

Functional Conservation of LncRNA *JPX* Despite Sequence and Structural Divergence

Heather Karner [1], Chiu-Ho Webb [1,†], Sarah Carmona [1], Yu Liu [2], Benjamin Lin [1], Micaela Erhard [1], Dalen Chan [3], Pierre Baldi [2], Robert C. Spitale [3] and Sha Sun [1]

1 - Department of Developmental and Cell Biology, *School of Biological Sciences, University of California Irvine, Irvine, CA 92697, USA*
2 - Department of Computer Science, *Institute for Genomics and Bioinformatics, University of California Irvine, Irvine, CA 92697, USA*
3 - Department of Pharmaceutical Sciences, *College of Health Sciences, University of California Irvine, Irvine, CA 92697, USA*

Correspondence to Sha Sun: *Department of Developmental and Cell Biology, University of California Irvine, 4232 McGaugh Hall, Irvine, CA 92697, USA. Fax: +1 949 824 4709. shasun@uci.edu*
https://doi.org/10.1016/j.jmb.2019.09.002
Edited by Philip C. Bevilacqua

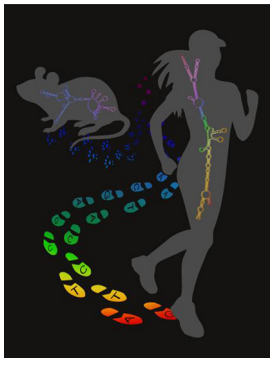
Abstract

Long noncoding RNAs (lncRNAs) have been identified in all eukaryotes and are most abundant in the human genome. However, the functional importance and mechanisms of action for human lncRNAs are largely unknown. Using comparative sequence, structural, and functional analyses, we characterize the evolution and molecular function of human lncRNA *JPX*. We find that human *JPX* and its mouse homolog, lncRNA *Jpx*, have deep divergence in their nucleotide sequences and RNA secondary structures. Despite such differences, both lncRNAs demonstrate robust binding to CTCF, a protein that is central to *Jpx*'s role in X chromosome inactivation. In addition, our functional rescue experiment using *Jpx*-deletion mutant cells shows that human *JPX* can functionally complement the loss of *Jpx* in mouse embryonic stem cells. Our findings support a model for functional conservation of lncRNAs independent from sequence and structural divergence. This study provides mechanistic insight into the evolution of lncRNA function.

© 2019 Elsevier Ltd. All rights reserved.



Heather Karner, Pierre Baldi, Robert C. Spitale and Sha Sun



Legend: *The structures of mouse lncRNA Jpx and human lncRNA JPX within their respective species. The footprints represent their evolutionary journey, with human JPX rapidly evolving away from mouse Jpx. Illustration by Lilly Li & Shannon J. Carmona.*

Introduction

Long noncoding RNAs (lncRNAs) are transcripts over 200 nucleotides in length that do not code for proteins. In contrast to protein-coding transcripts, which map to only about 1.5% of the human genome, DNA sequences for lncRNA transcripts are estimated to represent 70%–90% of the genome [1]. Owing to their low levels of expression and the general lack of functional information, lncRNAs were regarded as transcriptional noise. Only recently, the high frequency of their occurrence in the human genome and their direct relevance to various biological processes have been recognized [2–6]. Specifically, lncRNAs are known to be capable of scaffolding protein complexes and recruiting chromatin modifiers for transcriptional regulation [6–8]. Gene expression profiling has revealed highly tissue-specific transcription of lncRNAs, and a large number of lncRNAs that are active during animal development in humans, mice, flies, and farm animals [9–12]. Importantly, lncRNAs have been implicated in the evolution of new genes and associated with functions in sexual reproduction [11,13–15]. Furthermore, lncRNA functions have been shown to be conserved during embryonic development [16,17].

LncRNAs are present in all eukaryotes; however, the function and evolution of the vast majority of lncRNA genes still remain elusive [18–21]. Unlike protein-coding genes, in which functions are mostly defined by evolutionary conserved coding sequences and their flanking regulatory elements, lncRNAs are known to have poor sequence conservation [22–24]. Hence, it has been a challenge to uncover conserved features of lncRNAs and determine underlying mechanisms for function. RNA secondary structure is one molecular feature that has recently been recognized to be important for lncRNA function [25–30]. Yet, the direct connection between the structure and function of lncRNAs and their implications in molecular evolution remain unclear [29,31]. Defining this connection has been notoriously difficult because of the lack of structural and biochemical analysis of evolutionarily related lncRNAs and the complex nature of their interactions with protein factors.

In this article, we focus on lncRNAs involved in a mechanism of dosage compensation known as X chromosome inactivation (XCI) to determine whether function is conserved and if molecular features such as RNA sequence and secondary structure influence conservation. XCI is the evolutionary solution to the 1X:2X dosage imbalance between XY male and XX female mammals. Outside the lineage of modern mammals, different mechanisms are used to balance the sex chromosome gene dosage. Interestingly, dosage compensation has evolved independently in divergent species and frequently uses lncRNAs as key regulators. Examples include *roX1* and *roX2* for Drosophila, *Rsx* for opossum, and *Xist*/*XIST* for mouse and human [32–35]. This suggests that RNA, unlike protein, carries functional advantages that aid, and are sometimes prerequisites for, biological processes such as gene dosage controls in developing embryos. XCI is controlled by a genomic region known as the *X inactivation center* (*Xic*) that encodes a cluster of lncRNAs in both human and mouse genomes (Fig. S1) [36,37]. This gene cluster has evolved from a group of protein-coding genes during the divergence of eutherians and marsupials to become the home of all lncRNAs involved in XCI [38–41]. It is worthwhile to note that marsupials (e.g., opossums; Fig. S1B) do not have the master regulator lncRNA *Xist* for XCI. In the same chromosome locus, marsupials carry the protein-coding gene *Lnx3*, which does not possess dosage compensation functions or affect the sex chromosome [38,39]. However, a marsupial lncRNA, *Rsx*, was discovered to play the role of silencing the X chromosome in opossums. Although *Rsx* has no obvious sequence homology with *Xist*, the function of the two genes appears to be equivalent [35,42,43]. Thus, the use of lncRNAs for control of mammalian XCI represents convergent evolution of functions that may be independent of RNA nucleotide sequences.

We took advantage of a defined molecular mechanism in mouse XCI which involves the direct binding of a lncRNA known as *Jpx* with a specific chromatin insulator protein, CTCF, to initiate XCI (Fig. 1A) [44]. This model allowed us to identify the molecular features underlying the function and evolution of *Jpx*. In mice, *Jpx* has been shown to activate *Xist* [44–47]. By contrast, the function of its human homolog, *JPX*, is unknown [48,49], though we do know that it is expressed in early female human embryos (Fig. S2A) [50]. This indicates that the gene has a role in early embryogenesis and likely functions similarly to *Jpx*. Here we will compare mouse lncRNA *Jpx* with human lncRNA *JPX* and determine their homology at the levels of nucleotide sequences, RNA secondary structures, and molecular functions. Our results indicate that despite sequence and structural divergence, the two lncRNAs function through the same biochemical mechanism.

Results

Comparative sequence analysis suggests that human lncRNA *JPX* is functional

The molecular features and functional roles of human lncRNA *JPX* were not previously known. The

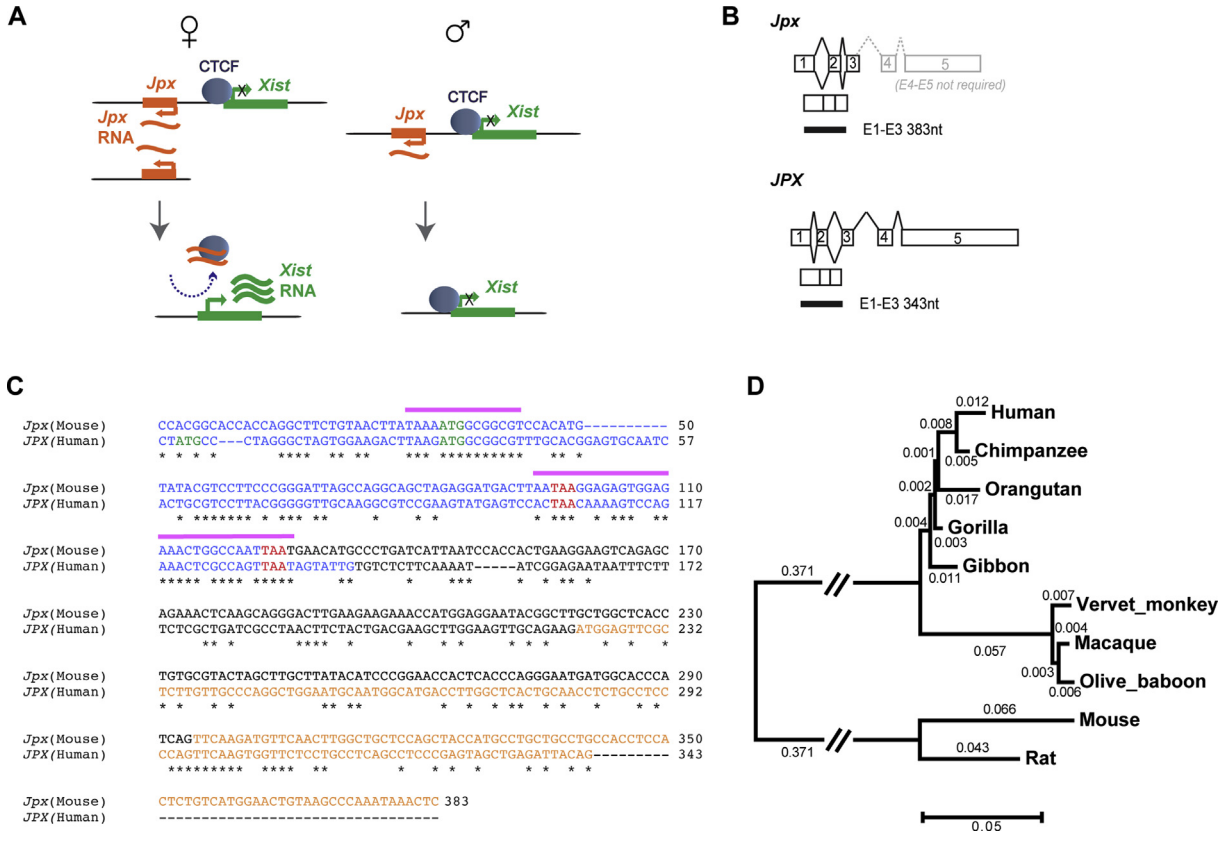


Fig. 1. Comparative sequence analysis of mouse lncRNA *Jpx* and human lncRNA *JPX*. (A) Model of mouse lncRNA *Jpx* mechanism. LncRNA *Jpx* is transcribed upstream of *Xist*, removes CTCF from the *Xist* promoter, and activates *Xist*. This activation is dose-dependent—*Jpx* titrates away CTCF only when present in twofold excess, such as in female cells, and is insufficient to activate *Xist* in male cells [44]. (B) Mouse *Jpx* (Top) and human *JPX* (Bottom) gene structures and transcript isoforms. The 383nt transcript of mouse *Jpx E1–E3* is required for function [44], which corresponds to the 343nt transcript of human *JPX E1–E3*. See also Figs. S1 and S2. (C) Sequence alignment of mouse *Jpx* (top) and human *JPX* (bottom) transcripts; exon 1 (blue)—exon 2 (black)—exon 3 (orange) analyzed by Clustal 2.1 [90] and the alignment was manually adjusted. Asterisks (*) mark matching nucleotides. Pink bars label the highly conserved regions between *Jpx* and its ancestral protein homolog *UspL* [21]. "ATG" in green and "TAA" in red mark potential start and stop codons, respectively. (D) Evolutionary relationships of taxa analyzed by MEGA7. Sequences were obtained from the UCSC whole genome assemblies and the evolutionary history was inferred using the Neighbor-Joining method [60]. The tree is drawn to scale, with branch lengths (next to the branches) in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the maximum composite likelihood method [61] and are in the units of the number of base substitutions per site.

