ANNUAL REVIEWS

Annual Review of Biochemistry

Long Noncoding RNAs: Molecular Modalities to Organismal Functions

John L. Rinn[1] and Howard Y. Chang[2,3]

[1]BioFrontiers Institute, Department of Biochemistry, University of Colorado, Boulder, Colorado 80303, USA; email: john.rinn@colorado.edu

[2]Center for Personal Dynamic Regulomes, Stanford University, Stanford, California 94305, USA

[3]Howard Hughes Medical Institute, Stanford University, Stanford, California 94305, USA

Annu. Rev. Biochem. 2020. 89:283–308

The *Annual Review of Biochemistry* is online at biochem.annualreviews.org

https://doi.org/10.1146/annurev-biochem-062917-012708

Keywords

long noncoding RNA, lncRNA, RNA–protein interactions, nuclear localization, RNA structure, massively parallel RNA reporter gene assays, MPRNA, RNA in vivo, RNA grammar

Abstract

We have known for decades that long noncoding RNAs (lncRNAs) can play essential functions across most forms of life. The maintenance of chromosome length requires an lncRNA (e.g., hTERC) and two lncRNAs in the ribosome that are required for protein synthesis. Thus, lncRNAs can represent powerful RNA machines. More recently, it has become clear that mammalian genomes encode thousands more lncRNAs. Thus, we raise the question: Which, if any, of these lncRNAs could also represent RNA-based machines? Here we synthesize studies that are beginning to address this question by investigating fundamental properties of lncRNA genes, revealing new insights into the RNA structure–function relationship, determining *cis*- and *trans*-acting lncRNAs in vivo, and generating new developments in high-throughput screening used to identify functional lncRNAs. Overall, these findings provide a context toward understanding the molecular grammar underlying lncRNA biology.

Contents

1. INTRODUCTION

One of the biggest surprises in the postgenomic era has been the discovery of a vast new landscape of regulatory elements in the human genome (1). This landscape includes hundreds of thousands of newfound DNA enhancer elements, hundreds of small peptides, and tens of thousands of long noncoding RNAs (lncRNAs) in excess of the number of protein-coding messenger RNAs (mRNAs)—thus, widespread transcription in noncoding regions leads to an explosion of lncRNA studies from laboratories around the world. As a result, several lines of evidence have pointed to the importance of some lncRNAs in human health and disease.

First, many lncRNA genes exhibit exquisite cell type–specific expression patterns in normal and disease states, even more so than protein-coding genes (2–10). Second, dozens of studies have demonstrated that gain and loss of function of select lncRNAs can influence cellular processes, development, and diseases (11–32). Third, well-established examples of lncRNAs that play important roles in gene regulation (e.g., *Xist*) have been found (15, 33–35). Moreover, the majority of haplotype blocks associated with human disease do not contain protein-coding genes, but most do encode lncRNAs (2, 36–39). In fact, mutations in several lncRNA loci are Mendelian—inherited in human disease (2, 37, 38). Collectively, these studies point to lncRNAs as a new horizon in our understanding of the regulatory logic encoded in the mammalian genome (28).

However, owing to the nascent field and rapid growth, our current knowledge base is fragmented and piecemeal. Thus, the challenge now is to have a holistic and focused understanding of lncRNAs from mouse models all the way to their RNA structure–function relationships. The goal of this review is to synthesize the emerging understanding of lncRNA properties, their structure–function relationships, and in turn the *cis*- and *trans*-acting mechanisms of lncRNAs in vivo.

Overall, we aim to have these studies set the context for a major future challenge to establish domains (words) and modalities (sentences) of lncRNA-based mechanisms—similar to what is known for protein-coding genes today (e.g., KH domain ~ RNA binding; helix-turn-helix ~ DNA

binding). Understanding the RNA-based vocabulary not only will be empowering but also could lead to new avenues for RNA-based therapeutics (33, 40).

2. DIFFERENCES BETWEEN lncRNA AND mRNA GENES

Most transcripts derived from polymerase II (Pol II) follow specific steps toward producing a mature RNA transcript: First assembly of factors and Pol II at promoters, followed by elongation and splicing out of introns, and then the addition of a methylated guanosine cap and polyadenylated tail. On a first principle, mRNAs and lncRNAs share all of these properties in their biogenesis. Yet, lncRNAs and mRNAs have very different outputs from this process, in which mRNAs are abundant and rapidly localized to the cytoplasm; in contrast, lncRNAs are often much less abundant and often reside in the nucleus. Thus, different regulatory properties must occur in otherwise seemingly similar production lines (41).

Several recent studies have unearthed distinct properties that distinguish the biogenesis of lncRNAs and mRNAs (**Figure 1**). We synthesize these findings along the life cycle of mRNAs and lncRNAs.

2.1. Promoter

The first step in transcriptional regulation is assembling the key factors on the DNA promoter and establishing the transcriptional start site. Although lncRNA transcripts are less conserved than those of mRNAs, their promoters are conserved to similar levels, underscoring the potential importance of regulation at their promoters (43, 44).

A recent study investigated the modifications to the C-terminal domain (CTD) of Pol II at intergenic lncRNAs (284 lincRNAs) and mRNA promoters (6,027) in one cell line (HeLa) (45). The authors found that Pol II has less efficient pausing on lncRNA promoters relative to that of mRNAs (**Figure 1*a***). This finding is an important checkpoint and could explain the less precise transcription of some lncRNAs. This work also revealed that the CTD modification threonine 4 phosphorylation (CTD-T4P), which is correlated with early termination, is more prevalent throughout lncRNA gene bodies ($n = 284$). This study has additional important implication for the splicing and termination of lincRNAs, as discussed below.

Another recent study took a more global approach to compare and contrast the chromatin and transcription factor (TF) binding properties of 5,196 lincRNA and 19,575 mRNA promoters across 7 and 11 cell lines, respectively (46). This evaluation of chromatin immunoprecipitation (ChIP) data for 370 TFs and 70 histone modifications revealed several properties of lncRNA promoters that are conserved across multiple cell lines. The lncRNA promoters in general had fewer TF binding events as well as fewer histone modifications (**Figure 1*b*,*c***). However, some TF binding events were actually more conserved at lncRNA promoters relative to mRNAs (e.g., GATA TFs; **Figure 1*b***). Overall, this study had a counterintuitive finding that large combinations of TFs do not result in a more specified expression, yet they rather robustly reinforce expression and result in a higher abundance of lncRNA transcripts (46).

More recently, this finding has been experimentally confirmed by examining the activity of 2,078 mRNA and lncRNA promoters using the massively parallel reporter assay (MPRA) (47). This validated that indeed mRNA promoters have intrinsically higher activity in MPRA. Interestingly, both lncRNAs and mRNAs organized in a divergent orientation, or bidirectionally promoted, were also more inherently active despite being tested independently. In contrast, enhancer lncRNA (eRNA) and lncRNA promoters had inherently lower activity in MPRA but were much more selectively expressed across the cell lines tested. Deeper exploration by

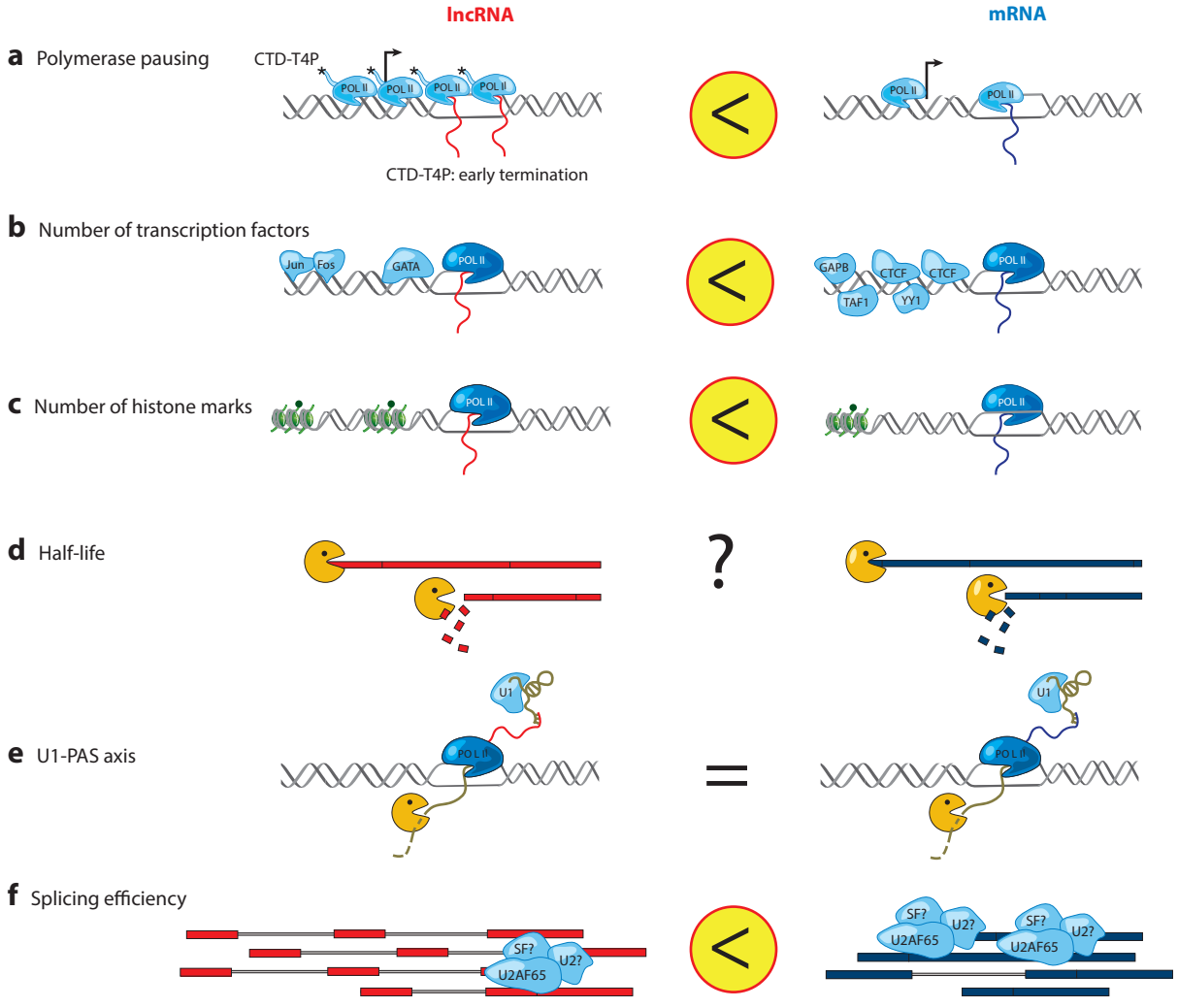


Figure 1

What is the difference between long noncoding RNAs (lncRNAs) and messenger RNAs (mRNAs)? (*a*) Polymerase C-terminal domain modifications [e.g., threonine 4 phosphorylation (CTD-T4P)] are enriched at lncRNA promoters, resulting in less polymerase pausing and earlier termination throughout lncRNA gene bodies. (*b*) The number of transcription factors bound at an lncRNA and mRNA promoter correlate with abundance and anticorrelate with tissue specificity. However, some transcription factors are specifically enriched in lncRNA promoters relative to mRNAs (e.g., GATA), whereas some prefer mRNA promoters (e.g., YY1). (*c*) In general, mRNA promoter regions contain more densely packaged histone modifications. Interestingly, histone H3 lysine 9 trimethylation (H3K9me3)—typically a silencing mark—is enriched at lncRNA promoter regions and gene bodies, even at highly abundant lncRNA loci (e.g., Tug1). (*d*) A definitive half-life for lncRNAs versus mRNAs remains uncertain. Several studies have demonstrated similar decay rates following destabilization of transcription (e.g., via an actinomycin-D time course). However, others using metabolic labeling of transcribed RNAs report a 10-fold decrease in lncRNA stability relative to that of mRNAs. This question remains open for further investigation. (*e*) The U1-polyadenylation sequence (PAS) structure is similar between lncRNAs and mRNAs, and thus it is unlikely to be responsible for the stability or splicing regulation differences observed between lncRNAs and mRNAs. (*f*) Splicing efficiency is the strongest distinguishing factor between lncRNAs and mRNAs with poor and efficient splicing, respectively. This can be attributed, in part, to far less binding of splicing regulatory factors to lncRNAs (e.g., U2AF65). Figure adapted with permission from Reference 42.

