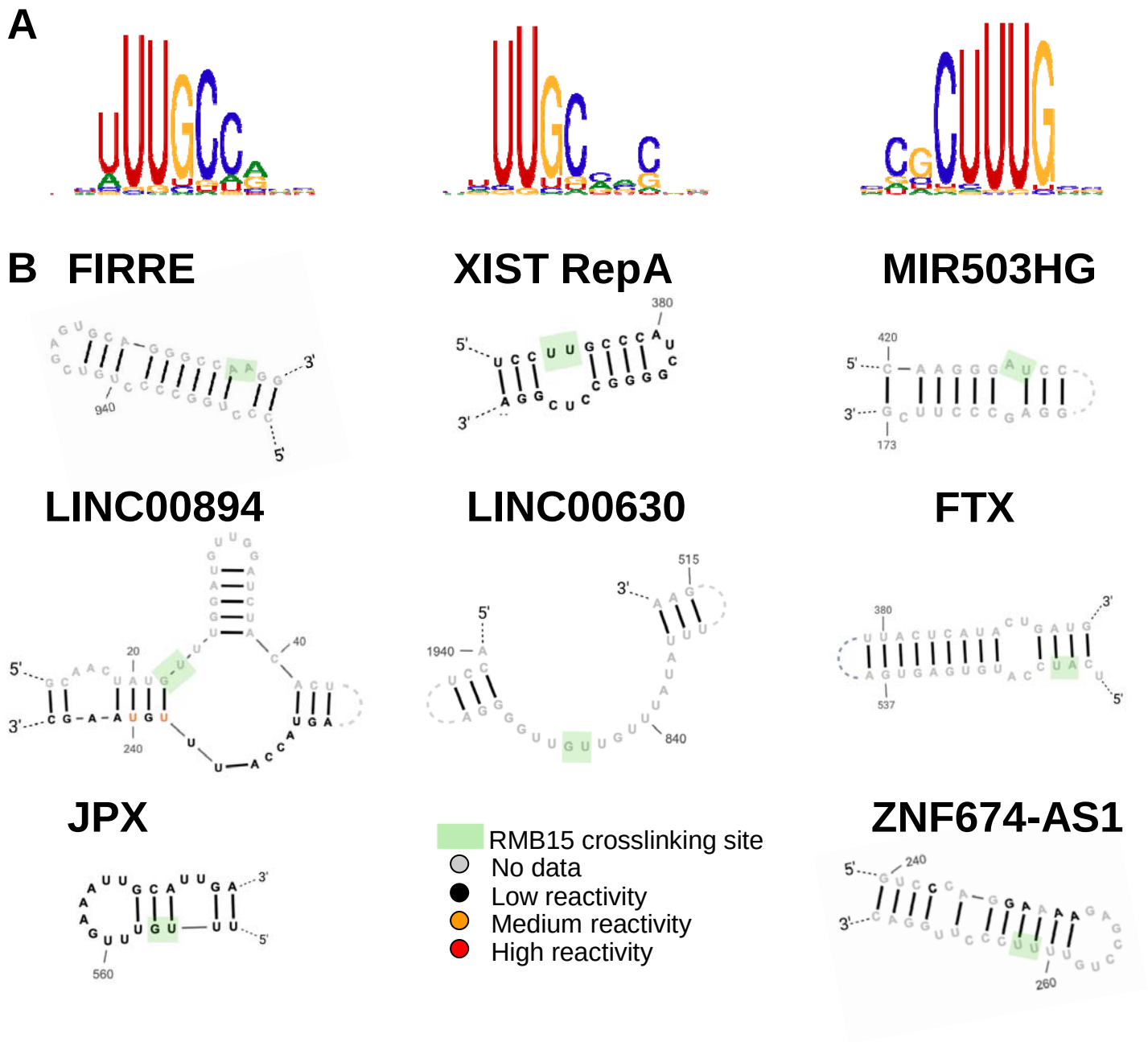



**Figure 7:** Representative RBM15 CLIP binding sites with lncRNAs. A) Binding sequence motif logo for RBM15 based on iCLIP data [16]. B) iCLIP RBM15 crosslinking RNA binding sites for RBM15 with FIRRE, FTX, JPX, LINC00630, LINC00894, MIR503HG, XIST RepA, and ZNF674-AS1. Crosslinking sites are highlighted in green. Chemical probing data is mapped to each structure (gray = no data, black = low reactivity, orange = medium reactivity, and red = high reactivity). See also **Supplementary Table 1**.

We then used 'fast predictions of RNA and protein interactions and domains at the Center for Genomic Regulation, Barcelona, Catalonia' (CatRAPID) signature to determine which region of RBM15 is predicted to interact with RNA[65]. 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 5**). CatRapid Graphic (for RNAs < 1,200 nt) and CatRAPID Fragments (for RNAs > 1,200 nt) [66,67] were then used to determine which amino acid residues of RBM15 interact with the

lncRNAs (**Figure 8)**. Interestingly, we observed RRMs 2 and 3 (amino acids 374 - 529) bind in the best capacity for the majority of lncRNAs.