GenBank annotation for human lncRNA *JPX* shows that the gene (Gene ID: 554203; RNA sequence: NR_024582) contains five exons, which are similar to its mouse homolog, *Jpx* (Gene ID: 70252) (Fig. 1B), which has also been shown to be composed of five exons, with the transcripts containing exons 4 and 5 being minor isoforms [45,51–53]. Using total RNA from human ovarian cancer SKOV3iP1 cells, we were able to isolate a primary *JPX* transcript spanning exons 1–3, *JPX E1–E3*. In mice, it has been shown that the corresponding *Jpx E1–E3* is the primary isoform responsible for the function of *Jpx* in XCI [44,54]. More precisely, nucleotides 1-383 of the *Jpx E1–E3* sequence are necessary and sufficient for mouse lncRNA *Jpx* binding to CTCF [44]. As lncRNA *JPX E1–E3* was detected in human cells, we suspected conservation of gene structure and nucleotide sequence between *JPX* and *Jpx*. To characterize their sequence homology, we performed a pairwise sequence alignment between the critical 383nt mouse *Jpx E1–E3* with the full-length (343nt) human *JPX E1–E3* (Fig. 1C). Despite an overall similarity of gene structure, including five exons in both *Jpx* and *JPX* (Fig. 1B), the exact sequence identity of exons 1–3 is approximately 40%, which is much less than the average nucleotide sequence identity (85%) for protein-coding sequences between humans and mice (one-tailed binomial test, $P < 10^{-6}$) [55,56]. Interestingly, exon 1 of *Jpx* and *JPX* both contain remnants of protein-coding sequences similar to the

chicken *UspL*, which have recently been reported as possible regulatory sequences for lncRNA function [21].

Although a lack of sequence conservation is not surprising for noncoding genes [57–59], it is unknown how lncRNAs evolve and how nucleotide changes affect the lncRNA function. Taking advantage of fully sequenced genomes of multiple vertebrates, we searched for other homologous sequences of *Jpx E1–E3* and *JPX E1–E3* in the UCSC genome assemblies. Appropriate exon sequences for alignment were obtained from primates and murine rodents, which allowed us to look into the phylogenetic history of *Jpx*. Using the Neighbor-Joining method [60], we constructed the evolutionary tree for *Jpx E1–E3* and *JPX E1–E3* in ten species (two rodents and eight primates), with evolutionary distances calculated from maximum composite likelihood method (Fig. 1D) [61]. The rate of *Jpx* sequence evolution between mouse and rat is 0.109 substitution per site, which is ~44% lower than the estimated neutral evolution rate of 0.196 between these two rodents (one-tailed binomial test, $P = 0.000014$) [56,62]. This is consistent with our understanding that *Jpx* has an important functional role in mice, and thus the sequence changes have been under substantial constraint in the rodent lineages. By contrast, the evolutionary distance between rodent *Jpx* and human *JPX* is 0.796 substitution per site, which is ~74% higher than the neutral rate of 0.457 between humans and rodents (one-tailed binomial test, $P < 10^{-6}$) [62]. A more rapid nucleotide substitution between human and rodent suggests positive selection acting on the sequence of human *JPX*.

We then compared our *Jpx/JPX* gene tree with the recently published species tree of primates. The neutral substitution rates in the lineages leading to the hominoid (human, chimpanzee, and orangutan), as estimated from the common ancestor between hominoids and Old World Monkeys (vervet monkey, macaque, and olive baboon), are within the range of 0.026–0.027 [63]. By contrast, the corresponding evolutionary rates of the *JPX* gene in the same lineages appear to be more variable, ranging from 0.009 to 0.027. Importantly in the human lineage, the nucleotide change rate is 0.012, which is twofold higher than the neutral substitution rate of 0.0058 for humans [63] (one-tailed binomial test, $P = 0.13$). A larger than twofold difference in the substitution rates is seen between the human and chimpanzee branches (bootstrap 80% over 500 replicates), which is notable given that the rates of evolution on these two lineages are estimated to be very similar with only a 1.9% difference [63]. Such observations suggest that adaptive nucleotide sequence changes have occurred in the hominoid lineages, which are supportive of a functional *JPX*, particularly in the human lineage.

Consistent with a possible role of *JPX* in regulating *XIST* within humans, reanalysis of available single-cell RNA-seq data revealed a positive correlation between *JPX* and *XIST* expression levels in human preimplantation embryos, especially in female cells of the epiblast lineage (Fig. S2A; raw data obtained from Ref. [50]). Such observations suggest that human lncRNA *JPX* likely functions as a positive regulator of *XIST*. In addition, data from the Genotype-Tissue Expression (GTEx) Project showed sex-dependent expression and a positive correlation between *JPX* and *XIST* expression across 51 female samples (Pearson's correlation $r = 0.783$, $P < 10^{-6}$) (Fig. S2B). Within specific tissue types, *JPX* and *XIST* activities in individual samples also showed positive correlations: $r = 0.784$ for the breast tissues ($n = 290$, $P < 10^{-6}$); $r = 0.690$ for the pituitary tissues ($n = 183$, $P < 10^{-6}$).

These results suggest that human lncRNA *JPX* is functionally important and that a detailed analysis of *Jpx/JPX* would provide a novel experimental model to understand conservation of function despite sequence diversity.

RNA structural probing reveals divergence of *Jpx/JPX* homologous lncRNAs

If human lncRNA *JPX* shares similar function with its mouse homolog, it is possible that there is conservation at the RNA structural level despite a nucleotide sequence divergence. Such conservation has been supported by previous RNA structure–function studies on well-characterized noncoding RNAs, such as ribozymes and riboswitches, but thus far there has been limited analysis for lncRNAs [24,26,64]. To explore how the sequence determines the secondary structure of RNA, we performed selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) RNA structural probing [65,66]. For mouse *Jpx*, we focused on the functional lncRNA transcript *Jpx 34-347* and designed the reverse primers spanning the 314nt sequence (Fig. 2A). Extension primers were also designed to probe the human lncRNA transcript *JPX 1-343* (Fig. 2B). We used the SHAPE reagent 2-methylnicotinic acid imidazolide (NAI) to modify structured *in vitro* transcribed RNA and map to residues that are accessible, such as unpaired or flexible bases (Fig. 2C). As has been reported [26,65], sites of nucleotide modification can be identified as stops to primer extension by reverse transcriptase. Using radiolabeled reverse primers to generate cDNAs from NAI-treated or DMSO-treated RNAs, we were able to resolve the modified unpaired bases by running the reverse transcribed cDNAs through denaturing gel electrophoresis for sequencing. The intensities of the gel bands are positively correlated with the NAI modification strengths and thus reveal features of the RNA secondary structures at single-base resolution. In the

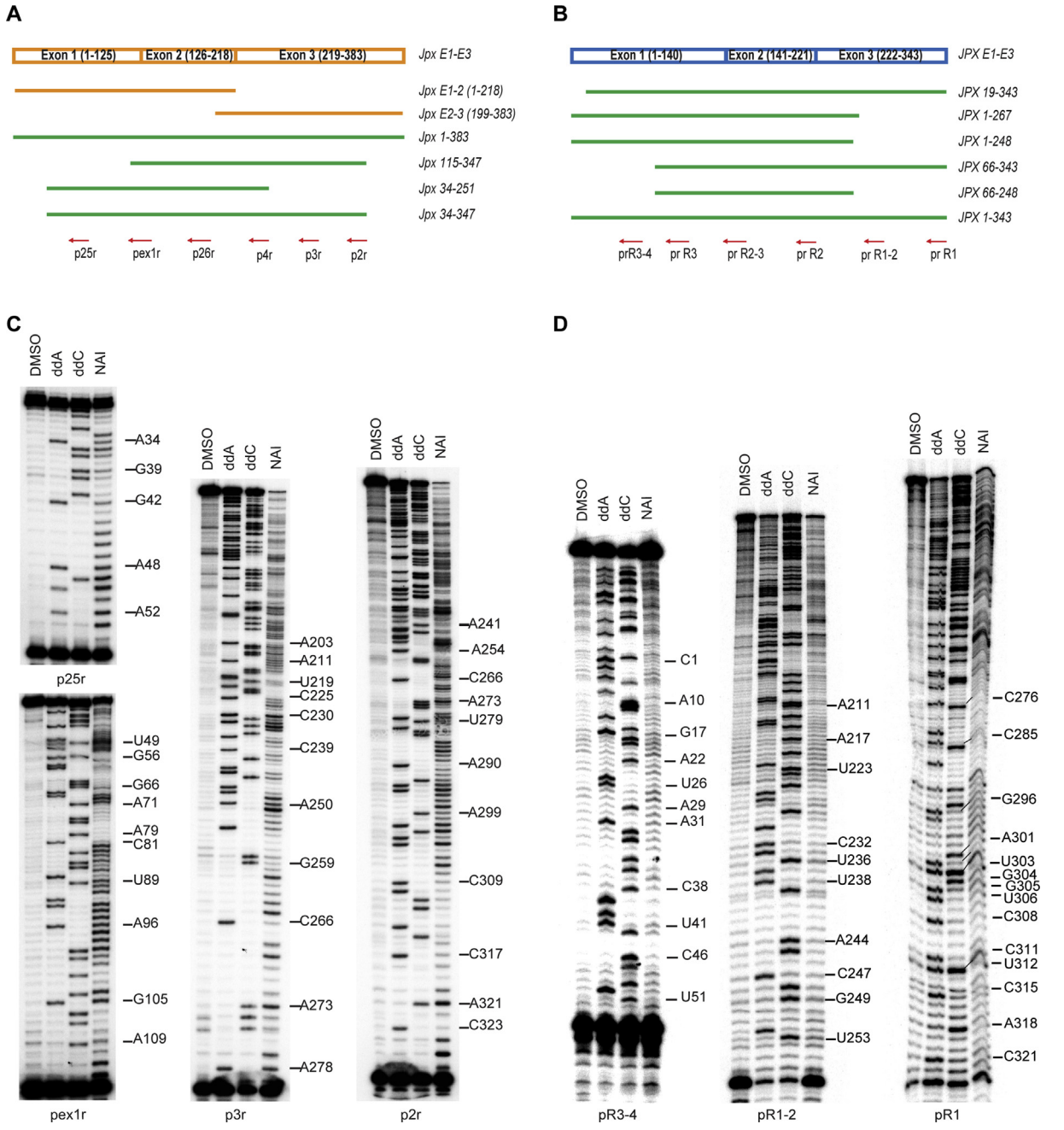


Fig. 2. Functional domain mapping and RNA structure probing. (A) Different length *in vitro* transcribed RNA corresponding to the coverage of *Jpx*. Mouse lncRNA 220nt E1−E2 or 183nt E2−E3 (orange) is not sufficient for protein binding *in vitro* [44]. The mouse 383nt functional *Jpx* transcript and its truncated forms (green) are assayed with RNA EMSA. Red arrows indicate positions of the reverse primers used in SHAPE. (B) Different length *in vitro* transcribed RNA corresponding to the coverage of *JPX*. A full-length 343nt *JPX* transcript and its truncated forms (green) are assayed with RNA EMSA. Red arrows indicate positions of the reverse primers used in SHAPE. (C) Mouse *Jpx* RNA structure probing by SHAPE: polyacrylamide gel electrophoresis (PAGE) resolves RNA footprint after treatment of RNA by either DMSO (control) or SHAPE modification reagent NAI [66], followed by RNA reverse transcription using primers indicated and RNA hydrolysis. (D) Human *JPX* RNA structure probing by SHAPE: as described in C. (C and D) At least two replicates were performed for each reaction and representative gel images are shown. Band intensity and corresponding nucleotide positions were integrated with SAFA software [88,89]. SHAPE reactivities reflect single-stranded (highly reactive) and double-stranded (not reactive) states at individual nucleotides. Nucleotide labels correspond with NAI-modified nucleotides.