saturation mutagenesis of dozens of lncRNA and mRNA promoters demonstrated that lncRNA promoters are driven by substantially fewer TF motifs than by mRNA promoters; mutations rarely effected expression of the latter owing to redundancy of TF binding events (47).

Collectively, these studies have demonstrated that lncRNA promoters have fewer TF binding events and fewer TF motifs that result in lower expression levels, yet they provide the advantage of highly specific promoter activity. Thus, specificity in expression can be accomplished by simplicity of promoter sequences, and robustness of expression is derived from highly complex promoter sequences and binding events—or robustness in motif redundancy (47).

2.2. Stability

It has become very clear that, in general, lncRNAs are expressed at a much lower rate in a given cell type relative to the expression of mRNAs (2, 3, 6). One possibility to explain the lower abundance of lncRNAs could be that their decay rates are faster than those of mRNAs. Several studies have tested this and reached opposite conclusions (46, 48, 49) (**Figure 1*d***). Two studies using transcriptional arrest by actinomycin-D (ActD) found that both lncRNAs and mRNAs exhibit a wide range of transcript half-lives, with some very stable and some rapidly turned over, but with no bias for lncRNAs to be less stable (46, 50). Similarly, another study that avoided the possible caveats of ActD termination and instead combined an inducible terminator with metabolic labeling (bromouridine pulse chase) found very little difference in the half-life profiles of mRNAs and lncRNAs (49). In contrast, a study using metabolic labeling (4-thiouracil pulse chase) found lncRNAs to be on average 10-fold less stable. Despite these divergent conclusions, one common thread is that decay alone cannot explain the lower abundance of lncRNAs (48) (**Figure 1*d***).

Although the majority of lncRNAs are lower in abundance, it is noteworthy that examples of very low abundance RNA molecules that play critical roles such as the telomerase RNA (TERC) have been identified. TERC is often present only as a handful of molecules despite its critical role in extending 46 telomeres in mammalian cells, yet TERC becomes highly abundant in cancer. Thus, in some cases pruning to lower abundance may be important.

2.3. Splicing

After TF loading, formation of the preinitiation complex at the promoters, and the initiation of lncRNA and mRNA transcription, the next step is splicing. It was previously found that splice sites and other features (exonic splicing enhancers) are as conserved in lncRNAs as they are in mRNAs (43). More recently, two independent studies have found that splicing efficiency (how fast an intron is removed) is the largest difference between mRNAs and lncRNAs (46, 48). A recent study determined that U1 binding interactions are more frequently observed in lncRNAs and thus could explain splicing inefficiency in lncRNAs (51). However, it was also observed (46) that the U1 binding sites, polyadenylation sites (the U1-PAS axis), are very similar between lncRNAs and mRNAs (**Figure 1*e***). Although motif usage and conservation of U1 binding sites are very similar between lncRNAs and mRNAs (46) (**Figure 1*e***), lncRNAs are much more poorly spliced in general (**Figure 1*f***). Collectively, these findings suggest that lncRNAs are inefficiently spliced, and this might be an underlying reason of their nuclear localization. Interestingly, two independent studies both found that efficiently spliced lncRNAs are more likely to be functional (46, 48) (**Figure 1*e*,*f***). This finding also has an important implication for lncRNA localization, as retained introns, containing U1 binding sites, may cause nuclear retention and differential RNA turnover (51–53).

Thus, the most discerning factor between lncRNAs and mRNAs is the efficiency in which introns are spliced out of lncRNAs versus mRNAs. However, some lncRNAs are as efficiently spliced

as mRNAs—yet are nuclear localized. In fact, the property of efficiently spliced lncRNAs correlates with lncRNAs that have been determined to be functional. In contrast, inefficient splicing could retain lncRNAs in the nucleus for hitherto unknown RNA-based regulatory roles.

3. STRUCTURE–FUNCTION RELATIONSHIPS OF lncRNAs

The relationships between structure and function in lncRNAs have challenged the field for some time, but recently progress has been made. In the beginning, investigators found that the primary sequences of many, but not all, lncRNAs are poorly conserved, leading some to suggest that many lncRNAs lacked function (54). However, the promoters and splice signals of lncRNA exons are evolutionarily conserved, indicating that transcription and RNA processing of lncRNAs could be important (43, 44, 55). Moreover, many lncRNAs exhibit syntenic conservation (in which an lncRNA resides in the same DNA neighborhood across species) but with widely varying lncRNA primary sequences (5, 56, 57). Notably, the telomerase lncRNA also has similar properties of syntenic conservation across many species yet is highly divergent in the primary sequence (58), owing to the evolutionary requirement of structures that can be derived from multiple primary sequences.

This apparent paradox started to be resolved with the recognition that the intrinsic capacity of RNA to base-pair transforms the primary sequence motifs into secondary and tertiary structures. RNA secondary structures may be the units of lncRNA words rather than the primary sequence (see **Figure 1**). Recent progress in RNA structure mapping includes improvements in mapping flexible single-stranded nucleotides in living cells (59–61), inferring protein–RNA interactions by footprint analysis of the RNA binding protein on the RNA (59, 62), and direct mapping of RNA base-pairing partners genome-wide by psoralen cross-linking (63–65) (**Figures 2** and **3**). Many of these transcriptome-wide maps, as well as targeted in vitro studies of single lncRNAs, have led to insights about functional motifs and their corresponding structures within lncRNAs.

3.1. Xist

Xist is an ~17-kb-long lncRNA that is critical for X chromosome inactivation in female mammals (66–71). Xist is a prime example of the modular organization of lncRNAs (40, 72). The primary sequence of Xist is notable for the presence of several repeats, named A-repeat through F-repeat, each of which contains clusters of multiple instances of a distinct short sequence. These repeats have been shown to mediate critical functions of Xist through genetic or blocking studies. Wutz and colleagues (73) showed that deletion of the A-repeat abrogated Xist's ability to silence transcription of the inactive X chromosome, but A-repeat was not required for Xist to accumulate and spread across the inactive X chromosome. A different region, the B-repeat and C-repeat, is needed for Xist to bind the RNA binding protein hnRNPK and nucleate polycomb repressive complex 1 on the inactive X chromosome (74, 75) (**Figure 2**). Finally, E-repeat is needed for Xist localization on the inactive X chromosome, through the interaction of Xist E-repeat with the nuclear matrix protein CIZ1 (76, 77).

The modularity of Xist organization was further demonstrated in lncRNA-directed proteomic studies. Chu et al. (78) and McHugh et al. (79) retrieved Xist RNA and associated proteins for mass spectrometry analysis. These studies showed that Xist interacts with ~81 proteins; 10 of these are direct RNA–protein interactions, and the remainder are inferred to be indirect protein–protein interactions. Chu et al. (78) further showed that deletion of A-repeat caused three proteins to cease association with Xist but did not affect the interaction of other interacting proteins. This analysis thus led to the identification of several proteins that are essential for Xist-mediated gene silencing—namely Spen, an adaptor protein that bridges Xist RNA to the histone deacetylase

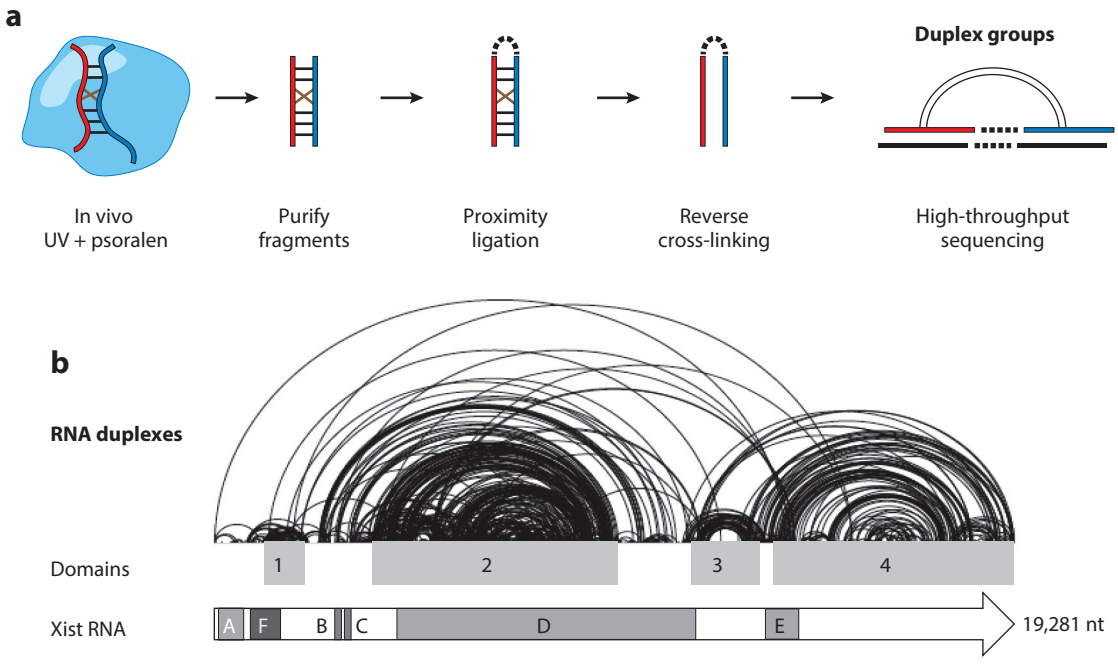


Figure 2

Structure and function in long noncoding RNAs (lncRNAs). (*a*) Psoralen-assisted RNA interaction and structure (PARIS). Psoralen cross-linking traps RNA duplexes in living cells. Proximity ligation transforms these structures into sequence libraries. Each gapped sequence read identifies portions of RNAs exhibiting intra- and intermolecular interactions, depicted as arcs. Panel *a* adapted from Reference 63 with permission. (*b*) RNA duplex map of human Xist as revealed by PARIS analysis. The multiple loops within loops highlight RNA structural domains 1–4. Xist repeats A–F in the primary sequence are shown below.

complex (78, 79), and WTAP, a subunit of the RNA N^6-methyladenosine (m^6A) methylation writer complex (78). Subsequent study showed that the A-repeat of Xist directly binds Spen and RBM15 proteins and that both a specific histone deacetylase (HDAC3) and m^6A modification of Xist are critical for Xist-mediated gene silencing (79, 80). In addition, a small region immediately 3′ to the A-repeat is important for Xist nucleation on the site from which Xist is transcribed (68) and also for binding to the lamin B receptor (LBR) that moves the inactive X chromosome to the nuclear periphery (75).

