Cell Systems

CellPress

Perspective

Ectopic expression to synthetic design: Deriving engineering principles of lncRNA-mediated epigenetic regulation

Shalley Sharma,[1] Seong Hu Kim,[1] Tian Hong,[2] Aaron M. Johnson,[3] Alisha Jones,[4] Keriayn N. Smith,[5,6,7] and Karmella A. Haynes[1,*]

[1]Wallace H. Coulter Department of Biomedical Engineering, Emory University, Atlanta, GA 30322, USA
[2]Department of Biological Sciences, The University of Texas at Dallas, Richardson, TX 75080, USA
[3]Department of Biochemistry and Molecular Genetics, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA
[4]Department of Chemistry, New York University, New York, NY 10003, USA
[5]School of Data Science and Society, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[6]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[7]RNA Discovery Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
*Correspondence: karmella.ann.haynes@emory.edu
https://doi.org/10.1016/j.cels.2025.101478

SUMMARY

A key challenge in synthetic biology is achieving durable amplification of low-level inputs in gene regulation systems. Current RNA-based tools primarily operate post-transcriptionally and often yield limited, transient responses. An underexplored feature of lowly expressed long non-coding RNAs (lncRNAs) is their ability to induce outsized effects on chromatin regulation across large genomic regions. Mechanistic insights from basic research are bringing the field closer to designing lncRNAs for epigenetic engineering. We review foundational studies on ectopic expression to uncover lncRNA-mediated epigenetic mechanisms and state-of-the-art transgenic systems for studying lncRNA-driven epigenetic regulation. We present perspectives on strategies for testing the composability of modular lncRNA elements to build rationally designed systems with programmable chromatin-modifying functions and potential biomedical applications such as gene dosage correction. Deepening mechanistic insights into lncRNA function, combined with the development of lncRNA-based technologies for genome regulation, will pave the way for significant advances in cell state control.

INTRODUCTION

Long non-coding RNAs (lncRNAs) are generally defined as RNA molecules that lack protein-coding capacity and exceed 200 nucleotides in length,[1] although a recent consensus statement has proposed a threshold of >500 nucleotides.[2] Once seen as transcriptional noise due to their low abundance and lack of sequence conservation,[2] lncRNAs are now recognized as critical components of cellular pathways, including gene expression regulation.[3] These include nuclear and cytoplasmic functions such as regulating transcriptional, post-transcriptional, and translational processes through base-pairing, scaffolding protein complexes, and acting as molecular decoys, among other mechanisms.[1]

Given their abundance and stoichiometry relative to their targets, several nuclear lncRNAs have been found to exert outsized effects through dynamic interactions that orchestrate protein localization and macromolecular crowding,[4,5] in contrast to stable regulatory mechanisms such as scaffolding mediated by ribosomal and telomerase RNA.[2] For instance, with only 100–200 molecules per cell,[6,7] human X-inactive specific transcript (*XIST*, ~19,000 nucleotides) is present at ~1 molecule per 1 megabase (Mb) of DNA, yet it regulates the silencing of 96% of ~800 genes across the inactive X chromosome (170 Mb).[8–11] Similarly, mouse *Airn*, *Kcnq1ot1*, and *Meg3* also regulate large genomic regions (~10.0, ~0.86, and ~1.0 Mb, respectively),[12] with few molecules of each being present in the cell. These examples highlight how a few nuclear lncRNAs can regulate extensive genomic territories despite low cellular abundance. Such large-scale regulatory influence arises from diverse modes of action where lncRNAs can act in *cis* or in *trans*. *Cis* effects are mediated either by the RNA product itself or by the act of transcription, often in combination, and their functional roles have largely been defined through extensive work in mouse models.[13,14]

Synthetic biology has used functional non-coding RNA as a design feature in transcription and translation control systems,[15,16] while lncRNA-mediated regulation of chromatin has only been recently leveraged in this field.[17] A major goal for engineered chromatin systems is to efficiently establish durable regulatory memory at multiple genes,[18,19] a feature exemplified by XIST's ability to induce chromosome-wide silencing.[20] More broadly, lncRNAs present a unique approach for signal amplification where a transient, low-level input can trigger a large-scale and persistent regulatory state. Considerable effort has been devoted to achieving this type of control for engineered biological systems.[21,22]

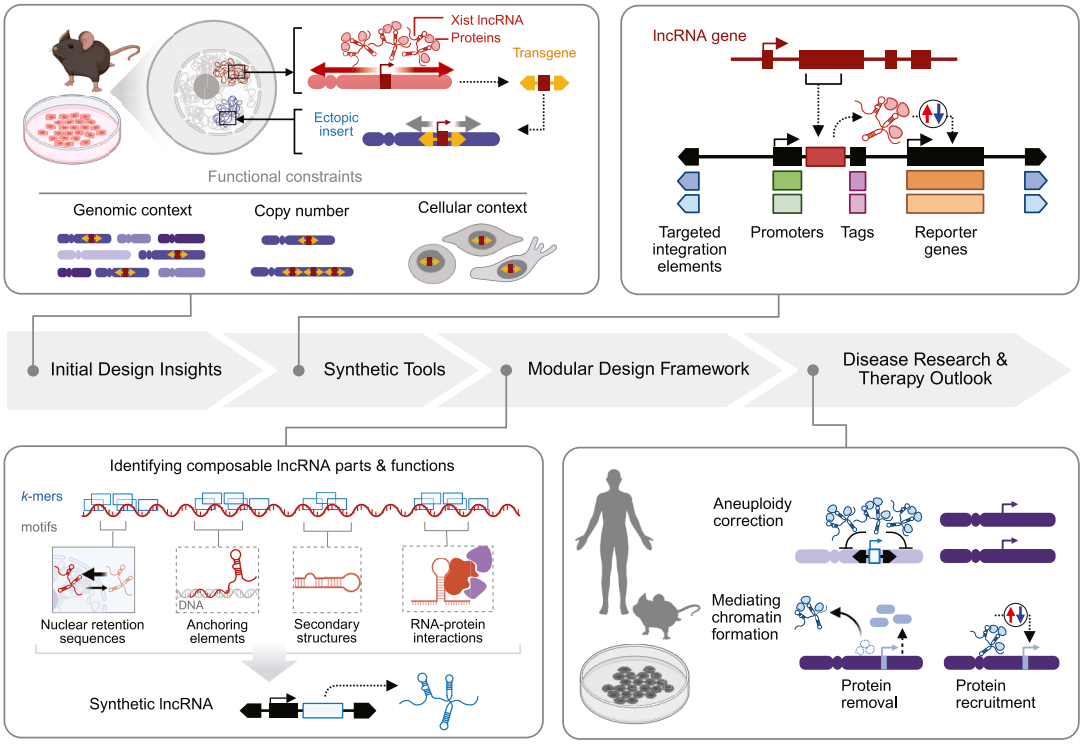


Figure 1. lncRNA engineering from foundational studies to rational design

Initial design insights: foundational discoveries include the research of *Xist*, where it was determined that ectopic silencing in *cis* could be accomplished by using YACs to transfer *Xist*-encoding DNA (the *XIC*) from the sex chromosome X to an autosome.[32–35] Results showed that lncRNA function depends on genomic context (integration site), copy number, and cellular context. Synthetic tools: the use of tools such as non-native promoters, targeted integration systems, RNA and protein tags, and reporter genes has enabled facile generation of ectopic lncRNA systems.[24,29,36] Modular design framework: experimental and computational studies of lncRNA sub-domains illuminate the potential, limitations, and open questions of using sub-domains as interchangeable parts.[29,37] Disease research and therapy outlook: engineered lncRNA systems have been used to explore the feasibility of gene dosage correction.[27] Studies of diverse chromatin-modulating lncRNAs in different disease contexts set the stage for further development.[38–41]

Critical gaps remain in understanding how lncRNAs can be harnessed as engineering components. Major challenges include deciphering their modular domains,[23,24] achieving reproducible activity across cellular and genomic contexts,[25] and designing lncRNAs with tunable activity.[9,17,26–29] Synthetic biology can address these gaps through iterative design, construction, and testing.[17,30,31] This review follows the development of lncRNA-based chromatin engineering from foundational studies to synthetic tools and a potential rational design framework, with medical translation considered in concluding remarks (Figure 1).

EMERGING DESIGN INSIGHTS FROM ECTOPIC lncRNA STUDIES

The ability of lowly expressed lncRNAs to induce robust epigenetic changes across large genomic domains presents an opportunity to engineer efficient genomic control systems.[36,42,43] Foundational studies, particularly those focused on the ectopic expression of the X-inactivation center (*XIC*), have demonstrated the feasibility of reprogramming chromatin states outside of the native context of *Xist* lncRNA.[33,44,45] Critically, these experiments, spanning from using large yeast artificial chromosomes (YACs) to minimal cosmid inserts, have revealed constraints related to cellular context, lncRNA dosage, and genomic position.[34,44,46] Revelations from attempts to reconstitute *XIC* activity have defined the conditions under which lncRNAs can or cannot reprogram chromatin, providing transferable design principles for synthetic lncRNA engineering.

The lncRNA *XIST*, the master regulator[47] of X chromosome inactivation in eutherian mammals,[48] may be considered a quintessential model for building large-scale epigenetic engineering, given its unparalleled capacity to regulate the silencing of approximately 96% of ~800 genes across the inactive X chromosome (~170 Mb).[24,36,49] Early efforts focused on transferring the full human *XIC* or its murine homolog (*Xic*) into autosomes, testing whether the complex, multigenic region spanning approximately 350–480 kb could initiate silencing elsewhere in the genome (Figure 2A). Autosomal insertion of the *Xic* resulted in only partial silencing, weaker than that observed on the X chromosome, suggesting that chromosomal/genomic context influences *Xist* expression and silencing efficiency.[32,35,44]

Genomic context influences lncRNA functionality

A central question in understanding lncRNA portability is why silencing by *XIST* appears to require its native chromosomal environment. One explanatory model proposes that *XIST*

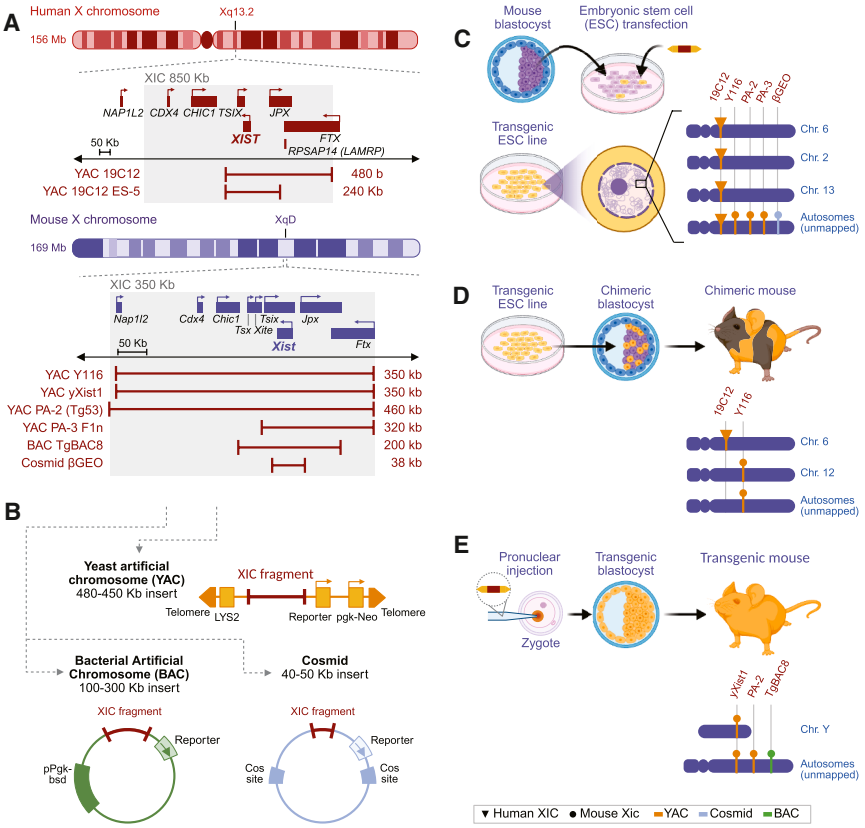


Figure 2. Transgene systems used to investigate *XIC*-mediated epigenetic regulation outside of the X chromosome

(A) Illustrations of the natural *XIC* at human cytoband Xq13.2 and *Xic* at mouse cytoband XqD (enlarged) show the *XIST* gene and surrounding genes. Regions that have been cloned into transgenes are marked with red bars. (B) General structures of YAC, a BAC,[50] or a cosmid vector used to deliver *XIC* fragments into the host cells.[46]

(C–E) Various transgenes have been used to generate ectopic inserts into autosomes or the Y chromosome. Their expression and *cis*-regulatory activity have been monitored in undifferentiated and differentiated ESCs[32,43,51,52] (C), chimeric mice generated from transgenic embryonic stem cells (ESCs) injected into blastocysts[33] (D), or transgenic mice[45] (E). Details and references for each transgene are provided in Table 1.

spreading depends on genomic "way stations" or "entry points," which are long interspersed nuclear element-1 (LINE-1) -enriched native DNA sequences along the X chromosome that act as docking hubs for *XIST* and its effector proteins.[11,53,54] This idea was tested by deleting a LINE-1 cluster near *Firre* on the mouse X chromosome. This deletion did not lead to detectable changes in local X-inactivation patterns *in vivo*.[55] Although this experiment did not demonstrate the necessity of LINE-1 elements, the model may still be applicable to other loci, such as ectopic *XIST* inserts that fail to reorganize chromatin effectively. The spread of lncRNA/chromatin-mediated regulatory states could also be limited by boundaries of topologically associating domains (TADs) demarcated by the CCCTC-binding factor (CTCF) protein.[56]

Relocation using YACs and bacterial artificial chromosomes (BACs) carrying the *Xic* suggests that the absence of extensive flanking sequences compromises distal silencing activity (Table 1).[25,34,50,57,58] For instance, ectopic inserts on autosomes and the Y chromosome expressed *Xist* transcripts yet failed to silence neighboring genes,[34,35] highlighting the need for local chromatin architecture. A cosmid containing *Xist* plus 6 and 9 kb flanking sequences silenced a reporter gene but failed to achieve broader repression (Figure 2B).[46] It is important to note that the effectiveness of *Xist* in silencing genes on autosomes is not uniform, and its success depends on the underlying structural organization of the chromosomes, i.e., genomic context.[59] Similarly, imprinting lncRNAs like *Airn* and *Kcnq1ot1* appear to require their native chromosomal neighborhoods.[60–63]

For engineering applications, genomic context can be approached as both a limitation and an opportunity. Including tethering elements in the transgene that mimic native entry points may enhance local silencing, while systematic testing of ectopic transgenes across the genome could reveal hidden or cryptic hubs where silencing is unexpectedly strong. At the same time, the inherent constraint of *cis*-limited regulation can be exploited as a design feature when the goal is to silence genes within a defined genomic range while leaving distal loci unaffected.

Copy number modulates but does not guarantee lncRNA-mediated silencing

In addition to genomic position, the copy number of lncRNA-expressing loci strongly influences *cis*-regulatory outcomes. Early studies demonstrated a dosage requirement for ectopic activity, *Xic* fragments of 460 and 210 kb could trigger silencing when present as multi-copy arrays but failed as single-copy inserts.[44] Therefore, insufficient expression dosage might constrain functionality. However, in some cases, high copy numbers were not sufficient to induce robust silencing. In *in vivo* studies, the *tg04* autosomal transgene carrying four copies showed no *Xist* expression and DNA hypermethylation in adult males, consistent with failure to establish an inactive-X-like state.[34,65] Other ectopic multi-copy inserts expressed *Xist* transcripts on autosomes and the Y chromosome, but with no evidence of local gene repression.[34,44] Together, these findings suggest that while increased lncRNA dosage can enhance ectopic silencing, genomic context may introduce another layer of regulation that overrides the dosage effect.