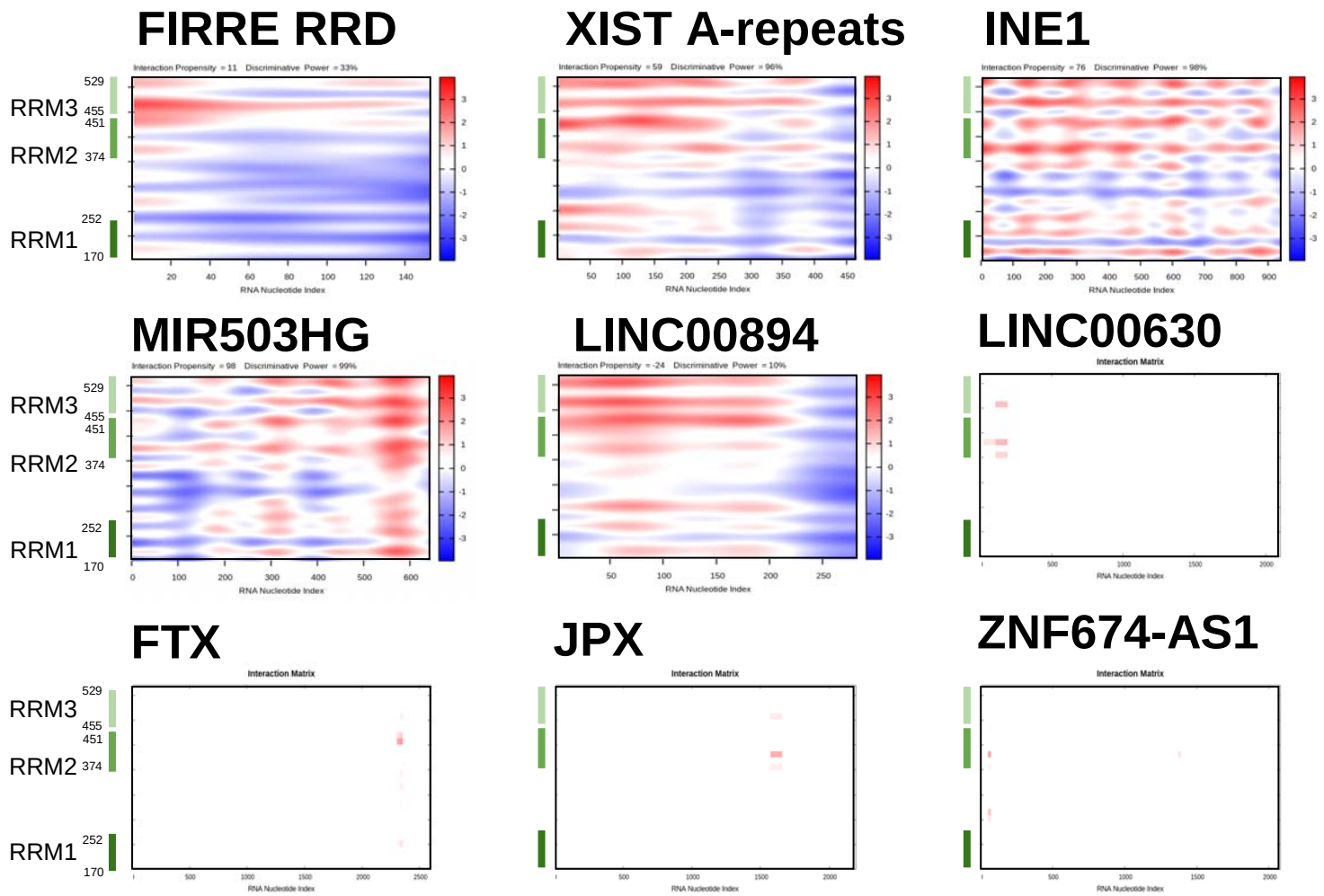



**Figure 8.** CatRAPID-predicted interaction domains for RBM15 and XCI-escaping lncRNAs. CatRAPID Graphic was used for RNAs shorter than 1,200 nts, including FIRRE RRD, XIST A-repeats, INE1, MIR503HG, and LINC00894. Red indicates a higher propensity for an interaction. CatRAPID Fragments was used for RNAs longer than 1,200 nts, including LINC00630, FTX, JPX, and ZNF674-AS1. The RRM domains of RBM15 are indicated with green vertical lines.

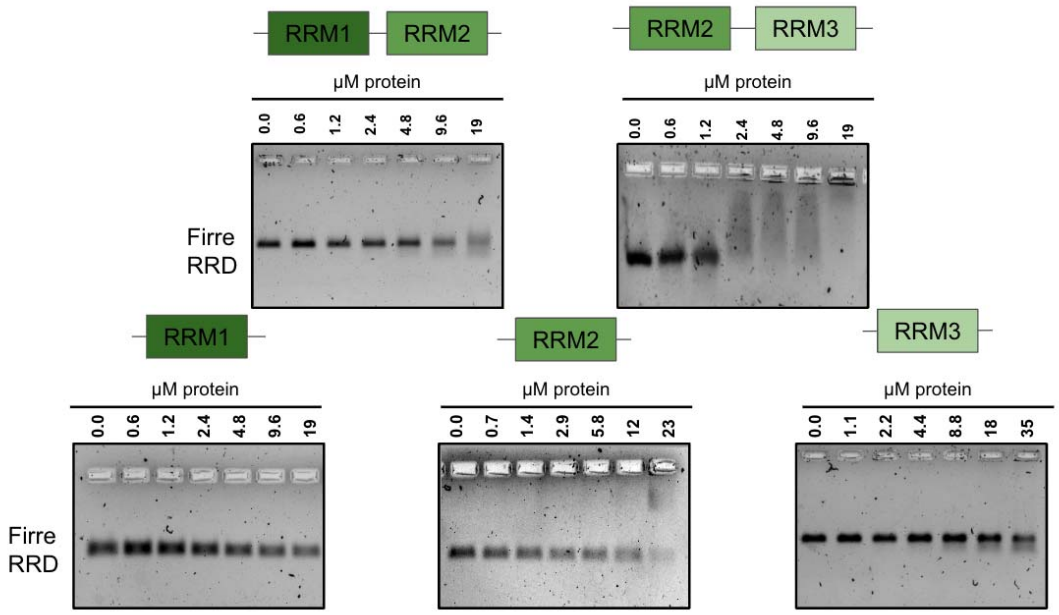
Encouraged by these findings, we carried out electrophoretic mobility shift assays (EMSA) for various RRM domains of RBM15 (RRM1, RRM2, RRM3, RRM12, RRM23; we were unable to obtain a soluble RRM123 construct of adequate concentration for EMSA experiments) with the FIRRE RRD and with the XIST A-repeats (**Figure 9A, B**).

RBM15 RRM12 shows weak binding with the FIRRE RRD RNA (**Figure 9A**). 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 Kd of 1.56 ± 0.46µM. 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.

Since the A-repeats fold independently of the rest of the XIST lncRNA [68], 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 9B**). 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 ± 0.21μM. 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.

Altogether we find that the RRD domain of the lncRNA FIRRE and the XIST A-repeats bind with low micromolar affinity with the RRM23 domain of RBM15. This is in excellent agreement with the CatRAPID predicted results (**Figure 8**).

# A: FIRRE RRD with RBM15




# B: XIST A-repeats with RBM15

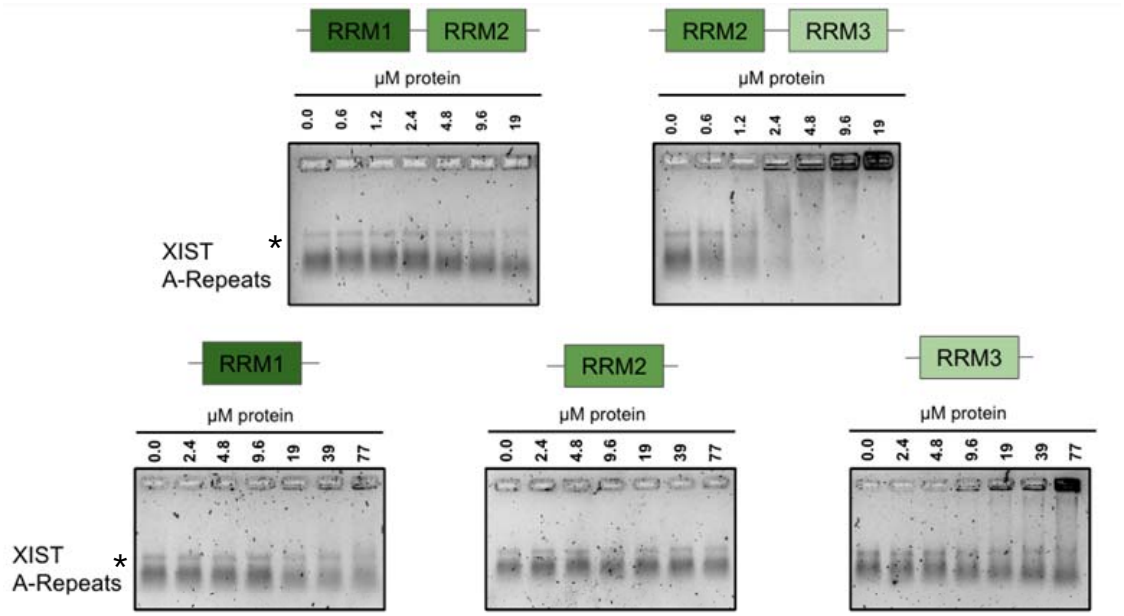



**Figure 9**. The interaction of RBM15 RRMs with the FIRRE RRD and XIST A-repeat lncRNAs. A) EMSA of RBM15 RRM12, 23, 1, 2, and 3 with the FIRRE RRD domain. B) EMSA of RBM15 RRM12, 23, 1, 2, and 3 with the XIST A-repeats. The asterisk indicates a second conformation sampled by the XIST A-repeats.