Recent advances in solving higher-order structure in living cells have started to provide an initial glimpse of the three-dimensional (3D) organization of these multiple functional elements on a long RNA. Three groups devised a psoralen-assisted method to cross-link and sequence two strands of RNA that are engaged in direct Watson-Crick base-pairing (63–65). After psoralen-assisted cross-linking, the two strands in an RNA duplex are ligated together, the psoralen cross-link reversed, and chimeric RNAs sequenced. Each sequence read provides single-molecule evidence of the RNA interactome within the cell (**Figure 2*a***). Applying such an approach to human XIST, Lu et al. (63) showed several notable features of the XIST structure. First, the secondary structure is modular, evidenced by loops within loops that segregate into four main domains (**Figure 2*b***). These domains only approximately correspond to the presence of some sequence repeats within Xist. Second, many RNA duplexes span long distances; the farthest one spans 7 kb and is conserved in evolution. Third, Xist and many other lncRNAs adopt multiple alternative

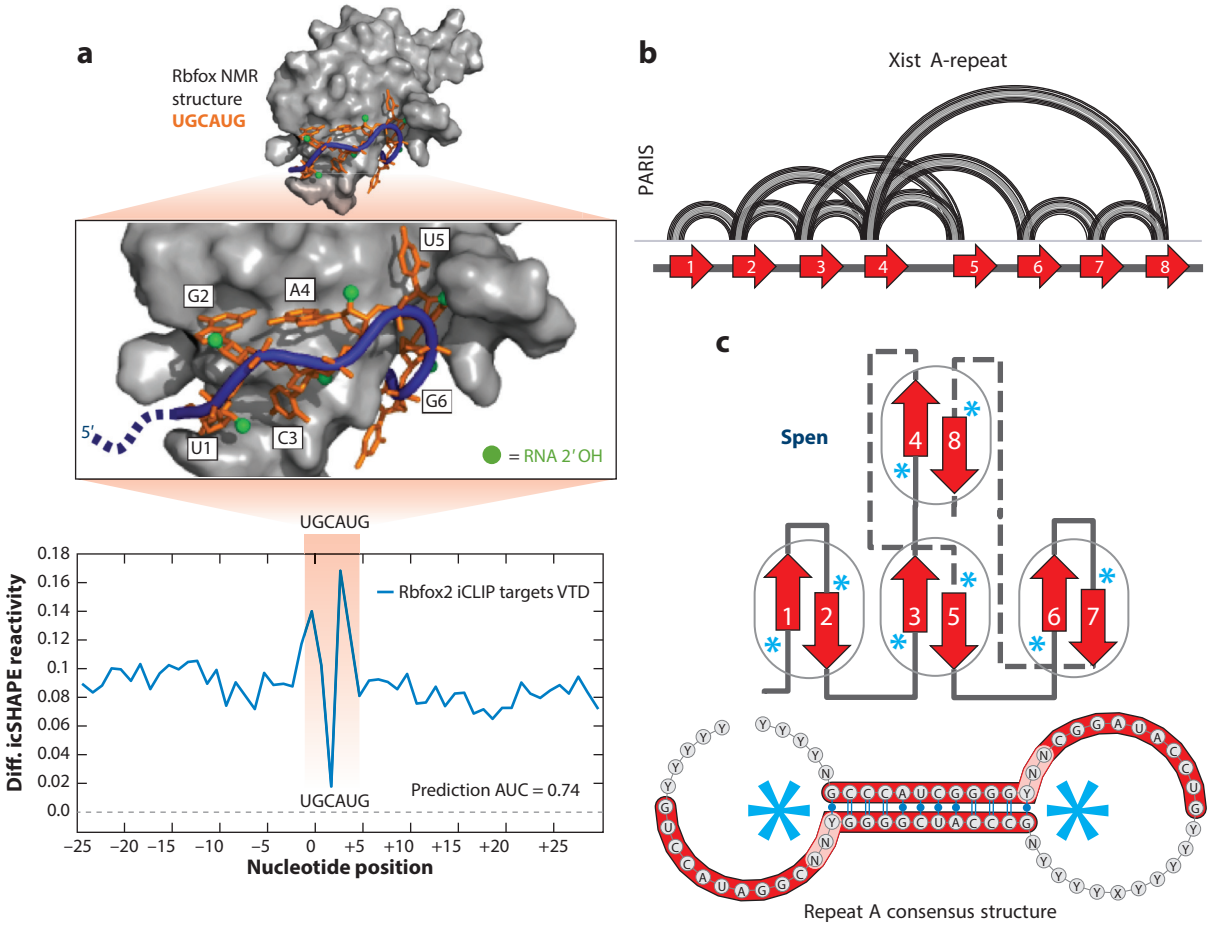


Figure 3

Noncoding RNA sequence and structure mediate RNA binding protein (RBP) interactions. (*a*) RNA structure imprint of RBP binding. In vivo icSHAPE (in vivo click selective 2-hydroxyl acylation and profiling experiment) shows that the Rbfox motif adopts a distinctive 2′ OH acylation profile as compared with identical sequences in vitro. VTD is the in vivo versus in vitro difference. Comparison of the NMR structure of the Rbfox2–RNA complex reveals the biochemical basis. For example, the 2′ OH of G2 and that of A4 are positioned outward in the bound complex, corresponding to a peak in icSHAPE signal in cells. The 2′ OH of C3 is protected by the RBP, corresponding to a trough in icSHAPE signal. This specific pattern of in vivo structural dynamics can be used to predict RBP binding transcriptome-wide (59). Panel *a* adapted with permission from Reference 59. (*b*) Xist A-repeat structure. The primary sequence of Xist A-repeat is shown; each arrow is an instance of the repeat. Eight copies of the repeat are shown for simplicity. Psoralen-assisted RNA interaction and structure (PARIS) data were used to identify multiple secondary structures that all have the inter-repeat staggered duplex, shown as the antiparallel staggered arrows. One of the most populated structures is shown. (*c*) Enlargement of an inter-repeat duplex. Two instances of the repeat are indicated in red, and the blue asterisks highlight the cross-linking sites of Xist to Spen. Panel *c* adapted with permission from Reference 63. Other abbreviation: iCLIP, individual-nucleotide-resolution cross-linking and immunoprecipitation.

structures, as evidenced by the fact that one nucleotide can be observed to participate in multiple, mutually exclusive secondary structures in the ensemble measurements.

The A-repeat contains 8.5 copies of an ~30-nucleotide sequence that is repeated in tandem. Due to its repetitive nature, it was challenging for one-dimensional (1D) RNA structure probing methods or computational modeling to identify a unique solution, and past studies suggested

several contradictory structures. Lu et al. (63) further clarified the structure of the A-repeat. The RNA duplex map clarified that the fundamental building block of the A-repeat structure is an inter-repeat staggered duplex (**Figure 3*b***). The 5′ portion of one repeat base-pairs with the 5′ portion of another repeat, and Spen binds to the junction of the single- and double-stranded regions of the staggered duplex. Importantly, although considerable flexibility exists in the choice of which repeats interact, the observed preference for repeat 4 and repeat 8 to interact mandates a fold back of the entire A-repeat structure, which creates a compact platform with a high-density Spen-associated silencing activity. RBM15, also known as Spenito in *Drosophila*, is a Spen-related RNA binding protein with RNA recognition motif (RRM) domains very similar to those of Spen. RBM15 also binds the A-repeat and is important for m^6A modification of Xist (80). The m^6A-mediated recruitment of a reader protein, named YTHDC1, also contributes to Xist-mediated gene silencing (80).

3.2. More and More lncRNA Structures: HOTAIR, COOLAIR, Braveheart, roX RNAs

Targeted structural analysis has now been applied to several lncRNAs, revealing their secondary structures. In general, lncRNAs are composed of modular domains. Each modular domain can be expressed individually and retains the structure found in the native full-length molecule (81). For example, Pyle and colleagues (81) showed that HOTAIR is modularly organized in vitro, with 5′ and 3′ domains that correspond to previously identified binding sites for chromatin modification complexes.

COOLAIR is an lncRNA in *Arabidopsis thaliana* that is important for vernalization, the process by which cold temperature epigenetically sets subsequent flowering time. Secondary structure mapping also identified specific domains and secondary structures important for its function, including a new right-hand turn motif (82, 83). Braveheart is a mouse lncRNA that functions in mesoderm and cardiac differentiation. A detailed dissection of the secondary structure of Braveheart identified a recurrent right-hand turn motif that is important for its interaction with YY1, a TF that mediates its differentiation function (83).

These examples illustrate the principle that RNA structure can highlight small functional motifs that are embedded in lncRNA sequences. This concept was also beautifully illustrated in the study of roX1 and roX2 RNAs, which are involved in X chromosome dosage compensation in *Drosophila melanogaster*. Despite the fact that they are genetically redundant, roX1 and roX2 differ greatly in size and primary sequence. Targeted SHAPE analysis showed that roX1 and roX2 share a common structural feature: the presence of multiple copies of a tandem stem-loop motif, which turned out to be the binding site of the male lethal (MLE) RNA helicase and additional components of the male sex lethal (MSL) dosage compensation complex (84, 85).

3.3. Nuclear Localization Sequences

A major distinction between many lncRNAs and mRNAs is their respective localization within the cell (86). Many lncRNAs have a predominant or sometimes exclusive nuclear localization. Consistent with this notion, investigators have identified the specific RNA motif within lncRNA that confers nuclear localization. Zhang and colleagues (87) performed systematic mutagenesis on the lncRNA Borg and identified a small pentamer (AGCCC) motif that is both necessary and sufficient to confer nuclear localization. Recent development of massively parallel RNA reporter gene assays (MPRNA), similar to MPRA used to identify regulatory DNA elements, has enabled systematic screens of RNA elements that direct subcellular localization (88, 89). In this approach, investigators fuse a library of sequences, typically tens of thousands derived from endogenous or

synthetic sequences, to a reporter RNA (88). The reporter library is transduced into cells, fractionation is performed to enrich for the nucleus or cytoplasm, and individual sequences that may drive RNA localization are identified by deep sequencing.