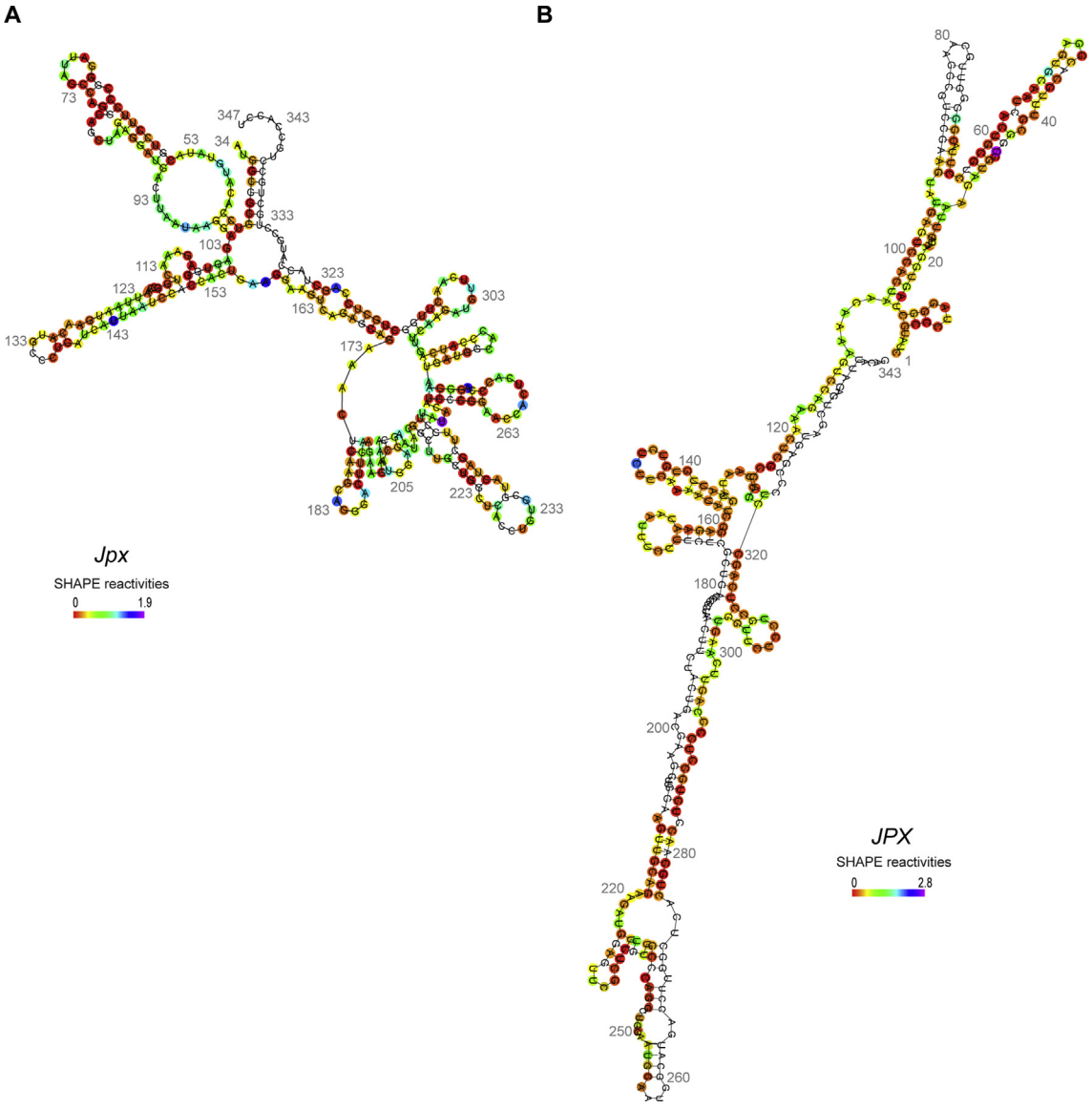


Fig. 3. RNA secondary structures of mouse *Jpx* and human *JPX* derived from SHAPE reactivity. (A) SHAPE reactivities at individual nucleotides of mouse *Jpx* were normalized to a scale of 0−1.9. (B) SHAPE reactivities of human *JPX* were normalized to a scale of 0−2.4. (A and B) Scales were denoted with color codes at individual nucleotides. The secondary structures were drawn with the RNAprobing web server (http://rna.tbi.univie.ac.at) based on the fold algorithms in Lorenz et al. and Washietl et al. [67,68]. See also Fig. S3.

SHAPE profiles for the 5′ and 3′ regions of *Jpx 34-347*, we observed segments of the base pairing in nucleotides G56-G66 and A71-C81 (Fig. 2C, pex1r panel), and in nucleotides C309-C317 (Fig. 2C, p2r panel), suggesting possible stem-loops at these sites. Similarly for the 5′ and 3′ regions of *JPX 1-343*, we observed the base pairing in nucleotides C38-C46 (Fig. 2D, pR3-4), and in nucleotides C276-C285 and U306-C311 (Fig. 2D, pR1), which suggest corresponding stem-loop features.

To systematically analyze the lncRNA SHAPE profiles, we calculated the SHAPE reactivity for each nucleotide after measuring band intensity, background (the DMSO lane) subtraction, and normalization [26]. By inputting all the SHAPE reactivity values into the ViennaRNA program for RNA secondary structure [67,68], we were able to derive the most likely structure for *Jpx 34-347* and *JPX 1-323* based on a linear log model for pairing probabilities [69]. As illustrated in Fig. 3A, mouse *Jpx 34*-347 RNA contains multiple stem-loops. Overall, about 50% of the nucleotides are base-paired, indicating that *Jpx 34*-347 RNA is highly structured. Both the 5′ and 3′ nucleotides, 34-114

and 252-347, respectively, are involved in stem-loop formation, suggesting possible secondary structural configurations necessary for function. We asked what structural features human *JPX* may share with mouse *Jpx*. As shown in Fig. 3B, *JPX 1*-343 RNA also contains stem-loops with more than 50% of paired bases. However, the overall structure is obviously different from the mouse *Jpx* RNA structure (Fig. 3).

To determine whether *in vitro* conditions may limit the RNA secondary structural probing, we performed SHAPE analysis *in vivo* with *JPX* lncRNA modified by NAI in human embryonic kidney (HEK293) cells. We note that *in vivo* RNA structural probing is particularly difficult on low abundance RNA for which results can be confounded because of the presence of more abundant RNA in the same sample [70,71]. To enrich *JPX* for *in vivo* SHAPE profiling, we adapted a cDNA amplification step using LMPCR (ligation-mediated polymerase chain reaction) which has shown to be instrumental for *in vivo* RNA structural probing [70,72]. We chose the sequence domain *JPX 104-172*, which shares 50% nucleotide identity and corresponds to the mouse *Jpx* sequence essential for function [44]. A direct comparison between *in vitro* and *in vivo* SHAPE profiles of *JPX 104-172* showed overlapping segments representing single-stranded RNA domains (Fig. S3A). There is an overall 70% exact matching between the *in vitro* and *in vivo* RNA structural predictions (Fig. S3B), with 76% (22 out of 29) single-stranded nucleotides from the *in vivo* structure falling into the loop regions predicted by *in vitro* SHAPE reactivities (Fig. S3C), supporting that *in vitro* SHAPE profiling is instructive to determine RNA secondary structural features. Based on the large differences revealed by the *in vitro* SHAPE reactivities for *JPX E1–E3* and *Jpx E1–E3* (Fig. 3), we conclude that human lncRNA *JPX* has diverged from mouse lncRNA *Jpx* in their overall secondary RNA structures.

LncRNA-protein binding *in vitro* demonstrates sequence requirements for RNA function

Given the low conservation of nucleotide sequences and RNA secondary structures, we asked whether specific domains or motifs might be

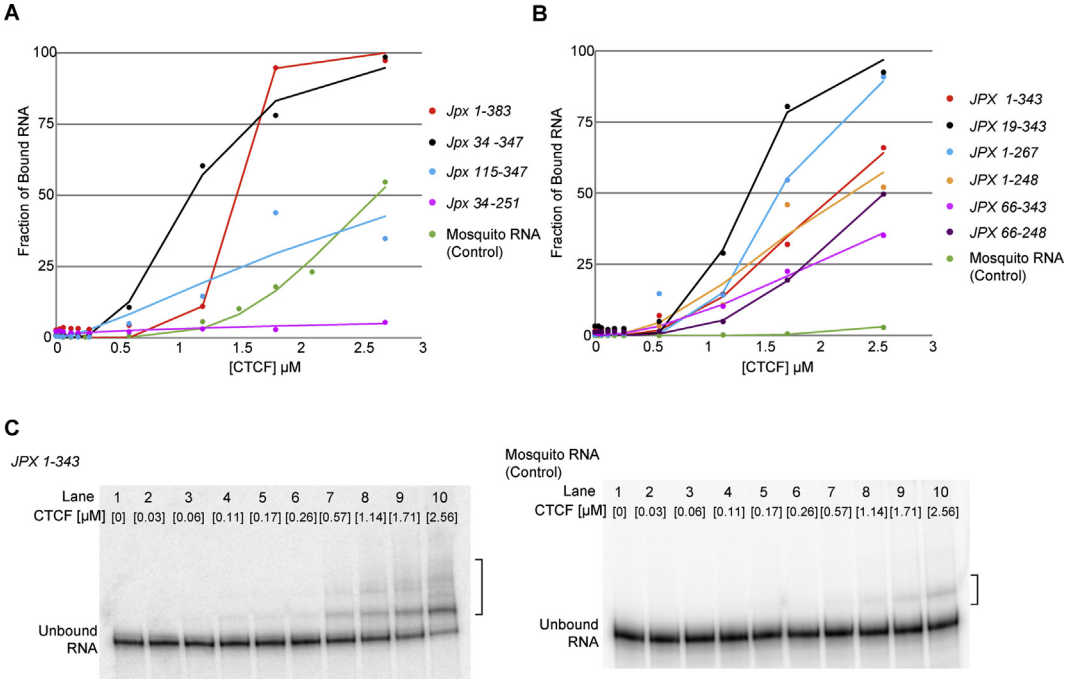
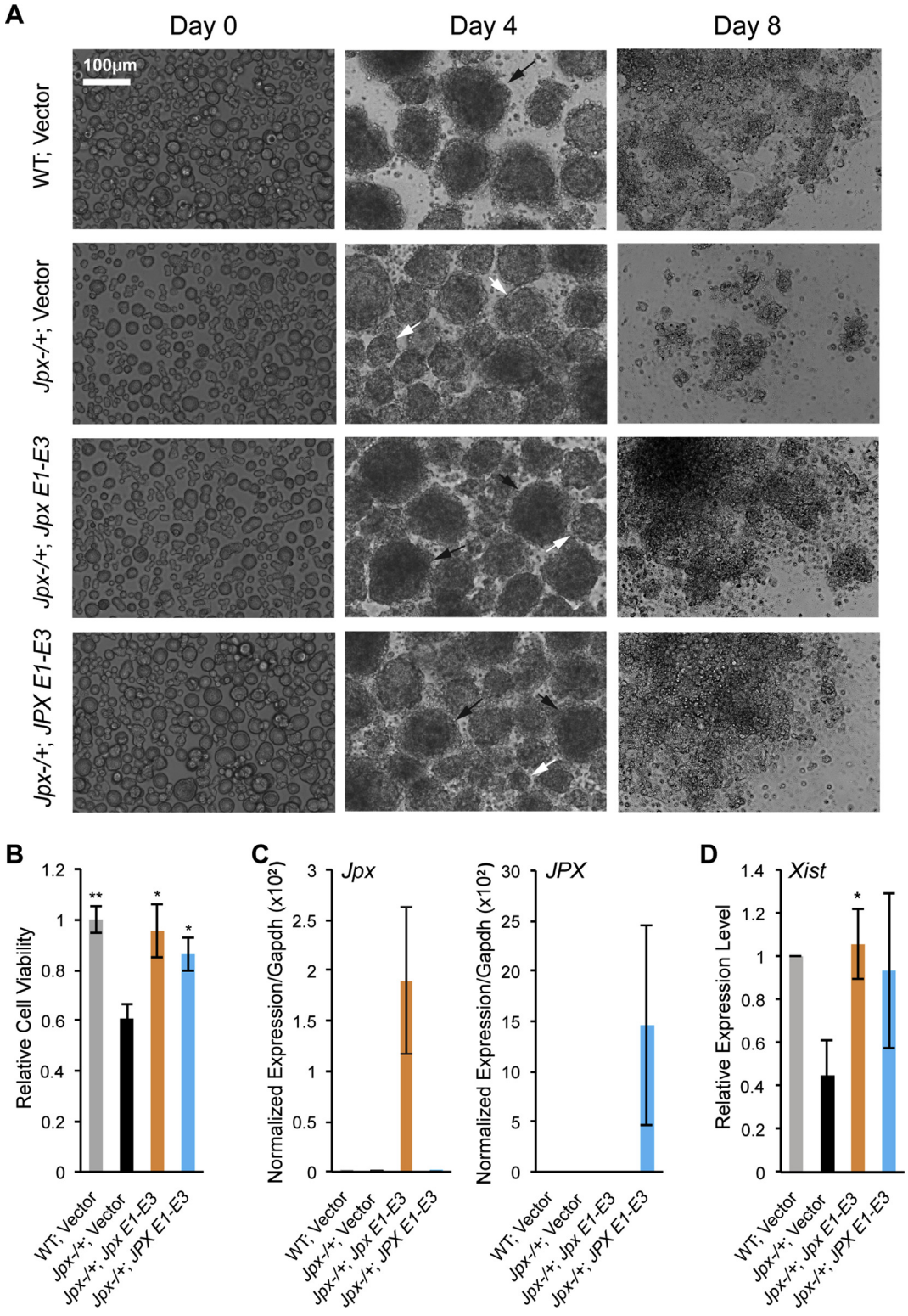


Fig. 4. Mouse *Jpx* RNA and human *JPX* RNA are capable of binding to CTCF. (A) Binding isotherm for CTCF–*Jpx* from RNA EMSA. Binding curve was plotted as the percent bound against CTCF concentration and was fit by a nonlinear regression to a binding isotherm. The *Jpx 1-383* (red) and the truncated *Jpx 34-347* (black) both show robust binding, whereas 5′-truncated *Jpx 115-347* (blue) and 3′-truncated *Jpx 34-251* (pink) show weak or nonspecific binding as compared with a 316nt control RNA (green) from the malaria mosquito, *Anopheles gambiae* [64]. (B) Binding isotherm for CTCF–*JPX* from RNA EMSA. All *JPX* RNA of different lengths demonstrated favorable binding as compared with the 316nt control RNA (green). The 5′-truncated *JPX 66-343* (pink) and *JPX 66-248* (purple) showed weaker binding than the full-length *JPX 1-343* (red). (C) Representative RNA EMSA gel image detecting direct binding of *JPX* RNA and CTCF protein *in vitro*. Left panel: binding of *JPX 1*-343 RNA with CTCF protein at increasing concentrations. Right panel: binding of a 316nt control RNA (from the malaria mosquito *Anopheles gambiae*) [64] with CTCF protein at increasing concentrations. RNA-protein shift is indicated by the bracket on the right side of the gel images. See also Fig. S4.