### FIRRE RRD and XIST A-repeats interact with RBM15 in structured regions

In addition to predicting which domains of RBM15 bind to the lncRNAs, CatRAPID also assesses the propensity for different regions of the RNA to interact with RBM15. CatRAPID predicted RBM15 binding occurs across the length of the A-repeats (**Figure 8**). Given the repetitive nature of the XIST A-repeats (a 24 nt element that is repeated 8.5 times in humans), it is not surprising that CatRAPID reveals a high propensity for

binding across much of the transcript. This is in agreement with crosslinking experiments carried out in a previous study [17] and as mentioned above (**Figure 7B**).

Because RASP chemical probing coverage for the FIRRE RRD is sparse (**Supplementary Figure 4**), we carried out SHAPE chemical probing on this domain to assess its secondary structure. This revealed that the RRD folds to form three stable stem loop motifs and contains a largely unstructured multiway junction (**Figure 10A**). It is unsurprising that this predicted structure, based on experimental restraints, is not in agreement with the structure predicted using sparse constraints (**Supplementary Figure 4**).

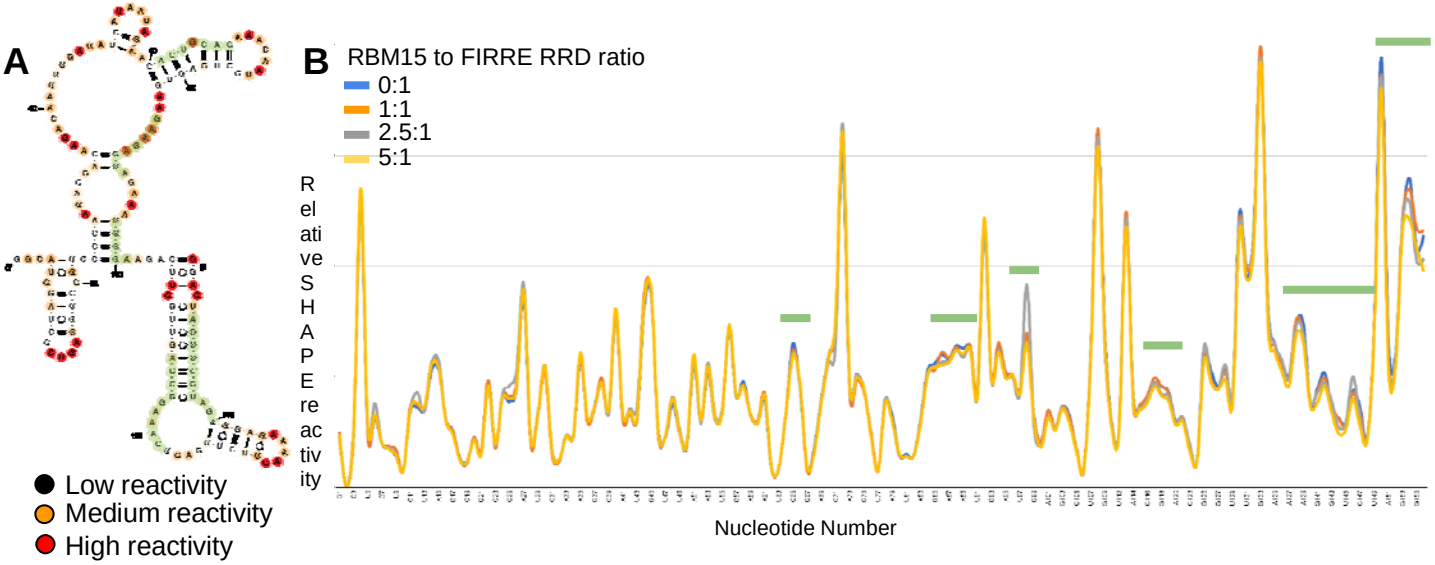



**Figure 10**. SHAPE analysis of the FIRRE RRD RNA. A) The secondary structure predicted for the FIRRE RRD based on 1M7 chemical probing. Chemical probing data is mapped to each structure (gray = no data, black = low reactivity, orange = medium reactivity, and red = high reactivity). B) Bar plot showing the SHAPE reactivity in the presence of RBM15 RRM123. Binding sites are indicated by green horizontal lines.

To determine where RBM15 interacts with the FIRRE RRD, we added RBM15 RRM123 at varying concentrations to the RNA prior to the addition of the SHAPE probe to identify sites of binding. Interestingly, we observe differences in SHAPE reactivity in several regions near the 3' end, which is an indication of binding (**Figure 10B**). These localize to partially structured regions containing A and U residues, which is in good agreement with the binding motif expected for RBM15 (**Figure 7**).

## 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 XCI-escaping 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 [3,69,70]. 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 [12].

We find that RBM15 binding sites occur in partially structured regions across each of the lncRNAs investigated. These regions typically contain uracil residues, which is in agreement with what has been previously reported [13–16]. 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 [17]. 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.

While further investigation is needed to fully comprehend how RBM15 is implicated in the regulatory roles of the nine 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.

## Methods

### Identifying lncRNAs and obtaining chemical probing data

The LNCipedia [71] was used to curate a list of all X-linked lncRNAs known to date. Genome-wide chemical probing datasets were downloaded from the RASP database [34]. An in-house generated script was used to identify lncRNAs possessing chemical probing data (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 [72] the secondary structure of each RNA was predicted (using chemical probing data as soft constraints) and the lowest energy structure was visualized using RNAcanvas [73].

### Identification of RNA-Protein Binding Partners

Potential protein binding partners for each lncRNA were identified with oRNAment [60]. 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 [63].

### *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) 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.

An overview of the approach is outlined in **Figure 11.**

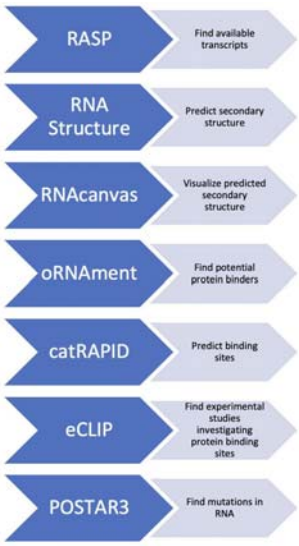



**Figure 11.** Overview of the approach used to investigate lncRNA structure and RBP interactions.

**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 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 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 6.9) 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 gel stained with SYBR Safe (Invitorgen). Gels were run at 60V for 45 minutes, followed by visualization on a GelDoc imager. Band quantification was carried out in ImageLab and $K_d$s 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.