Using this approach, Lubelsky & Ulitsky (89) determined that a short sequence derived from Alu repeats is sufficient for RNA nuclear localization. This RNA element bound to hnRNPK, a known interactor of nuclear lncRNAs such as Xist (78). Similarly, Shukla et al. (88) dissected more than 10,000 RNA elements from 38 human lncRNAs by MPRA and identified more than 100 RNA elements that suffice to enrich for nuclear localization of an otherwise cytosolic transcript. These authors further found that the putative RNA nuclear localization signals were larger domains that were validated by orthogonal methods such as single-molecule RNA fluorescence in situ hybridization (smRNA-FISH). Interestingly, both of these studies identified a C-rich motif (Sirloin) that was overrepresented in RNA nuclear enrichment signals relative to input sequences. However, this motif was not sufficient to localize reporters, whereas larger domains of RNA had a larger effect (88).

Another important piece to this puzzle was identified by Yin et al. (51). They performed a similar screen with complementary DNAs from fragmented human transcriptome and identified interaction with U1 snRNA as a mechanism for nuclear localization of several RNAs. Moreover, Chen et al. (90) identified a short sequence on Xist, near the A-repeat, that mediates interaction with LBR, which recruits Xist to the nuclear lamina (75). Collectively, these studies have identified a myriad of sequences and domains that facilitate the nuclear localization of lncRNAs.

Because these related screens identified nonoverlapping RNA elements, it is clear that the space of RNA nuclear retention elements is far from saturated. Nonetheless, the emerging principles are that (*a*) lncRNAs possess address codes—short RNA elements that direct their localization—and (*b*) these RNA address codes are sites of RNA–protein interaction, where the RNA binding protein has a large role in RNP localization. Another important consideration is that larger domains identified in these and earlier studies (e.g., Malat1 nuclear localization signal) (91) may adopt similar secondary structures despite divergent primary sequences.

These studies revealed that functional RNA elements in lncRNAs and mRNAs are systematically demarcated by distinct RNA structural profiles. Furthermore, the structural imprint of RNA binding proteins interacting with specific motifs on RNAs can be directly read out and therefore predicted based on the in vivo RNA structurome (59). Moreover, new approaches to study RNA function with much greater throughput and comprehensiveness. Martin et al. (92) and Buenrostro et al. (93) have established novel platforms for parallel synthesis and functional assays across thousands to millions of RNA species. Using this mutate-and-test strategy, investigators can delineate every nucleotide and base pair required for regulatory RNA function and, indeed, infer its evolutionary path (92, 93). Collectively, these new tools enable investigators to achieve a comprehensive understanding of the information content of diverse RNA species.

4. GENETIC APPROACHES TO UNRAVELING lncRNA REGULATORY ROLES IN VIVO

A general mantra in biology is to identify genes that contribute to cellular and organismal phenotypes and ultimately those genes that are critical for human health and disease. For lncRNAs, this presents several new challenges. First and foremost, there are vast catalogs of lncRNAs, with some estimates as high as 100,000 and some as low as 5,000 (1, 2, 6, 44). Thus, it is akin to finding a needle in a haystack—lncRNA genes that will have profound roles in development and disease. Classically, premium candidates implicated in development and disease are tested in mouse models to identify potential phenotypes and relevance to human health.

Yet, it is important to remember the fundamental importance of genetic model systems, as Edward Lewis is famously quoted:

> The laws of genetics had never depended upon knowing what the genes were chemically and would hold true even if they were made of green cheese. (94, p. 5)

This quote underscores the importance of removing DNA elements associated with human disease and remembering that, if a phenotype arises, something important is encoded within the DNA.

Based on this advice, we first discuss some case studies underscoring the importance of using multiple mouse models to identify the underlying mechanisms of DNA versus RNA regulatory elements and their combinatorial actions. Although many important studies have found lncRNA roles in cell lines, we first focus on those using genetic approaches to unravel underlying *cis* regulatory roles in vivo. We use the word "*cis*" as a proxy for "local gene regulation" of either neighboring or nearby genes, yet it is important to note that the true definition of *cis* would require further analyses to determine if the regulatory affects are on the same allele. Next, we discuss similar genetic approaches that have revealed *trans*-acting lncRNAs (**Figures 4** and **5**).

4.1. Genetic Identification of *Cis* Regulatory Elements Harbored in lncRNA Loci In Vivo

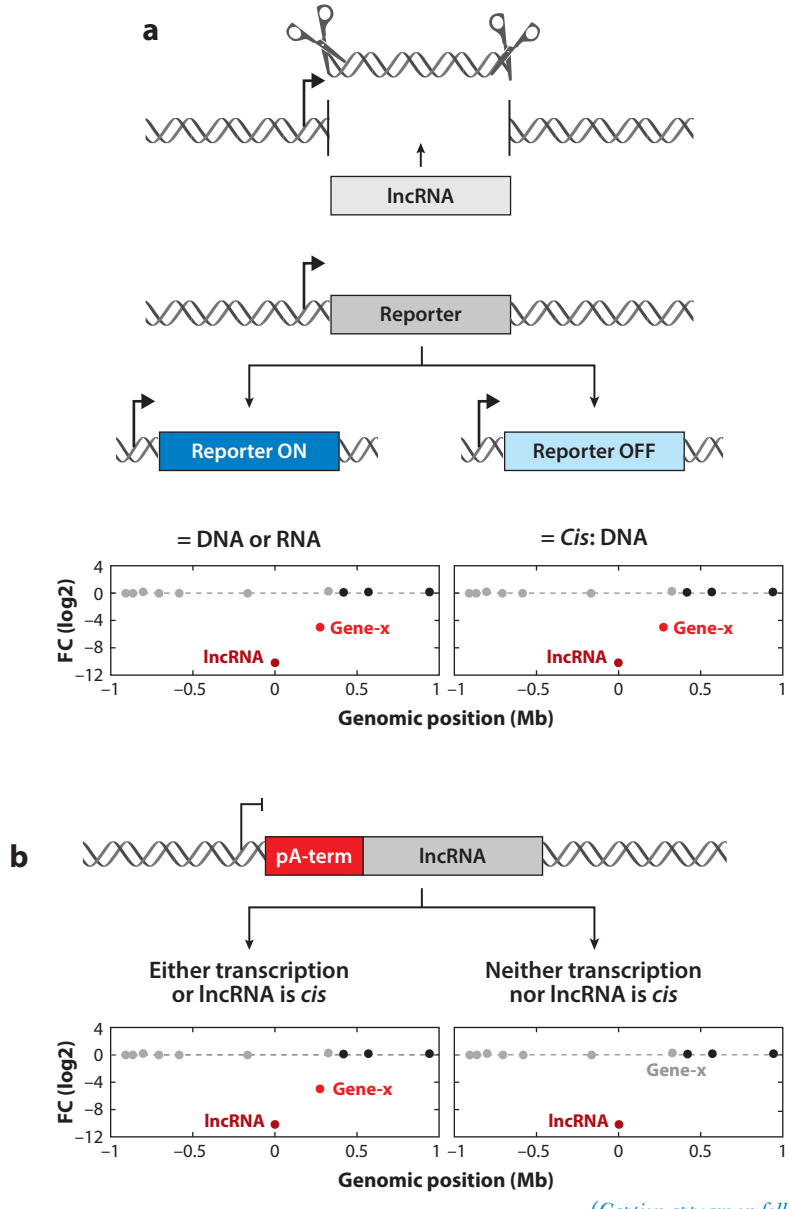
The intricate complexity of mammalian genomes with interleaved and overlapping gene regulatory elements and transcripts presents several challenges in using genetic approaches to understand lncRNAs. This goal requires multiple genetic approaches to disentangle potential overlapping mechanisms, such as DNA elements contained within gene bodies, the act of transcription, and the mature RNA product.

A fundamental example of this problem is the *Dync1i1* mouse knockout model. Removal of a large piece of the gene locus results in digit formation defects, which could be attributed to this gene (95). However, it was later discovered that exon 15 of *Dync1i1* is actually a DNA enhancer for the neighboring DLX5/6 region, which phenocopies the *Dync1i1* (split hand and foot malformations). Thus, at first pass the phenotype of *Dync1i1* protein-coding product would be misclassified and the DNA enhancers missed (95).

Both mRNA and lncRNA loci have overlapping complexity that has required multiple loss-of function (LOF) and gain-of function (GOF) approaches to unravel the role of the locus. An example similar to that of *Dync1i1* is the lncRNA p21 locus (96). Several studies have shown that the gene product regulates many genes in *trans* and also in *cis* (the neighboring p21) (96–98). These findings were determined by oligo-based knockdown of the RNA (in mouse embryonic fibroblasts), which suggests a role of the mature RNA product in these processes (98). However, a gene ablation with reporter knock in (that controls for the act of transcription) identified an overlapping DNA element that regulates numerous surrounding genes, including p21. Importantly, the *cis* regulation was observed in both tissues expressing linc-p21 and those that do not, thus definitively determining the *cis*-like effect could not be due to either the act of transcription (reporter as a control) or the mature RNA product (because effect was seen in the presence and absence of expression). Nonetheless, this finding does not rule out other potential *trans* roles for the mature lncRNA product, but it does define the DNA as a *cis*-acting element (**Figure 4*a***). This does not rule out a role for the lncRNA product in other biological pathways—for example, it is noted in many studies that *linc-p21* expression is highly upregulated in several cancers.

Thus, full gene ablation with reporter knock in is a salient approach to rule out a role for the mature RNA for *cis* gene-regulatory events. If a *cis* effect is observed in the presence or absence of the reporter, it must be due to a DNA element (11, 96, 99, 100). Together, these studies have

unearthed densely packed noncoding regulatory elements based on DNA and RNA molecular modalities of the linc-p21 locus (**Figure 4*a***). More recently, another lncRNA locus, *Peril*, was determined to function through DNA enhancer elements similar to linc-p21 (100). Surprisingly, *Peril* was previously considered a super enhancer for Sox2 (the adjacent mRNA locus), yet in vivo it has no effect on Sox2; instead, it regulates two distal genes, *Mccc1* and *Exosc9*, that phenocopy the early lethality phenotype observed in *Peril* knockout mouse models (11). Thus, it is worth noting how misleading neighboring genes can be in identifying targets of *cis* DNA-encoded enhancers in lncRNAs.