Cellular context as a design constraint

A central question for lncRNA engineering is the extent to which cellular context shapes regulatory outcomes. A leading idea from *XIST* studies is that silencing is only effective within an

Table 1. Specific applications of ectopic systems used to investigate *XIC*- and *Xist*-mediated epigenetic regulation outside of the X chromosome

Citation	Cloned locus	Size (Mb)	Transgene system	Integration site(s) and context	Key outcome(s)
Lee et al.[32]	mouse *Xic*	0.45	YAC: Y116, β-gal reporter	autosomes; male mESCs (J1 cell line)	increased *Xist* expression during differentiation, coating of autosome in *cis* (FISH)
				autosomes; chimeric male mouse fibroblasts	lacZ *cis*-silencing
Matsuura et al.[34]	mouse *Xic*	0.35	YAC: yXist1	autosome and chr. Y; transgenic mice	*Xist* hypermethylation and no expression on the autosome
Heard et al.[35]	mouse *Xic*	0.21, 0.46	YAC: PA-2	autosomes; transgenic mice	*Xist* RNA expression, but no *cis*-silencing
Lee and Jaenisch[33]	mouse *Xic*	0.45	YAC: Y116 (20 copies)	chr. 12 (pericentromere); chimeric male mouse fibroblasts	*Xist* coating of chr. 12 in *cis*, endogenous gene repression, histone hypoacetylation
Herzing et al.[46]	mouse *Xist* (+9 kb upstream, 6 kb downstream)	0.04	cosmid, β-gal reporter	autosome; male m ESCs (CCE cell line)	*Xist* sufficient for β-gal silencing, no distal silencing
Heard et al.[44]	mouse *Xic*	0.46, 0.32	YAC: PA-2, PA-3	autosome; male mouse ESCs	multi-copy arrays but not single-copy *Xist* transgenes induced repression
Heard et al.[25]	human *XIC*	0.48	YAC: 19C12	chr. 2, 13, other unmapped autosome sites; male mESCs	*XIST* shows pre-differentiation coating and unstable silencing
Migeon et al.[45,57]	human *XIC*	0.24	YAC: 19C12 truncated (ES-5)	chr. 6; male mESCs (differentiated)	*XIST* fails to induce silencing
				chr.6; chimeric male mouse fibroblasts	high *XIST* expression but no *cis*-repression in somatic cells
Augui et al.[50]	mouse *Xic*	0.136, 0.20	BAC: BAC5, BAC8	autosomes; male mESCs	*Xist* region (BAC8) and distal *Xpct* (BAC5) physically pair with endogenous *Xic*
Sun et al.[64]	mouse *Xic*	0.20	BAC: tgBAC8[50]	autosomes; transgenic mice	ectopic, unpaired, paternal *Xist* shows imprinting in female progeny
Loda et al.[58]	mouse *Xic*	0.30	BAC:CH26-171B21	X chromosome and autosome (chr 6); male mESCs	efficient X-linked silencing and chromatin changes; limited gene silencing from autosomal insertions due to lower LINE-1 density and insulating chromatin domains

mESCs, mouse embryonic stem cells; YAC, yeast artificial chromosome; BAC, bacterial artificial chromosome; Tg, transgene; XCR, X chromosome repeat; *XIC*, X-inactivation center; FISH, fluorescence *in situ* hybridization.

early developmental window when the global cellular state, including transcriptional programs and the availability of RNA-binding proteins (RBPs), is primed to engage in X inactivation.[51,66] Therefore, lncRNAs might display reduced or altered activity depending on host cell type, differentiation stage, and species-specific genomic and protein variations. To illustrate, in mouse embryonic stem cells (mESCs), *Xist* expressed ectopically from an autosomal transgene formed nuclear clouds but silenced neighboring genes only after differentiation, demonstrating a direct before/after cell-state dependency (Figure 2C).[43,51,52] Likewise, an autosomal 450-kb *Xic* transgene was inactive in undifferentiated mESCs but induced *Xist* and *cis* repression upon differentiation.[32]

Similar effects were observed in chimeras, a mouse *Xic* insert induced *Xist*-mediated repression detectable in mouse fibroblasts,[33] whereas a low-copy human *XIC* inserted on chromosome 6 was transcriptionally active in mESCs but failed to induce *cis*-inactivation in somatic cells of chimeric mice (Figures 2D and 2E),[45] revealing design challenges across species. Host-specific context dependency was further supported when human *XIST was engineered to* coat the host chromosome (in mESCs) in *cis* and partially silenced nearby genes,[25] yet its accumulation occurred prematurely before differentiation, and the replication kinetics of the coated autosomes deviated from the late-replicating *Xi* pattern, contrasting native mouse *Xist*.[25,33] In another cross-species study, marsupial *Rsx* exhibiting XIC-like properties displayed only partial *cis*-based repression in mESCs.[26,36] Together, these findings suggest that cellular context can dampen the efficacy of ectopic lncRNA function, but silencing remains achievable under certain conditions, albeit often with altered dynamics or incomplete fidelity.

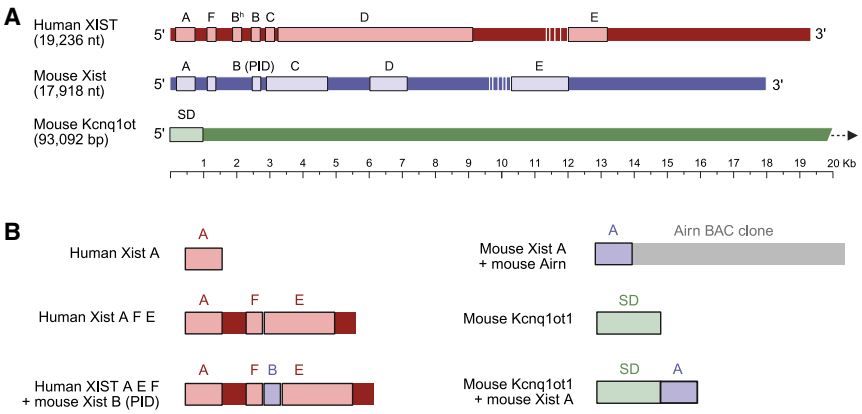


Figure 3. Synthetic constructs derived from full-length lncRNAs

(A) Maps of full-length processed transcripts for lncRNAs from which isolated sub-domains were tested in studies highlighted in this review. Exon boundaries are shown as gaps. *XIST* annotations are from Raposo et al.[71] Database IDs (NCBI and Ensembl) are as follows: human *XIST*, NR_001564, *XIST*-204; mouse *Xist*, NR_001463, *Xist*-201; mouse *Kcnq1ot1*, NR_001461, *Kcnq1ot1*-201; mouse *Airn*, NR_027772, *Airn*-206.

(B) lncRNA sub-domains that have been tested in isolation and as hybrid constructs.[29,37] The *Xist* A/*Airn* hybrid was generated by inserting the *Xist* A sequence upstream of *Airn* exon 1 in a BAC clone.

Discrete lncRNA sequences serve as transferable units for synthetic regulation

While ectopic studies using large *XIC* transgenes underscore the importance of genomic context, dosage, and cellular environment for broad-scale silencing, subsequent work revealed that lncRNAs also contain smaller domains with transferable regulatory activity. Indeed, dissection of lncRNA sub-regions revealed features that can be repurposed for epigenetic engineering, demonstrating that silencing and localization functions are encoded within discrete RNA sequences.[29,59,67] However, previous studies using these fragments showed sufficient but limited activity outside of their native context. For example, fragments of human *XIST* containing the ~600 nt 5′ repeat A (Rep A) domain alone, or in combination with domains F and E, were sufficient to silence an adjacent *GFP* reporter when expressed from an autosomal locus.[24,29] This domain is additive, and just two tandem consensus repeats initiated gene silencing, whereas nine copies achieved the full silencing efficacy comparable to the native 5′ region.[24] Furthermore, adding the mouse *Xist* Polycomb-interacting domain (PID) enhanced repression, showing that modular silencing domains (SDs) can be successfully recombined across species.[29]

Analogous results were found for imprinting lncRNAs. The ~890 nt 5′ SD of *Kcnq1ot1* was validated as sufficient to recruit DNA (cytosine-5) methyltransferase 1 (DNMT1) and repress flanking reporter genes.[68] The modular activities of these lncRNAs allow for mixing and matching, and combining the *Kcnq1ot1* SD with the 5′ *Xist* Rep A in an episomal system resulted in improved repression.[69] Similarly, a hybrid construct combining *Xist* Rep A with full-length *Airn* at an ectopic locus successfully induced Polycomb accumulation and gene silencing across chromosome 6, an activity that *Airn* alone could not achieve, suggesting these modules mediate distinct, sequential steps.[37] Silencing by the *Airn* domain has been shown to correlate with RNA abundance, indicating that expression level may modulate its repressive potency.[70] In summary, experiments where sub-domains of lncRNAs, like *XIST*, *Kcnq1ot1*, and *Airn* are isolated and combined (as illustrated in Figure 3), mark a turning point where synthetic approaches treat lncRNA sequences as composable parts rather than indivisible units.

Overall, efforts to recreate stable lncRNA-mediated epigenetic states outside their native context have revealed persistent challenges. Transgene position, insufficient LINE-1 density, and insulating chromatin boundaries may have limited the spreading of facultative heterochromatin from ectopic sites.[54,58,72] These outcomes and their mechanistic underpinnings are summarized in Box 1, which distills the general design principles, risks, and failure modes that have emerged from ectopic *XIST*/*Xic* systems that have been reported so far. An important consideration is that these studies have not yet sampled the genome or cell states in a comprehensive way. YAC, BAC, and cosmid experiments remain low throughput and have been largely restricted to fibroblasts and ESCs, leaving open the possibility that broader, systematically designed synthetic approaches could uncover permissive contexts and achieve more durable, programmable epigenetic states.

SYNTHETIC TOOLS TO EXPRESS, TRACK, AND DEPLOY lncRNA FUNCTIONS

Previous research on *XIC* function often employed YACs to transfer large *XIC* sequences into mESC genomes via random integration. Although informative, these YAC-based studies were limited in resolution and flexibility.[25,34,44,74] Streamlined transgenes with customizable promoters and pre-determined integration sites surpass YAC-based studies by enabling customization of genetic regulatory elements, offering a more controlled system to rigorously investigate lncRNA-mediated chromatin organization (Figure 4A).

Programmable promoters for controlling lncRNA expression magnitude and timing

Promoters and enhancers are non-coding DNA sequences that recruit the transcriptional machinery and modulate the initiation, rate, and magnitude of lncRNA production.[75–77] In studies that used large regions from the *XIC*, *XIST* transcript levels increased during ESC differentiation, suggesting the importance of dynamic regulation by adjacent promoter and enhancer elements.[32] Several studies have demonstrated the efficacy of different RNA polymerase II (RNA Pol II) promoters for lncRNA expression, including constitutive and engineered inducible systems. Constitutive promoters continuously drive the expression of a gene without external stimuli (Figure 4B). In *Drosophila*, non-native constitutive promoters *Hsp83* or *Hsp70* were used to express transgenic lncRNAs *roX1* or *roX2*, respectively, which partially restored dosage compensation observed as survival

Box 1. Design principles, risks, and failure modes for synthetic lncRNA systems

Ectopic *XIC*/*XIST* studies reveal that custom lncRNA-mediated epigenetic regulation is constrained by genomic position, dosage, and cellular environment. Genomic context: lncRNA activity is often influenced by flanking regulatory DNA and chromatin architecture. Inserts lacking X-specific repetitive elements (e.g., LINE-1 way stations) or placed within restrictive chromatin might fail to propagate silencing, yielding unstable, or partial repression.[25,34,35,50,58] Dosage effects: *Xist* fragments induced repression mostly as multi-copy arrays, whereas single-copy or low-copy inserts were inactive.[44] In some cases multiple copies of *Xist* at autosomal and Y chromosomes (tg04 and tg15, respectively) failed to induce repression, suggesting that genomic context can override high dosage.[34] Cellular context: host-specific factors can influence activity. Human *XIST* in mESCs caused premature accumulation and incomplete silencing, while mouse *Xist* functioned more as expected in their native context.[25,33,45] Sequence modularity: functional domains (e.g., Rep A and PID) retain partial activity when transplanted outside their native locus, enabling modular design, but the activity is *cis*-limited unless supported by broader genomic context.[24,29,37,46,68,73] Key failure modes: improper chromosomal context, unbalanced expression dosage, and mismatched host species. Design insights: (1) optimize lncRNA dosage (expression levels), (2) ensure appropriate developmental windows, and (3) integrate within receptive chromatin to ensure robust lncRNA function.

(28% and 47%) of *roX1/2*-null males.[78] A major limitation of constitutive promoters is that lncRNA overexpression can cause hyperactivity and gene dosage-associated lethality, hindering analysis of many otherwise informative cell and mouse lines.[57]

Engineered inducible promoters, typically activated by a transcription factor that is allosterically controlled with a ligand such as doxycycline (dox), have been used to precisely control the timing and levels of lncRNA transcripts. Wutz et al. performed experiments with full-length *Xist* under the control of a dox-inducible promoter[36] where dox was added to transgenic ESCs and then washed out.[36,51] They discovered that repression required constant induction of *Xist* in early stem cell stages and became stable and independent of dox at 72 h of differentiation. Engreitz et al. used a similar inducible system[36] inserted at the endogenous *Xist* in male ES cells, along with RNA antisense purification (RAP) and RNA-FISH (fluorescence *in situ* hybridization) to investigate the spatial dynamics of *Xist* spreading. From 1 to 6 h, *Xist* formed RNA clouds near the transcription site before expanding and spreading to distant regions, guided by 3D chromosome architecture.[11] Minks et al. used a dox-inducible pcDNA5/FRT/TO system to express *XIST* constructs from an autosome in human HT1080 cells. This inducible system's cytomegalovirus promoter (pCMV) enabled tight control over both the timing and level of *XIST* transcript production. Upon addition of dox, *XIST* RNA rapidly accumulated at the site of transcription as visualized by RNA-FISH. They monitored the levels of expression of the downstream reporter (*EGFP*), which decreased with the increasing levels of *XIST* transcripts until 5 days after induction.[24] Overall, these studies demonstrate that non-native constitutive and inducible RNA Pol II promoters are effective tools for investigating lncRNA activity (Table 2).[24,36,78,79]

Targeted chromosomal insertion strategies to control lncRNA context and effects

The insertion of lncRNA-expressing transgenes into different chromosomal sites allows scientists to systematically test the influence of local chromatin context on lncRNA activity (Table 2). The targeted approach integrates the lncRNA transgene at a single genomic site, guided by homology between a sequence in the transfected plasmid and a pre-installed landing pad (*FRT* or *loxP*).[11,24,29,36,49,51,80,86] This method reduces the occurrence of multiple insertions and avoids position-dependent variations in lncRNA expression and activity. Wutz et al. used such a targeted strategy in the lncRNA installation system by first inserting a *loxP* site and fluorescent reporter (Hprt-pBI-EGFP-lox-neo) at the *Hprt* locus, then inserting various small CMV-*Xist* constructs via Cre recombinase (Figure 4C).[36] This streamlined strategy allows for efficient and flexible generation of *Xist* transgenes with controlled expression. FRT landing pads have been introduced into human autosomes, chromosomes 3 and 8, and were used to generate single-copy insertions of *XIST* sequence-expressing transgenes in HT1080 cells.[24,29,49]

Artificial reporter genes to visualize lncRNA-mediated regulation dynamics

Traditionally, *cis*-restricted effects have been determined by changes in expression levels of endogenous chromosomal genes near the transgene insertion site. Artificial reporter genes, such as luciferase, EGFP, and β-galactosidase, have provided a means to consistently quantify the impact of lncRNAs on gene expression (Figure 4D). Minks et al. used an elegant system in human HT1080 cells, where an *EGFP* reporter was inserted ~7.7 kb downstream of the *XIST* transgene integrated via FRT sites.[24] Upon dox-induced expression of full-length or truncated *XIST* constructs, *EGFP* expression was efficiently silenced, enabling quantitative comparisons between full-length and truncated versions of *XIST*. In *Drosophila*, *roX1* and *roX2* (*Hsp83* driven) were placed upstream of an insulated *lacZ* reporter gene to assess dosage compensation.[85] The *roX*-expressing transgenes functioned as chromatin entry sites by recruiting the male-specific lethal (MSL) complex and activated the lacZ reporter and nearby genes in a sex-specific, *cis*-acting manner. Collectively, these reporter-based systems allow researchers to consistently and sensitively measure the regulatory potential of lncRNAs, providing a versatile platform for dissecting structure-function relationships and validating synthetic constructs.