A
Day 0
Day 4
Day 8
100μm
WT; Vector
Jpx-/+; Vector
Jpx-/+; Jpx E1-E3
Jpx-/+; JPX E1-E3
B
Relative Cell Viability
C
Jpx
Normalized Expression/Gapdh (x10²)
JPX
Normalized Expression/Gapdh (x10²)
D
Xist
Relative Expression Level
WT; Vector
Jpx-/+; Vector
Jpx-/+; Jpx E1-E3
Jpx-/+; JPX E1-E3

required for the molecular functions of *Jpx* and *JPX* lncRNAs. Using the known molecular interaction between mouse *Jpx* and CTCF, we performed an *in vitro* RNA electrophoretic mobility shift assay (EMSA) and tested the binding capacity of *Jpx* RNA with regard to various truncation forms (Fig. 2A). It has been reported that truncated *Jpx* RNAs, 220nt *Jpx E1−E2* and 183nt *Jpx E2−E3*, failed to bind CTCF [44], suggesting that both halves of *Jpx E1−E3* are needed for it to function. We then removed segments of the 5′ and 3′ ends of *Jpx E1−E3* and characterized the binding kinetics of mutant *Jpx* RNAs in comparison to the 383nt *Jpx E1−E3*. For negative control, we used a 316nt *drz-Agam-2-1* ribozyme RNA from *Anopheles gambiae* [64]. As shown in Fig. 4A, with increasing CTCF concentration, the full-length *Jpx 1-383* (red) and the truncated *Jpx 34-347* (black) both showed robust binding. By contrast, the truncated *Jpx 115-347* (blue) and *Jpx 34-251* (pink) failed to bind CTCF. We conclude that both the 5′ sequence of nucleotides 34-114 and the 3′ sequence of nucleotides 252-347 are required for *Jpx* to bind CTCF, and that these regions may be responsible for its function.

We then asked whether the molecular function of human *JPX* has diverged from mouse *Jpx* and performed EMSA on *JPX 1*-343 RNA for its CTCF-binding capacity *in vitro* (Figs. 4B−C and S4). We used purified recombinant CTCF with the full-length mouse CTCF protein of 736 amino acids, which is a 98% match with the human CTCF protein. In vertebrates, CTCF is highly conserved and functioning as a global transcriptional regulator in all cell types [73,74]. Therefore, whether human *JPX* binds CTCF the same as mouse *Jpx* would inform functional importance and conservation at the molecular level. As shown in EMSA with increasing concentration of CTCF protein, *JPX 1*-343 RNA was capable of binding the CTCF protein and was robustly shifted by CTCF (Fig. 4C, left panel). By contrast, the 316nt control RNA of mosquito ribozyme showed weak interaction only at the highest concentration of CTCF (Fig. 4C, right panel). We looked further into the binding kinetics of CTCF against various truncation forms of human lncRNA *JPX* (Figs. 2B & 4B) with the goal of identifying RNA sequence domains critical for binding. As shown in Fig. 4B, all *JPX* RNA truncations were able to bind CTCF in comparison with the control RNA (green). Interestingly, the 5′ truncation form, *JPX 19-343* (black), and the 3′ truncation form, *JPX 1-267* (blue), showed stronger binding than the full-length *JPX 1-343* (red). Our protein-binding assays therefore indicate that human *JPX* RNA is capable of binding CTCF, and that such a CTCF−*JPX* interaction can be robust against removal of *JPX* 5′ or 3′ RNA sequences. Overall, the EMSA results demonstrate that human *JPX* RNA has maintained, or may have even reinforced, its molecular binding capacity with CTCF protein.

Consistent with our observation of direct RNA-protein binding of *Jpx*−CTCF and *JPX*−CTCF *in vitro*, genome-wide studies have reported CTCF-RNA interactions in both mouse and human cells [75,76]. Notably, mouse *Jpx* RNA was identified as one of the locus-specific interacting RNAs of CTCF in mouse embryonic stem cells by CLIP-seq [76]. In addition, we also found human *JPX* present as one of the RNA transcripts pulled down with human CTCF by PAR-CLIP from human bone osteosarcoma U2OS cells (Supplementary data from Ref. [75]). Moreover, RNA-binding regions in CTCF have been recently reported and shown to be essential for the molecular interaction and function of CTCF in mouse and human cells [77,78], which confirms CTCF binding to endogenous RNAs and the functional importance of CTCF-RNA binding in gene regulation. Our evidence of direct binding of CTCF to human *JPX* RNA, similar to CTCF−*Jpx* RNA interaction in the mouse, thus supports the functional importance and conservation of human *JPX* to its mouse homolog.

Rescue of *Jpx*−/+ mES cell viability and morphology by a human *JPX* transgene indicates functional complementation

Since mouse lncRNA *Jpx* activates XCI through its binding to CTCF (Fig. 1A), we therefore asked whether human lncRNA *JPX* could function

Fig. 5. Rescue of *Jpx*-deletion in mouse ES cells by human lncRNA *JPX*. (A) Overexpression of *Jpx* and *JPX* RNA rescues outgrowth defect in *Jpx*−/+ female mutant mES cells. Wild-type (WT) control and *Jpx*−/+ female mES cells transfected with vector only, Tg(EF1α:JpxE1−E3), or Tg(EF1α:JPXE1−E3). Representative brightfield images are shown of cultures on day 0, 4, and 8 of mES differentiation. Black arrows indicate normal EBs present in cultures. White arrows indicate disintegrating EBs in cultures. (B) Rescue of the cell viability defect in the *Jpx*−/+ mutant cells. At least three independent transfections were performed and average viability ± SEM is shown (*, $P < 0.05$ and **, $P < 0.01$ from one-tail paired Student *t*-tests in comparison to "*Jpx*−/+; Vector"). (C) Overexpression of *Jpx* and *JPX* achieved by Tg(EF1α:JpxE1−E3) and Tg(EF1α:JPXE1−E3), respectively, in the *Jpx*−/+ mutant mES cells. At least three independent transfections were performed and qRT-PCR of *Jpx* expression (Left panel) and *JPX* expression (Right panel) was normalized to *Gapdh* mRNA. Average expression ± SEM is shown. (D) *Xist* RNA expression rescued by *Jpx* RNA overexpression. The qRT-PCR of *Xist* expression was normalized to *Gapdh* mRNA and is shown relative to the WT level (set to "1"). Average expression ± SEM is shown from at least three independent transfections (*, $P < 0.05$ and **, $P < 0.01$ from one-tail paired Student *t*-tests in comparison with "*Jpx*−/+; Vector"). See also Figs. S5−S7.

equivalently for XCI. Using a functional complementation test, we addressed whether the *in vivo* function of human lncRNA *JPX* is equivalent to mouse lncRNA *Jpx*. We used *Jpx*−/+ female mouse embryonic stem (mES) cells for this purpose, an established *Jpx*−/+ heterozygous knockout cell line. As previously reported, *Jpx*−/+ female mES cells die during cell differentiation because of failed XCI, which is associated with morphology defects and a loss of *Xist* transcription [45]. By introducing Tg(EF1α:JPXE1−E3) transgene expressing *JPX E1−E3* transcript into *Jpx*−/+ mutant female mES cells, we asked whether the cell viability defects caused by loss of mouse lncRNA *Jpx* could be rescued by expression of human lncRNA *JPX*. As a reference in parallel, we also introduced Tg(EF1α:JpxE1−E3) transgene expressing *Jpx E1−E3* into *Jpx*−/+ female mES cells. As shown in Fig. 5A (second row), *Jpx*−/+ female mES cells receiving the vector-only transgene were dying during mES differentiation, displaying irregular and disaggregated embryoid bodies (EBs) at Day 4 (white arrows). These phenotypes were rescued by transiently transfected transgenes overexpressing *Jpx E1−E3* (third row) or *JPX E1−E3* (bottom row), with which the mutant cells formed better EBs at Day 4 and showed less dissociated cells. At Day 8 of mES differentiation, the rescue effects were most obvious. Although wild-type female mES cells receiving the vector-only transgene showed EB attachment and cell outgrowth, the mutant *Jpx*−/+ had no attached EBs or cell outgrowth, and instead, had mostly disintegrated EBs and floating cells in the media. By contrast, *Jpx*−/+ female mES cells receiving either *Jpx E1−E3* or *JPX E1−E3* showed clearly attached EBs and cell outgrowth comparable with the wild-type control cells, suggesting complete reversal of cell lethality.

To validate the observed differences in EB morphology, we quantified the cell viability at Day 4. As shown in Fig. 5B, mutant *Jpx*−/+ mES cells showed viability defect with ~40% reduction as compared with wild-type female control cells during ES cell differentiation (one-tailed Student *t*-test, $P < 0.01$). In comparison, mutant *Jpx*−/+ mES cells receiving *Jpx E1−E3* or *JPX E1−E3* were rescued, and cell viabilities in both cases were significantly elevated (one-tailed Student *t*-test, $P < 0.05$), reaching 96% and 86% of the wild-type level, respectively. To confirm that the phenotypic rescue of *Jpx*−/+ mutant female cells was a response to the transgene expression, we assayed the *Jpx* and *JPX* RNA quantities in these transfected *Jpx*−/+ mutant female mES cells (Fig. 5C, Day 8 shown). Mutant *Jpx*−/+ mES cells receiving *Jpx E1−E3* strongly expressed mouse *Jpx* but not human *JPX*, and cells that had received *JPX E1−E3* were the only ones strongly expressing human *JPX* RNA (Fig. 5C). We did not observe any adverse defects when the same transgenes were expressed in wild-type female mES cells (Fig. S5). EB morphology and outgrowth were normal, and cell viabilities were comparable with wild-type control cells carrying the empty vector (Fig. S5A−B). At the molecular level, the *Jpx E1−E3* transgene in wild-type female mES cells induced higher *Xist* expression (Fig. S5C, one-tailed Student *t*-test, $P < 0.01$), consistent with previous reports on the *trans* activation role of mouse *Jpx* on *Xist* [44,47]. By contrast, the *JPX E1−E3* transgene did not enhance *Xist* expression in wild-type female mES cells, presumably because of the presence of intact endogenous mouse *Jpx* in these cells (Fig. S5C).

A heterozygous *Jpx* deletion in the mouse female ES cell compromises overall *Jpx* expression leading to reduced *Xist* expression during ES differentiation [45]. As shown in Fig. 5D, mutant *Jpx*−/+ mES cells at differentiation Day 8 had an overall lower level of *Xist* transcripts (45% of the wild-type level). Expression of *Jpx E1−E3* transgene fully rescued *Xist* expression to 105% of the wild-type level in *Jpx*−/+ mES cells (one-tailed paired Student *t*-test, $P = 0.002$), consistent with the *trans* activation role of mouse *Jpx* on *Xist* [44,47]. Expression of the human *JPX E1−E3* transgene in *Jpx*−/+ mES cells rescued *Xist* expression to 93% of the wild-type level (one-tailed paired Student *t*-test, $P = 0.108$). This is above and beyond the confidence interval of *Xist* expression in *Jpx*−/+ female mES cells (13%−76% of the wild-type level), indicating that human *JPX E1−E3* RNA is capable of activating *Xist* to complement the loss of mouse *Jpx* RNA in the mES cells. Taken together, our results support that human lncRNA *JPX* is functionally homologous to mouse lncRNA *Jpx* in their molecular roles affecting XCI.

Discussion

Our analyses comparing human lncRNA *JPX* with its mouse homolog lncRNA *Jpx* reveal large differences in their nucleotide sequences and RNA structures, but nevertheless a conservation in their molecular functions—both human *JPX* and mouse *Jpx* lncRNAs bind CTCF, and human *JPX* can rescue *Jpx*-deletion defects in mouse embryonic stem cells. Therefore, evolutionary constraints on the sequence−structure−function linkage appear to be more relaxed and complex, consistent with current views on the genome-wide sequence evolution of noncoding RNA genes [18,24,59]. A signature of positive selection acting on the human *JPX* sequence supports adaptive evolution of functional lncRNAs, which have been described in diverse organisms from *Drosophila* to mammalian species [11,13,14,57,79].