(*Caption appears on following page*)

Figure 4 (*Figure appears on preceding page*)

In vivo identification of *cis* regulatory elements harbored in long noncoding RNA (lncRNA) loci. (*a*) Full-gene ablation models can definitively determine if any DNA-based *cis* regulatory elements are contained within an lncRNA locus [the same would hold true for messenger RNA (mRNA) loci]. The act of transcription at this locus can be tracked by replacing the lncRNA with a reporter gene (e.g., GFP or LacZ), indicating where the lncRNA locus is active and, importantly, inactive in vivo. If the reporter allele affects neighboring gene expression in a tissue where the reporter gene is active, this result could be attributed to the function of either deleted DNA element(s) or the lncRNA product itself. Because the act of transcription at the lncRNA reporter locus is not abolished, it cannot be responsible for the phenotype (*left*). However, if molecular or tissue phenotypes are observed in tissues or cells where the reporter is silent (*right*), then this suggests that underlying DNA elements are unambiguously responsible, for example, by regulating neighboring gene(s). (*b*) *Cis*-like mediated effects can also be determined using polyadenylation termination (pA-term) sites inserted immediately downstream of an lncRNA (or mRNA) promoter. If the lncRNA is depleted and a *cis*-like effect is observed, this suggests that either the act of transcription or the lncRNA product may regulate neighboring genes. However, if only the lncRNA is depleted and not neighboring genes, this suggests that DNA is not serving as a local regulatory element. If other genes change outside this local neighborhood, it would suggest a *trans*-acting mechanism for the lncRNA product. Several studies have estimated that ~40% of lncRNA loci contain *cis* DNA elements but have not rule out other roles for the lncRNA product in *trans*. It is important to note that many mRNA loci are known to contain *cis* DNA regulatory roles separate from the function of the protein product (e.g., Dync1i1).

Haunt (previously termed linc-HOXA1) further exemplifies the importance of looking for multiple overlapping functions within a gene locus (101, 102). The Haunt lncRNA is almost exclusively expressed in mouse embryonic stem cells (mESC) and upon differentiation is no longer expressed (17, 101, 102). A study using smRNA-FISH originally found that the *HAUNT* locus is abundantly expressed when *HOXA* gene expression is low and vice versa, consistent with a role in maintaining pluripotency (102). A more recent study found that indeed increased expression of *Haunt* resulted in depletion of *HOXA* gene expression (101). This study also revealed that a *cis*-acting DNA activator is encoded in the *HAUNT* locus as well. By performing larger and larger deletions of the *HAUNT* locus, they found reduced expression of *HOXA* genes, whereas depletion of the lncRNA activated *HOXA* genes. Thus, this study further demonstrated the diametrically opposing (e.g., activation, repression) roles encoded in the DNA elements of gene loci: mRNAs and lncRNAs alike. In summary, examples such as the *HAUNT* locus underscore the importance of the act of transcription itself in regulating gene expression. Perhaps simply switching between heterochromatic and euchromatic states, as described (103), and associating with nuclear architectural proteins such as Lamina establish cell-specific nuclear organization and, in turn, gene-expression (33, 103–106).

Another possibility is the lncRNA does not function at all despite being expressed, being spliced, and having gene-expression patterns that correlate with the neighboring genes. A good example of this is the lncRNA locus termed *LOCKD* (lincRNA downstream of CDKN1b) (107). Here the authors performed two LOF approaches: full gene ablation and polyadenylation termination (pA-term) at the promoter to prevent the act of transcription yet leave the DNA intact (**Figure 4*b***). In the full gene ablation, they found that only *Lockd* and its neighboring gene *Cdkn1b* were downregulated genome-wide. Further examination of the pA-term mutant revealed the same transcriptional effect of regulating only *Cdkn1b*. Together, these two experiments concretely determine that the lncRNA acts solely on its own gene neighborhood. This finding raises an interesting question: Why would it be transcribed in the first place? It also means that depletion of *Cdkn1b* through this mechanism has no effect on any other genes, despite a key role for *Cdkn1b* in cell growth control.

Beyond enhancer DNA elements in lncRNA gene loci, lncRNA promoters are also potential long-range gene regulatory elements in *cis* (108). In fact, approximately 50% of lncRNA promoters

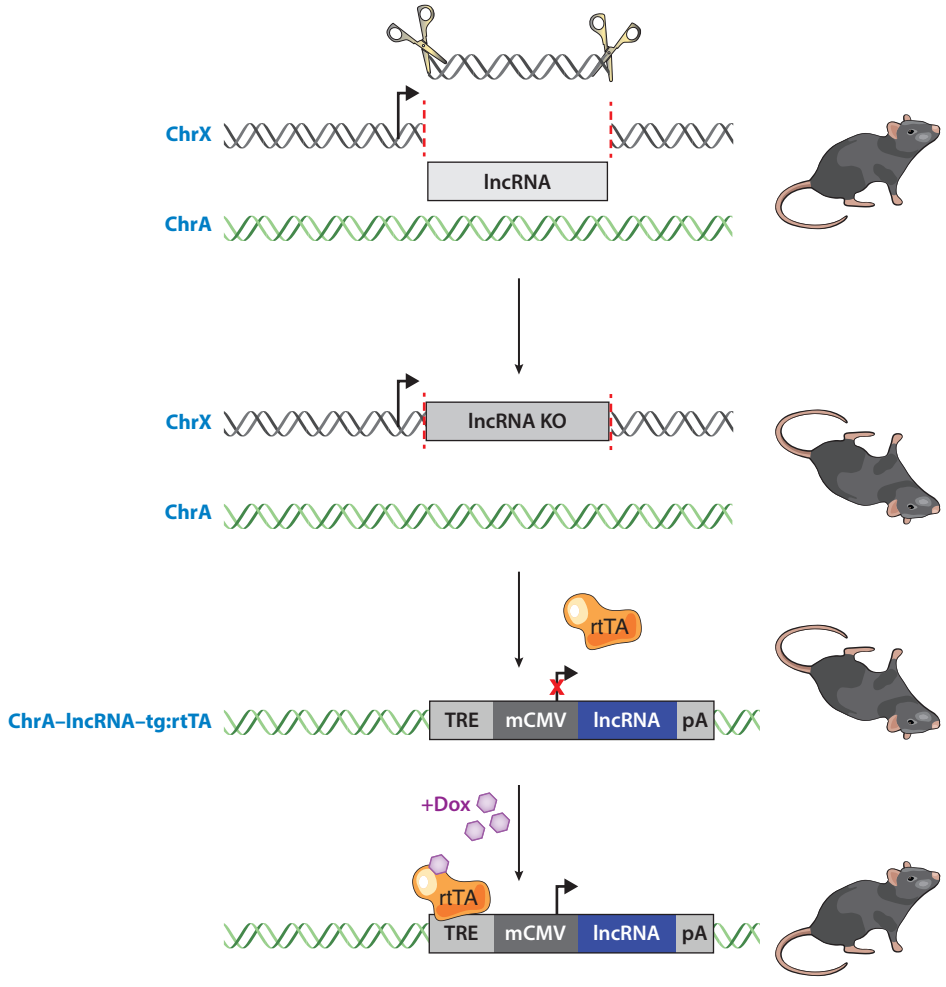


Figure 5

In vivo identification of *trans*-acting lncRNAs. Outline of a *trans*-gene rescue model requires generation of a full lncRNA KO. Then, a cDNA of the lncRNA is inserted on a separate chromosome (ChrA) with an inducible promoter strain containing the tetracycline-controlled transactivator (rtTA) element to be induced by Dox to bind to the TRE in the promoter of the transgene lncRNA (tg). If a phenotype is found in the lncRNA KO (ruling out *cis* elements above), then a direct role of an RNA-based mechanism can be tested to rescue this phenotype. For example, with FENDRR, the lncRNA KO results in lethality (or other phenotypes), but the lncRNA rescues this viability defect. This has also been the case for phenotypes in the immune system where the RNA can rescue the defect in *trans* (e.g., FIRRE, PNKY, FENDRR, linc-EPS). Using an inducible system is ideal, because if the phenotype occurs early in development, pups can be born with normal lncRNA expression. Similarly, if phenotypes are found in adulthood, the transgene lncRNA can be induced at different developmental time points to determine when and where a phenotype is rescued by the lncRNA. Abbreviations: cDNA, complementary DNA of an RNA; ChrA, chromosome A; Dox, doxycycline; KO, knockout; lncRNAs, long noncoding RNAs; mCMV, minimal cytomegalovirus; pA, polyadenylation; TRE, tet-responsive element.

function in *cis*, similar to what has been found for lncRNA gene bodies (~40%) (108, 109). Beyond the examples shown here, several studies have shed important light by making inversion mutants (110, 111) for lncRNA promoters and genes bodies that in turn have similar or different regulatory affects than gene ablations (11, 110–112).

More recently, it was found that the promoter and multiple DNA elements of the *PVT1* lncRNA locus can play important roles, resulting in a dynamic DNA regulatory module (113–115). *PVT1* is one of the first lncRNA genes identified by human cancer genetics and lies at the breakpoint of the t(8:14) translocation in Burkitt's lymphoma. *PVT1* is 55 kb away from the *MYC* oncogene on human chromosome 8. Cho et al. (113) found that silencing of the *PVT1* promoter promotes *MYC* transcriptional activation and drives cancer cell growth in a manner independent of the PVT1 lncRNA transcript. The authors discovered that the *PVT1* locus harbors four intragenic enhancers that normally contact the *PVT1* promoter. But when the *PVT1* promoter is not active, the *PVT1* enhancers can loop over to the *MYC* oncogene and drive *MYC* activation. Targeted gene editing of the *PVT1* promoter and somatic mutation patterns in human cancers confirm the importance of this mechanism. Thus, lncRNA promoters can act as inducible DNA boundary elements, segregating gene regulatory activities into proper neighborhoods. The ability of the promoter to capture nearby enhancers is critical for this newly recognized function, rather than the resulting RNA transcript that may possess or lack additional activities.

Collectively, these studies demonstrate the importance of *cis* regulatory roles of lncRNAs via RNA, DNA, and combinations therein. Combined with many other eRNAs reviewed elsewhere, it is clear that important RNA-based roles regulate neighboring genes (116, 117).

4.2. *Trans*-Acting lncRNAs In Vivo

Above, we discussed just a few of the many lncRNAs that appear to function in *cis*, but identifying *trans*-based mechanisms is even more difficult and further requires a rescue experiment to reinstate the expression of an lncRNA and determine if regulatory events are restored. Although many *trans*-acting lncRNAs have been identified in cell lines, here we focus on only recent studies that have identified *trans*-acting lncRNAs in vivo. Moreover, we focus on those that have clear connections to human health and disease.

The *PNKY* locus is a great example of how multiple genetic models can define an RNA-based mechanism for an lncRNA (32). Specifically, a *PNKY* locus knockout resulted in a dramatic cellular and in vivo phenotype—blocking neuronal stem cells to differentiate into projection neurons in vivo. Surprisingly, deletion of the *PNKY* locus did not have *cis* regulatory effects despite being bidirectionally transcribed from Pou3f2 (OCT7). The authors performed the ultimate test by adding an ectopic *PNKY* locus in the knockout background. Remarkably, this restored *Pnky* expression rescued the neurogenesis phenotypes in vivo. Thus, these findings genetically define PNKY as a *trans*-acting lncRNA (32).