RNA and protein tagging methods for tracking lncRNA abundance and localization

Another customizable feature is the addition of nucleic acid and protein tags onto nuclear lncRNAs and proteins to quantify their expression levels and track their localization in living cells (Figure 4E). Compared with fluorescence cytology that uses nucleic acid probes and antibodies in fixed cells, tagging readily

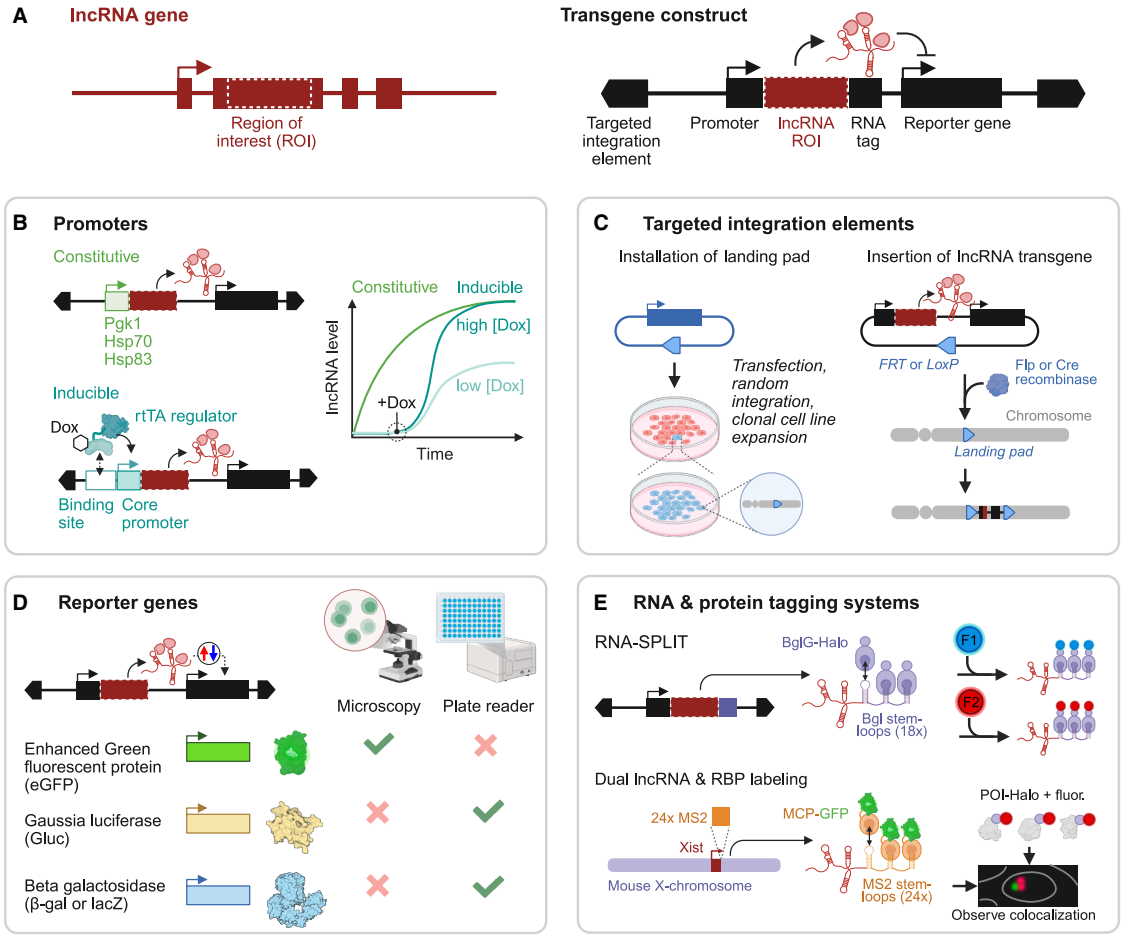


Figure 4. Customizable elements of streamlined lncRNA transgenic systems

(A) General structure of lncRNA-expressing constructs containing a promoter, lncRNA or region of interest (ROI) within the full-length sequence, an RNA sequence tag, and a reporter gene, with targeted insertion elements at either end.

(B) Promoter systems for lncRNA expression. Top: constitutive promoters drive continuous expression of the lncRNA. Bottom: a dox-inducible system (Tet-On) uses the CMV-TetO promoter activated by the reverse tetracycline transactivator (rtTA) in the presence of dox, which binds to the Tet operator sites (binding region), allowing temporal control of lncRNA expression.

(C) Targeted integration of a transgene in cells. A vector carrying FRT or *loxP* sites is randomly integrated into the host genome to create a genomic landing pad. These recombination sites enable targeted insertion of an expression vector containing an lncRNA transgene flanked by matching FRT and *loxP* sites. Site-specific integration is facilitated by Flp or Cre recombinase.

(D) Reporter gene systems to monitor lncRNA expression or activity. Fluorescent (*eGFP*), luminescent (*GLuc*), or enzymatic (β-galactosidase) reporters enable detection through microscopy, luminometry, or colorimetric assays, respectively.

(E) RNA and protein tagging systems: RNA sequential pulse localization imaging over time (RNA-SPLIT), based on Bgl stem loops,[9] and an MS2 stem-loop system that was used in an lncRNA and protein colocalization study.[6]

enables time-resolved studies of nucleoprotein complex behavior. To label lncRNAs, stem-loop arrays such as Bgl and MS2 have been inserted into the endogenous lncRNA gene or into a transgenic lncRNA locus. A stem-loop-specific protein fused to a fluorescent tag is allowed to bind, enabling visualization of the amount and location of lncRNA. For instance, a system called RNA sequential pulse localization imaging over time (RNA-SPLIT) and super-resolution 3D structured illumination microscopy (3D-SIM) has been used to visualize *Xist* RNA dynamics in mESCs.[9] An 18× Bgl stem-loop array was inserted into exon 7 of an *Xist* transgene, and expressed lncRNA was detected with a stem-loop-binding BglG protein fused to a HaloTag-fluorophore conjugate.[87] Pulsed addition of different colored fluorophores was used to distinguish early versus late transcribed RNA. The data revealed that upon dox induction, *Xist* RNA initially spreads but later becomes confined into well-defined "*Xist* territories," consistent with a two-phase model of spreading and turnover.[9] In a similar strategy, MS2 RNA hairpin arrays were inserted into the endogenous *Xist* gene, and an MS2 coat protein (MCP)-GFP fusion protein was used for high-resolution tracking of *Xist* foci.[6] RBPs or histone H2B were tagged with Halo and conjugated with a far-red fluorophore. Dual-tagging of *Xist* and H2B enabled quantitative spatial analyses revealing ~50 discrete foci, called supramolecular complexes (SMACs), which contained two *Xist* transcripts each and were significantly more compact near *Xist* RNA than in the surrounding area. These results support a

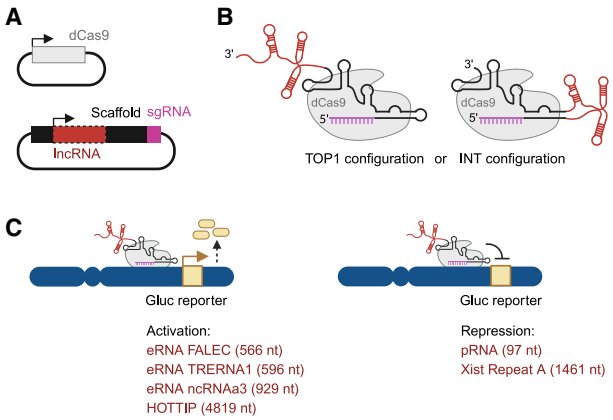


Figure 5. CRISP-Disp for targeted recruitment of functional lncRNA domains

(A) Expression cassettes for CRISP-Disp include a vector that expresses dCas9 and one that encodes the engineered lncRNA fused to a dCas9-binding scaffold and a genome-targeting sgRNA.

(B) Two different configurations were used to test the engineered sgRNAs with lncRNA extensions. TOP1 carries the lncRNA sequence at the 3′ of the sgRNA scaffold, while the INT configuration carries it internally.

(C) Activation-associated *eRNAs FALEC*, *TRERNA1*, and *ncRNAa3*, and the lncRNA *HOTTIP*, and repression-associated lncRNAs *pRNA* and *Xist* Rep A were targeted to a chromosomally inserted reporter gene (*GLuc*) to compare *cis*-regulatory effects.

mechanistic model where local RNA concentration drives efficient recruitment and confinement of epigenetic silencers.

CRISP-Disp platforms to deploy lncRNAs for site-directed *trans*-regulation

The transgenic methods discussed in the previous sections recapitulate the *cis*-regulatory activity of lncRNAs in the vicinity of the lncRNA transcription site. CRISPR-display (CRISP-Disp) provides a means to decouple lncRNA function from its transcription site by tethering lncRNAs to pre-defined genomic loci for *trans*-regulation. Experimental evidence suggests that tethering of lncRNA molecules at some genomic sites is mediated by the proteins Yin Yang 1 (YY1), CTCF and heterogeneous nuclear ribonucleoproteins (hnRNPs) hnRNPU and hnRNPK.[1,88] In CRISP-Disp, lncRNA anchoring is mediated by a complex containing a dead Cas9 protein (dCas9) and a modified single guide RNA (sgRNA).[17] The sgRNA consists of a dCas9-binding stem-loop scaffold, a 20-nt DNA targeting region, and a 3′ extension that includes the lncRNA sequence (Figure 5A). Two different topologies were tested for the modified sgRNAs, where the lncRNA sequence of interest was either inserted internally within the engineered loop of the sgRNA (INT) or appended to its 3′ end (TOP1) (Figure 5B). These hybrids, co-delivered with dCas9, were targeted to a promoter-proximal region of a chromosomally integrated Gaussia luciferase reporter (*GLuc*) in HEK293FT cells. The *Xist* Rep A (1,461 nt), a conserved silencing module, when fused to sgRNAs and directed to the luciferase promoter, induced partial repression (~30%–50%), demonstrating RepA's intrinsic silencing capacity (Figure 5C).[17]

The repressive nucleolar remodeling complex (NoRC)-binding pRNA stem-loop (*pRNA*) also induced transcriptional silencing when targeted using CRISP-Disp. Putative transcription-activating lncRNAs have also been tested in this system. *HOTTIP* lncRNA (~4,799 nucleotides) significantly activated transcription,[89] achieving ~2- to 3-fold upregulation. Enhancer RNAs (*eRNAs*), including *ncRNA-a3* (921 nt) and *TRERNA1* (~596 nt),[90,91] also promoted significant activation of *GLuc*, while *FALEC* (~566 nt) showed no significant effect, suggesting position-dependent activity.[17,92] Potentially, co-expression of several site-specific sgRNA-lncRNA hybrids could support multiplexing, enabling simultaneous targeting of distinct genomic sites for complex regulatory studies.[91]

Considerations when using synthetic lncRNA expression tools

One key design insight from the work summarized here is that tuning transcript dosage has major consequences for lncRNA function and provides a useful handle for probing regulatory dynamics. Second, the intrinsic properties of the affected promoter can influence sensitivity to lncRNA-mediated regulation: in an episomal transgene system, the 5′ repeat region of mouse *Xist* induced *cis*-silencing of luciferase reporters under the control of promoters derived from X-linked genes *Pgk1*, *Hprt*, and *G6pd*, but not autosomal promoters from *Aprt*, *Ins*, or viral promoter SV40,[79] consistent with silencing-resistant escape genes.[93,94] Additionally, extrinsic factors such as cell state, tissue type, and age may influence target gene sensitivity.[95,96] Third, targeted landing pads and tethering strategies allow "plug-in" style testing of various lncRNA sequences at a single locus. Landing pads could also be installed at different loci to test lncRNAs in different genomic contexts. Finally, dynamic reporters and molecular tags allow spatiotemporal tracking of regulation, while CRISP-Disp demonstrates that lncRNA modules can be decoupled from their transcription site and redeployed in *trans*. Together, these tools lay the groundwork for systematic testing of lncRNA modularity and the rational construction of synthetic regulators.

TOWARD A MODULAR DESIGN FRAMEWORK: INTERCHANGEABLE PARTS DERIVED FROM lncRNA-PROTEIN COMPLEXES

Approaching synthetic lncRNAs as substrates for rational design requires the identification of modular parts that can be mixed, matched, and redeployed in new contexts.[97] In the following sections, we describe lncRNA composition as a structural hierarchy[98] that reflects how an engineer might build a synthetic construct from first principles. At the most fundamental level, sequence composition encodes baseline functional potential. Layered onto this are functional motifs that dictate subcellular positioning, such as nuclear retention sequences, and anchoring elements that tether lncRNAs to chromatin or RNA targets. Higher-order RNA structures provide scaffolding frameworks that stabilize interactions, particularly with proteins. Finally, RNA-protein interfaces function as recruitable effectors that execute regulatory outcomes. This progression and/or complexity, which includes primary sequence information through to multi-molecular interactions, illustrates how natural lncRNA features might be repurposed as interchangeable parts in programmable regulators of chromatin.

K-mer composition: A programmable sequence code?

One challenge with forward design of synthetic lncRNA at the nucleotide level is that lncRNA function is evolutionarily maintained

Table 2. Transgenic systems for lncRNA functional dissection

Transgene system[a]	lncRNA promoter[b]	lncRNA integration[c]	Reporter or tag	Host cells	lncRNA expressed
Hprt-pBI-EGFP-lox-neo (custom)[36,51,80]	inducible: *CMV*, rtTA, dox(+)	targeted: Cre/*loxP* at *Hprt* chr. X	EGFP fluorescence	mESCs	full-length and truncated m*Xist*[11,36]
pcDNA5/FRT/TO (ThermoFisher)	inducible: *CMV*, rtTA, dox(+)	targeted: Flp/*FRT* chr. 3	EGFP fluorescence	human HT1080	h*XIST* (full length and Rep A)[24,49]
		targeted: Flp/*FRT* chr. 8	none	human HT1080	hybrid m*Xist* (PID) h*XIST* (AFE)[29]
pTRE-tight vector (Takara Bio Europe/ Clontech)[9,59,81]	inducible: *core CMV (TRE-tight)*,[82] rtTA, dox(+)	targeted: Cre/*loxP* chr. 15 and X	halo-tag fluorescence	mESCs	m*Xist* with BglI stem loops[9]
		targeted: Cre/*loxP* chr. 3, 12 and 17	none	mESCs	m*Xist* exons 1–8[59]
ptetOP-Xist-PA (custom)[51]	inducible: rtTA3, dox(+)	targeted: Cre/*loxP* chr. 11	halo-tag fluorescence	mESCs	m*Xist* (ΔA, ΔB/C mutants)[6]
pCaSpeR-h83T3, pCaSpeR-hs[78,83]	constitutive: *Hsp83* and *Hsp70*	random; P-element autosome	none	*Drosophila*	*roX1* or *roX2*[78,83]
pCaSpeR, pCaspeR-hs[84,85]	constitutive: *Hsp83*	random: P-element chr. 3	lacZ β-galactosidase assay	*Drosophila*	*roX1* or *roX2*[84,85]
pGL3-Pgk1 (custom)	constitutive: *Pgk1*	non-integrated: episomal	luciferase luminescence	mouse BALB/3T3 cell line	m*Xist* conserved 5′ repeats[79]

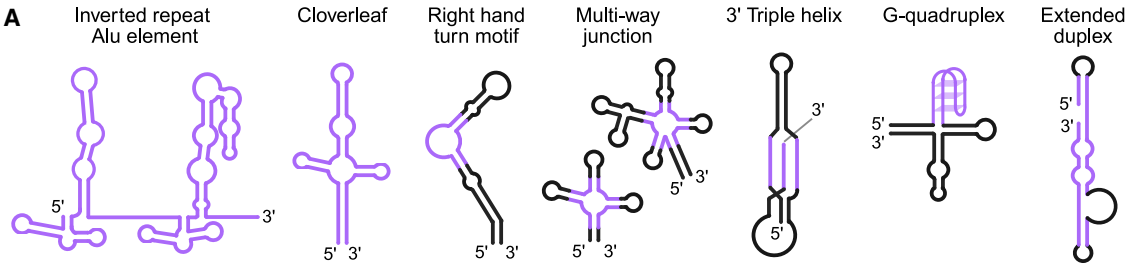
[a]Transgene systems are plasmids into which lncRNA-encoding DNAs have been introduced. Custom-built (custom) plasmids are listed as named in the cited report. Vendors are shown for commercially available plasmids.
[b]Inducible promoters are listed as the promoter symbol followed by the artificial regulator protein (e.g., rtTA) and its chemical modulator (e.g., dox). CMV, cytomegalovirus promoter; rtTA, bacterial TetR (reverse tetracycline-controlled transactivator), VP16 fusion protein; dox(+), transcription activated by dox; tet, tetracycline; m*Xist*, mouse *Xist*; *hXIST*, human *XIST*; halo-tag, synthetic protein fusion tag that binds to chemical ligands (detectable).
[c]lncRNA integration sites listing either random site or targeted integration, Cre/*loxP* or Flp/FRT, indicate the integration system used in the study and are followed by the chromosome number. Random insertion was done via P-elements (transposon-mediated random insertion), and non-integrated expression was done via episomal/extrachromosomal plasmid.

without obvious linear sequence conservation.[99] Instead, conservation has been observed as frequencies of *k*-mers, nucleotide substrings of length k that can appear in any order and may overlap one another.[100] *K*-mer analysis provides a global statistical signature, where enrichments of *k*-mers correlate with repressive or activating activity. For instance, the sequence evaluation from *K*-mer representation (SEEKR) method showed that lncRNAs with related functions, such as *cis*-repressors or *cis*-activators, share distinct *k*-mer frequency profiles.[101] This principle was validated using synthetic lncRNAs that lacked linear homology to the repressive murine *Xist* gene. The synthetic lncRNAs' ability to silence a reporter gene correlated directly with their *k*-mer similarity to *Xist*, establishing that k-mer composition itself is a key functional determinant independent of primary sequence alignment.[101] Many RBPs, which we discuss under "RNA-protein interactions as recruitable effectors," bind to consensus *k*-mers, as well as non-consensus sequences. *K*-mers could be applied to synthetic lncRNA design by guiding sequence composition, i.e., for desired protein recruitment, but they cannot yet substitute for mechanistic motifs (which we describe in the following sections). Therefore, *k*-mers are more suited as predictive design constraints rather than stand-alone, composable building blocks.