Early studies have reported that noncoding RNA structure can be retained across species and could

possibly play a role in functional conservation, but similar analyses on lncRNAs have yet to be comprehensively performed [26,64]. To address whether structural conservation contributes into the functional conservation of *Jpx/JPX*, *in vitro* SHAPE analysis of both mouse lncRNA *Jpx* and human lncRNA *JPX* was performed. Upon comparison, the two structures appear divergent, yet highly structured, with both having stem-loop formations and ~50% of their nucleotides base-paired. The divergence seen between these two lncRNAs may be the result of human *JPX* evolving because of positive selection [80]. This is supported by our protein-binding assays, which show robust interactions between CTCF and the human *JPX* RNA. A comparison of *in vivo* and *in vitro* SHAPE analyses on human *JPX 104-172* indicates a high percentage of RNA nucleotides reactive to NAI, independently of the cellular microenvironment. This region corresponds to the E1−E2 junction that is relatively more conserved in sequence and shown to bind with CTCF, thus representing a possible lncRNA domain important for function. Consistent stem-loop features obtained from *in vivo* and *in vitro* predictions suggest that this RNA domain is structurally stable and support the overall structural divergence observed *in vitro* for human *JPX E1−E3* and mouse *Jpx E1−E3*. RNA domains with conservation in both sequence and secondary structure have been reported for lncRNAs such as *roX*, *XIST*, *RepA*, and *MEG3* [26,29,81,82]. We have identified regions of high sequence identity in *Jpx/JPX* sequence alignment (Fig. 1C), which correspond to nucleotides with similar SHAPE reactivities indicative of similar secondary structures. However, a direct comparison of the RNA secondary structures for *Jpx* and *JPX* (Fig. 3) shows no obvious correlation between nucleotide conservation and secondary structure match. It is possible that conserved nucleotides are important for a higher-order structure involving tertiary interactions between RNA domains. Further work combining mutagenesis of the conserved nucleotides in *Jpx/JPX* with structural and functional characterizations should help define the RNA motif conservation.

The lncRNA-binding capability of CTCF is supported by recent characterizations of CTCF as an RNA-binding protein with specific functions in mammalian cells [75−78]. Our analyses of *Jpx/JPX* lncRNA-CTCF binding are also consistent with the earlier report that *Jpx* RNA directly binds CTCF in activating *Xist* expression in mouse embryonic stem cells [44]. To determine whether specific RNA domains are responsible for binding to CTCF, we compare the binding affinities for the full-length *Jpx E1−E3* RNA and its mutant versions with 5′ and/or 3′ truncations, which show that both the 5′ (bases 34-114) and 3′ (bases 252-347) are necessary for CTCF-binding. Human *JPX E1−E3* RNA also binds to CTCF, and its binding capacity appears more robust against sequence deletions—the 5′ truncation mutant *JPX 19-343* and the 3′ truncation mutant *JPX 1-257* show even stronger binding than the full-length *JPX 1-343*. Together, our RNA EMSA results argue that CTCF-lncRNA binding with *Jpx/JPX* does not directly depend on sequence specificity.

Functional homology between human *JPX* and mouse *Jpx* is further supported by the complementary test in the mouse *Jpx*−/+ mutant ES cells. Deleting a single copy of the *Jpx* gene in female mES cells disrupts *Xist* upregulation and leads to cell death during ES differentiation. Exogenous expression of either mouse *Jpx* or human *JPX* lncRNA in the mutant cells rescues *Xist* expression and cell viability. The mouse ES cell system is most suitable for the functional complementation experiment because differentiation of mES cells faithfully recapitulates the upregulation of *Xist* and the XCI process, whereas human embryonic stem cells so far are not showing the establishment of random XCI or the change of XCI state during human ES cell differentiation [83−85]. Transient transfection of mouse *Jpx E1−E3* or human *JPX E1−E3* was sufficient to increase *Xist* expression and rescue the phenotypic defects of *Jpx*−/+ female mES cells, consistent with the *trans*-acting role of *Jpx/JPX* RNA on activating *Xist*. In contrast, transgenic mES cell lines established from stable integrations of mouse *Jpx E1−E3* or human *JPX E1−E3* transgenes show poor EB morphologies during ES cell differentiation (Fig. S6). Expression of the transgenic *Jpx* or *JPX* RNA appears less efficient and variable between independent clones, which leads to insufficient rescue of cell viability in *Jpx*−/+ mutant female mES cells (Fig. S7). The differences we observed between transiently transfected mES cells versus stable transgenic clones likely reflect the regulatory mechanism of *Jpx* that involves the *trans*-localization of lncRNA molecules and the quantitative threshold needed for activating *Xist* [46,47].

It is interesting to note that human *JPX E1−E3* is capable of rescuing *Jpx*−/+ mES cell viability similarly to mouse *Jpx E1−E3* (Fig. 5B; two-tailed paired Student *t*-test, $P = 0.469$), but the efficiency of activating *Xist* is more variable with human *JPX E1−E3* than it is with mouse *Jpx E1−E3* (Fig. 5D). This suggests that human *JPX* is complementary to mouse *Jpx* with regard to essential cellular functions. The genetic and structural divergence seen between the two may contribute to the differences in the specificity of molecular interactions, which affects the regulatory efficiency on the target gene (i.e., *Xist*). This is also consistent with results from the *in vitro* protein-binding assays using CTCF.

In conclusion, through comparative sequence and functional analyses involving the homologous human *JPX*, our results have demonstrated a convergent function of *Jpx/JPX* between mice and humans despite a rapid divergence in the nucleotide

sequences and a change of the RNA secondary structures. Our findings suggest that lncRNAs are capable of maintaining essential roles in embryogenesis, and such lncRNA functions may be resistant to evolutionary constraints at both RNA sequence and structural levels.

Materials and Methods

Plasmid preparation

The Tg(EF1α:JpxE1−E3) construct was generated by PCR-cloning the *Jpx* transcript out of the cDNA prepared from the total RNA of differentiated mES cells. The *Jpx E1−E3* isoform was cloned into the pEF1/V5-His mammalian expression vector (Invitrogen Cat# V92020), which contains an EF-1α promoter for mammalian expression and a T7 promoter for *in vitro* transcription. In parallel, the Tg(EF1α:JPXE1−E3) construct was generated by PCR-cloning the *JPX* transcript out of the cDNA prepared from the total RNA of human SKOV3iP1 ovarian cancer cells. The *JPX E1−E3* isoform was cloned into the pEF1/V5-His mammalian expression vector the same way as *Jpx E1−E3*.

RNA electrophoretic mobility shift assay

RNA EMSA was carried out as previously described [44,86,87] using *in vitro* transcribed RNAs uniformly labeled with ATP[α-32p] and purified recombinant CTCF protein. Specifically, *Jpx E1−E3 1-383* and truncated RNAs, or *JPX E1−E3* full-length and truncated RNAs, were *in vitro* transcribed using T7 polymerase and DNA templates PCR-amplified from Tg(EF1α:JpxE1−E3) or Tg(EF1α:JPXE1−E3) cDNA plasmid. The control RNA, a 316nt *drz-Agam-2-1* ribozyme RNA, was prepared using genomic DNA of *Anopheles gambiae* [64]. Primers used are as follows:

For *Jpx E1−E3 1-383* and truncated RNAs

JW_1F—TTCCCGCGAAATTAATACGACTCACTATAGG
GAGATGGCGGCGTCCACATGTAT
JW_2R—AGGTGGCAGGCAGCAGGCAT
JW_4R—ATAAGCAAGCTAGTACGCAC
JW_5F—TTCCCGCGAAATTAATACGACTCACTATAGG
GAGTGGCCAATTAATGAACAT
JW_21F—TTCCCGCGAAATTAATACGACTCACTATAG
GGAGCCACGGCACCACCAGGCTTC
JW_22R—GAGTTTATTTGGGCTTACAG

For *JPX E1−E3 1-343* and truncated RNAs

h J P X - E M S A -
T7+F2—TTCCCGCGAAATTAATACGACTCACTATAGGGA
GGGAAGACTTAAGATGGCGGC
h J P X - E M S A -
T7+F3—TTCCCGCGAAATTAATACGACTCACTATAGGGA
GCTTACGGGGGTTGCAAG
hJPX-EMSA-R1—CTGTAATCTCAGCTACTCGGGAG
hJPX-EMSA-R2—GGTCATGCCATTGCATTCC
hJPX-EMSA-R3—AGCCTGGGCAACAAGAG
hJPX-EMSA-R4—TCGTCAGTAGAAGTTAGGCG

For *drz-Agam-2-1* ribozyme RNA (control)

JW_14F—TTCCCGCGAAATTAATACGACTCACTATA
GCTCTGCAAATGGGGTAGGA
JW_24R—GTTTTTTCGTTTGCCGTTGAAGG

Recombinant CTCF protein was prepared and purified as previously described [44]. Mouse CTCF cDNA corresponding to the full-length 736 amino acids was cloned with C-terminal 6xHis tag into pFLAG-2 (Sigma). FLAG-CTCF-6xHis protein was induced in Rosetta-Gami B-cells (EMD Millipore) with 0.2 M of IPTG at room temperature and was then purified with Ni-NTA resin (Qiagen) with 50 mM sodium phosphate (pH 8.0), 300 mM NaCl, and 250 mM imidazole. Eluates were dialyzed against 50 mM Tris-HCl (pH 7.5), 2.5 mM MgCl2, 150 mM NaCl, 0.1 mM $ZnSO_4$, 1 mM DTT, 0.1% Tween-20, and 10% glycerol.

For gel shift (EMSA), RNAs were incubated with CTCF protein and the complexes were resolved in a 5% acrylamide gel. The gel was then exposed to a phosphorimager screen (Molecular Dynamics), scanned with Typhoon phosphorimager (GE Healthcare), and analyzed using ImageQuant software (GE Healthcare). The fraction of the RNA-protein complex was plotted against the concentration of CTCF and fit with a binding equation.

In vitro and *in vivo* RNA SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension)

RNA SHAPE was performed as previously described [65,66]. Specifically, RNAs were *in vitro* transcribed from a plasmid expressing either *Jpx* or *JPX* (Tg(EF1α:JpxE1−E3) or Tg(EF1α:JPXE1−E3). Then 2 pmol of purified RNA was denatured, folded, and modified with 1 μmol of NAI. Subsequently, the RNA was reverse transcribed followed by primer extension with γ-32P-ATP 5′-labeled reverse primers. Four reactions were set up for each primer extension: (1) DMSO-only negative control, (2) dideoxy-ATP (ddA) in DMSO, (3) dideoxy-CTP (ddC) in DMSO, and (4) NAI in DMSO. The γ-32P-ATP end-labeled primers and ~2 pmol of RNA from the modification step were added and incubated at 95 °C for the 2-min annealing step followed by a 2 °C/s step-down cooling to 4 °C. Reverse transcription was performed using first-strand cDNA synthesis kit containing SuperScriptIII® (2 units/μL; Invitrogen, Life Technologies).

In vivo SHAPE was performed as previously described with some modifications [66,70,72]. Briefly, HEK293T cells transiently transfected with Tg(EF1α:JPXE1−E3) for 24 h were collected and incubated at 37 °C with NAI at a 10% concentration for every 4×10^6 cells. RNA was extracted using TRIzol® (Life Technologies) according to manufacturer's instructions. Reverse transcription was performed with nonradiolabeled reverse primers. In order to enrich the presence of *JPX*, cDNA fragments generated during primer extension with cold primers were amplified using LMPCR [70,72]. A linker sequence with a 5′ phosphate and 3′ 3-carbon spacer group was added to the RNA using CircLigase™ ssDNA ligase (epicenter) according to manufacturer's instructions. PCR amplification was performed with a nonradiolabeled forward primer targeting the linker sequence and γ-32P-ATP end-labeled reverse primers.