**SHAPE chemical probing**

SHAPE was carried out on the FIRRE RRD using the 1-methyl-7-nitroisatoic anhydride (1M7) chemical probe as previously described [17,74]. Briefly, RNA was snap-cooled (95°C for 5 minutes, followed by 5 minutes on ice) in buffer containing 200 mM NaCl, 100 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 8, 0.2 mM EDTA pH 8 and refolded at 37°C in buffer containing 100 mM NaCl, 50 mM HEPES pH 8, 16.5 mM $MgCl_2$ for 30 minutes. RNA was then incubated with 1.5 mM 1M7 for 15 minutes at room temperature, followed by quenching with 0.5 M sodium 2-(*N*-morpholino)ethanesulfonic acid (NaMES). To identify sites of RBM15 binding, the RNA was snap-cooled and folded as previously mentioned, then incubated with up to 1 µM RBM15 123 (for 5 min) prior to the addition of the chemical probe: 1.5 mM 1M7. Following an ethanol wash, the RNA was reverse transcribed with Superscript III (Invitrogen) according to the manufacturer's instructions. cDNA was sequenced on an ABI 3730XL instrument (Cornell Genomics Facility). Reads were analyzed and the secondary structure was predicted using HiTRACE [75].

## Supplementary Tables

**Supplementary Table 1. CLIP RBM15 binding sites**

| RNA | RBM15 Binding by X-chromosome position |
|---|---|
| FTX (-) | 1. 74280445-74280446<br>2. 74288671-74288672<br>3. 74289673-74289674<br>4. 74281025-74281026<br>5. 74280445-74280446 |
| INE1 (+) | No data |
| JPX (+) | 1. 73975480-73975481<br>2. 73999046-73999047<br>3. 73969605-73969606<br>4. 73985596-73985597<br>5. 73966807-73966808<br>6. 73961008-73961009<br>7. 73947331-73947332<br>8. 73999848-73999849<br>9. 73999848-73999849<br>10. 73969605-73969606<br>11. 73999532-73999533<br>12. 73999257-73999258<br>13. 74000101-74000102<br>14. 74000146-74000147<br>15. 74000339-74000340<br>16. 73993862-73993863<br>17. 73993862-73993863<br>18. 73947342-73947343 |
| MIR503HG (-) | 1. 134544676-134544677<br>2. 134544627-134544628<br>3. 134545649-134545650<br>4. 134544531-134544532 |
| ZNF674-AS1 (+) | 1. 46546591-46546592 |
| LINC00630 (+) | 1. 102834739-102834740<br>2. 102826990-102826991<br>3. 102842481-102842482 |

| | |
|---|---|
| | 4. 102805047-102805048<br>5. 102879633-102879634<br>6. 102834732-102834752<br>7. 102834739-102834740<br>8. 102823001-102823002<br>9. 102796115-102796116<br>10. 102834703-102834704 |
| LINC00894 (+) | 1. 149960902-149960903 |
| FIRRE (-) | 1. 131795764-131795765<br>2. 131830367-131830368<br>3. 131807947-131807948<br>4. 131830312-131830332<br>5. 131830443-131830444<br>6. 131830367-131830368<br>7. 131795724-131795725 |
| *XIST (-) | 1. 73843036-73843037<br>2. 73839268-73839269<br>3. 73847384-73847385<br>4. 73824437-73824438<br>5. 73835061-73835062<br>6. 73825061-73825062<br>7. 73824636-73824637<br>8. 73824027-73824028<br>9. 73824802-73824803<br>10. 73824226-73824227<br>11. 73841229-73841230<br>12. 73845874-73845875<br>13. 73830712-73830713<br>14. 73839119-73839120<br>15. 73830855-73830856<br>16. 73850037-73850038<br>17. 73823762-73823763<br>18. 73846838-73846839<br>19. 73822884-73822885<br>20. 73828450-73828451<br>21. 73821746-73821747<br>22. 73852125-73852145<br>23. 73852365-73852705<br>24. 73851845-73852105<br>25. 73852165-73852345<br>26. 73850585-73850625<br>27. 73851265-73851325<br>28. 73850845-73850865<br>29. 73851525-73851605<br>30. 73829165-73829205<br>31. 73850505-73850525<br>32. 73841465-73841485<br>33. 73827665-73827685<br>34. 73851625-73851665<br>35. 73826785-73826805<br>36. 73841865-73841885<br>37. 73843865-73843885<br>38. 73851345-73851425 |

| | |
|---|---|
| | 39. 73827045-73827065<br>40. 73851025-73851145<br>41. 73851165-73851185<br>42. 73820945-73820965<br>43. 73841405-73841425<br>44. 73826465-73826485<br>45. 73850885-73851005<br>46. 73833265-73833285<br>47. 73841545-73841565<br>48. 73842105-73842125<br>49. 73831085-73831125<br>50. 73827165-73827205<br>51. 73850645-73850685<br>52. 73823705-73823725<br>53. 73826145-73826165<br>54. 73851685-73851825<br>55. 73820765-73820785<br>56. 73850385-73850465<br>57. 73820865-73820885<br>58. 73821005-73821045<br>59. 73845225-73845245<br>60. 73820685-73820705<br>61. 73847745-73847765<br>62. 73851205-73851245<br>63. 73826265-73826285<br>64. 73848325-73848345<br>65. 73827805-73827825<br>66. 73827365-73827385<br>67. 73846345-73846365<br>68. 73825005-73825025<br>69. 73829105-73829125<br>70. 73842185-73842225<br>71. 73845005-73845025<br>72. 73826085-73826105<br>73. 73827245-73827265<br>74. 73842285-73842305<br>75. 73848445-73848465<br>76. 73850785-73850825<br>77. 73826045-73826065<br>78. 73846785-73846805<br>79. 73846385-73846405<br>80. 73850245-73850265<br>81. 73826565-73826605<br>82. 73843785-73843805<br>83. 73845525-73845545<br>84. 73821965-73821985<br>85. 73826405-73826425<br>86. 73827765-73827785<br>87. 73845825-73845845<br>88. 73851445-73851465<br>89. 73826665-73826685<br>90. 73841605-73841625<br>91. 73846085-73846105<br>92. 73849225-73849245<br>93. 73850285-73850325 |