Nuclear retention sequences as localization tags

Nuclear accumulation of many lncRNAs is a regulated outcome of specific *cis*-acting sequence elements that function as nuclear retention signals.[102] These sequences can be thought of as localization tags, modular motifs that, when appended or deleted, directly determine whether an RNA remains nuclear or diffuses into the cytoplasm. Classic examples include the *MALAT1* lncRNA which contains two defined regions (E and M), whose loss led to cytoplasmic redistribution and dispersal from nuclear speckles.[103] Likewise, tandem repeat domains (RRDs) within *FIRRE* can redirect a normally cytoplasmic mRNA into the nucleus, and this retention is lost upon depletion of the RBP hnRNPU.[104] More compact motifs can function in a similar manner: AGCCC pentamers within *BORG* serve as discrete nuclear retention signals, with even a single copy sufficient to anchor a reporter RNA in the nucleus.[105] In another case, SIRLOIN elements, pyrimidine-rich sequences often derived from short interspersed nuclear element (SINE) repeats, have been shown in *JPX* and *PVT1* to confer nuclear localization through hnRNPK recruitment, with either motif mutation or RBP knockdown resulting in cytoplasmic relocalization.[106,107] Transposable element-derived sequences within lncRNAs, such as mammalian-wide interspersed repeat (MIR) and L2b, also function as nuclear retention domains. Their mutation within specific lncRNAs results in a significant cytoplasmic shift.[108] Together, these studies demonstrate that nuclear retention can be engineered through short motifs, repeat elements, or tethering domains, often acting via RBPs such as hnRNPU and hnRNPK.

From a design perspective, nuclear-targeting motifs operate like address labels for synthetic lncRNAs, offering modular handles to control nuclear residency. However, the modularity of nuclear retention sequences should be viewed with caution: their activity can depend on RNA context, unexpected secondary structure formation, and cell-specific differences in the



B

Structured motif	lncRNA host	Cellular function
Inverted repeat alu element (IRAlu)	p21 lncRNA (2 isoforms)	Cellular localization
Cloverleaf	Cyrano (mouse, human, zebrafish)	microRNA degradation
Right-hand turn (RHT), Multi-way junction	Braveheart, COOLAIR (3 isoforms), SRA	Long range 3-D interactions, protein binding
3' end triple helix	MALAT1, NEAT1 (2 isoforms)	Stability
G-quadruplex (G4)	TERRA	Stability, protein binding, splicing regulation
Extended duplex	XIST	Transcriptional silencing

Figure 6. lncRNA structured motifs
(A) The seven primary structural motifs identified in lncRNAs and (B) their corresponding lncRNA hosts and functions.

availability of interacting RBPs. As a result, localization tags that appear portable in reporter assays may not always function predictably in diverse cellular environments.

Triplexes, R-loops, and RNA-RNA interactions as anchoring elements

Anchoring lncRNAs to specific genomic sites can be achieved through sequence-encoded interactions with DNA or RNA that form RNA-DNA triplexes, R-loops, and RNA-RNA interactions.[109,110] RNA-DNA triplexes are structures where an RNA strand associates with duplex DNA via Hoogsteen base-pairing. These structures often arise at purine-pyrimidine-rich tracts and can be predicted computationally and validated experimentally using RNase H insensitivity, mutational controls, and biophysical methods such as circular dichroism and NMR.[111,112] Triplex formation has been described for the lncRNA *Fendrr*, which regulates cardiac development potentially through polycomb repressive complex (PRC2) recruitment.[113,114] Other mechanisms include recruiting p300, histone methyltransferase, and other chromatin regulators to target genes (*Sarrah*,[115] *GAU1*,[116] *MEG3*, and *HOTAIR*[117,118]) and recruiting the HUSH complex under hypoxia (Hif1alpha-AS[119]). *Khps1* contains a triplex-forming sequence (TFS) that localizes to the sphingosine kinase 1 (SPHK1) enhancer, and this TFS is portable to other RNAs.[120]

R-loops, by contrast, represent a distinct three-stranded structure in which an RNA strand invades duplex DNA and displaces one DNA strand.[121,122] These hybrids are formed primarily co-transcriptionally, where nascent RNAs reanneal to the template DNA in *cis*.[123] They are characterized by RNase H sensitivity and can be detected by S9.6 antibody binding, which specifically recognizes DNA:RNA hybrids rather than Hoogsteen triplexes.[121,122] The vimentin antisense RNA1 (*VIM-AS1)* lncRNA exemplifies this mechanism, forming an R-loop at the VIM promoter. Disruption of this structure diminishes its genomic association.[124] It is important to note that evidence for *trans*-acting R-loops is lacking, limiting the potential use of R-loops in synthetic systems to *cis*-regulatory configurations. RNA-RNA interactions involve inter- and intra-molecular pairing that can reshape lncRNA structure to hinder or promote protein binding to the lncRNA.[125,126] *HOTAIR* forms double-stranded RNA (dsRNA) with a target nascent RNA. Evidence suggests that *HOTAIR* interacts with PRC2 proteins, which bridges PRC2 with the chromatin near the nascent RNA to induce repression.[125]

Systematic engineering of such interactions is still nascent, representing an underexplored route to programmable anchoring. From a design standpoint, triplexes, R-loops, and RNA-RNA interactions have been defined at the sequence level via genetic perturbation studies, suggesting potential utility as modules for targeting lncRNAs to chromatin or transcripts. Their portability, however, may be hampered by sequence context-dependent function, unknown interactions with DNA- and RBPs, and inconsistent structural stability in diverse cellular contexts. These are important limitations to consider when deploying these elements in synthetic constructs.

Stem-loop structures as modular scaffolds

RNA structure is hierarchical in nature: primary sequences that comprise *k*-mers, for example, fold into secondary structured elements, which in turn govern tertiary and quaternary structure and interactions.[127] Among lncRNAs, several structured motifs have been identified, including inverted repeat Alu element stem loops, cloverleaf structures, right-hand turn motifs, 3′ end triple helices, G quadruplexes, and extended duplexes (Figure 6A).[128–135] These motifs facilitate interactions with other nucleic acids, ligands, and proteins to carry out regulatory roles in the cell (Figure 6B). Because many lncRNA-protein interactions depend on the formation of specific secondary or tertiary RNA structures, structural motifs often serve as scaffolds for effector recruitment.

Among these, RNA G quadruplexes (rG4s) form a distinct class of higher-order motifs stabilized by guanine tetrads and

monovalent cations such as K^+. Recent studies have identified rG4 elements within several nuclear lncRNAs, including *XIST*, *MALAT1*, and *NEAT1*, that mediate specific protein interactions.[136–138] In particular, PRC2 was shown to recognize folded rG4 structures within *XIST* with higher affinity than non-G4 regions,[138,139] underscoring how discrete RNA folds can influence recruitment of chromatin-modifying complexes.

Given that RNA function is fundamentally driven by structure, these folded motifs represent potential modules that could be rationally concatenated into arrays to tune the function of synthetic lncRNAs. Little work has been reported on the systematic construction of synthetic lncRNAs using structure as the primary design principle.[24,29,36,129,140,141] This structure-centric approach holds considerable promise. If we can definitively map which structured domains confer specific functions, we could in principle concatenate them together to engineer RNAs with predictable, tunable activities.

Several key limitations must be addressed when considering the concatenation of RNA structural motifs. Constructs designed and screened *in vitro* may fold and behave differently in a cellular environment: RBPs, ribonucleases, and even ionic conditions may modulate the stability of individual structured domains within a larger synthetic construct, leading to context-dependent function that differs from the behavior of isolated motifs. Therefore, we must ensure that the intended secondary and tertiary structures actually fold as expected via, for instance, in-cell 2D structural probing. Additionally, the spatial arrangement and linker regions between motifs may significantly influence both domain folding and inter-domain interactions, potentially creating unexpected regulatory crosstalk or steric hindrance. In-cell structural validation is critical for addressing these limitations.[129,142,143]

RNA-protein interactions as recruitable effectors

LncRNAs exert much of their regulatory control through interfaces with chromatin proteins that can either recruit effectors to genomic loci or block their activity to prevent repression. In transcriptionally plastic domains, the best-studied mechanism is Polycomb's involvement with *XIST*, *HOTAIR*, and possibly taurine up-regulated gene 1 (TUG1)'s function, which involves PRC2/enhancer of zeste homolog 2 (EZH2) mediated H3K27me3 accumulation and gene silencing.[144–146] Other lncRNAs, including *MEG3* and Chaer, have been linked to binding-mediated degradation of EZH2 and loss of PRC from chromatin to allow gene activation.[38,40,147] However, Polycomb engagement remains controversial as several studies argue that PRC2 exhibits largely promiscuous or low-specificity RNA binding *in vitro*, that lncRNA-mediated silencing can be decoupled from PRC, and that many in-cell claims of direct PRC2-RNA binding are artifactual.[148–151]

Constitutive silencing-associated chromatin protein SUV39H1 (which generates H3K9me3) is recruited to chromatin by a 200-nucleotide domain within lncRNAs *hOCT4P3* (human) and *mOct4P4* (mouse).[152] In transcriptionally active chromatin, the lncRNA *HOTTIP* promotes transcription by recruiting mixed-lineage leukemia protein (MLL)/WD repeat-containing protein 5 (WDR5) complexes.[153,154] Others, such as *MALAT1* and *NEAT1*, use repeat elements or stem-loop structures to scaffold RBPs like hnRNPs, non-POU domain-containing octamer-binding protein (NONO), and serine/arginine-rich splicing factor (SRSF) proteins to organize nuclear bodies and splicing assemblies.[41,155–157] These studies highlight the potential to harness RNA-protein interfaces as modular switches to direct distinct transcriptional outcomes, including repression, activation, and nuclear organization.

Interactomes of individual lncRNAs dictate functional specificity using interactions that confer distinct activities.[158] These include *XIST* interactions with split ends (SPEN) and RNA-binding motif protein (RBM15) via its RepA domain to modulate transcriptional silencing, *NEAT1*'s interactions with NONO, paraspeckle component 1 (PSPC1), and splicing factor proline- and glutamine-rich (SFPQ) to support paraspeckle nucleation, and HOTAIR's interactions that mediate its regulation of gene clusters in *trans*, via lysine-specific histone demethylase 1A (LSD1) (CoREST) to couple H3K4 demethylation with PRC2-mediated H3K27 trimethylation. Individual lncRNAs frequently converge on the same proteins (e.g., hnRNPK and hnRNPU) and may produce similar or divergent outcomes depending on sequence motifs, RNA structures, or protein partners engaged.[73,158–160] For example, hnRNPU binds *Firre* to guide its *trans*-chromosomal interactions,[161] whereas hnRNPK binds *lincRNA-p21* to facilitate p53-dependent repression[162] and also binds the B-repeat of *Xist* to help establish H3K27me3.[163] Therefore, while some protein-lncRNA interactions are distinct and may support insulated recruitment, others involve shared proteins whose activities may create crosstalk, underscoring the need for strategies that can selectively insulate or rewire these interfaces.

Data from efforts to characterize and predict protein recruitment provide useful information for rational design. Technologies such as RNA ImmunoPrecipitation followed by deep sequencing (RIP-seq), CrossLinking and ImmunoPrecipitation followed by deep sequencing (CLIP-seq), Split and Pool Identification of RBP targets (SPIDR), Oligonucleotide-mediated proximity-interactome MAPping (O-MAP), and Comprehensive Identification of RNA-binding Proteins by Mass Spectrometry (ChIRP-MS) have mapped thousands of lncRNA-protein interactions, providing empirical data for predictive modeling.[145,163–167] Algorithms like Catalonia fast predictions of RNA And Protein Interactions and Domains (catRAPID) predict binding based on sequence features and physicochemical properties, validated with interactions between *XIST*, *HOTAIR*, and PRC2,[168] with single-cell transcriptome-based RAPID (scRAPID) integrating single-cell regulatory networks to refine predictions.[169] Machine learning models such as Higher-Order Nucleotide Encoding Convolutional Neural Network-Based Method (HOCNNLB) and Bidirectional Encoder Representations from Transformers for RNA-Binding Proteins (BERT-RBP), including those that use k-mer encoding or transformer attention, have emerged as a powerful approach to predict protein binding sites with high accuracy.[170–173] Snail family transcriptional repressor 1 (SNAIL)-binding motifs that have been identified within *HOTAIR* were experimentally validated as sufficient to redirect protein occupancy and modulate histone modification.[23,172,174] Finally, high-throughput approaches have expanded the resolution and scale at which RNA-protein binding modules can be identified.[107] These tools frame lncRNA-protein binding as a design space that can be computationally explored and experimentally tested.[166,168,169]

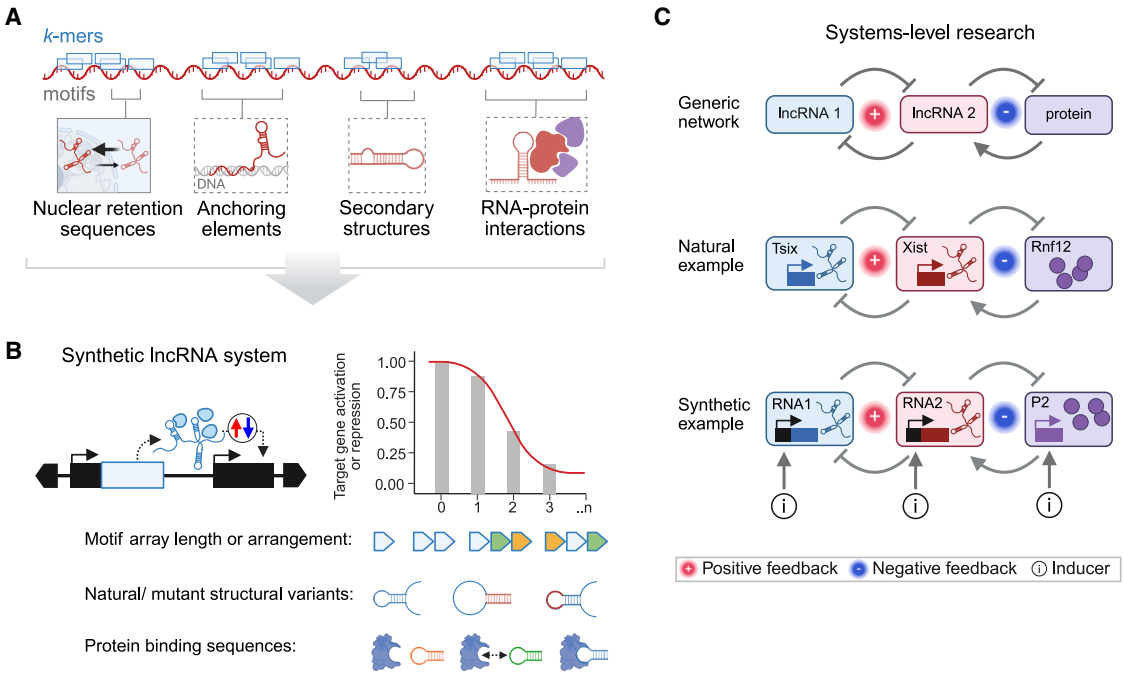


Figure 7. lncRNA structure-guided experimental design

(A) A hierarchical assembly view of lncRNAs, with *k*-mer sequence codes underlying modular motifs for localization, anchoring, structural scaffolding, and protein-recruiting effector functions.