Similarly, a recent study genetically defined the lncRNA Firre as a *trans*-acting RNA in vivo. Full gene ablation of the *FIRRE* locus from the X chromosome resulted in physiological and molecular hematopoietic defects in the common lymphoid progenitors (CLPs). Importantly, these defects could be rescued by ectopic expression of *Firre* from a transgene (118). Gene-expression analysis of the knockout cells did not identify any *cis*-acting DNA elements encoded in the locus despite the large genetic deletion, including the promoter region of the *FIRRE* locus. In contrast, the vast majority of genes (65/71) that were downregulated in *FIRRE* knockout CLPs were restored upon induction of the transgene in vivo (and vice versa). Thus, the Firre RNA product regulates cellular differentiation and gene expression in *trans*. Consistent with this notion, another recent study found that the Firre RNA product is required for histone H3 lysine 27 trimethylation

(H3K27me3) deposition on the inactive X chromosome in *trans* (119). However, Froberg et al. (120) did not observe this defect in allele-specific deletions of the *FIRRE* locus in cell lines. Collectively, these studies using genetic approaches have two important take-home messages: (*a*) Firre functions via an RNA-based mechanism in *trans* in several biological processes (118, 119), and (*b*) Firre has very little effect, in *cis*, on X chromosome inactivation (118, 121, 122), despite being required for the largest conserved chromosomal structure (122).

The *FENDRR* (FOXF1 adjacent noncoding developmental regulatory RNA) locus (11, 123, 124) was also found to function as a *trans*-acting lncRNA. The *FENDRR* locus was originally identified by comparisons of numerous mutations leading to the fatal human disease AVCD-MPV (alveolar capillary dysplasia and misalignment of pulmonary veins) (124). What was interesting about the *FENDRR* locus was that, although mutations to the neighboring *FOXF1* were thought to be causal, it was determined that mutations to the *FENDRR* lncRNA locus were the sole mutations in some patients, not the neighboring *FOXF1* (124).

Two studies generated mutant mouse models recapitulating mutations found in AVCD-MPV (11, 123). Indeed, both studies found a perinatal lethality phenotype similar to that in human disease. One mouse model identified a phenotype in the lung that phenocopies the human disease (11), and the other mouse model found defects in the heart (123). These findings could provide hitherto unknown insights into AVCD-MPV. The latter study went a step further by restoring *Fendrr* lncRNA expression through a *trans* gene (**Figure 5**). Remarkably, this rescued the phenotype and demonstrates that Fendrr can exert its regulatory role in *trans*. These studies underscore the importance of using mouse models to unravel the molecular modalities of lncRNAs. A critical first step is determining whether an lncRNA works in *cis* or in *trans*, so that therapies based on a *cis* or a *trans* mechanism can be developed.

Several *trans*-acting mechanisms for lncRNAs have been identified as key players in the immune system, both adaptive and native. In the native immune system, linc-EPS (erythroid prosurvival) is a key example of a *trans*-acting lncRNA. linc-EPS is highly expressed in macrophages and rapidly downregulated upon TLR4 stimulation and in turn upregulation of proinflamatory genes [or immune regulation genes (IRGs)]; this finding suggests a *trans*-based mechanism, as these genes are not located in proximity to linc-EPS (19). Importantly, ectopic expression of linc-EPS restores many of these IRGs, confirming a *trans*-acting role. Interestingly, mouse models of full gene ablation of linc-EPS result in endotoxic lethality when challenged, even at typically nonlethal doses. Similar to linc-EPS, linc-Cox2 has been found to have both *cis* and *trans* roles in innate immunity (125, 126). Wang et al. (127) further discovered that viruses can hijack host cell lncRNAs to facilitate viral replication. Infection by several viruses induces the expression of lncRNA ACOD1, which directly binds to and activates the metabolic enzyme glutamic-oxaloacetic transaminase to reprogram host cell metabolism in a manner that favors viral replication. Together, these and other examples such as THRIL, lnc13, and AS-IL-1a have all been implicated in orchestrating the exquisite responses needed in the native immune system.

Outside of the immune system, NORAD is a recently characterized lncRNA that is important to genomic stability (29). *Norad* knockout mice exhibit premature aging and mitochondrial dysfunction (128). NORAD is an abundant cytoplasmic lncRNA that contains multiple binding sites for the RNA binding protein Pumilio, thus titrating away Pumilio activity from mRNAs. Transgenic overexpression of *Pumilio* in vivo phenocopies *Norad* knockout mice, and the genomic instability of *Norad* knockout cells can be rescued by reexpression in *trans* of a NORAD RNA containing the Pumilio binding sites but not by NORAD mutants lacking the Pumilio binding elements (129).

In the brain, Ang et al. (130) found that lnc-NR2F1 is a developmentally regulated proneurogenic lncRNA. The *lnc-NR2F1* locus is recurrently mutated in patients with autism spectrum

disorder or intellectual disability, including a family with a chromosomal translocation that truncates the 3′ end of the lncRNA. These investigators inserted a premature polyadenylation signal into the *lnc-Nr2f1* locus in either the reverse (control) or the forward (knockout) orientation, a powerful strategy to control for changes in the DNA regulatory landscape. As expected, only the forward orientation knocked out *lnc-Nr2f1* expression and showed that lnc-N2f1 is required to promote the expression of multiple neurogenic genes. The knockout phenotype was further rescued by reexpression of the lncRNA in *trans*, and the full-length lncRNA, but not the disease-associated short isoform, can associate with chromatin and promote neuronal maturation. Overall, these recent studies have genetically defined several functional *trans*-acting lncRNAs in vivo that could represent novel RNA machines.

5. lncRNA SCREENS AND THERAPEUTICS

Functional studies of lncRNAs have moved into an exciting period owing to the advent of new technologies. The first approaches to interrogating lncRNA function used small interfering RNAs (siRNAs), which engaged the RNA interference pathway to knock down lncRNAs. The advantage of using siRNAs is that they interrogated RNA function independently of the genomic locus. The disadvantage of using siRNAs is that they have limited efficacy against nuclear lncRNA and substantial off-target effects (131, 132). Over the last several years, investigators have employed antisense oligonucleotides (ASO), typically with locked nucleic acid chemistry, because ASO have excellent efficacy to knock down nuclear lncRNAs and are efficacious in vivo, including in the brain (133). ASO continue to represent the best therapeutic approach to target lncRNAs, as exemplified by the success of targeting MALAT1 lncRNA in breast cancers in vivo (23, 24), although different mouse models have suggested alternative interpretations (134, 135).

Over the last two years, the development of CRISPR (clustered regularly interspaced short palindromic repeat) genome and epigenome editing has also transformed the ability to interrogate lncRNA loci function (31). Liu et al. (31) developed a CRISPR interference (CRISPRi) library, constructing more than 160,000 synthetic guide RNAs that target a dead Cas9-KRAB domain fusion to target promoters of more than 16,000 human lncRNAs. CRISPRi targets an ∼1-kb heterochromatin to the promoter of interest, silencing the targeted DNA locus and RNA transcript. The CRISPRi library can be transduced into target cells as a pool, and member single guide RNAs (sgRNAs, which direct CRISPRi targeting) that confer growth advantage, disadvantage, or other phenotypes (such as expression of a cell surface marker or reporter gene) can be rapidly identified on the basis of composition of the sgRNA library after selection.

Liu et al. (31) conducted a cell-growth screen of nearly 16,000 human lncRNA loci in seven human cell types. These authors discovered ∼500 lncRNA loci that are essential for cell growth or survival. Notably, most lncRNA hits are essential in only one cell type, in contrast to the promiscuous hits of most cell-essential protein-coding genes (31). This result is concordant with the exquisite cell type–specific selectivity of lncRNA expression and suggests a number of lncRNA targets with excellent therapeutic index to target diseased cell types (**Figure 6**).

Similarly, Konermann et al. (136) devised a CRISPR activation (CRISPRa) screen, in which a library of sgRNAs targeted dCas9-VPR, a synthetic gene activation construct, to the promoters of human genes. The authors conducted a screen for genes, including lncRNAs, that can confer resistance mutant B-RAF inhibition by vermurafenib in melanoma cells (136). Both CRISPRi and CRISPRa interrogate lncRNA loci function without distinguishing potential functions from the DNA elements in the lncRNA locus versus the lncRNA transcript or the act of transcription itself (104). Another CRISPR-Cas9–mediated screen used a dual guide strategy to delete lncRNAs toward identifying any loci that may have functional contributions (137). Church and colleagues

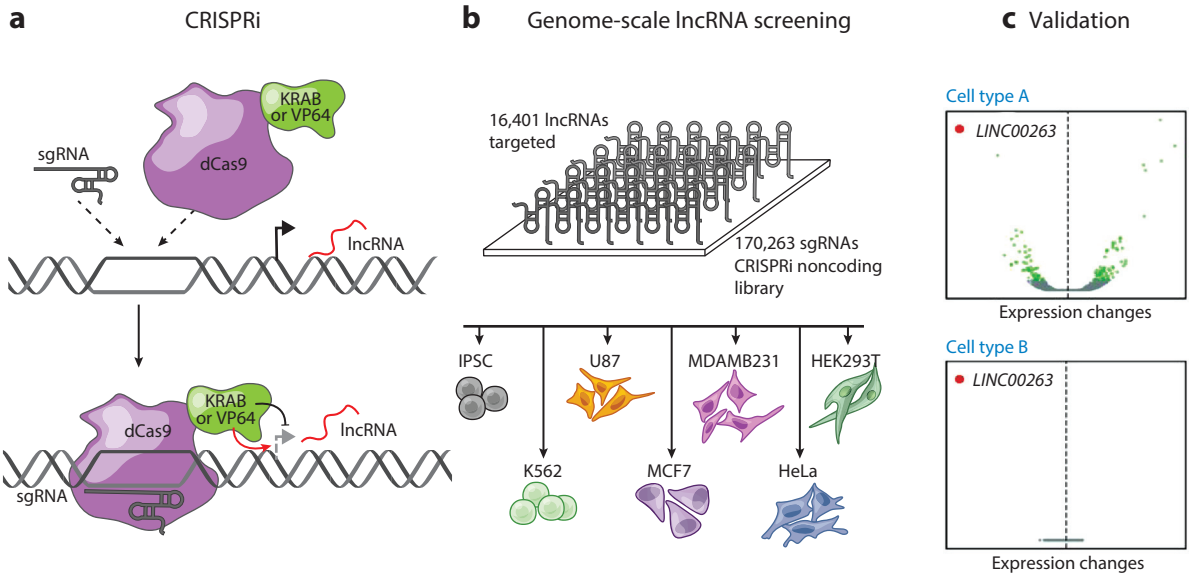


Figure 6

Genome-wide functional analysis of long noncoding RNAs (lncRNAs) by CRISPR interference (CRISPRi). (*a*) CRISPRi (KRAB) and CRISPR activation (CRISPRa) (VP64) design. (*b*) Pooled CRISPRi screen to knock down expression of 16,401 lncRNAs in seven human cell types. (*c*) This example RNA-seq analysis following CRISPRi screen reveals that *LINC00263* induces significant expression changes in cell growth genes in cell type A, but not in cell type B, thus confirming cell-type specificity. The red dots indicate reduced expression of *LINC00263*. Figure adapted with permission from Reference 31.