| | |
|---|---|
| | 94. 73826925-73826945<br>95. 73842985-73843005<br>96. 73821165-73821185<br>97. 73823985-73824005<br>98. 73842065-73842085<br>99. 73827405-73827425<br>100. 73841705-73841725<br>101. 73850205-73850225<br>102. 73824745-73824765<br>103. 73843925-73843945<br>104. 73844705-73844725<br>105. 73827885-73827905<br>106. 73845265-73845285<br>107. 73825865-73825905<br>108. 73826985-73827005<br>109. 73848525-73848545<br>110. 73849985-73850005<br>111. 73841745-73841765<br>112. 73841805-73841825<br>113. 73849805-73849825<br>114. 73822165-73822185<br>115. 73839225-73839245<br>116. 73852709-73852710<br>117. 73829148-73829149<br>118. 73820981-73820982<br>119. 73827289-73827290<br>120. 73821746-73821747<br>121. 73822044-73822045<br>122. 73845272-73845273<br>123. 73833306-73833307<br>124. 73850689-73850690<br>125. 73845569-73845570<br>126. 73826590-73826591<br>127. 73844342-73844343<br>128. 73844626-73844627<br>129. 73846140-73846141<br>130. 73841770-73841771<br>131. 73841510-73841511<br>132. 73842150-73842151<br>133. 73846838-73846839<br>134. 73821901-73821902<br>135. 73827973-73827974<br>136. 73824802-73824803<br>137. 73825061-73825062<br>138. 73846511-73846512<br>139. 73847804-73847805<br>140. 73822325-73822326<br>141. 73847157-73847158<br>142. 73826189-73826190<br>143. 73837495-73837496<br>144. 73849272-73849273<br>145. 73844754-73844755<br>146. 73848367-73848368<br>147. 73845874-73845875<br>148. 73848881-73848882 |

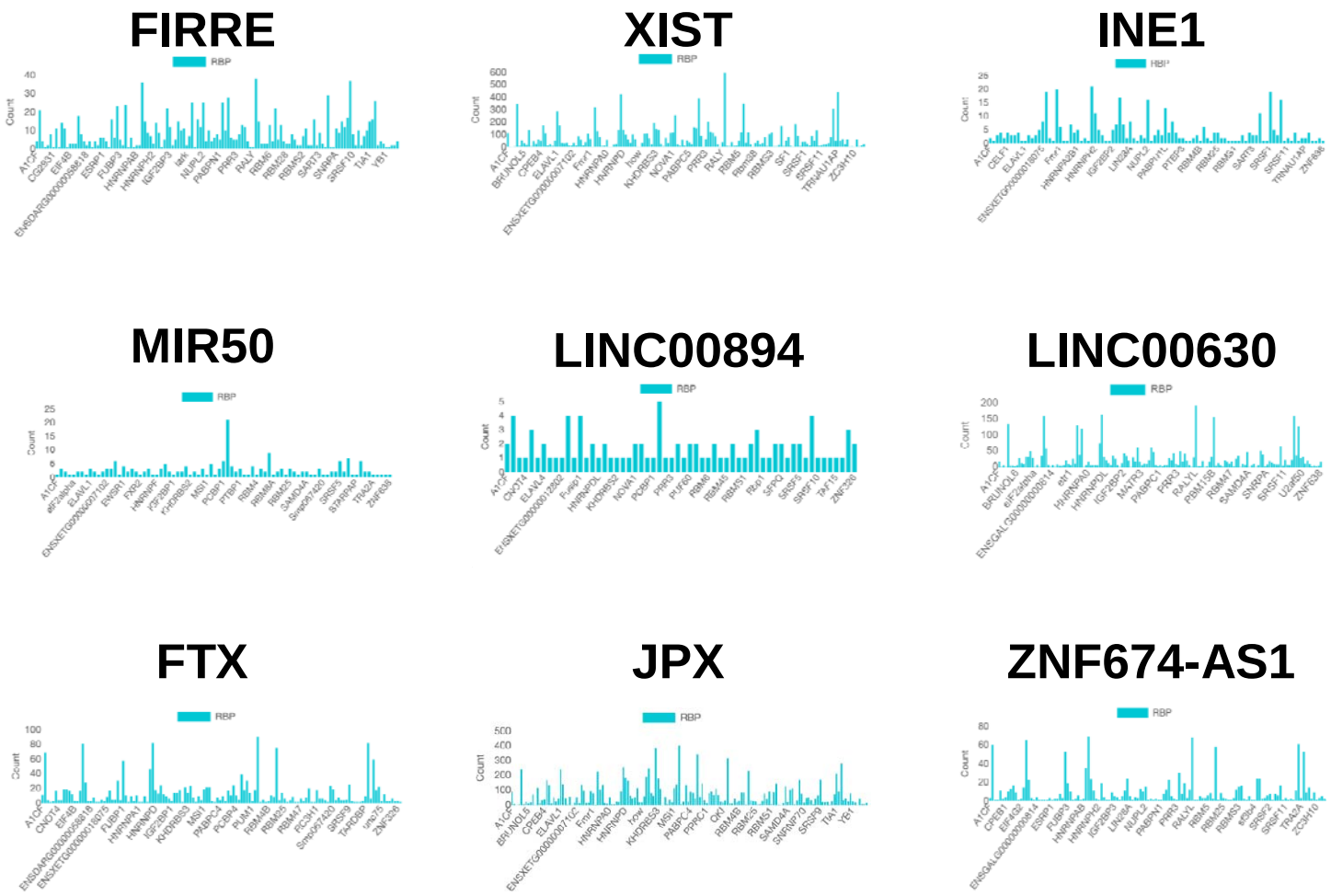
| | |
|---|---|
| | 149. 73848756-73848757<br>150. 73823650-73823651<br>151. 73848668-73848669<br>152. 73848575-73848576<br>153. 73824027-73824028<br>154. 73823578-73823579<br>155. 73845169-73845170<br>156. 73849209-73849210<br>157. 73842895-73842896<br>158. 73849551-73849552<br>159. 73849858-73849859<br>160. 73850037-73850038<br>161. 73843925-73843926<br>162. 73839119-73839120<br>163. 73843677-73843678<br>164. 73847287-73847288<br>165. 73843107-73843108<br>166. 73825874-73825875<br>167. 73842726-73842727<br>168. 73842605-73842606<br>169. 73821869-73821870<br>170. 73824437-73824438<br>171. 73839268-73839269<br>172. 73847894-73847895<br>173. 73849722-73849723<br>174. 73841330-73841331<br>175. 73823411-73823412<br>176. 73821344-73821345<br>177. 73845064-73845065<br>178. 73843276-73843277<br>179. 73841229-73841230<br>180. 73847719-73847720<br>181. 73824636-73824637<br>182. 73825650-73825651<br>183. 73847999-73848000<br>184. 73843493-73843494<br>185. 73825518-73825519<br>186. 73848412-73848413<br>187. 73821508-73821509<br>188. 73831134-73831135<br>189. 73823677-73823678<br>190. 73823298-73823299<br>191. 73823267-73823268<br>192. 73849376-73849377<br>193. 73845366-73845367<br>194. 73841688-73841689<br>195. 73825760-73825761<br>196. 73824905-73824906<br>197. 73843583-73843584<br>198. 73833889-73833890<br>199. 73848191-73848192<br>200. 73849407-73849408<br>201. 73822706-73822707<br>202. 73824098-73824099<br>203. 73844916-73844917 |