(B) Schematic of how synthetic biology and quantitative analysis can be used to refine understanding of the biological function of chromatin-modulating lncRNAs.

(C) An example of emergent dynamical behaviors illustrated as a generic lncRNA circuit with feedback loops. This topology represents the regulatory relationships between the lncRNAs *Xist* and *Tsix* and the protein Rnf12. Potentially, such a network could be built with a pair of mutually repressive lncRNAs (RNA1 and RNA2) and a protein activator of RNA2 (P2). Inducible promoters could enable the control of component levels.

Synthetic and systems approaches to testing regulators built from modular parts

Synthetic and systems-level strategies represent complementary paths for advancing lncRNA-mediated epigenetic engineering. At present, the synthetic constructs and design frameworks discussed here are conceptual and largely theoretical, representing forward-looking ideas that could guide future experimental realization. Synthetic biology defines and assembles modular parts into custom regulators (Figures 7A and 7B), while systems biology investigates how these parts interact dynamically within gene networks to generate emergent behaviors (Figure 7C). On the synthetic side, systematic testing of lncRNA functional modules moves beyond the study of native transcripts to the rational engineering of artificial effectors for programmable epigenetic control. For instance, a repressor could be constructed by first encoding an SD based on the *k*-mer frequency profile of a known repressive lncRNA like *Xist*,[101] or by using tandem repeats of key motifs to achieve additive and potent silencing effects.[24] This functional domain could then be linked to a genomic targeting module, such as a guanine and adenine (GA)-rich, TFS designed to bind a specific promoter,[117,118] ensuring the repressive activity is directed to the intended locus. To guarantee the synthetic lncRNA operates in the correct subcellular compartment, a potent nuclear retention sequence, such as the RRDs from *FIRRE* or the AGCCC motifs from *BORG*, could be appended to the transcript, ensuring its accumulation in the nucleus where it can engage with chromatin.[104,105,175,176] Reporter gene regulation assays could be designed to systematically test regulatory tunability, for instance, to determine the additive effect of motif arrays, the impact of different RNA folds, or the effects of weak or strong RBP affinity (Figure 7B).

At the systems level, functions of synthetic or natural lncRNAs can be treated as modular components within larger regulatory circuits, where their combinatorial activities drive emergent dynamics. Mathematical models have provided insights into such behaviors of endogenous lncRNA networks. For example, *Xist*-mediated repression of the antisense lncRNA *Tsix* may generate positive feedback (double-negative in this case), whereas *Xist*-mediated repression of its activator *Rnf12* may provide the link for negative feedback (Figure 7C).[177,178] Mutzel et al. showed that the combination of a negative feedback loop and a positive feedback loop, both involving *Xist*, can produce a bistable switch ensuring mono-allelic upregulation of *Xist*.[179] While their study validated this architecture in the context of X chromosome inactivation, an open question is whether such feedback circuits can be reconstituted as independent, portable modules that toggle large-scale chromatin repression. A second key question is whether complex lncRNA circuits can be decomposed into smaller functional modules, each contributing distinct dynamical features such as bistability or oscillation.[180] These two questions are well-suited for synthetic platforms where modular components are used to build regulator lncRNAs, and synthetic promoters and protein regulators are used as control points to manipulate the system (Figure 7C). Ultimately, a combined

synthetic and systems approach is needed to successfully utilize lncRNA-based regulation systems for practical applications.

APPLICATION OUTLOOK: DISEASE RESEARCH AND THERAPY

The ability to engineer lncRNA structure and expression presents a new avenue for transformative biomedical applications. The most compelling strategy is using ectopic *XIST* expression to epigenetically correct chromosomal aneuploidies. By integrating the *XIST* gene into chromosome 21, researchers successfully silenced the extra chromosome copy in trisomy 21 (T21) induced pluripotent stem cells (iPSCs), rescuing deficits and accelerating neural rosette formation.[27,28,181–183] For this approach to be viable, efficient delivery (e.g., via exosomes or nanoparticles) is paramount.[184,185] However, efficacy is generally highest in pluripotent cells, and induction in differentiated somatic cells often yields only partial silencing, highlighting a key translational challenge.[182] Engineered variants with enhanced activity, or multisite tethering approaches such as CRISP-Disp, may help address this limitation, though in practice the appropriate therapeutic window for T21 is likely restricted to early development. Furthermore, X-linked disorders driven by gene dysregulation, such as Rett syndrome, highlight the therapeutic potential of reprogramming X-chromatin to, for instance, restore *MECP2* expression.[186–191]

Studies that have manipulated stress-responsive lncRNAs, such as *MEG3*, *Chaer*, *MALAT1*, *PAPAS*, *DRAIR*, *HOTAIR*, *H19*, and *NEAT1*, suggest that these molecules can modulate chromatin complexes at target genes.[38–40,192–198] While these lncRNAs have not yet been integrated into the synthetic systems discussed in this review, the research findings point to a potential therapeutic avenue for altering chromatin organization through *cis*- or *trans*-regulation in diseases such as cancer and neurodegeneration.[147,186,193]

CONCLUSIONS

LncRNAs have emerged as potent epigenetic regulators and programmable tools for synthetic biology. Foundational studies on *XIST*, *Airn*, and *KCNQ1OT1* demonstrated their capacity to induce widespread chromatin silencing through long-range *cis*-regulatory effects.[47,60–62] Early efforts to study these functions relied on megabase-scale BACs and YACs encoding full-length *XIC*, which were technically cumbersome and context-dependent.[25,44,58] Advances in synthetic biology have since enabled rational dissection of lncRNAs into smaller functional modules, including repeat domains, triplex-forming motifs, and even minimal *k*-mers predictive of silencing activity.[24,101,117] These insights have shifted the field toward constructing compact systems capable of replicating essential lncRNA functions, allowing for ectopic expression and modular recombination across contexts. Platforms incorporating inducible promoters, targeted genomic integration (e.g., Cre/*loxP* and Flp/FRT), and live-cell tagging now allow precise control and visualization of lncRNA behavior in time and space.[9,24,36,80,86] The CRISP-Disp system further extends this framework by enabling programmable tethering of lncRNA-derived domains to specific loci, revealing modular silencing or activating capacities of lncRNA sub-regions.[17,29,101] Looking ahead, the convergence of modular lncRNA design with genome-targeting offers a path toward programmable RNA-based regulators that not only deepen mechanistic insight into chromatin biology but also lay the groundwork for innovative therapeutic strategies.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the many important contributions to the field that could not be cited in this review due to space limitations. This work was supported by the National Science Foundation (NSF-2243665 to K.A.H., NSF-2243667 to A.M.J., and NSF-2243666 to K.N.S.) and the National Institutes of Health (GM144358 to A.M.J. and R35GM149531 to T.H.). We thank our collaborators, including A. Corbett, and lab members for their valuable input, discussions, and suggestions throughout the development of this work.

AUTHOR CONTRIBUTIONS

S.S., S.H.K., and K.A.H. conceptualized and wrote the review. K.N.S. and A.M.J. contributed to writing and editing. A.J. and T.H. also provided additional writing contributions.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors S.S., K.A.H., and S.H.K. used ChatGPT 5 to streamline text as needed and to find additional relevant literature that may have been overlooked. After using this tool/service, the authors reviewed and manually edited the content and take full responsibility for the content of the published article.

REFERENCES

1. Statello, L., Guo, C.-J., Chen, L.-L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. Nat. Rev. Mol. Cell Biol. *22*, 96–118. https://doi.org/10.1038/s41580-020-00315-9.

2. Mattick, J.S., Amaral, P.P., Carninci, P., Carpenter, S., Chang, H.Y., Chen, L.-L., Chen, R., Dean, C., Dinger, M.E., Fitzgerald, K.A., et al. (2023). Long non-coding RNAs: definitions, functions, challenges and recommendations. Nat. Rev. Mol. Cell Biol. *24*, 430–447. https://doi.org/10.1038/s41580-022-00566-8.

3. Guttman, M., and Rinn, J.L. (2012). Modular regulatory principles of large non-coding RNAs. Nature *482*, 339–346. https://doi.org/10.1038/nature10887.

4. Ulitsky, I., and Bartel, D.P. (2013). lincRNAs: genomics, evolution, and mechanisms. Cell *154*, 26–46. https://doi.org/10.1016/j.cell.2013.06.020.

5. Zacco, E., Broglia, L., Kurihara, M., Monti, M., Gustincich, S., Pastore, A., Plath, K., Nagakawa, S., Cerase, A., Sanchez de Groot, N., et al. (2024). RNA: The unsuspected conductor in the orchestra of macromolecular crowding. Chem. Rev. *124*, 4734–4777. https://doi.org/10.1021/acs.chemrev.3c00575.

6. Markaki, Y., Chong, J.G., Wang, Y., Jacobson, E.C., Luong, C., Tan, S.Y.X., Jachowicz, J.W., Strehle, M., Maestrini, D., Banerjee, A.K., et al. (2021). Xist nucleates local protein gradients to propagate silencing across the X chromosome. Cell *184*, 6212. https://doi.org/10.1016/j.cell.2021.11.028.

7. Sunwoo, H., Wu, J.Y., and Lee, J.T. (2015). The Xist RNA-PRC2 complex at 20-nm resolution reveals a low Xist stoichiometry and suggests a hit-and-run mechanism in mouse cells. Proc. Natl. Acad. Sci. USA *112*, E4216–E4225. https://doi.org/10.1073/pnas.1503690112.

8. Di Michele, F., Chillón, I., and Feil, R. (2023). Imprinted Long Non-Coding RNAs in Mammalian Development and Disease. Int. J. Mol. Sci. *24*, 13647. https://doi.org/10.3390/ijms241713647.

9. Rodermund, L., Coker, H., Oldenkamp, R., Wei, G., Bowness, J., Rajkumar, B., Nesterova, T., Susano Pinto, D.M., Schermelleh, L., and Brockdorff, N. (2021). Time-resolved structured illumination microscopy reveals key principles of Xist RNA spreading. Science *372*, eabe7500. https://doi.org/10.1126/science.abe7500.

10. Marks, H., Chow, J.C., Denissov, S., Françoijs, K.-J., Brockdorff, N., Heard, E., and Stunnenberg, H.G. (2009). High-resolution analysis of epigenetic changes associated with X inactivation. Genome Res. *19*, 1361–1373. https://doi.org/10.1101/gr.092643.109.

11. Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., et al. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science *341*, 1237973. https://doi.org/10.1126/science.1237973.

12. Andergassen, D., Dotter, C.P., Wenzel, D., Sigl, V., Bammer, P.C., Muckenhuber, M., Mayer, D., Kulinski, T.M., Theussl, H.-C., Penninger, J.M., et al. (2017). Mapping the mouse Allelome reveals tissue-specific regulation of allelic expression. eLife *6*, e25125. https://doi.org/10.7554/eLife.25125.

13. Lewandowski, J.P., Dumbović, G., Watson, A.R., Hwang, T., Jacobs-Palmer, E., Chang, N., Much, C., Turner, K.M., Kirby, C., Rubinstein, N.D., et al. (2020). The Tug1 lncRNA locus is essential for male fertility. Genome Biol. *21*, 237. https://doi.org/10.1186/s13059-020-02081-5.

14. Andergassen, D., and Rinn, J.L. (2022). From genotype to phenotype: genetics of mammalian long non-coding RNAs in vivo. Nat. Rev. Genet. *23*, 229–243. https://doi.org/10.1038/s41576-021-00427-8.

15. Zhao, E.M., Mao, A.S., de Puig, H., Zhang, K., Tippens, N.D., Tan, X., Ran, F.A., Han, I., Nguyen, P.Q., Chory, E.J., et al. (2022). RNA-responsive elements for eukaryotic translational control. Nat. Biotechnol. *40*, 539–545. https://doi.org/10.1038/s41587-021-01068-2.

16. Park, S.V., Yang, J.-S., Jo, H., Kang, B., Oh, S.S., and Jung, G.Y. (2019). Catalytic RNA, ribozyme, and its applications in synthetic biology. Biotechnol. Adv. *37*, 107452. https://doi.org/10.1016/j.biotechadv.2019.107452.

17. Shechner, D.M., Hacisuleyman, E., Younger, S.T., and Rinn, J.L. (2015). Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. Nat. Methods *12*, 664–670. https://doi.org/10.1038/nmeth.3433.

18. Cano-Rodriguez, D., and Rots, M.G. (2016). Epigenetic editing: On the verge of reprogramming gene expression at will. Curr. Genet. Med. Rep. *4*, 170–179. https://doi.org/10.1007/s40142-016-0104-3.

19. Rots, M.G., and Jeltsch, A. (2024). Development of locus-directed editing of the epigenome from basic mechanistic engineering to first clinical applications. Methods Mol. Biol. *2842*, 3–20. https://doi.org/10.1007/978-1-0716-4051-7_1.

20. Perotti, I., Broglia, L., Tartaglia, G.G., and Cerase, A. (2025). Xist condensates: perspectives for therapeutic intervention. Genome Biol. *26*, 215. https://doi.org/10.1186/s13059-025-03666-8.

21. Wang, B., Barahona, M., and Buck, M. (2014). Engineering modular and tunable genetic amplifiers for scaling transcriptional signals in cascaded gene networks. Nucleic Acids Res. *42*, 9484–9492. https://doi.org/10.1093/nar/gku593.

22. Cao, W., Liu, L., Sun, Q., Shan, Y., and Chen, Y. (2025). A framework for complex signal processing via synthetic biological operational amplifiers. Nat. Commun. *16*, 7037. https://doi.org/10.1038/s41467-025-62464-9.

23. Battistelli, C., Garbo, S., Riccioni, V., Montaldo, C., Santangelo, L., Vandelli, A., Strippoli, R., Tartaglia, G.G., Tripodi, M., and Cicchini, C. (2021). Design and functional validation of a mutant variant of the LncRNA HOTAIR to counteract snail function in epithelial-to-mesenchymal transition. Cancer Res. *81*, 103–113. https://doi.org/10.1158/0008-5472.CAN-20-1764.

24. Minks, J., Baldry, S.E.L., Yang, C., Cotton, A.M., and Brown, C.J. (2013). XIST-induced silencing of flanking genes is achieved by additive action of repeat a monomers in human somatic cells. Epigenetics Chromatin *6*, 23. https://doi.org/10.1186/1756-8935-6-23.

25. Heard, E., Mongelard, F., Arnaud, D., Chureau, C., Vourc'h, C., and Avner, P. (1999). Human XIST yeast artificial chromosome transgenes show partial X inactivation center function in mouse embryonic stem cells. Proc. Natl. Acad. Sci. USA *96*, 6841–6846. https://doi.org/10.1073/pnas.96.12.6841.

26. Grant, J., Mahadevaiah, S.K., Khil, P., Sangrithi, M.N., Royo, H., Duckworth, J., McCarrey, J.R., VandeBerg, J.L., Renfree, M.B., Taylor, W., et al. (2012). Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation. Nature *487*, 254–258. https://doi.org/10.1038/nature11171.

27. Jiang, J., Jing, Y., Cost, G.J., Chiang, J.-C., Kolpa, H.J., Cotton, A.M., Carone, D.M., Carone, B.R., Shivak, D.A., Guschin, D.Y., et al. (2013). Translating dosage compensation to trisomy 21. Nature *500*, 296–300. https://doi.org/10.1038/nature12394.

28. Chiang, J.-C., Jiang, J., Newburger, P.E., and Lawrence, J.B. (2018). Trisomy silencing by XIST normalizes Down syndrome cell pathogenesis demonstrated for hematopoietic defects in vitro. Nat. Commun. *9*, 5180. https://doi.org/10.1038/s41467-018-07630-y.

29. Navarro-Cobos, M.J., Morales-Guzman, S.I., Baldry, S.E.L., and Brown, C.J. (2023). Derivation of a minimal functional XIST by combining human and mouse interaction domains. Hum. Mol. Genet. *32*, 1289–1300. https://doi.org/10.1093/hmg/ddac285.