(138) have reported that the fusion of both KRAB and MeCP2 domains to dCas9 enhances CRISPRi efficacy, including toward lncRNAs. More recently, the advent of Cas13, an RNA-targeting CRISPR system, has offered the possibility of knocking down nuclear and cytoplasmic RNAs with greater specificity than can be achieved with siRNAs (139, 140). It remains to be seen whether Cas13 can be adapted into high-throughput pooled screens for lncRNA function. These new prospective and genome-wide tools complement reverse genetics studies of individual lncRNAs.

6. CONCLUSION: TOWARD A MOLECULAR GRAMMAR FOR lncRNAs

It is classically known that RNA plays central roles in biology. For example, all of our protein-coding genes are generated from the ribosome, an RNA machine. Moreover, TERC is an essential RNA component required to rejuvenate chromosomal ends. The deep knowledge of these two lncRNAs is a model for the approaches discussed above, all aiming to identify new RNA machines in a vast sea of lncRNA annotations. New progress and the approaches discussed above are leading to important discoveries that will further guide the identification of other lncRNA machines.

However, we also want to begin understanding how these principles translate more globally to thousands of lncRNA annotations. This effort has remained challenging without the molecular grammar of lncRNAs. For example, for protein-coding genes we know the letters (amino acids), words (domains), and sentences (structures). Thus, to near fluency, we can read a sequence and make some inference of function.

In contrast, we have almost no vernacular for lncRNAs. This has been a challenge, because noncoding RNA conveys information in a fundamentally different way: by disparate sequences

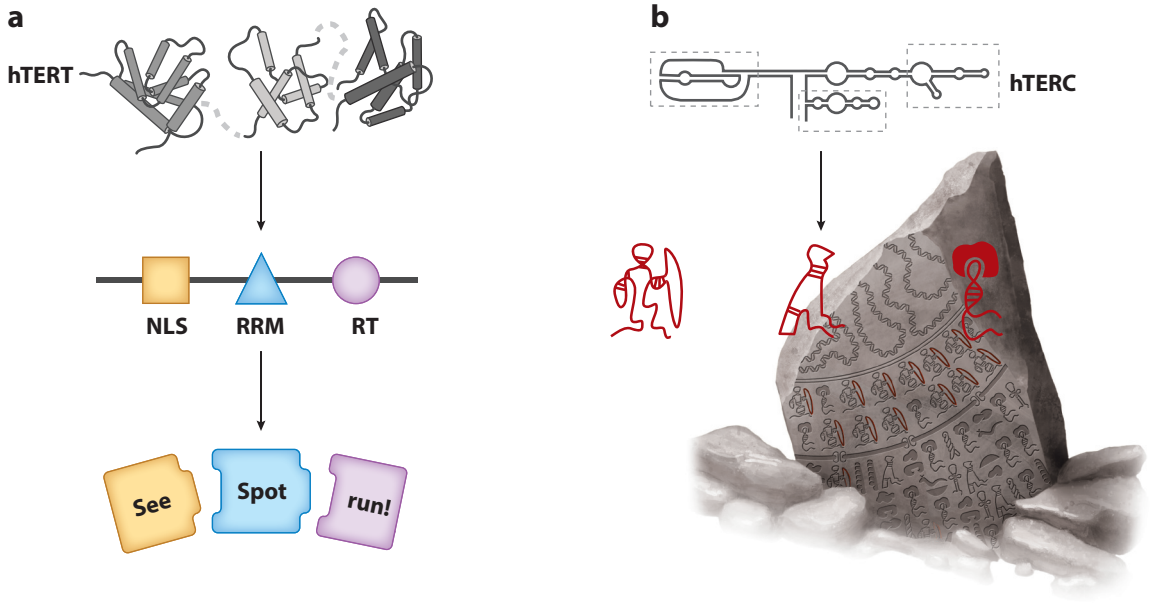


Figure 7

Toward a molecular grammar for long noncoding RNAs (lncRNAs). The telomerase lncRNA ribonucleic complex (lncRNP) is used as a model for how to move forward to understand the molecular grammar of lncRNAs. (*a*) hTERT is a protein component of telomerase and contains a nuclear localization sequence (NLS), an RNA recognition motif (RRM), and the enzymatic reverse-transcriptase (RT) domain telomere extensions. These protein domains serve as words that immediately infer information for a hypothesis to their functional roles. These protein-based domains are highly conserved and can often be identified by primary sequences throughout evolution. By way of analogy, these words result in a familiar lexicon that can almost be translated into a sentence (e.g., "See Spot run!"). (*b*) hTERC is the telomerase lncRNA and often contains three structural domains. In sharp contrast to TERT, the primary sequence of TERC domains is usually not conserved, yet orthologous TERC often adopts a similar secondary structure throughout evolution. Thus, RNA speaks a different language akin to hieroglyphics, in which a picture (or secondary structure) represents a word. The ultimate goal moving forward is to identify more hieroglyphics and understand their structure–function roles. With more and more examples, understanding the RNA vernacular like we can for protein-coding genes today could unveil a new genomic, RNA-based language.

that can form similar structure–function relationships. By way of analogy, proteins transmit information through a lexicon, and RNA transmits through a symbolic language like hieroglyphics.

A good example of bringing these two languages together is the ribonucleic–protein complex telomerase that is composed of the lncRNA TERC and several protein-coding components (58, 72, 141) (**Figure 7**). For example, three protein components of telomerase are readily recognizable as containing a nuclear localization domain, RNA binding domain (RRM), and reverse-transcriptase (RT) domain (to replicate telomere ends). On the basis of these components, we can almost read these words or domains within proteins as "See Spot run" (**Figure 7**) and identify putative functions on the basis of this sentence. Similarly, the human TERC has three domains; however, these sequences can widely vary but result in similar secondary structures (58). In this way, RNA communicates in words that are defined by secondary structure, similar to hieroglyphics. It is important to note that TERC, like many lncRNAs, is not highly conserved across eukaryotes yet adopts conserved structural elements (56, 57).

A first step toward this goal is the deep evolutionary analysis of lncRNA genes followed by efforts to construct synthetic lncRNAs de novo (56). Quinn et al. (85) used a step-wise approach of micro- and macrohomologies to identify orthologs of roX1 and roX2 lncRNAs across *Drosophila*

phylogeny. These authors further succeeded in grafting the roX box stem loop to a random sequence, bacterial LacZ mRNA, and showing that this construct suffices to endow the synthetic lncRNA with detectable, albeit weak, dosage compensation activity in vivo (85).

Future efforts will be important to resolve structure–function relationships in order to better understand RNA words or identify more structural hieroglyphics, toward developing an lncRNA vocabulary. Thus, the studies reviewed here inspire the ultimate goal. The future challenge is to predict regulatory domains from lncRNA primary sequences and, in turn, their functional contributions in vivo.

DISCLOSURE STATEMENT

H.Y.C. is a co-founder of Accent Therapeutics and Boundless Bio, and he is an adviser to 10x Genomics, Arsenal Biosciences, and Spring Discovery.

ACKNOWLEDGMENTS

We thank all the scientists for their contributions reviewed here and apologize to colleagues whose work was not reviewed at length due to space constraints. We thank Sigrid Knemeyer for illustrations. This review is supported by the National Institutes of Health (R35-CA209919, RM1-HG007735, R01-HG004361 to H.Y.C., and U01-DA040612, P01-GM099117, R01 MH102416 to J.L.R.). H.Y.C. is an Investigator of the Howard Hughes Medical Institute (HHMI), and J.L.R. is an HHMI Faculty Scholar.

LITERATURE CITED

1. Hon C-C, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJL, et al. 2017. An atlas of human long non-coding RNAs with accurate 5′ ends. *Nature* 543(7644):199–204
2. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, et al. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25(18):1915–27
3. Molyneaux BJ, Goff LA, Brettler AC, Chen H-H, Brown JR, et al. 2015. DeCoN: genome-wide analysis of in vivo transcriptional dynamics during pyramidal neuron fate selection in neocortex. *Neuron* 85(2):275–88
4. Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, et al. 2012. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* 22(3):577–91
5. Ulitsky I, Bartel DP. 2013. lincRNAs: genomics, evolution, and mechanisms. *Cell* 154(1):26–46
6. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22(9):1775–89
7. Du Z, Fei T, Verhaak RGW, Su Z, Zhang Y, et al. 2013. Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nat. Struct. Mol. Biol.* 20(7):908–13
8. Kelley D, Rinn J. 2012. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol.* 13(11):R107
9. Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, et al. 2008. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* 18(9):1433–45
10. Sarropoulos I, Marin R, Cardoso-Moreira M, Kaessmann H. 2019. Developmental dynamics of lncRNAs across mammalian organs and species. *Nature* 571:510–14
11. Sauvageau M, Goff LA, Lodato S, Bonev B, Groff AF, et al. 2013. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife* 2:e01749
12. Grote P, Wittler L, Hendrix D, Koch F, Währisch S, et al. 2013. The tissue-specific lncRNA *Fendrr* is an essential regulator of heart and body wall development in the mouse. *Dev. Cell* 24(2):206–14