Samples were resolved on a 10% denaturing polyacrylamide SHAPE Gel (0.4 mm). The gel was dried and

placed into a phosphor-imaging cassette for exposure overnight and was scanned using a Typhoon phosphorimager. Band intensities and SHAPE reactivities were calculated using SAFA software [88,89]. After subtracting the DMSO background from the NAI band intensities, the average of the top 10% minus the top 2% reactivities was calculated and set to 1. This was then used to normalize the SHAPE reactivities [26]. When aligning SHAPE band positions to the transcript, the ladders (ddA and ddC) generated at the reverse transcription step are 1 nucleotide longer than the corresponding DMSO negative control and NAI samples [65]. RNA secondary structures are predicted with ViennaRNA Software (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAprobing.cgi) [67,68]. The extension primers used are as follows:

For *Jpx*

p2r—AGGTGGCAGGCAGCAGGCAT
p3r—CTTGAACTGATGGGTGCCAT
p4r—ATAAGCAAGCTAGTACGCAC
pex1r—GGGCATGTTCATTAATTGGCCAG
p25r—TGGCTAATCCCGGGAAGGAC
p26r—CTTCAAGTCCCTGCTTGAGTTTC

For *JPX*

pr R1—CTGTAATCTCAGCTACTCGG
pr R1-2—AGTGAGCCAAGGTCATGCCA
pr R2—GAAGTTAGGCGATCAGCGAG
pr R2-3—GAGACACAATACTATTAACTGGC
pr R3—CATACTTCGGACGCCTTGCAAC
prR3-4—CCCCGTAAGGACGCAGTGAT
LMPCR Linker Sequence—AGATCGGAAGAGCGGTTCAGC
AGGAATGCCGAGACCGATCTCGTATGCCGTCTTCTGCT
TG
Linker Primer F—GGAAGAGCGGTTCAGCAGGA

Transfection of *JpxE1−E3* or *JPXE1−E3* constructs in mouse ES (mES) cells

For transient transfection, wild-type 16.7 female mES cells and *Jpx*−/+ mutant female mES cells [45] were cultured on feeder cells in media containing LIF (leukemia inhibitory factor); the cells were collected on differentiation Day 0 and ~1 × 10^6 cells were seeded per well of a nontissue culture treated 6-well plate for transfection. The mES cells were then transfected with 2 μg of GFP plasmid (for visual confirmation of transfection) and 2.5 μg of either empty vector, Tg(EF1α:JpxE1−E3), or Tg(EF1α:JPXE1−E3) plasmid per well of a 6-well plate according to the Lipofectamine® 2000 (Invitrogen) manufacturer's instructions. Cells were cultured in feeder-free and LIF-free conditions for differentiation. On Day 1 of mES differentiation, 2 mL of the transfection media was gently refreshed with regular mES media. Cells were viewed on a fluorescent microscope to assess GFP fluorescence and confirmation of successful transfection. Once transfections were deemed successful using Lipofectamine® 2000, the GFP plasmid was no longer included during the transient transfection. For analyses, cells were transferred to tissue culture−treated plates on Day 4 for EB outgrowth and allowed to grow until Day 8 with the cells being collected on Day 4 and Day 8 to assess viability and gene expression.

To obtain stable transgenic mES cell lines with transfection, wild-type 16.7 female mES cells and *Jpx*−/+ mutant female mES cells [45] were seeded on feeder cells in 6-well plates at 3 × 10^5 cells per well in media containing LIF. The next day, cells were transfected with 2.5 μg of either empty vector, Tg(EF1α:JpxE1−E3) or Tg(EF1α:JPXE1−E3), plasmid per well of a 6-well plate according to the Lipofectamine® 2000 (Invitrogen) manufacturer's instructions while still in the stem cell state. After 24 h, an entire well of cells were treated with 0.25% trypsin-EDTA and transferred to a 10 cm dish of Neo-resistant feeder cells. Because the plasmids contain the Neomycin-resistance gene, media was switched to G418 (400ug/mL) + LIF selection media to select for cells that had stable integration of the transgenes. The cell medium was changed every day; after 19 days, colonies were picked and treated with 0.25% trypsin-EDTA to break up the colonies and then transferred to a 24-well plate of Neo-resistant mEFs and grown out. When confluent, half of a 24-well was stocked and the other half was used for PCR screening of successful integration of plasmid. For differentiation, ~5 × 10^5 cells were seeded per well of a non-tissue culture treated 6-well plate without LIF and allowed to grow until Day 4, when they were transferred to tissue culture treated 6-well plates for EB outgrowth. Cells were collected on Day 4 and Day 8 to assess viability and qPCR analysis.

Cell death assay

On Day 4 of mES differentiation, supernatant and EBs were collected, spun down, and broken up with 0.05% Trypsin-EDTA. Cells were then resuspended in 1 mL of mES media, and 20 μL was taken from each sample for staining with trypan blue. Cells were counted on a Countess™ II FL Automated Cell Counter, and cell viability was recorded. For Day 8 of mES differentiation, supernatant and attached cells from EB outgrowth were collected, spun down, and treated with 0.05% Trypsin-EDTA before resuspended in 1 mL of mES media for cell count.

Quantitative real-time PCR

Cells from one well of a 6-well plate were spun down and resuspended in 1 mL TRIzol® (Life Technologies) for each respective sample. RNA was extracted, and residual genomic DNA was removed with TURBO™ DNase (Ambion, Life Technologies) treatment according to manufacturer's instructions. Reverse transcription of RNA to cDNA was performed using SuperScriptIII® (Invitrogen, Life Technologies) according to manufacturer's instructions. Real-time PCR for *Jpx*, *JPX*, *Xist*, and *Gapdh* RNA expression was performed using FS Universal SYBR Green Master (Rox) (Sigma-Aldrich) according to manufacturer's instructions under the following conditions: 95 °C for 10 min, 95 °C for 15 s, 58 °C for 30 s, 72 °C for 30 s, then repeat steps 2 through 4 for 39 cycles. Primers used for PCR were as follows:

mJpx 76 (e1)-F—TTAGCCAGGCAGCTAGAGGA
mJpx 225 (ex2)-R—AGCCGTATTCCTCCATGGTT
hJPX E1-F—AATCACTGCGTCCTTACGGG
hJPX E3-R—GCAGGAGAACCACTTGAACT
XistNS33-F—CAGAGTAGCGAGGACTTGAAGAG
XistBP2F-R—CCCGCTGCTGAGTGTTTGATA

Gapdh-F—ATGAATACGGCTACAGCAACAGG
Gapdh-R—GAGATGCTCAGTGTTGGGGG

Acknowledgments

The authors thank Dr. R. Warrior for suggesting the nucleotide substitution analysis, Drs. G. MacGregor, A. Luptak, and S. Atwood for equipment, Dr. O. Razorenova for SKOV3iP1 cells, Dr. C. Suetterlin for critique of the manuscript, and S. Karner for proofreading the manuscript. H.K. was supported by GAANN Fellowship from the U.S. Department of Education Grant P200A120207. B.L. and M.E. were supported by UCI Undergraduate Research Opportunity Fellowships. UCI start-up funds, Council on Research Single Investigator Foundation Grant SIIG-2014-2015-49, Hellman Fellowship, and American Cancer Society Institutional Research Award (IRG-16-187-13) to S.S. funded this work.

Author Contributions

H.K., C.W., and S.S. designed research; H.K., C.W., Y.L., P.B., and S.S performed nucleotide sequence analyses; H.K. and C.W. performed and analyzed RNA structure and RNA-protein binding experiments; H.K. performed and analyzed mES cell experiments; D.C. and R.C.S. provided new reagents and analytic tools for RNA structure experiments; C.W., S.C., B.L., and M.E. contributed new reagents for RNA-protein binding and mES cell experiments; H.K, C.W., and S.S. wrote the manuscript with input from all authors.

Declaration of Interests

The authors declare no competing interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2019.09.002.

Received 5 July 2019;
Received in revised form 29 August 2019;
Accepted 2 September 2019
Available online 10 September 2019

Keywords:
Long noncoding RNA;
Sequence–structure–function;
Molecular evolution;
Comparative analysis;
X chromosome inactivation

† Present address: Omigen, Inc., Irvine, CA 92618, USA.

Abbreviations used:
LncRNA, long noncoding RNA; *Jpx*, just proximal to *Xist*; *Xist*, X-inactive specific transcript; XCI, X chromosome inactivation; SHAPE, selective 2′-hydroxyl acylation analyzed by primer extension; EMSA, RNA electrophoretic mobility shift assay; mES, mouse embryonic stem.

References

[1] P. Kapranov, G. St Laurent, T. Raz, F. Ozsolak, C.P. Reynolds, P.H. Sorensen, G. Reaman, P. Milos, R.J. Arceci, J.F. Thompson, T.J. Triche, The majority of total nuclear-encoded non-ribosomal RNA in a human cell is "dark matter" un-annotated RNA, BMC Biol. 8 (2010) 149, https://doi.org/10.1186/1741-7007-8-149.
[2] J.L. Rinn, H.Y. Chang, Genome regulation by long noncoding RNAs, Annu. Rev. Biochem. 81 (2012) 145–166, https://doi.org/10.1146/annurev-biochem-051410-092902.
[3] A. Fatica, I. Bozzoni, Long non-coding RNAs: new players in cell differentiation and development, Nat. Rev. Genet. 15 (2014) 7–21, https://doi.org/10.1038/nrg3606.
[4] R.B.-T. Perry, I. Ulitsky, The functions of long noncoding RNAs in development and stem cells, Development 143 (2016) 3882–3894, https://doi.org/10.1242/dev.140962.
[5] H. Wu, L. Yang, L.L. Chen, The diversity of long noncoding RNAs and their generation, Trends Genet. 33 (2017) 540–552, https://doi.org/10.1016/j.tig.2017.05.004.
[6] F. Kopp, J.T. Mendell, Functional classification and experimental dissection of long noncoding RNAs, Cell 172 (2018) 393–407, https://doi.org/10.1016/j.cell.2018.01.011.
[7] T.R. Mercer, J.S. Mattick, Structure and function of long noncoding RNAs in epigenetic regulation, Nat. Struct. Mol. Biol. 20 (2013) 300–307, https://doi.org/10.1038/nsmb.2480.
[8] J.D. Ransohoff, Y. Wei, P.A. Khavari, The functions and unique features of long intergenic non-coding RNA, Nat. Rev. Mol. Cell Biol. 19 (2017) 143–157, https://doi.org/10.1038/nrm.2017.104.
[9] T. Derrien, R. Johnson, G. Bussotti, A. Tanzer, S. Djebali, H. Tilgner, G. Guernec, D. Martin, A. Merkel, D.G. Knowles, J. Lagarde, L. Veeravalli, X. Ruan, Y. Ruan, T. Lassmann, P. Carninci, J.B. Brown, L. Lipovich, J.M. Gonzalez, M. Thomas, C.A. Davis, R. Shiekhattar, T.R. Gingeras, T.J. Hubbard, C. Notredame, J. Harrow, R. Guigó, The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression, Genome Res. 22 (2012) 1775–1789, https://doi.org/10.1101/gr.132159.111.
[10] J. Sun, Y. Lin, J. Wu, Long non-coding RNA expression profiling of mouse testis during postnatal development, PLoS One 8 (2013) e75750, https://doi.org/10.1371/journal.pone.0075750.
[11] K. Wen, L. Yang, T. Xiong, C. Di, D. Ma, M. Wu, Z. Xue, X. Zhang, L. Long, W. Zhang, J. Zhang, X. Bi, J. Dai, Q. Zhang, Z.J. Lu, G. Gao, Critical roles of long noncoding RNAs in Drosophila spermatogenesis, Genome Res. (2016) 1233–1244, https://doi.org/10.1101/gr.199547.115.
[12] C. Kern, Y. Wang, J. Chitwood, I. Korf, M. Delany, H. Cheng, J.F. Medrano, A.L. Van Eenennaam, C. Ernst, P. Ross,

H. Zhou, Genome-wide identification of tissue-specific long non-coding RNA in three farm animal species, BMC Genom. 19 (2018) 684, https://doi.org/10.1186/s12864-018-5037-7.

[13] H. Dai, Y. Chen, S. Chen, Q. Mao, D. Kennedy, P. Landback, A. Eyre-Walker, W. Du, M. Long, The evolution of courtship behaviours through the origination of a new gene in Drosophila, Proc. Natl. Acad. Sci. 105 (2008) 7478–7483, https://doi.org/10.1073/pnas.0800693105.

[14] T.J.A.J. Heinen, F. Staubach, D. Häming, D. Tautz, Emergence of a new gene from an intergenic region, Curr. Biol. 19 (2009) 1527–1531, https://doi.org/10.1016/J.CUB.2009.07.049.

[15] G. Gao, M.D. Vibranovski, L. Zhang, Z. Li, M. Liu, Y.E. Zhang, X. Li, W. Zhang, Q. Fan, N.W. VanKuren, M. Long, L. Wei, A long-term demasculinization of X-linked intergenic noncoding RNAs in Drosophila melanogaster, Genome Res. 24 (2014) 629–638, https://doi.org/10.1101/gr.165837.113.

[16] I. Ulitsky, A. Shkumatava, C.H. Jan, H. Sive, D.P. Bartel, Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution, Cell 147 (2011) 1537–1550, https://doi.org/10.1016/j.cell.2011.11.055.