30. Motosugi, N., Sugiyama, A., Okada, C., Otomo, A., Umezawa, A., Akutsu, H., Hadano, S., and Fukuda, A. (2022). De-erosion of X chromosome dosage compensation by the editing of XIST regulatory regions restores the differentiation potential in hPSCs. Cell Rep. Methods *2*, 100352. https://doi.org/10.1016/j.crmeth.2022.100352.

31. Ma, Q., Yang, L., Tolentino, K., Wang, G., Zhao, Y., Litzenburger, U.M., Shi, Q., Zhu, L., Yang, C., Jiao, H., et al. (2022). Inducible lncRNA transgenic mice reveal continual role of HOTAIR in promoting breast cancer metastasis. eLife *11*, e79126. https://doi.org/10.7554/eLife.79126.

32. Lee, J.T., Strauss, W.M., Dausman, J.A., and Jaenisch, R. (1996). A 450 kb transgene displays properties of the mammalian X-inactivation center. Cell *86*, 83–94. https://doi.org/10.1016/S0092-8674(00)80079-3.

33. Lee, J.T., and Jaenisch, R. (1997). Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. Nature *386*, 275–279. https://doi.org/10.1038/386275a0.

34. Matsuura, S., Episkopou, V., Hamvas, R., and Brown, S.D. (1996). Xist expression from an Xist YAC transgene carried on the mouse Y chromosome. Hum. Mol. Genet. *5*, 451–459. https://doi.org/10.1093/hmg/5.4.451.

35. Heard, E., Kress, C., Mongelard, F., Courtier, B., Rougeulle, C., Ashworth, A., Vourc'h, C., Babinet, C., and Avner, P. (1996). Transgenic mice carrying an Xist-containing YAC. Hum. Mol. Genet. *5*, 441–450. https://doi.org/10.1093/hmg/5.4.441.

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

37. Trotman, J.B., Abrash, E.W., Murvin, M.M., Braceros, A.K., Li, S., Boyson, S.P., Salcido, R.T., Cherney, R.E., Bischoff, S.R., Kaufmann, K., et al. (2025). Isogenic comparison of Airn and Xist reveals core principles of Polycomb recruitment by lncRNAs. Mol. Cell *85*, 1117–1133.e14. https://doi.org/10.1016/j.molcel.2025.02.014.

38. Wang, Z., Zhang, X.-J., Ji, Y.-X., Zhang, P., Deng, K.-Q., Gong, J., Ren, S., Wang, X., Chen, I., Wang, H., et al. (2016). The long noncoding RNA Chaer defines an epigenetic checkpoint in cardiac hypertrophy. Nat. Med. *22*, 1131–1139. https://doi.org/10.1038/nm.4179.

39. Reddy, M.A., Amaram, V., Das, S., Tanwar, V.S., Ganguly, R., Wang, M., Lanting, L., Zhang, L., Abdollahi, M., Chen, Z., et al. (2021). lncRNA DRAIR is downregulated in diabetic monocytes and modulates the inflammatory phenotype via epigenetic mechanisms. JCI Insight *6*, e143289. https://doi.org/10.1172/jci.insight.143289.

40. Zou, D., Liu, L., Zeng, Y., Wang, H., Dai, D., and Xu, M. (2022). LncRNA MEG3 up-regulates SIRT6 by ubiquitinating EZH2 and alleviates nonalcoholic fatty liver disease. Cell Death Discov. *8*, 103. https://doi.org/10.1038/s41420-022-00889-7.

41. Taiana, E., Ronchetti, D., Todoerti, K., Nobili, L., Tassone, P., Amodio, N., and Neri, A. (2020). LncRNA NEAT1 in paraspeckles: A structural scaffold

for cellular DNA damage response systems? Noncoding. RNA *6*, 26. https://doi.org/10.3390/ncrna6030026.

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

43. (1996). Requirement for Xist in X chromosome inactivation. Trends Genet. *12*, 89. https://doi.org/10.1016/0168-9525(96)81414-4.

44. Heard, E., Mongelard, F., Arnaud, D., and Avner, P. (1999). Xist yeast artificial chromosome transgenes function as X-inactivation centers only in multicopy arrays and not as single copies. Mol. Cell. Biol. *19*, 3156–3166. https://doi.org/10.1128/MCB.19.4.3156.

45. Migeon, B.R., Chowdhury, A.K., Dunston, J.A., and McIntosh, I. (2001). Identification of TSIX, encoding an RNA antisense to human XIST, reveals differences from its murine counterpart: implications for X inactivation. Am. J. Hum. Genet. *69*, 951–960. https://doi.org/10.1086/324022.

46. Herzing, L.B., Romer, J.T., Horn, J.M., and Ashworth, A. (1997). Xist has properties of the X-chromosome inactivation centre. Nature *386*, 272–275. https://doi.org/10.1038/386272a0.

47. Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature *349*, 38–44. https://doi.org/10.1038/349038a0.

48. Vallot, C., and Rougeulle, C. (2013). Long non-coding RNAs and human X-chromosome regulation: a coat for the active X chromosome. RNA Biol. *10*, 1262–1265. https://doi.org/10.4161/rna.25802.

49. Chow, J.C., Hall, L.L., Baldry, S.E.L., Thorogood, N.P., Lawrence, J.B., and Brown, C.J. (2007). Inducible XIST-dependent X-chromosome inactivation in human somatic cells is reversible. Proc. Natl. Acad. Sci. USA *104*, 10104–10109. https://doi.org/10.1073/pnas.0610946104.

50. Augui, S., Filion, G.J., Huart, S., Nora, E., Guggiari, M., Maresca, M., Stewart, A.F., and Heard, E. (2007). Sensing X chromosome pairs before X inactivation via a novel X-pairing region of the Xic. Science *318*, 1632–1636. https://doi.org/10.1126/science.1149420.

51. Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol. Cell *5*, 695–705. https://doi.org/10.1016/S1097-2765(00)80248-8.

52. Marahrens, Y., Panning, B., Dausman, J., Strauss, W., and Jaenisch, R. (1997). Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. *11*, 156–166. https://doi.org/10.1101/gad.11.2.156.

53. Popova, B.C., Tada, T., Takagi, N., Brockdorff, N., and Nesterova, T.B. (2006). Attenuated spread of X-inactivation in an X;autosome translocation. Proc. Natl. Acad. Sci. USA *103*, 7706–7711. https://doi.org/10.1073/pnas.0602021103.

54. Chow, J.C., Ciaudo, C., Fazzari, M.J., Mise, N., Servant, N., Glass, J.L., Attreed, M., Avner, P., Wutz, A., Barillot, E., et al. (2016). LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. Cell *166*, 782. https://doi.org/10.1016/j.cell.2016.07.013.

55. Hasenbein, T.P., Hoelzl, S., Smith, Z.D., Gerhardinger, C., Gonner, M.O.C., Aguilar-Pimentel, A., Amarie, O.V., Becker, L., Calzada-Wack, J., Dragano, N.R.V., et al. (2024). X-linked deletion of Crossfirre, Firre, and Dxz4 in vivo uncovers diverse phenotypes and combinatorial effects on autosomes. Nat. Commun. *15*, 10631. https://doi.org/10.1038/s41467-024-54673-5.

56. Bauer, M., Vidal, E., Zorita, E., Üresin, N., Pinter, S.F., Filion, G.J., and Payer, B. (2021). Chromosome compartments on the inactive X guide TAD formation independently of transcription during X-reactivation. Nat. Commun. *12*, 3499. https://doi.org/10.1038/s41467-021-23610-1.

57. Migeon, B.R., Winter, H., Kazi, E., Chowdhury, A.K., Hughes, A., Haisley-Royster, C., Morrison, H., and Jeppesen, P. (2001). Low-Copy-Number Human Transgene Is Recognized as an X Inactivation Center in Mouse ES Cells, but Fails to Induce cis-Inactivation in Chimeric Mice. Genomics *71*, 156–162. https://doi.org/10.1006/geno.2000.6421.

58. Loda, A., Brandsma, J.H., Vassilev, I., Servant, N., Loos, F., Amirnasr, A., Splinter, E., Barillot, E., Poot, R.A., Heard, E., et al. (2017). Genetic and epigenetic features direct differential efficiency of Xist-mediated silencing at X-chromosomal and autosomal locations. Nat. Commun. *8*, 690. https://doi.org/10.1038/s41467-017-00528-1.

59. Tang, Y.A., Huntley, D., Montana, G., Cerase, A., Nesterova, T.B., and Brockdorff, N. (2010). Efficiency of Xist-mediated silencing on autosomes is linked to chromosomal domain organisation. Epigenetics Chromatin *3*, 10. https://doi.org/10.1186/1756-8935-3-10.

60. Sleutels, F., Zwart, R., and Barlow, D.P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature *415*, 810–813. https://doi.org/10.1038/415810a.

61. Wutz, A., Theussl, H.C., Dausman, J., Jaenisch, R., Barlow, D.P., and Wagner, E.F. (2001). Non-imprinted Igf2r expression decreases growth and rescues the Tme mutation in mice. Development *128*, 1881–1887. https://doi.org/10.1242/dev.128.10.1881.

62. Mancini-Dinardo, D., Steele, S.J.S., Levorse, J.M., Ingram, R.S., and Tilghman, S.M. (2006). Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. Genes Dev. *20*, 1268–1282. https://doi.org/10.1101/gad.1416906.

63. Shin, J.-Y., Fitzpatrick, G.V., and Higgins, M.J. (2008). Two distinct mechanisms of silencing by the KvDMR1 imprinting control region. EMBO J. *27*, 168–178. https://doi.org/10.1038/sj.emboj.7601960.

64. Sun, S., Payer, B., Namekawa, S., An, J.Y., Press, W., Catalan-Dibene, J., Sunwoo, H., and Lee, J.T. (2015). Xist imprinting is promoted by the hemizygous (unpaired) state in the male germ line. Proc. Natl. Acad. Sci. USA *112*, 14415–14422. https://doi.org/10.1073/pnas.1519528112.

65. Norris, D.P., Patel, D., Kay, G.F., Penny, G.D., Brockdorff, N., Sheardown, S.A., and Rastan, S. (1994). Evidence that random and imprinted Xist expression is controlled by preemptive methylation. Cell *77*, 41–51. https://doi.org/10.1016/0092-8674(94)90233-X.

66. Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., and Wutz, A. (2004). A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. PLoS Biol. *2*, E171. https://doi.org/10.1371/journal.pbio.0020171.

67. McHugh, C.A., Chen, C.-K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., et al. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature *521*, 232–236. https://doi.org/10.1038/nature14443.

68. Mohammad, F., Mondal, T., Guseva, N., Pandey, G.K., and Kanduri, C. (2010). Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. Development *137*, 2493–2499. https://doi.org/10.1242/dev.048181.

69. Mohammad, F., Pandey, R.R., Nagano, T., Chakalova, L., Mondal, T., Fraser, P., and Kanduri, C. (2008). Kcnq1ot1/Lit1 noncoding RNA mediates transcriptional silencing by targeting to the perinucleolar region. Mol. Cell. Biol. *28*, 3713–3728. https://doi.org/10.1128/MCB.02263-07.

70. Schertzer, M.D., Braceros, K.C.A., Starmer, J., Cherney, R.E., Lee, D.M., Salazar, G., Justice, M., Bischoff, S.R., Cowley, D.O., Ariel, P., et al. (2019). LncRNA-induced spread of Polycomb controlled by genome architecture, RNA abundance, and CpG island DNA. Mol. Cell *75*, 523–537.e10. https://doi.org/10.1016/j.molcel.2019.05.028.

71. Raposo, A.C., Casanova, M., Gendrel, A.-V., and da Rocha, S.T. (2021). The tandem repeat modules of Xist lncRNA: a swiss army knife for the control of X-chromosome inactivation. Biochem. Soc. Trans. *49*, 2549–2560. https://doi.org/10.1042/BST20210253.

72. Lyon, M.F. (1998). X-Chromosome inactivation: a repeat hypothesis. Cytogenet. Cell Genet. *80*, 133–137. https://doi.org/10.1159/000014969.

73. Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol. Cell *32*, 232–246. https://doi.org/10.1016/j.molcel.2008.08.022.

74. Lafrenière, R.G., Brown, C.J., Rider, S., Chelly, J., Taillon-Miller, P., Chinault, A.C., Monaco, A.P., and Willard, H.F. (1993). 2.6 Mb YAC contig of

the human X inactivation center region in Xq13: physical linkage of the RPS4X, PHKA1, XIST and DXS128E genes. Hum. Mol. Genet. *2*, 1105–1115. https://doi.org/10.1093/hmg/2.8.1105.

75. Chen, L.-L. (2016). Linking long noncoding RNA localization and function. Trends Biochem. Sci. *41*, 761–772. https://doi.org/10.1016/j.tibs.2016.07.003.

76. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461. https://doi.org/10.1038/nature12787.

77. Hon, C.-C., Ramilowski, J.A., Harshbarger, J., Bertin, N., Rackham, O.J.L., Gough, J., Denisenko, E., Schmeier, S., Poulsen, T.M., Severin, J., et al. (2017). An atlas of human long non-coding RNAs with accurate 5′ ends. Nature *543*, 199–204. https://doi.org/10.1038/nature21374.

78. Meller, V.H., and Rattner, B.P. (2002). The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. EMBO J. *21*, 1084–1091. https://doi.org/10.1093/emboj/21.5.1084.

79. Allaman-Pillet, N., Djemaï, A., Bonny, C., and Schorderet, D.F. (2000). The 5′ repeat elements of the mouse Xist gene inhibit the transcription of X-linked genes. Gene Expr. *9*, 93–101. https://doi.org/10.3727/000000001783992632.

80. Gossen, M., Freundlieb, S., Bender, G., Müller, G., Hillen, W., and Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. Science *268*, 1766–1769. https://doi.org/10.1126/science.7792603.

81. Moindrot, B., Cerase, A., Coker, H., Masui, O., Grijzenhout, A., Pintacuda, G., Schermelleh, L., Nesterova, T.B., and Brockdorff, N. (2015). A Pooled shRNA Screen Identifies Rbm15, Spen, and Wtap as Factors Required for Xist RNA-Mediated Silencing. Cell Rep. *12*, 562–572. https://doi.org/10.1016/j.celrep.2015.06.053.

82. Liu, B., Wang, S., Brenner, M., Paton, J.F.R., and Kasparov, S. (2008). Enhancement of cell-specific transgene expression from a Tet-Off regulatory system using a transcriptional amplification strategy in the rat brain. J. Gene Med. *10*, 583–592. https://doi.org/10.1002/jgm.1178.

83. Meller, V.H., Gordadze, P.R., Park, Y., Chu, X., Stuckenholz, C., Kelley, R.L., and Kuroda, M.I. (2000). Ordered assembly of roX RNAs into MSL complexes on the dosage-compensated X chromosome in Drosophila. Curr. Biol. *10*, 136–143. https://doi.org/10.1016/S0960-9822(00)00311-0.

84. Kelley, R.L., Meller, V.H., Gordadze, P.R., Roman, G., Davis, R.L., and Kuroda, M.I. (1999). Epigenetic spreading of the Drosophila dosage compensation complex from roX RNA genes into flanking chromatin. Cell *98*, 513–522. https://doi.org/10.1016/S0092-8674(00)81979-0.

85. Henry, R.A., Tews, B., Li, X., and Scott, M.J. (2001). Recruitment of the Male-specific Lethal (MSL) Dosage Compensation Complex to an Autosomally Integrated roXChromatin Entry Site Correlates with an Increased Expression of an Adjacent Reporter Gene in Male Drosophila. J. Biol. Chem. *276*, 31953–31958. https://doi.org/10.1074/jbc.M103008200.

86. Fukushige, S., and Sauer, B. (1992). Genomic targeting with a positive-selection lox integration vector allows highly reproducible gene expression in mammalian cells. Proc. Natl. Acad. Sci. USA *89*, 7905–7909. https://doi.org/10.1073/pnas.89.17.7905.

87. Benink, H.A., and Urh, M. (2015). HaloTag technology for specific and covalent labeling of fusion proteins. Methods Mol. Biol. *1266*, 119–128. https://doi.org/10.1007/978-1-4939-2272-7_8.

88. Werner, M.S., and Ruthenburg, A.J. (2015). Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes. Cell Rep. *12*, 1089–1098. https://doi.org/10.1016/j.celrep.2015.07.033.