| | |
|---|---|
| | 204. 73822785-73822786<br>205. 73822734-73822735<br>206. 73845996-73845997<br>207. 73842793-73842794<br>208. 73837185-73837186<br>209. 73843707-73843708<br>210. 73845757-73845758<br>211. 73842461-73842462<br>212. 73844588-73844589<br>213. 73822522-73822523<br>214. 73843409-73843410<br>215. 73823836-73823837<br>216. 73844084-73844085<br>217. 73821649-73821650<br>218. 73822945-73822946<br>219. 73824729-73824730<br>220. 73830834-73830835<br>221. 73846223-73846224<br>222. 73830668-73830669<br>223. 73847590-73847591<br>224. 73840532-73840533<br>225. 73823234-73823235<br>226. 73842946-73842947<br>227. 73825689-73825690<br>228. 73825346-73825347<br>229. 73846273-73846274<br>230. 73821394-73821395<br>231. 73845115-73845116<br>232. 73822586-73822587<br>233. 73825444-73825445<br>234. 73845216-73845217<br>235. 73845662-73845663<br>236. 73847846-73847847<br>237. 73822986-73822987<br>238. 73849619-73849620<br>239. 73849506-73849507<br>240. 73847635-73847636<br>241. 73849984-73849985<br>242. 73830603-73830604<br>243. 73836646-73836647<br>244. 73843435-73843436<br>245. 73824064-73824065<br>246. 73841194-73841195<br>247. 73821433-73821434<br>248. 73847383-73847384<br>249. 73821577-73821578<br>250. 73839056-73839057<br>251. 73824274-73824275<br>252. 73839237-73839238<br>253. 73846784-73846785<br>254. 73845965-73845966<br>255. 73848090-73848091<br>256. 73835021-73835022<br>257. 73843009-73843010<br>258. 73825289-73825290 |

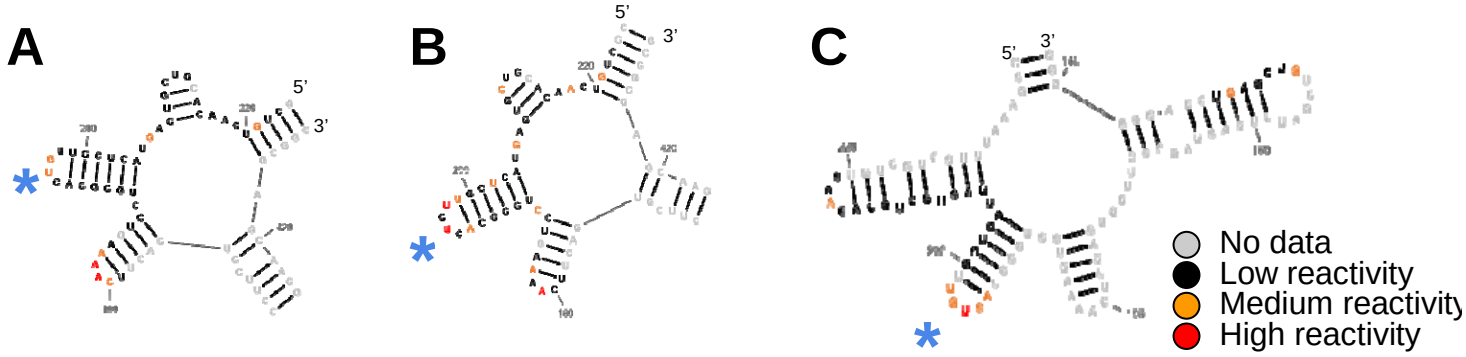
| | |
|---|---|
| | 259. 73844983-73844984<br>260. 73824191-73824192<br>261. 73843165-73843166<br>262. 73843361-73843362<br>263. 73843077-73843078<br>264. 73833855-73833856<br>265. 73844427-73844428<br>266. 73823376-73823377<br>267. 73822459-73822460<br>268. 73844874-73844875<br>269. 73825583-73825584<br>270. 73839024-73839025<br>271. 73823346-73823347<br>272. 73843331-73843332 |

*These iCLIP sites correspond to the full length XIST RNA
(+) corresponds to transcripts from the positive strand
(-) corresponds to transcripts from the negative strand