[17] A. Kapusta, C. Feschotte, Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications, Trends Genet. 30 (2014) 439–452, https://doi.org/10.1016/J.TIG.2014.08.004.

[18] W. Haerty, C.P. Ponting, No gene in the genome makes sense except in the light of evolution, Annu. Rev. Genom. Hum. Genet. 15 (2014) 71–92, https://doi.org/10.1146/annurev-genom-090413-025621.

[19] A. Necsulea, M. Soumillon, M. Warnefors, A.A.A. Liechti, T. Daish, U. Zeller, J.C. Baker, F. Grutzner, H. Kaessmann, F. Grützner, H. Kaessmann, F. Grutzner, H. Kaessmann, The evolution of lncRNA repertoires and expression patterns in tetrapods, Nature 505 (2014) 635–640, https://doi.org/10.1038/nature12943. http://www.nature.com/nature/journal/v505/n7485/abs/nature12943.html#supplementary-information.

[20] H. Ling, K. Vincent, M. Pichler, R. Fodde, I. Berindan-Neagoe, F.J. Slack, G.A. Calin, Junk DNA and the long non-coding RNA twist in cancer genetics, Oncogene 34 (2015) 5003–5011, https://doi.org/10.1038/onc.2014.456.

[21] H. Hezroni, R. Ben-Tov Perry, Z. Meir, G. Housman, Y. Lubelsky, I. Ulitsky, A subset of conserved mammalian long non-coding RNAs are fossils of ancestral protein-coding genes, Genome Biol. 18 (2017) 1–15, https://doi.org/10.1186/s13059-017-1293-0.

[22] M. Cabili, C. Trapnell, L. Goff, M. Koziol, B. Tazon-Vega, A. Regev, J.L. Rinn, Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses, Genes Dev. 25 (2011) 1915–1927, https://doi.org/10.1101/gad.17446611.

[23] H. Hezroni, D. Koppstein, M.G. Schwartz, A. Avrutin, D.P. Bartel, I. Ulitsky, Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species, Cell Rep. 11 (2015) 1110–1122, https://doi.org/10.1016/J.CELREP.2015.04.023.

[24] J.M. Kirk, S.O. Kim, K. Inoue, M.J. Smola, D.M. Lee, M.D. Schertzer, J.S. Wooten, A.R. Baker, D. Sprague, D.W. Collins, C.R. Horning, S. Wang, Q. Chen, K.M. Weeks, P.J. Mucha, J.M. Calabrese, Functional classification of long non-coding RNAs by k-mer content, Nat. Genet. 50 (2018) 1474–1482, https://doi.org/10.1038/s41588-018-0207-8.

[25] I. V Novikova, S.P. Hennelly, K.Y. Sanbonmatsu, Structural architecture of the human long non-coding RNA, steroid receptor RNA activator, Nucleic Acids Res. 40 (2012) 5034–5051, https://doi.org/10.1093/nar/gks071.

[26] I.A. Ilik, J.J. Quinn, P. Georgiev, F. Tavares-Cadete, D. Maticzka, S. Toscano, Y. Wan, R.C. Spitale, N. Luscombe, R. Backofen, H.Y. Chang, A. Akhtar, Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in Drosophila, Mol. Cell 51 (2013) 156–173, https://doi.org/10.1016/j.molcel.2013.07.001.

[27] R. Fang, W.N. Moss, M. Rutenberg-Schoenberg, M.D. Simon, Probing Xist RNA structure in cells using targeted structure-seq, PLoS Genet. 11 (2015), https://doi.org/10.1371/journal.pgen.1005668.

[28] M.J. Smola, T.W. Christy, K. Inoue, C.O. Nicholson, M. Friedersdorf, J.D. Keene, D.M. Lee, J.M. Calabrese, K.M. Weeks, SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells, Proc. Natl. Acad. Sci. 113 (2016) 10322–10327, https://doi.org/10.1073/pnas.1600008113.

[29] F. Liu, S. Somarowthu, A.M. Pyle, Visualizing the secondary and tertiary architectural domains of lncRNA RepA, Nat. Chem. Biol. 13 (2017) 282–289, https://doi.org/10.1038/nchembio.2272.

[30] R. Delli Ponti, A. Armaos, S. Marti, G.G. Tartaglia, A method for RNA structure prediction shows evidence for structure in lncRNAs, Front. Mol. Biosci. 5 (2018) 111, https://doi.org/10.3389/fmolb.2018.00111.

[31] P. Johnsson, K.V. Morris, Expanding the functional role of long noncoding RNAs, Cell Res. 24 (2014) 1284–1285, https://doi.org/10.1038/cr.2014.104.

[32] B. Payer, J.T. Lee, X chromosome dosage compensation: how mammals keep the balance, Annu. Rev. Genet. 42 (2008) 733–772, https://doi.org/10.1146/annurev.genet.42.110807.091711.

[33] T. Straub, P.B. Becker, Dosage compensation: the beginning and end of generalization, Nat. Rev. Genet. 8 (2007) 47–57, https://doi.org/10.1038/nrg2013.

[34] A. Wutz, T.P. Rasmussen, R. Jaenisch, Chromosomal silencing and localization are mediated by different domains of Xist RNA, Nat. Genet. 30 (2002) 167–174, https://doi.org/10.1038/ng820.

[35] J. Grant, S.K. Mahadevaiah, P. Khil, M.N. Sangrithi, H. Royo, J. Duckworth, J.R. McCarrey, J.L. Vandeberg, M.B. Renfree, W. Taylor, G. Elgar, R.D. Camerini-Otero, M.J. Gilchrist, J.M.A. Turner, Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation, Nature 487 (2012) 254–258, https://doi.org/10.1038/nature11171.

[36] N. Brockdorff, A. Ashworth, G.F. Kay, V.M. McCabe, D.P. Norris, P.J. Cooper, S. Swift, S. Rastan, 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 (1992) 515–526, https://doi.org/10.1016/0092-8674(92)90519-I.

[37] C.J. Brown, B.D. Hendrich, J.L. Rupert, R.G. Lafrenière, Y. Xing, J. Lawrence, H.F. Willard, The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus, Cell 71 (1992) 527–542, https://doi.org/10.1016/0092-8674(92)90520-M.

[38] L. Duret, C. Chureau, S. Samain, J. Weissanbach, P. Avner, The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene, Science 312 (2006) 1653–1655, https://doi.org/10.1126/science.1126316 (80-.).
[39] E.A. Elisaphenko, N.N. Kolesnikov, A.I. Shevchenko, I.B. Rogozin, T.B. Nesterova, N. Brockdorff, S.M. Zakian, A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements, PLoS One 3 (2008) e2521, https://doi.org/10.1371/journal.pone.0002521.
[40] J.E. Horvath, C.B. Sheedy, S.L. Merrett, A.B. Diallo, D.L. Swofford, E.D. Green, H.F. Willard, P. Nisc Comparative Sequencing, E.D. Green, H.F. Willard, Comparative analysis of the primate X-inactivation center region and reconstruction of the ancestral primate XIST locus, Genome Res. 21 (2011) 850–862, https://doi.org/10.1101/gr.111849.110.
[41] M. Casanova, T.M. Liyakat Ali, C. Rougeulle, Enlightening the contribution of the dark matter to the X chromosome inactivation process in mammals, Semin. Cell Dev. Biol. 56 (2016) 48–57, https://doi.org/10.1016/J.SEMCDB.2016.05.003.
[42] J.T. Lee, M.S. Bartolomei, X-inactivation, imprinting, and long noncoding RNAs in health and disease, Cell 152 (2013) 1308–1323, https://doi.org/10.1016/j.cell.2013.02.016.
[43] T. Sado, N. Brockdorff, Advances in understanding chromosome silencing by the long non-coding RNA Xist, Philos. Trans. R. Soc. B Biol. Sci. 368 (2013), https://doi.org/10.1098/rstb.2011.0325.
[44] S. Sun, B.C. Del Rosario, A. Szanto, Y. Ogawa, Y. Jeon, J.T. Lee, Jpx RNA activates Xist by evicting CTCF, Cell 153 (2013) 1537–1551, https://doi.org/10.1016/j.cell.2013.05.028.
[45] D. Tian, S. Sun, J.T. Lee, The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation, Cell 143 (2010) 390–403, https://doi.org/10.1016/j.cell.2010.09.049.
[46] C. Li, T. Hong, C.H. Webb, H. Karner, S. Sun, Q. Nie, A self-enhanced transport mechanism through long noncoding RNAs for X chromosome inactivation, Sci. Rep. 6 (2016), https://doi.org/10.1038/srep31517.
[47] S. Carmona, B. Lin, T. Chou, K. Arroyo, S. Sun, LncRNA Jpx induces Xist expression in mice using both trans and cis mechanisms, PLoS Genet. 14 (2018) 1–21, https://doi.org/10.1371/journal.pgen.1007378.
[48] B.R. Migeon, The single active X in human cells: evolutionary tinkering personified, Hum. Genet. 130 (2011) 281–293, https://doi.org/10.1007/s00439-011-1016-7.
[49] B. de Hoon, E. Splinter, B. Eussen, J.C.W. Douben, E. Rentmeester, M. van de Heijning, J.S.E. Laven, J.E.M.M. de Klein, J. Liebelt, J. Gribnau, X chromosome inactivation in a female carrier of a 1.28 mb deletion encompassing the human X inactivation centre, Philos. Trans. R. Soc. B Biol. Sci. 372 (2017), https://doi.org/10.1098/rstb.2016.0359.
[50] S. Petropoulos, D. Edsgard, B. Reinius, Q. Deng, S.P. Panula, S. Codeluppi, A. Plaza Reyes, S. Linnarsson, R. Sandberg, F. Lanner, Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos, Cell 165 (2016) 1012–1026, https://doi.org/10.1016/j.cell.2016.03.023.
[51] K. Tsuritani, T. Irie, R. Yamashita, Y. Sakakibara, H. Wakaguri, A. Kanai, J. Mizushima-Sugano, S. Sugano, K. Nakai, Y. Suzuki, Distinct class of putative "non-conserved" promoters in humans: comparative studies of alternative promoters of human and mouse genes, Genome Res. 17 (2007) 1005–1014, https://doi.org/10.1101/gr.6030107.
[52] N.N. Kolesnikov, E.A. Elisafenko, Comparative organization and the origin of noncoding regulatory RNA genes from X-chromosome inactivation center of human and mouse, Genetika 46 (2010) 1386–1391, https://doi.org/10.1134/S1022795410100200.
[53] C.M. Johnston, A.E.T. Newall, N. Brockdorff, T.B. Nesterova, Enox, a novel gene that maps 10 kb upstream of Xist and partially escapes X inactivation, Genomics 80 (2002) 236–244, https://doi.org/10.1006/geno.2002.6819.
[54] J.T. Lee, N. Lu, Y. Han, Genetic analysis of the mouse X inactivation center defines an 80-kb multifunction domain, Proc. Natl. Acad. Sci. 96 (1999) 3836–3841, https://doi.org/10.1073/pnas.96.7.3836.
[55] W. Makalowski, J. Zhang, M.S. Boguski, Comparative analysis of 1196 orthologous mouse and human full-length mRNA and protein sequences, Genome Res. 6 (1996) 846–857, https://doi.org/10.1101/gr.6.9.846.
[56] R.A. Gibbs, G.M. Weinstock, M.L. Metzker, D.M. Muzny, E.J. Sodergren, S. Scherer, G. Scott, D. Steffen, K.C. Worley, P.E. Burch, G. Okwuonu, S. Hines, L. Lewis, C. Deramo, O. Delgado, S. Dugan-Rocha, G. Miner, M. Morgan, A. Hawes, R. Gill, R.A. Holt, M.D. Adams, P.G. Amanatides, H. Baden-Tillson, M. Barnstead, S. Chin, C.A. Evans, S. Ferriera, C. Fosler, A. Glodek, Z. Gu, D. Jennings, C.L. Kraft, T. Nguyen, C.M. Pfannkoch, C. Sitter, G.G. Sutton, J.C. Venter, T. Woodage, D. Smith, H.M. Lee, E. Gustafson, P. Cahill, A. Kana, L. Doucette-Stamm, K. Weinstock, K. Fechtel, R.B. Weiss, D.M. Dunn, E.D. Green, R.W. Blakesley, G.G. Bouffard, P.J. de Jong, K. Osoegawa, B. Zhu, M. Marra, J. Schein, I. Bosdet, C. Fjell, S. Jones, M. Krzywinski, C. Mathewson, A. Siddiqui, N. Wye, J. McPherson, S. Zhao, C.M. Fraser, J. Shetty, S. Shatsman, K. Geer, Y. Chen, S. Abramzon, W.C. Nierman, R.A. Gibbs, G.M. Weinstock, P.H. Havlak, R. Chen, K.J. Durbin, A. Egan, Y. Ren, X.Z. Song, B. Li, Y. Liu, X. Qin, S. Cawley, G.M. Weinstock, K.C. Worley, A.J. Cooney, R.A. Gibbs, L.M. D'Souza, K. Martin, J.Q. Wu, M.L. Gonzalez-Garay, A.R. Jackson, K.J. Kalafus, M.P. McLeod, A. Milosavljevic, D. Virk, A. Volkov, D.A. Wheeler, Z. Zhang, J.A. Bailey, E.E. Eichler, E. Tuzun, E. Birney, E. Mongin, A. Ureta-Vidal, C. Woodwark, E. Zdobnov, P. Bork, M. Suyama, D. Torrents, M. Alexandersson, B.J. Trask, J.M. Young, D. Smith, H. Huang, K. Fechtel, H. Wang, H. Xing, K. Weinstock, S. Daniels, D. Gietzen, J. Schmidt, K. Stevens, U. Vitt, J. Wingrove, F. Camara, M.M. Albà, J.F. Abril, R. Guigo, A. Smit, I. Dubchak, E.M. Rubin, O. Couronne, A. Poliakov, N. Hübner, D. Ganten, C. Goesele, O. Hummel, T. Kreitler, Y.A. Lee, J. Monti, H. Schulz, H. Zimdahl, H. Himmelbauer, H. Lehrach, H.J. Jacob, S. Bromberg, J. Gullings-Handley, M.I. Jensen-Seaman, A.E. Kwitek, J. Lazar, D. Pasko, P.J. Tonellato, S. Twigger, C.P. Ponting, J.M. Duarte, S. Rice, L. Goodstadt, S.A. Beatson, R.D. Emes, E.E. Winter, C. Webber, P. Brandt, G. Nyakatura, M. Adetobi, F. Chiaromonte, L. Elnitski, P. Eswara, R.C. Hardison, M. Hou, D. Kolbe, K. Makova, W. Miller, A. Nekrutenko, C. Riemer, S. Schwartz, J. Taylor, S. Yang, Y. Zhang, K. Lindpaintner, T.D. Andrews, M. Caccamo, M. Clamp, L. Clarke, V. Curwen, R. Durbin, E. Eyras, S.M. Searle,