89. Rezaei, O., Tamizkar, K.H., Sharifi, G., Taheri, M., and Ghafouri-Fard, S. (2020). Emerging Role of Long Non-Coding RNAs in the Pathobiology of Glioblastoma. Front. Oncol. *10*, 625884. https://doi.org/10.3389/fonc.2020.625884.

90. Cech, T.R., and Steitz, J.A. (2014). The noncoding RNA revolution-trashing old rules to forge new ones. Cell *157*, 77–94. https://doi.org/10.1016/j.cell.2014.03.008.

91. Perez-Pinera, P., Jones, M.F., Lal, A., and Lu, T.K. (2015). Putting Noncoding RNA on Display with CRISPR. Mol. Cell *59*, 146–148. https://doi.org/10.1016/j.molcel.2015.07.002.

92. Thiel, D., Conrad, N.D., Ntini, E., Peschutter, R.X., Siebert, H., and Marsico, A. (2019). Identifying lncRNA-mediated regulatory modules via ChIA-PET network analysis. BMC Bioinform. *20*, 292. https://doi.org/10.1186/s12859-019-2900-8.

93. Berletch, J.B., Yang, F., Xu, J., Carrel, L., and Disteche, C.M. (2011). Genes that escape from X inactivation. Hum. Genet. *130*, 237–245. https://doi.org/10.1007/s00439-011-1011-z.

94. Peeters, S., Leung, T., Fornes, O., Farkas, R.A., Wasserman, W.W., and Brown, C.J. (2023). Refining the genomic determinants underlying escape from X-chromosome inactivation. NAR Genom. Bioinform. *5*, lqad052. https://doi.org/10.1093/nargab/lqad052.

95. Berletch, J.B., Ma, W., Yang, F., Shendure, J., Noble, W.S., Disteche, C.M., and Deng, X. (2015). Escape from X inactivation varies in mouse tissues. PLoS Genet. *11*, e1005079. https://doi.org/10.1371/journal.pgen.1005079.

96. Hoelzl, S., Hasenbein, T.P., Engelhardt, S., and Andergassen, D. (2025). Aging promotes reactivation of the Barr body at distal chromosome regions. Nat. Aging *5*, 984–996. https://doi.org/10.1038/s43587-025-00856-8.

97. Grabow, W., and Jaeger, L. (2013). RNA modularity for synthetic biology. F1000Prime Rep. *5*, 46. https://doi.org/10.12703/P5-46.

98. Sanbonmatsu, K. (2022). Getting to the bottom of lncRNA mechanism: structure-function relationships. Mamm. Genome *33*, 343–353. https://doi.org/10.1007/s00335-021-09924-x.

99. Olazagoitia-Garmendia, A., Senovilla-Ganzo, R., García-Moreno, F., and Castellanos-Rubio, A. (2023). Functional evolutionary convergence of long noncoding RNAs involved in embryonic development. Commun. Biol. *6*, 908. https://doi.org/10.1038/s42003-023-05278-z.

100. Jenike, K.M., Campos-Domínguez, L., Boddé, M., Cerca, J., Hodson, C.N., Schatz, M.C., and Jaron, K.S. (2025). k-mer approaches for biodiversity genomics. Genome Res. *35*, 219–230. https://doi.org/10.1101/gr.279452.124.

101. Kirk, J.M., Kim, S.O., Inoue, K., Smola, M.J., Lee, D.M., Schertzer, M.D., Wooten, J.S., Baker, A.R., Sprague, D., Collins, D.W., et al. (2018). Functional classification of long non-coding RNAs by k-mer content. Nat. Genet. *50*, 1474–1482. https://doi.org/10.1038/s41588-018-0207-8.

102. Guo, C.-J., Xu, G., and Chen, L.-L. (2020). Mechanisms of Long Noncoding RNA Nuclear Retention. Trends Biochem. Sci. *45*, 947–960. https://doi.org/10.1016/j.tibs.2020.07.001.

103. Miyagawa, R., Tano, K., Mizuno, R., Nakamura, Y., Ijiri, K., Rakwal, R., Shibato, J., Masuo, Y., Mayeda, A., Hirose, T., et al. (2012). Identification of cis- and trans-acting factors involved in the localization of MALAT-1 noncoding RNA to nuclear speckles. RNA *18*, 738–751. https://doi.org/10.1261/rna.028639.111.

104. Hacisuleyman, E., Shukla, C.J., Weiner, C.L., and Rinn, J.L. (2016). Function and evolution of local repeats in the Firre locus. Nat. Commun. *7*, 11021. https://doi.org/10.1038/ncomms11021.

105. Zhang, B., Gunawardane, L., Niazi, F., Jahanbani, F., Chen, X., and Valadkhan, S. (2014). A novel RNA motif mediates the strict nuclear localization of a long noncoding RNA. Mol. Cell. Biol. *34*, 2318–2329. https://doi.org/10.1128/MCB.01673-13.

106. Lubelsky, Y., and Ulitsky, I. (2018). Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. Nature *555*, 107–111. https://doi.org/10.1038/nature25757.

107. Lubelsky, Y., Zuckerman, B., and Ulitsky, I. (2021). High-resolution mapping of function and protein binding in an RNA nuclear enrichment sequence. EMBO J. *40*, e106357. https://doi.org/10.15252/embj.2020106357.

108. Carlevaro-Fita, J., Polidori, T., Das, M., Navarro, C., Zoller, T.I., and Johnson, R. (2019). Ancient exapted transposable elements promote nuclear enrichment of human long noncoding RNAs. Genome Res. *29*, 208–222. https://doi.org/10.1101/gr.229922.117.

109. Leisegang, M.S., Warwick, T., Stötzel, J., and Brandes, R.P. (2024). RNA-DNA triplexes: molecular mechanisms and functional relevance. Trends Biochem. Sci. *49*, 532–544. https://doi.org/10.1016/j.tibs.2024.03.009.

110. Wu, S.K., Roberts, J.T., Balas, M.M., and Johnson, A.M. (2020). RNA matchmaking in chromatin regulation. Biochem. Soc. Trans. *48*, 2467–2481. https://doi.org/10.1042/BST20191225.

111. Warwick, T., Brandes, R.P., and Leisegang, M.S. (2023). Computational Methods to Study DNA:DNA:RNA Triplex Formation by lncRNAs. Non-coding. RNA *9*, 10. https://doi.org/10.3390/ncrna9010010.

112. Bou-Nader, C., Bothra, A., Garboczi, D.N., Leppla, S.H., and Zhang, J. (2022). Structural basis of R-loop recognition by the S9.6 monoclonal antibody. Nat. Commun. *13*, 1641. https://doi.org/10.1038/s41467-022-29187-7.

113. Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., Macura, K., Bläss, G., Kellis, M., Werber, M., et al. (2013). The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. Dev. Cell *24*, 206–214. https://doi.org/10.1016/j.devcel.2012.12.012.

114. Ali, T., Rogala, S., Krause, N.M., Bains, J.K., Melissari, M.-T., Währisch, S., Schwalbe, H., Herrmann, B.G., and Grote, P. (2023). Fendrr synergizes with Wnt signalling to regulate fibrosis related genes during lung development via its RNA:dsDNA triplex element. Nucleic Acids Res. *51*, 6227–6237. https://doi.org/10.1093/nar/gkad395.

115. Trembinski, D.J., Bink, D.I., Theodorou, K., Sommer, J., Fischer, A., van Bergen, A., Kuo, C.-C., Costa, I.G., Schürmann, C., Leisegang, M.S., et al. (2020). Aging-regulated anti-apoptotic long non-coding RNA Sarrah augments recovery from acute myocardial infarction. Nat. Commun. *11*, 2039. https://doi.org/10.1038/s41467-020-15995-2.

116. Zhang, X., Ding, T., Yang, F., Zhang, J., Xu, H., Bai, Y., Shi, Y., Yang, J., Chen, C., Zhu, C., et al. (2024). Peptidylprolyl isomerase A guides SENP5/GAU1 DNA-lncRNA triplex generation for driving tumorigenesis. Nat. Commun. *15*, 9068. https://doi.org/10.1038/s41467-024-53493-x.

117. Mondal, T., Subhash, S., Vaid, R., Enroth, S., Uday, S., Reinius, B., Mitra, S., Mohammed, A., James, A.R., Hoberg, E., et al. (2015). MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA-DNA triplex structures. Nat. Commun. *6*, 7743. https://doi.org/10.1038/ncomms8743.

118. Kalwa, M., Hänzelmann, S., Otto, S., Kuo, C.-C., Franzen, J., Joussen, S., Fernandez-Rebollo, E., Rath, B., Koch, C., Hofmann, A., et al. (2016). The lncRNA HOTAIR impacts on mesenchymal stem cells via triple helix formation. Nucleic Acids Res. *44*, 10631–10643. https://doi.org/10.1093/nar/gkw802.

119. Leisegang, M.S., Bains, J.K., Seredinski, S., Oo, J.A., Krause, N.M., Kuo, C.-C., Günther, S., Sentürk Cetin, N., Warwick, T., Cao, C., et al. (2022). HIF1α-AS1 is a DNA:DNA:RNA triplex-forming lncRNA interacting with the HUSH complex. Nat. Commun. *13*, 6563. https://doi.org/10.1038/s41467-022-34252-2.

120. Blank-Giwojna, A., Postepska-Igielska, A., and Grummt, I. (2019). lncRNA KHPS1 Activates a Poised Enhancer by Triplex-Dependent Recruitment of Epigenomic Regulators. Cell Rep. *26*, 2904–2915.e4. https://doi.org/10.1016/j.celrep.2019.02.059.

121. Pfeiffer, V., Crittin, J., Grolimund, L., and Lingner, J. (2013). The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening. EMBO J. *32*, 2861–2871. https://doi.org/10.1038/emboj.2013.217.

122. Luo, H., Zhu, G., Eshelman, M.A., Fung, T.K., Lai, Q., Wang, F., Zeisig, B.B., Lesperance, J., Ma, X., Chen, S., et al. (2022). HOTTIP-dependent R-loop formation regulates CTCF boundary activity and TAD integrity in leukemia. Mol. Cell *82*, 833–851.e11. https://doi.org/10.1016/j.molcel.2022.01.014.

123. Hegazy, Y.A., Fernando, C.M., and Tran, E.J. (2020). The balancing act of R-loop biology: The good, the bad, and the ugly. J. Biol. Chem. *295*, 905–913. https://doi.org/10.1074/jbc.REV119.011353.

124. Boque-Sastre, R., Soler, M., Oliveira-Mateos, C., Portela, A., Moutinho, C., Sayols, S., Villanueva, A., Esteller, M., and Guil, S. (2015). Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. Proc. Natl. Acad. Sci. USA *112*, 5785–5790. https://doi.org/10.1073/pnas.1421197112.

125. Balas, M.M., Hartwick, E.W., Barrington, C., Roberts, J.T., Wu, S.K., Bettcher, R., Griffin, A.M., Kieft, J.S., and Johnson, A.M. (2021). Establishing RNA-RNA interactions remodels lncRNA structure and promotes PRC2 activity. Sci. Adv. *7*, eabc9191. https://doi.org/10.1126/sciadv.abc9191.

126. Dunn-Davies, H., Dudnakova, T., Nogara, A., Rodor, J., Thomas, A.C., Parish, E., Gautier, P., Meynert, A., Ulitsky, I., Madeddu, P., et al. (2024). Control of endothelial cell function and arteriogenesis by MEG3:EZH2 epigenetic regulation of integrin expression. Mol. Ther. Nucl. Acids *35*, 102173. https://doi.org/10.1016/j.omtn.2024.102173.

127. Tinoco, I., Jr., and Bustamante, C. (1999). How RNA folds. J. Mol. Biol. *293*, 271–281. https://doi.org/10.1006/jmbi.1999.3001.

128. Chillón, I., and Pyle, A.M. (2016). Inverted repeat Alu elements in the human lincRNA-p21 adopt a conserved secondary structure that regulates RNA function. Nucleic Acids Res. *44*, 9462–9471. https://doi.org/10.1093/nar/gkw599.

129. Jones, A.N., Pisignano, G., Pavelitz, T., White, J., Kinisu, M., Forino, N., Albin, D., and Varani, G. (2020). An evolutionarily conserved RNA structure in the functional core of the lincRNA Cyrano. RNA *26*, 1234–1246. https://doi.org/10.1261/rna.076117.120.

130. Hawkes, E.J., Hennelly, S.P., Novikova, I.V., Irwin, J.A., Dean, C., and Sanbonmatsu, K.Y. (2016). COOLAIR Antisense RNAs Form Evolutionarily Conserved Elaborate Secondary Structures. Cell Rep. *16*, 3087–3096. https://doi.org/10.1016/j.celrep.2016.08.045.

131. Brown, J.A., Valenstein, M.L., Yario, T.A., Tycowski, K.T., and Steitz, J.A. (2012). Formation of triple-helical structures by the 3′-end sequences of MALAT1 and MENβ noncoding RNAs. Proc. Natl. Acad. Sci. USA *109*, 19202–19207. https://doi.org/10.1073/pnas.1217338109.

132. Ross, C.J., Rom, A., Spinrad, A., Gelbard-Solodkin, D., Degani, N., and Ulitsky, I. (2021). Uncovering deeply conserved motif combinations in rapidly evolving noncoding sequences. Genome Biol. *22*, 29. https://doi.org/10.1186/s13059-020-02247-1.

133. Jones, A.N., and Sattler, M. (2019). Challenges and perspectives for structural biology of lncRNAs-the example of the Xist lncRNA A-repeats. J. Mol. Cell Biol. *11*, 845–859. https://doi.org/10.1093/jmcb/mjz086.

134. Jones, A.N., Tikhaia, E., Mourão, A., and Sattler, M. (2022). Structural effects of m6A modification of the Xist A-repeat AUCG tetraloop and its recognition by YTHDC1. Nucleic Acids Res. *50*, 2350–2362. https://doi.org/10.1093/nar/gkac080.

135. Duszczyk, M.M., Zanier, K., and Sattler, M. (2008). A NMR strategy to unambiguously distinguish nucleic acid hairpin and duplex conformations applied to a Xist RNA A-repeat. Nucleic Acids Res. *36*, 7068–7077. https://doi.org/10.1093/nar/gkn776.

136. Mou, X., Liew, S.W., and Kwok, C.K. (2022). Identification and targeting of G-quadruplex structures in MALAT1 long non-coding RNA. Nucleic Acids Res. *50*, 397–410. https://doi.org/10.1093/nar/gkab1208.

137. Simko, E.A.J., Liu, H., Zhang, T., Velasquez, A., Teli, S., Haeusler, A.R., and Wang, J. (2020). G-quadruplexes offer a conserved structural motif for NONO recruitment to NEAT1 architectural lncRNA. Nucleic Acids Res. *48*, 7421–7438. https://doi.org/10.1093/nar/gkaa475.

138. Lee, Y.W., Weissbein, U., Blum, R., and Lee, J.T. (2024). G-quadruplex folding in Xist RNA antagonizes PRC2 activity for stepwise regulation of X chromosome inactivation. Mol. Cell *84*, 1870–1885.e9. https://doi.org/10.1016/j.molcel.2024.04.015.

139. Wang, X., Goodrich, K.J., Gooding, A.R., Naeem, H., Archer, S., Paucek, R.D., Youmans, D.T., Cech, T.R., and Davidovich, C. (2017). Targeting of Polycomb repressive complex 2 to RNA by short repeats of consecutive guanines. Mol. Cell *65*, 1056–1067.e5. https://doi.org/10.1016/j.molcel.2017.02.003.

140. Fanucchi, S., Fok, E.T., Dalla, E., Shibayama, Y., Börner, K., Chang, E.Y., Stoychev, S., Imakaev, M., Grimm, D., Wang, K.C., et al. (2019). Publisher Correction: Immune genes are primed for robust transcription by proximal long noncoding RNAs located in nuclear compartments. Nat. Genet. *51*, 364. https://doi.org/10.1038/s41588-018-0341-3.

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

142. Smola, M.J., and Weeks, K.M. (2018). In-cell RNA structure probing with SHAPE-MaP. Nat. Protoc. *13*, 1181–1195. https://doi.org/10.1038/nprot.2018.010.

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

144. Somarowthu, S., Legiewicz, M., Chillón, I., Marcia, M., Liu, F., and Pyle, A.M. (2015). HOTAIR forms an intricate and modular secondary structure. Mol. Cell *58*, 353–361. https://doi.org/10.1016/j.molcel.2015.03.006.

145. Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic discovery of Xist RNA binding proteins. Cell *161*, 404–416. https://doi.org/10.1016/j.cell.2015.03.025.

146. Hu, M., Wang, Y., Meng, Y., Hu, J., Qiao, J., Zhen, J., Liang, D., and Fan, M. (2022). Hypoxia induced-disruption of lncRNA TUG1/PRC2 interaction impairs human trophoblast invasion through epigenetically activating Nodal/ALK7 signalling. J. Cell. Mol. Med. *26*, 4087–4100. https://doi.org/10.1111/jcmm.17450.

147. Jin, L., Cai, Q., Wang, S., Wang, S., Mondal, T., Wang, J., and Quan, Z. (2018). Long noncoding RNA MEG3 regulates LATS2 by promoting the ubiquitination of EZH2 and inhibits proliferation and invasion in gallbladder cancer. Cell Death Dis. *9*, 1017. https://doi.org/10.1038/s41419-018-1064-1.

148. Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. Nat. Struct. Mol. Biol. *20*, 1250–1257. https://doi.org/10.1038/nsmb.2679.

149. Portoso, M., Ragazzini, R., Brenčič, Ž., Moiani, A., Michaud, A., Vassilev, I., Wassef, M., Servant, N., Sargueil, B., and Margueron, R. (2017). PRC2 is dispensable for HOTAIR-mediated transcriptional repression. EMBO J. *36*, 981–994. https://doi.org/10.15252/embj.201695335.

150. Hall Hickman, A., and Jenner, R.G. (2024). Apparent RNA bridging between PRC2 and chromatin is an artifact of non-specific chromatin precipitation upon RNA degradation. Cell Rep. *43*, 113856. https://doi.org/10.1016/j.celrep.2024.113856.

151. Guo, J.K., Blanco, M.R., Walkup, W.G., 4th, Bonesteele, G., Urbinati, C.R., Banerjee, A.K., Chow, A., Ettlin, O., Strehle, M., Peyda, P., et al. (2024). Denaturing purifications demonstrate that PRC2 and other widely reported chromatin proteins do not appear to bind directly to RNA in vivo. Mol. Cell *84*, 1271–1289.e12. https://doi.org/10.1016/j.molcel.2024.01.026.

152. Scarola, M., Comisso, E., Rosso, M., Del Sal, G., Schneider, C., Schoeftner, S., and Benetti, R. (2020). FUS-dependent loading of SUV39H1 to OCT4 pseudogene-lncRNA programs a silencing complex with OCT4 promoter specificity. Commun. Biol. *3*, 632. https://doi.org/10.1038/s42003-020-01355-9.

153. Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature *472*, 120–124. https://doi.org/10.1038/nature09819.

154. Yang, Y.W., Flynn, R.A., Chen, Y., Qu, K., Wan, B., Wang, K.C., Lei, M., and Chang, H.Y. (2014). Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. eLife *3*, e02046. https://doi.org/10.7554/eLife.02046.

155. Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A., et al. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol. Cell *39*, 925–938. https://doi.org/10.1016/j.molcel.2010.08.011.

156. Miao, H., Wu, F., Li, Y., Qin, C., Zhao, Y., Xie, M., Dai, H., Yao, H., Cai, H., Wang, Q., et al. (2022). MALAT1 modulates alternative splicing by cooperating with the splicing factors PTBP1 and PSF. Sci. Adv. *8*, eabq7289. https://doi.org/10.1126/sciadv.abq7289.

157. Yamazaki, T., Souquere, S., Chujo, T., Kobelke, S., Chong, Y.S., Fox, A.H., Bond, C.S., Nakagawa, S., Pierron, G., and Hirose, T. (2018). Functional domains of NEAT1 architectural lncRNA induce paraspeckle assembly through phase separation. Mol. Cell *70*, 1038–1053.e7. https://doi.org/10.1016/j.molcel.2018.05.019.

158. Sauvageau, M. (2019). Diverging RNPs: Toward understanding lncRNA-protein interactions and functions. Adv. Exp. Med. Biol. *1203*, 285–312. https://doi.org/10.1007/978-3-030-31434-7_10.

159. Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A., and Panning, B. (2002). Xist RNA and the mechanism of X chromosome inactivation. Annu. Rev. Genet. *36*, 233–278. https://doi.org/10.1146/annurev.genet.36.042902.092433.

160. Lee, Y.H., Hass, E.P., Campodonico, W., Lee, Y.K., Lasda, E., Shah, J.S., Rinn, J.L., and Hwang, T. (2024). Massively parallel dissection of RNA in RNA-protein interactions in vivo. Nucleic Acids Res. *52*, e48. https://doi.org/10.1093/nar/gkae334.

161. Hacisuleyman, E., Goff, L.A., Trapnell, C., Williams, A., Henao-Mejia, J., Sun, L., McClanahan, P., Hendrickson, D.G., Sauvageau, M., Kelley, D.R., et al. (2014). Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. Nat. Struct. Mol. Biol. *21*, 198–206. https://doi.org/10.1038/nsmb.2764.

162. Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M., et al. (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell *142*, 409–419. https://doi.org/10.1016/j.cell.2010.06.040.

163. Pintacuda, G., Wei, G., Roustan, C., Kirmizitas, B.A., Solcan, N., Cerase, A., Castello, A., Mohammed, S., Moindrot, B., Nesterova, T.B., et al. (2017). HnRNPK recruits PCGF3/5-PRC1 to the Xist RNA B-repeat to establish Polycomb-mediated chromosomal silencing. Mol. Cell *68*, 955–969.e10. https://doi.org/10.1016/j.molcel.2017.11.013.

164. Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol. Cell *40*, 939–953. https://doi.org/10.1016/j.molcel.2010.12.011.

165. Bousard, A., Raposo, A.C., Żylicz, J.J., Picard, C., Pires, V.B., Qi, Y., Gil, C., Syx, L., Chang, H.Y., Heard, E., et al. (2019). The role of Xist-mediated Polycomb recruitment in the initiation of X-chromosome inactivation. EMBO Rep. *20*, e48019. https://doi.org/10.15252/embr.201948019.

166. Wolin, E., Guo, J.K., Blanco, M.R., Goronzy, I.N., Gorhe, D., Dong, W., Perez, A.A., Keskin, A., Valenzuela, E., Abdou, A.A., et al. (2025). SPIDR enables multiplexed mapping of RNA-protein interactions and uncovers a mechanism for selective translational suppression upon cell stress. Cell *188*, 5384–5402.e25. https://doi.org/10.1016/j.cell.2025.06.042.

167. Tsue, A.F., Kania, E.E., Lei, D.Q., Fields, R., McGann, C.D., Marciniak, D.M., Hershberg, E.A., Deng, X., Kihiu, M., Ong, S.-E., et al. (2024). Multiomic characterization of RNA microenvironments by oligonucleotide-mediated proximity-interactome mapping. Nat. Methods *21*, 2058–2071. https://doi.org/10.1038/s41592-024-02457-6.

168. Bellucci, M., Agostini, F., Masin, M., and Tartaglia, G.G. (2011). Predicting protein associations with long noncoding RNAs. Nat. Methods *8*, 444–445. https://doi.org/10.1038/nmeth.1611.

169. Fiorentino, J., Armaos, A., Colantoni, A., and Tartaglia, G.G. (2024). Prediction of protein-RNA interactions from single-cell transcriptomic data. Nucleic Acids Res. *52*, e31. https://doi.org/10.1093/nar/gkae076.

170. Zhang, S.-W., Wang, Y., Zhang, X.-X., and Wang, J.-Q. (2019). Prediction of the RBP binding sites on lncRNAs using the high-order nucleotide encoding convolutional neural network. Anal. Biochem. *583*, 113364. https://doi.org/10.1016/j.ab.2019.113364.

171. Ji, Y., Zhou, Z., Liu, H., and Davuluri, R.V. (2021). DNABERT: pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome. Bioinformatics *37*, 2112–2120. https://doi.org/10.1093/bioinformatics/btab083.

172. Yamada, K., and Hamada, M. (2022). Prediction of RNA-protein interactions using a nucleotide language model. Bioinform. Adv. *2*, vbac023. https://doi.org/10.1093/bioadv/vbac023.

173. Kratz, M.B., and Smita, K.N. (2024). Predicting conserved functional interactions for long noncoding RNAs via deep learning. Front. RNA Res. *2*.

174. Huang, W., Xiong, T., Zhao, Y., Heng, J., Han, G., Wang, P., Zhao, Z., Shi, M., Li, J., Wang, J., et al. (2024). Computational prediction and experimental validation identify functionally conserved lncRNAs from zebrafish to human. Nat. Genet. *56*, 124–135. https://doi.org/10.1038/s41588-023-01620-7.

175. Ron, M., and Ulitsky, I. (2022). Context-specific effects of sequence elements on subcellular localization of linear and circular RNAs. Nat. Commun. *13*, 2481. https://doi.org/10.1038/s41467-022-30183-0.

176. Zuckerman, B., Ron, M., Mikl, M., Segal, E., and Ulitsky, I. (2020). Gene architecture and sequence composition underpin selective dependency of nuclear export of long RNAs on NXF1 and the TREX complex. Mol. Cell *79*, 251–267.e6. https://doi.org/10.1016/j.molcel.2020.05.013.

177. Barakat, T.S., Loos, F., van Staveren, S., Myronova, E., Ghazvini, M., Grootegoed, J.A., and Gribnau, J. (2014). The trans-activator RNF12 and cis-acting elements effectuate X chromosome inactivation independent of X-pairing. Mol. Cell *53*, 965–978. https://doi.org/10.1016/j.molcel.2014.02.006.

178. Schulz, E.G., Nora, E.P., and Heard, E. (2011). Rnf12–a jack of all trades in X inactivation? PLoS Genet. *7*, e1002002. https://doi.org/10.1371/journal.pgen.1002002.

179. Mutzel, V., Okamoto, I., Dunkel, I., Saitou, M., Giorgetti, L., Heard, E., and Schulz, E.G. (2019). A symmetric toggle switch explains the onset of random X inactivation in different mammals. Nat. Struct. Mol. Biol. *26*, 350–360. https://doi.org/10.1038/s41594-019-0214-1.

180. Nordick, B., Yu, P.Y., Liao, G., and Hong, T. (2022). Nonmodular oscillator and switch based on RNA decay drive regeneration of multimodal gene expression. Nucleic Acids Res. *50*, 3693–3708. https://doi.org/10.1093/nar/gkac217.

181. Gupta, K., Czerminski, J.T., and Lawrence, J.B. (2024). Trisomy silencing by XIST: translational prospects and challenges. Hum. Genet. *143*, 843–855. https://doi.org/10.1007/s00439-024-02651-8.

182. Czermiński, J.T., and Lawrence, J.B. (2020). Silencing trisomy 21 with XIST in neural stem cells promotes neuronal differentiation. Dev. Cell *52*, 294–308.e3. https://doi.org/10.1016/j.devcel.2019.12.015.

183. Moon, J.E., and Lawrence, J.B. (2022). Chromosome silencing in vitro reveals trisomy 21 causes cell-autonomous deficits in angiogenesis and early dysregulation in Notch signaling. Cell Rep. *40*, 111174. https://doi.org/10.1016/j.celrep.2022.111174.

184. Davodabadi, F., Farasati Far, B., Sargazi, S., Fatemeh Sajjadi, S., Fathi-Karkan, S., Mirinejad, S., Ghotekar, S., Sargazi, S., and Rahman, M.M. (2024). Nanomaterials-based targeting of long non-coding RNAs in cancer: A cutting-edge review of current trends. ChemMedChem *19*, e202300528. https://doi.org/10.1002/cmdc.202300528.

185. Tamblin-Hopper, P., Kiss-Toth, E., Sudbery, I., Young, D., and Wilkinson, J.M. (2024). The potential therapeutic applications of long non-coding RNAs. J. Transl. Genet. Genom. *8*, 225–243. https://doi.org/10.20517/jtgg.2024.12.

186. Giovenino, C., Trajkova, S., Pavinato, L., Cardaropoli, S., Pullano, V., Ferrero, E., Sukarova-Angelovska, E., Carestiato, S., Salmin, P., Rinninella, A., et al. (2023). Skewed X-chromosome inactivation in unsolved neurodevelopmental disease cases can guide re-evaluation For X-linked genes. Eur. J. Hum. Genet. *31*, 1228–1236. https://doi.org/10.1038/s41431-023-01324-w.

187. Carrette, L.L.G., Blum, R., Ma, W., Kelleher, R.J., 3rd, and Lee, J.T. (2018). Tsix-Mecp2 female mouse model for Rett syndrome reveals that low-level MECP2 expression extends life and improves neuromotor function. Proc. Natl. Acad. Sci. USA *115*, 8185–8190. https://doi.org/10.1073/pnas.1800931115.

188. Carrette, L.L.G., Wang, C.-Y., Wei, C., Press, W., Ma, W., Kelleher, R.J., 3rd, and Lee, J.T. (2018). A mixed modality approach towards Xi reactivation for Rett syndrome and other X-linked disorders. Proc. Natl. Acad. Sci. USA *115*, E668–E675. https://doi.org/10.1073/pnas.1715124115.

189. Migeon, B.R. (2020). X-linked diseases: susceptible females. Genet. Med. *22*, 1156–1174. https://doi.org/10.1038/s41436-020-0779-4.

190. Grimm, N.-B., and Lee, J.T. (2022). Selective Xi reactivation and alternative methods to restore MECP2 function in Rett syndrome. Trends Genet. *38*, 920–943. https://doi.org/10.1016/j.tig.2022.01.007.

191. Adrianse, R.L., Smith, K., Gatbonton-Schwager, T., Sripathy, S.P., Lao, U., Foss, E.J., Boers, R.G., Boers, J.B., Gribnau, J., and Bedalov, A. (2018). Perturbed maintenance of transcriptional repression on the inactive X-chromosome in the mouse brain after Xist deletion. Epigenetics Chromatin *11*, 50. https://doi.org/10.1186/s13072-018-0219-8.

192. El Said, N.H., Della Valle, F., Liu, P., Paytuví-Gallart, A., Adroub, S., Gimenez, J., and Orlando, V. (2021). Malat-1-PRC2-EZH1 interaction supports adaptive oxidative stress dependent epigenome remodeling in skeletal myotubes. Cell Death Dis. *12*, 850. https://doi.org/10.1038/s41419-021-04082-z.

193. Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.-C., Hung, T., Argani, P., Rinn, J.L., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature *464*, 1071–1076. https://doi.org/10.1038/nature08975.

194. Viereck, J., Bührke, A., Foinquinos, A., Chatterjee, S., Kleeberger, J.A., Xiao, K., Janssen-Peters, H., Batkai, S., Ramanujam, D., Kraft, T., et al. (2020). Targeting muscle-enriched long non-coding RNA H19 reverses pathological cardiac hypertrophy. Eur. Heart J. *41*, 3462–3474. https://doi.org/10.1093/eurheartj/ehaa519.

195. Butler, A.A., Johnston, D.R., Kaur, S., and Lubin, F.D. (2019). Long noncoding RNA NEAT1 mediates neuronal histone methylation and age-related memory impairment. Sci. Signal. *12*, eaaw9277. https://doi.org/10.1126/scisignal.aaw9277.

196. Kukharsky, M.S., Ninkina, N.N., An, H., Telezhkin, V., Wei, W., Meritens, C.R. de, Cooper-Knock, J., Nakagawa, S., Hirose, T., Buchman, V.L., et al. (2020). Long non-coding RNA Neat1 regulates adaptive behavioural response to stress in mice. Transl. Psychiatry *10*, 171. https://doi.org/10.1038/s41398-020-0854-2.

197. Shihabudeen Haider Ali, M.S., Cheng, X., Moran, M., Haemmig, S., Naldrett, M.J., Alvarez, S., Feinberg, M.W., and Sun, X. (2019). LncRNA Meg3 protects endothelial function by regulating the DNA damage response. Nucleic Acids Res. *47*, 1505–1522. https://doi.org/10.1093/nar/gky1190.

198. Zhao, Z., Sentürk, N., Song, C., and Grummt, I. (2018). lncRNA PAPAS tethered to the rDNA enhancer recruits hypophosphorylated CHD4/NuRD to repress rRNA synthesis at elevated temperatures. Genes Dev. *32*, 836–848. https://doi.org/10.1101/gad.311688.118.