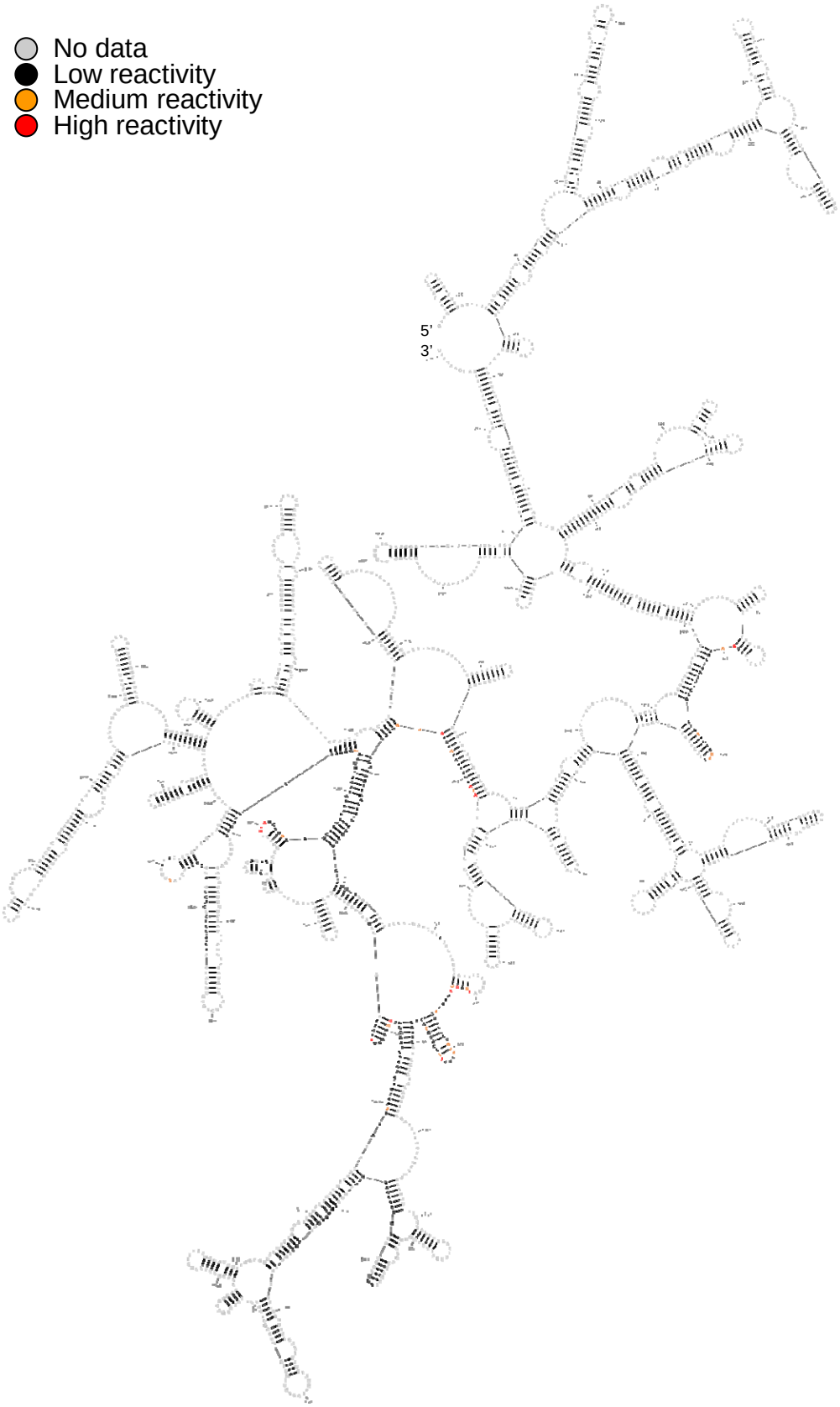
## Supplementary Figures



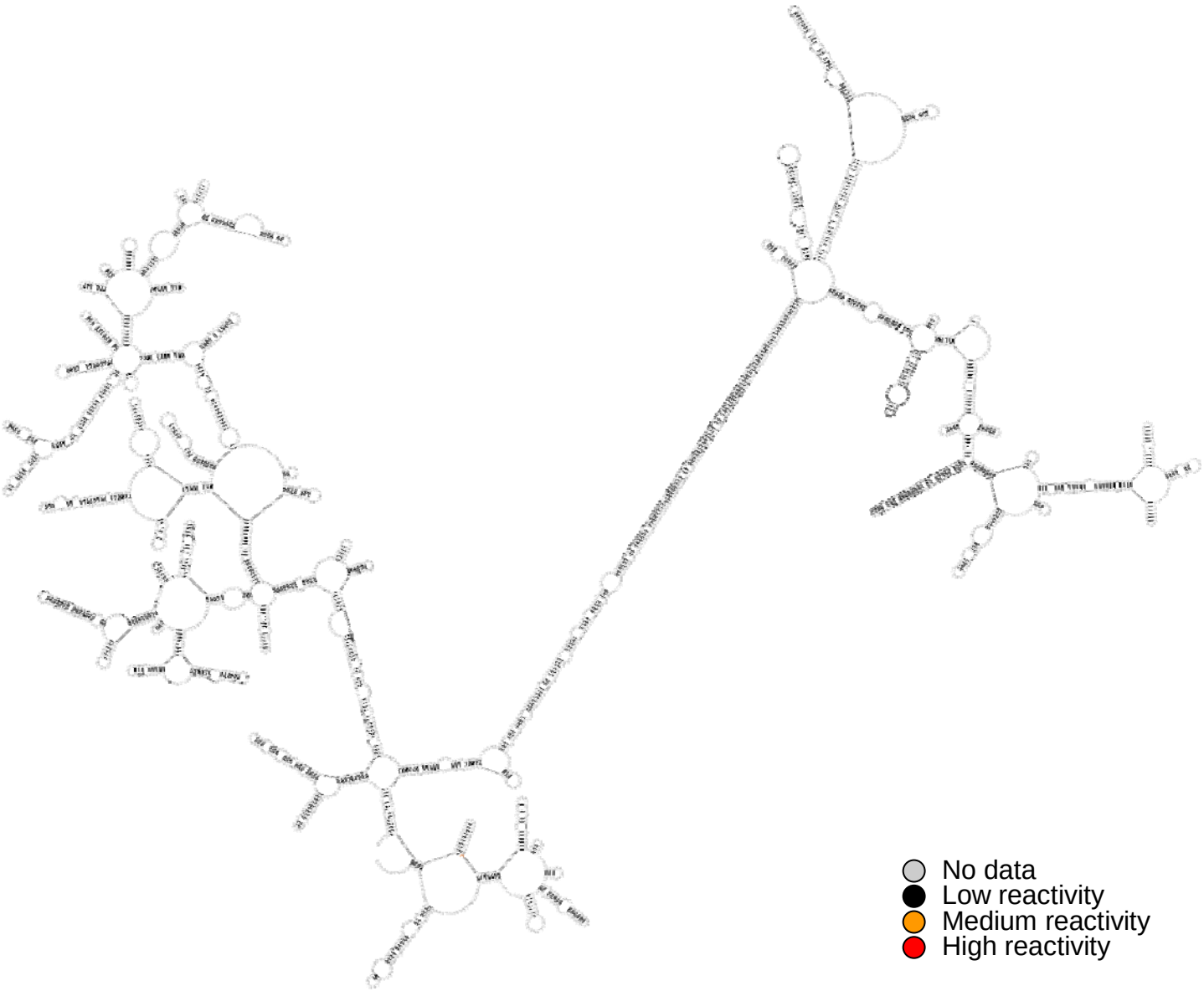

**Supplementary Figure 1.** oRNAment bar plots for RBPs predicted to interact with XCI-escaping lncRNAs.



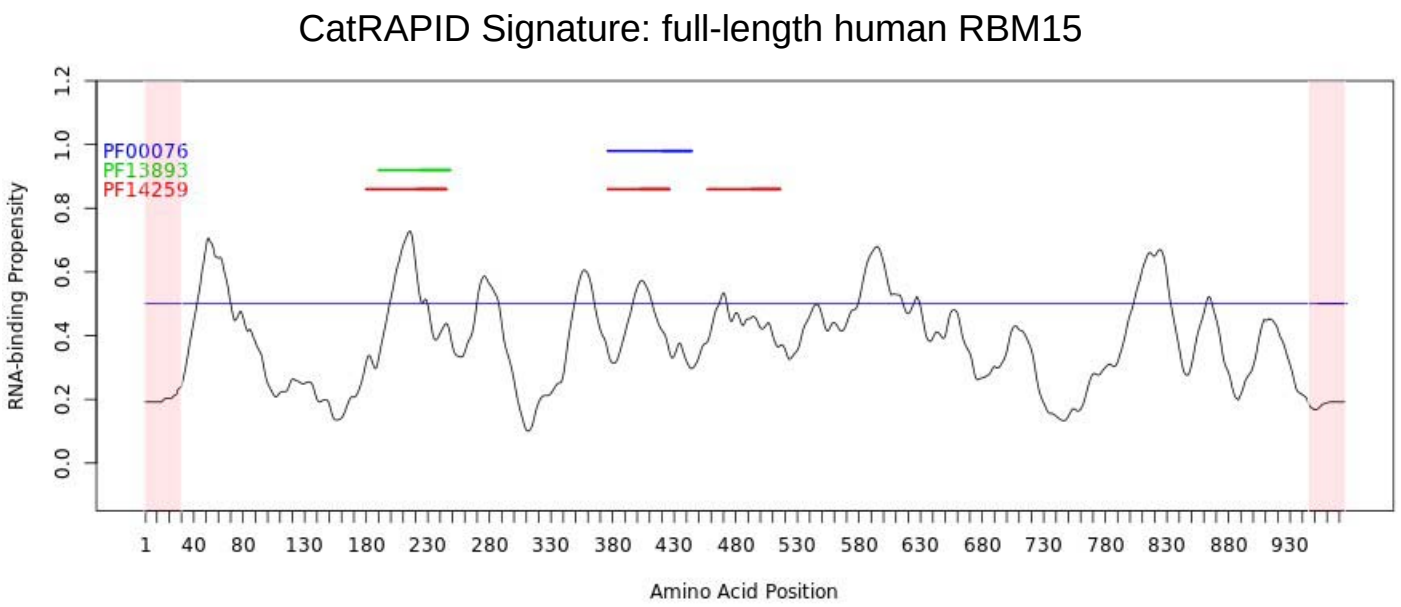

**Supplementary Figure 2.** Secondary structure of the 5-way multiway junction supported by NAI-N3 SHAPE chemical probing data for the lncRNA MIR503HG. The 5-way junction A) in chromatin *in vitro* and B) in chromatin *in vivo* are structurally identical. The 5-way junction predicted for MIR503HG in the nucleoplasm C) only shares one helical domain (blue asterisk) with structures derived from MIR503HG when localized in chromatin. Chemical probing data is mapped to each structure (gray = no data, black = low reactivity, orange = medium reactivity, and red = high reactivity).