G.M. Cooper, S. Batzoglou, M. Brudno, A. Sidow, E.A. Stone, J.C. Venter, B.A. Payseur, G. Bourque, C. López-Otín, X.S. Puente, K. Chakrabarti, S. Chatterji, C. Dewey, L. Pachter, N. Bray, V.B. Yap, A. Caspi, S.D.G. Tesler, P.A. Pevzner, S.C.D. Haussler, K.M. Roskin, R. Baertsch, H. Clawson, T.S. Furey, A.S. Hinrichs, D. Karolchik, W.J. Kent, K.R. Rosenbloom, H. Trumbower, M. Weirauch, D.N. Cooper, P.D. Stenson, B. Ma, M. Brent, M. Arumugam, D. Shteynberg, R.R. Copley, M.S. Taylor, H. Riethman, U. Mudunuri, J. Peterson, M. Guyer, A. Felsenfeld, S. Old, S. Mockrin, F. Collins, Genome sequence of the Brown Norway rat yields insights into mammalian evolution, Nature 428 (2004) 493–521, https://doi.org/10.1038/nature02426.

[57] C.P. Ponting, G. Lunter, Signatures of adaptive evolution within human non-coding sequence, Hum. Mol. Genet. 15 (2006) R170–R175, https://doi.org/10.1093/hmg/ddl182.

[58] J. Ponjavic, C.P. Ponting, G. Lunter, Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs, Genome Res. 17 (2007) 556–565, https://doi.org/10.1101/gr.6036807.

[59] A. Necsulea, H. Kaessmann, Evolutionary dynamics of coding and non-coding transcriptomes, Nat. Rev. Genet. 15 (2014) 734–748, https://doi.org/10.1038/nrg3802.

[60] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425, https://doi.org/10.1093/oxfordjournals.molbev.a040454.

[61] K. Tamura, M. Nei, S. Kumar, Prospects for inferring very large phylogenies by using the neighbor-joining method, Proc. Natl. Acad. Sci. 101 (2004) 11030–11035, https://doi.org/10.1073/pnas.0404206101.

[62] G.M. Cooper, M. Brudno, E.A. Stone, I. Dubhack, S. Batzoglu, A. Sidow, Characterization of evolutionary rates and constraints in three mammalian genomes, Genome Res. 14 (2004) 539–548, https://doi.org/10.1101/gr.2034704.

[63] P. Moorjani, C. Eduardo, G. Amorim, P.F. Arndt, M. Przeworski, C.E.G. Amorim, P.F. Arndt, M. Przeworski, Variation in the molecular clock of primates, Proc. Natl. Acad. Sci. 113 (2016) 10607–10612, https://doi.org/10.1073/pnas.1600374113.

[64] C.-H.T. Webb, N.J. Riccitelli, D.J. Ruminski, A. Luptak, Widespread occurrence of self-cleaving ribozymes, Science 326 (2009), https://doi.org/10.1126/science.1178084 (80-.) 953–953.

[65] K.A. Wilkinson, E.J. Merino, K.M. Weeks, Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution, Nat. Protoc. 1 (2006) 1610–1616, https://doi.org/10.1038/nprot.2006.249.

[66] R.C. Spitale, P. Crisalli, R.A. Flynn, E.A. Torre, E.T. Kool, H.Y. Chang, RNA SHAPE analysis in living cells, Nat. Chem. Biol. 9 (2013) 18–20, https://doi.org/10.1038/nchembio.1131.

[67] R. Lorenz, S.H. Bernhart, C. Honer Zu Siederdissen, H. Tafer, C. Flamm, P.F. Stadler, I.L. Hofacker, ViennaRNA package 2.0, Algorithms Mol. Biol. 6 (2011) 26, https://doi.org/10.1186/1748-7188-6-26.

[68] S. Washietl, I.L. Hofacker, P.F. Stadler, M. Kellis, RNA folding with soft constraints: reconciliation of probing data and thermodynamic secondary structure prediction, Nucleic Acids Res. 40 (2012) 4261–4272, https://doi.org/10.1093/nar/gks009.

[69] K. Zarringhalam, M.M. Meyer, I. Dotu, J.H. Chuang, P. Clote, Integrating chemical footprinting data into RNA secondary structure prediction, PLoS One 7 (2012) e45160, https://doi.org/10.1371/journal.pone.0045160.

[70] C.K. Kwok, Y. Ding, Y. Tang, S.M. Assmann, P.C. Bevilacqua, Determination of in vivo RNA structure in low-abundance transcripts, Nat. Commun. 4 (2013) 2971, https://doi.org/10.1038/ncomms3971.

[71] C. Xue, F. Li, Finding noncoding RNA transcripts from low abundance expressed sequence tags, Cell Res. 18 (2008) 695–700, https://doi.org/10.1038/cr.2008.59.

[72] J.B. Lucks, S.A. Mortimer, C. Trapnell, S. Luo, S. Aviran, G.P. Schroth, L. Pachter, J.A. Doudna, A.P. Arkin, Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq), Proc. Natl. Acad. Sci. 108 (2011) 11063–11068, https://doi.org/10.1073/pnas.1106501108.

[73] R. Ohlsson, R. Renkawitz, V. Lobanenkov, CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease, Trends Genet. 17 (2001) 520–527, https://doi.org/10.1016/S0168-9525(01)02366-6.

[74] J.E. Phillips, V.G. Corces, CTCF: master weaver of the genome, Cell 137 (2009) 1194–1211, https://doi.org/10.1016/j.cell.2009.06.001.

[75] R. Saldaña-Meyer, E. González-Buendía, G. Guerrero, V. Narendra, R. Bonasio, F. Recillas-Targa, D. Reinberg, CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53, Genes Dev. 28 (2014) 723–734, https://doi.org/10.1101/gad.236869.113.

[76] J.T.T. Kung, B. Kesner, J.Y.Y. An, J.Y.Y. Ahn, C. Cifuentes-Rojas, D. Colognori, Y. Jeon, A. Szanto, B.C. Del Rosario, S.F.F. Pinter, J.A.A. Erwin, J.T.T. Lee, B.C. del Rosario, S.F.F. Pinter, J.A.A. Erwin, J.T.T. Lee, B.C. Del Rosario, S.F.F. Pinter, J.A.A. Erwin, J.T.T. Lee, Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF, Mol. Cell 57 (2015) 361–375, https://doi.org/10.1016/j.molcel.2014.12.006.

[77] A.S. Hansen, T.-H.S. Hsieh, C. Cattoglio, I. Pustova, X. Darzacq, R. Tjian, An RNA-binding region regulates CTCF clustering and chromatin looping, BioRxiv (2018), https://doi.org/10.1101/495432.

[78] R. Saldana-Meyer, J. Rodriguez-Hernaez, M. Nishana, K. Jacome-Lopez, E.P. Nora, B.G. Bruneau, M. Furlan-Magaril, J. Skok, D. Reinberg, RNA interactions with CTCF are essential for its proper function, BioRxiv (2019), https://doi.org/10.1101/530014.

[79] C. Kutter, S. Watt, K. Stefflova, M.D. Wilson, A. Goncalves, C.P. Ponting, D.T. Odom, A.C. Marques, Rapid turnover of long noncoding RNAs and the evolution of gene expression, PLoS Genet. 8 (2012) e1002841, https://doi.org/10.1371/journal.pgen.1002841.

[80] P. Johnsson, L. Lipovich, D. Grandér, K.V. Morris, Evolutionary conservation of long non-coding RNAs; sequence, structure, function, Biochim. Biophys. Acta 1840 (2014) 1063–1071, https://doi.org/10.1016/j.bbagen.2013.10.035.

[81] Z. Lu, Q.C. Zhang, B. Lee, R.A. Flynn, M.A. Smith, J.T. Robinson, C. Davidovich, A.R. Gooding, K.J. Goodrich, J.S. Mattick, J.P. Mesirov, T.R. Cech, H.Y. Chang, RNA duplex map in living cells reveals higher-order transcriptome structure, Cell 165 (2016) 1267–1279, https://doi.org/10.1016/j.cell.2016.04.028.

[82] T. Uroda, E. Anastasakou, A. Rossi, J.-M. Teulon, J.-L. Pellequer, P. Annibale, O. Pessey, A. Inga, I. Chillón, M. Marcia, Conserved pseudoknots in lncRNA MEG3 are

essential for stimulation of the p53 pathway, Mol. Cell (2019), https://doi.org/10.1016/j.molcel.2019.07.025.
[83] S. Patel, G. Bonora, A. Sahakyan, R. Kim, C. Chronis, J. Langerman, S. Fitz-Gibbon, L. Rubbi, R.J.P. Skelton, R. Ardehali, M. Pellegrini, W.E. Lowry, A.T. Clark, K. Plath, Human embryonic stem cells do not change their X inactivation status during differentiation, Cell Rep. 18 (2017) 54–67, https://doi.org/10.1016/J.CELREP.2016.11.054.
[84] S.A. Khan, P.N.C.B. Audergon, B. Payer, X-chromosome activity in naive human pluripotent stem cells—are we there yet? Stem Cell Investig. 4 (2017) https://doi.org/10.21037/sci.2017.06.03, 54–54.
[85] A. Sahakyan, Y. Yang, K. Plath, The role of Xist in X-chromosome dosage compensation, Trends Cell Biol. 28 (2018) 999–1013, https://doi.org/10.1016/J.TCB.2018.05.005.
[86] C. Cifuentes-Rojas, K. Kannan, L. Tseng, D.E. Shippen, Two RNA subunits and POT1a are components of Arabidopsis telomerase, Proc. Natl. Acad. Sci. 108 (2011) 73–78, https://doi.org/10.1073/pnas.1013021107.
[87] L.M. Hellman, M.G. Fried, Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions, Nat. Protoc. 2 (2007) 1849–1861, https://doi.org/10.1038/nprot.2007.249.
[88] A. Laederach, R. Das, Q. Vicens, S.M. Pearlman, M. Brenowitz, D. Herschlag, R.B. Altman, Semiautomated and rapid quantification of nucleic acid footprinting and structure mapping experiments, Nat. Protoc. 3 (2008) 1395–1401, https://doi.org/10.1038/nprot.2008.134.
[89] R. Das, A. Laederach, S.M. Pearlman, D. Herschlag, R.B. Altman, SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments, RNA 11 (2005) 344–354, https://doi.org/10.1261/rna.7214405.
[90] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, Clustal W and Clustal X Version 2.0, vol. 23, Bioinformatics, 2007, pp. 2947–2948, https://doi.org/10.1093/bioinformatics/btm404.