**Supplementary Figure 3**. Secondary structure of LINC00630 (*in vitro*) based on NAI-N3 SHAPE chemical probing data. Chemical probing data is mapped to each structure (gray = no data, black = low reactivity, orange = medium reactivity, and red = high reactivity).




**Supplementary Figure 4**. Secondary structure of lncRNA FIRRE (*in vivo*) based on DMS-MaPseq chemical probing data. Chemical probing data is mapped to each structure (gray = no data, black = low reactivity, orange = medium reactivity, and red = high reactivity).




**Supplementary Figure 5**. CatRAPID Signature binding profile for RBM15. RNA binding propensity above 0.5 (horizontal line) is considered significant.

## Acknowledgments

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

## Conflicts of interest

The authors declare no conflicts of interest.

## 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 CatRAPID analysis. ANJ, EB, LE, CB, and ST carried out binding shift and SHAPE experiments. GW wrote scripts for assessing data. ANJ wrote the manuscript and ANJ, CM, SX, and EB revised the manuscript.

## 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. Fang, S. *et al.* NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res.* **46**, D308–D314 (2018).
9. Jones, A. N. *et al.* An evolutionarily conserved RNA structure in the functional core of the lincRNA Cyrano. *RNA* **26**, 1234–1246 (2020).
10. Jones, A. N., Tikhaia, E., Mourão, A. & Sattler, M. Structural effects of m6A modification of the Xist A-repeat AUCG tetraloop and its recognition by YTHDC1. *Nucleic Acids Res.* **50**, 2350–2362 (2022).
11. Kim, D. N. *et al.* Zinc-finger protein CNBP alters the 3-D structure of lncRNA Braveheart in solution. *Nat. Commun.* **11**, 148 (2020).
12. 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).
13. 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).
14. Van Nostrand, E. L. *et al.* A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 (2020).
15. Dominguez, D. *et al.* Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. *Mol. Cell* **70**, 854–867.e9 (2018).
16. 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).
17. Patil, D. P. *et al.* m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537**, 369–373 (2016).
18. 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).

19. Zolotukhin, A. S. *et al.* Nuclear export factor RBM15 facilitates the access of DBP5 to mRNA. *Nucleic Acids Res.* **37**, 7151–7162 (2009).
20. Graindorge, A. *et al.* In-cell identification and measurement of RNA-protein interactions. *Nat. Commun.* **10**, 5317 (2019).
21. Berletch, J. B., Yang, F., Xu, J., Carrel, L. & Disteche, C. M. Genes that escape from X inactivation. *Hum. Genet.* **130**, 237–245 (2011).
22. 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).
23. 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).
24. Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404–416 (2015).
25. 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).
26. 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.
27. 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).
28. Hacisuleyman, E. *et al.* Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat. Struct. Mol. Biol.* **21**, 198–206 (2014).
29. Lewandowski, J. P. *et al.* The Firre locus produces a trans-acting RNA molecule that functions in hematopoiesis. *Nat. Commun.* **10**, 5137 (2019).
30. Sun, L. *et al.* Long noncoding RNAs regulate adipogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 3387–3392 (2013).
31. 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).
32. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

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

34. 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).

35. 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).

36. Wan, Y. *et al.* Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* **505**, 706–709 (2014).

37. Wu, X. & Bartel, D. P. Widespread Influence of 3'-End Structures on Mammalian mRNA Processing and Stability. *Cell* **169**, 905–917.e11 (2017).

38. Lu, Z. *et al.* RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* **165**, 1267–1279 (2016).

39. Sun, L. *et al.* RNA structure maps across mammalian cellular compartments. *Nat. Struct. Mol. Biol.* **26**, 322–330 (2019).

40. Guo, J. U. & Bartel, D. P. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **353**, (2016).

41. Weng, X. *et al.* Keth-seq for transcriptome-wide RNA structure mapping. *Nat. Chem. Biol.* **16**, 489–492 (2020).

42. Zubradt, M. *et al.* DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods* **14**, 75–82 (2017).

43. 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).

44. Li, J. *et al.* Identification of potential autophagy-associated lncRNA in prostate cancer. *Aging* **13**, 13153–13165 (2021).

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

46. 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).

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

48. 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).

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

50. 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).

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

52. 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).

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

54. 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).

55. 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).

56. 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).

57. 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).

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

59. 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).

60. 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).

61. Karner, H. *et al.* Functional Conservation of LncRNA JPX Despite Sequence and Structural Divergence. *J. Mol. Biol.* **432**, 283–300 (2020).

62. Liu, F., Somarowthu, S. & Pyle, A. M. Visualizing the secondary and tertiary architectural domains of lncRNA RepA. *Nat. Chem. Biol.* **13**, 282–289 (2017).

63. 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).

64. Huppertz, I. *et al.* iCLIP: protein-RNA interactions at nucleotide resolution. *Methods* **65**, 274–287 (2014).

65. Bellucci, M., Agostini, F., Masin, M. & Tartaglia, G. G. Predicting protein associations with long noncoding RNAs. *Nat. Methods* **8**, 444–445 (2011).

66. 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).

67. 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).

68. 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).

69. Bridges, M. C., Daulagala, A. C. & Kourtidis, A. LNCcation: lncRNA localization and function. *J. Cell Biol.* **220**, (2021).

70. Guo, C.-J. *et al.* Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. *Cell* **181**, 621–636.e22 (2020).

71. Volders, P.-J. *et al.* LNCipedia 5: towards a reference set of human long non-coding RNAs. *Nucleic Acids Res.* **47**, D135–D139 (2019).

72. Reuter, J. S. & Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 1–9 (2010).

73. Johnson, P. Z. & Simon, A. E. RNAcanvas: interactive drawing and exploration of nucleic acid structures. *Nucleic Acids Res.* **51**, W501–W508 (2023).

74. Cheong, H.-K., Hwang, E. & Cheong, C. Rapid preparation of RNA samples using DNA-affinity chromatography and DNAzyme methods. *Methods Mol. Biol.* **941**, 113–121 (2012).

75. Yoon, S. *et al.* HiTRACE: high-throughput robust analysis for capillary electrophoresis. *Bioinformatics* **27**,

1798–1805 (2011).