13. Huarte M, Rinn JL. 2010. Large non-coding RNAs: Missing links in cancer? *Hum. Mol. Genet.* 19(R2):R152–61
14. Sun L, Goff LA, Trapnell C, Alexander R, Lo KA, et al. 2013. Long noncoding RNAs regulate adipogenesis. *PNAS* 110(9):3387–92
15. Hacisuleyman E, Goff LA, Trapnell C, Williams A, Henao-Mejia J, et al. 2014. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat. Struct. Mol. Biol.* 21(2):198–206
16. Loewer S, Cabili MN, Guttman M, Loh Y-H, Thomas K, et al. 2010. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat. Genet.* 42(12):1113–17
17. Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, et al. 2011. LincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477(7364):295–300
18. Henry WS, Hendrickson DG, Beca F, Glass B, Lindahl-Allen M, et al. 2016. LINC00520 is induced by Src, STAT3, and PI3K and plays a functional role in breast cancer. *Oncotarget* 7(50):81981–94
19. Atianand MK, Hu W, Satpathy AT, Shen Y, Ricci EP, et al. 2016. A long noncoding RNA lincRNA-EPS acts as a transcriptional brake to restrain inflammation. *Cell* 165(7):1672–85
20. Marín-Béjar O, Marchese FP, Athie A, Sánchez Y, González J, et al. 2013. Pint lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2. *Genome Biol.* 14(9):R104
21. Mercer TR, Mattick JS. 2013. Structure and function of long noncoding RNAs in epigenetic regulation. *Nature* 20(3):300–7
22. Marchese FP, Grossi E, Marín-Béjar O, Bharti SK, Raimondi I, et al. 2016. A long noncoding RNA regulates sister chromatid cohesion. *Mol. Cell* 63(3):397–407
23. Arun G, Diermeier S, Akerman M, Chang K-C, Wilkinson JE, et al. 2016. Differentiation of mammary tumors and reduction in metastasis upon *Malat1* lncRNA loss. *Genes Dev.* 30(1):34–51
24. Gutschner T, Hämmerle M, Eissmann M, Hsu J, Kim Y, et al. 2013. The noncoding RNA *MALAT1* is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res.* 73(3):1180–89
25. Mao YS, Sunwoo H, Zhang Bin, Spector DL. 2010. Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat. Cell Biol.* 13(1):95–101
26. Gong C, Li Z, Ramanujan K, Clay I, Zhang Y, et al. 2015. A long non-coding RNA, *LncMyoD*, regulates skeletal muscle differentiation by blocking IMP2-mediated mRNA translation. *Dev. Cell* 34(2):181–91
27. Morris KV, Mattick JS. 2014. The rise of regulatory RNA. *Nat. Rev. Genet.* 15(6):423–37
28. Lin N, Chang K-Y, Li Z, Gates K, Rana ZA, et al. 2014. An evolutionarily conserved long noncoding RNA *TUNA* controls pluripotency and neural lineage commitment. *Mol. Cell* 53(6):1005–19
29. Lee S, Kopp F, Chang T-C, Sataluri A, Chen B, et al. 2016. Noncoding RNA *NORAD* regulates genomic stability by sequestering PUMILIO proteins. *Cell* 164(1–2):69–80
30. Leucci E, Vendramin R, Spinazzi M, Laurette P, Fiers M, et al. 2016. Melanoma addiction to the long non-coding RNA *SAMMSON*. *Nature* 531(7595):518–22
31. Liu SJ, Horlbeck MA, Cho SW, Birk HS, Malatesta M, et al. 2017. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* 355(6320):aah7111
32. Andersen RE, Hong SJ, Lim JJ, Cui M, Harpur BA, et al. 2019. The long noncoding RNA *Pnky* is a *trans*-acting regulator of cortical development in vivo. *Dev. Cell* 49(4):632–37
33. Rinn J, Guttman M. 2014. RNA and dynamic nuclear organization. *Science* 345(6202):1240–41
34. Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, et al. 2013. The *Xist* lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 341(6147):1237973
35. Lee JT. 2009. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev.* 23(16):1831–42
36. Briggs JA, Wolvetang EJ, Mattick JS, Rinn JL, Barry G. 2015. Mechanisms of long non-coding RNAs in mammalian nervous system development, plasticity, disease, and evolution. *Neuron* 88(5):861–77

37. Abe Y, Kikuchi A, Kobayashi S, Wakusawa K, Tanaka S, et al. 2014. Xq26.1–26.2 gain identified on array comparative genomic hybridization in bilateral periventricular nodular heterotopia with overlying polymicrogyria. *Dev. Med. Child Neurol.* 56(12):1221–24
38. Maass PG, Rump A, Schulz H, Stricker S, Schulze L, et al. 2012. A misplaced lncRNA causes brachydactyly in humans. *J. Clin. Investig.* 122(11):3990–4002
39. Sánchez Y, Huarte M. 2013. Long non-coding RNAs: challenges for diagnosis and therapies. *Nucleic Acid Ther.* 23(1):15–20
40. Guttman M, Rinn JL. 2012. Modular regulatory principles of large non-coding RNAs. *Nature* 482(7385):339–46
41. Quinn JJ, Chang HY. 2016. Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.* 17(1):47–62
42. Melé M, Mattioli K, Mallard W, Shechner DM, Gerhardinger C, Rinn JL. 2017. Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Res.* 27:27–37
43. Ponjavic J, Ponting CP, Lunter G. 2007. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.* 17(5):556–65
44. Guttman M, Amit I, Garber M, French C, Lin MF, et al. 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458(7235):223–27
45. Schlackow M, Nojima T, Gomes T, Dhir A, Carmo-Fonseca M, Proudfoot NJ. 2017. Distinctive patterns of transcription and RNA processing for human lincRNAs. *Mol. Cell* 65(1):25–38
46. Melé M, Mattioli K, Mallard W, Shechner DM, Gerhardinger C, Rinn JL. 2017. Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Res.* 27(1):27–37
47. Mattioli K, Volders P-J, Gerhardinger C, Lee JC, Maass PG, et al. 2019. High-throughput functional analysis of lncRNA core promoters elucidates rules governing tissue specificity. *Genome Res.* 29(3):344–55
48. Mukherjee N, Calviello L, Hirsekorn A, de Pretis S, Pelizzola M, Ohler U. 2016. Integrative classification of human coding and noncoding genes through RNA metabolism profiles. *Nat. Struct. Mol. Biol.* 24(1):86–96
49. Tani H, Mizutani R, Salam KA, Tano K, Ijiri K, et al. 2012. Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res.* 22(5):947–56
50. Clark MB, Johnston RL, Inostroza-Ponta M, Fox AH, Fortini E, et al. 2012. Genome-wide analysis of long noncoding RNA stability. *Genome Res.* 22(5):885–98
51. Yin Y, Lu JY, Zhang X, Shao W, Xu Y, et al. 2018. U1 snRNP regulates chromatin retention of noncoding RNAs. bioRxiv 310433. **https://doi.org/10.1101/310433**
52. Zuckerman B, Ulitsky I. 2019. Predictive models of subcellular localization of long RNAs. *RNA* 25(5):557–72
53. Wang KC, Chang HY. 2011. Molecular mechanisms of long noncoding RNAs. *Mol. Cell* 43(6):904–14
54. Ponting CP, Oliver PL, Reik W. 2009. Evolution and functions of long noncoding RNAs. *Cell* 136(4):629–41
55. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, et al. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* 38(6):626–35
56. Ulitsky I. 2016. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat. Rev. Genet.* 17(10):601–14
57. Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I. 2015. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep.* 11(7):1110–22
58. Podlevsky JD, Chen JJ-L. 2016. Evolutionary perspectives of telomerase RNA structure and function. *RNA Biol.* 13(8):720–32
59. Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, et al. 2015. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* 519(7544):486–90
60. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS. 2014. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 505(7485):701–5

61. Zubradt M, Gupta P, Persad S, Lambowitz AM, Weissman JS, Rouskin S. 2017. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods* 14(1):75–82
62. Smola MJ, Calabrese JM, Weeks KM. 2015. Detection of RNA-protein interactions in living cells with SHAPE. *Biochemistry* 54(46):6867–75
63. Lu Z, Zhang QC, Lee B, Flynn RA, Smith MA, et al. 2016. RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* 165(5):1267–79
64. Aw JGA, Shen Y, Wilm A, Sun M, Lim XN, et al. 2016. In vivo mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol. Cell* 62(4):603–17
65. Sharma E, Sterne-Weiler T, O'Hanlon D, Blencowe BJ. 2016. Global mapping of human RNA-RNA interactions. *Mol. Cell* 62(4):618–26
66. Sahakyan A, Yang Y, Plath K. 2018. The role of *Xist* in X-chromosome dosage compensation. *Trends Cell Biol*. 28:999–1013
67. Balaton BP, Dixon-McDougall T, Peeters SB, Brown CJ. 2018. The eXceptional nature of the X chromosome. *Hum. Mol. Genet*. 27(R2):242–49
68. da Rocha ST, Heard E. 2017. Novel players in X inactivation: insights into *Xist*-mediated gene silencing and chromosome conformation. *Nature* 24(3):197–204
69. Furlan G, Rougeulle C. 2016. Function and evolution of the long noncoding RNA circuitry orchestrating X-chromosome inactivation in mammals. *Wiley Interdiscip. Rev. RNA* 7(5):702–22
70. Jégu T, Aeby E, Lee JT. 2017. The X chromosome in space. *Nat. Rev. Genet.* 18(6):377–89
71. Monfort A, Wutz A. 2017. Progress in understanding the molecular mechanism of *Xist* RNA function through genetics. *Philos. Trans. R. Soc. A* 372(1733):20160368
72. Zappulla DC, Cech TR. 2004. Yeast telomerase RNA: a flexible scaffold for protein subunits. *PNAS* 101(27):10024–29
73. Wutz A, Rasmussen TP, Jaenisch R. 2002. Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nat. Genet*. 30(2):167–74
74. Pintacuda G, Wei G, Roustan C, Kirmizitas BA, Solcan N, et al. 2017. hnRNPK recruits PCGF3/5-PRC1 to the Xist RNA B-repeat to establish polycomb-mediated chromosomal silencing. *Mol. Cell* 68(5):955–69.e10
75. Colognori D, Sunwoo H, Kriz AJ, Wang C-Y, Lee JT. 2019. *Xist* deletional analysis reveals an interdependency between *Xist* RNA and Polycomb complexes for spreading along the inactive X. *Mol. Cell* 74(1):101–10
76. Ridings-Figueroa R, Stewart ER, Nesterova TB, Coker H, Pintacuda G, et al. 2017. The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. *Genes Dev*. 31(9):876–88
77. Sunwoo H, Colognori D, Froberg JE, Jeon Y, Lee JT. 2017. Repeat E anchors Xist RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1). *PNAS* 114(40):10654–59
78. Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, et al. 2015. Systematic discovery of Xist RNA binding proteins. *Cell* 161(2):404–16
79. McHugh CA, Chen C-K, Chow A, Surka CF, Tran C, et al. 2015. The *Xist* lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* 521:232–36
80. Patil DP, Chen C-K, Pickering BF, Chow A, Jackson C, et al. 2016. M^6A RNA methylation promotes *XIST*-mediated transcriptional repression. *Nature* 537(7620):369–73
81. Somarowthu S, Legiewicz M, Chillón I, Marcia M, Liu F, Pyle AM. 2015. HOTAIR forms an intricate and modular secondary structure. *Mol. Cell* 58(2):353–61
82. Hawkes EJ, Hennelly SP, Novikova IV, Irwin JA, Dean C, Sanbonmatsu KY. 2016. COOLAIR antisense RNAs form evolutionarily conserved elaborate secondary structures. *Cell Rep*. 16(12):3087–96
83. Xue Z, Hennelly S, Doyle B, Gulati AA, Novikova IV, et al. 2016. A G-rich motif in the lncRNA *Braveheart* interacts with a zinc-finger transcription factor to specify the cardiovascular lineage. *Mol. Cell* 64(1):37–50
84. Ilik IA, Quinn JJ, Georgiev P, Tavares-Cadete F, Maticzka D, et al. 2013. Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in *Drosophila*. *Mol. Cell* 51(2):156–73

85. Quinn JJ, Zhang QC, Georgiev P, Ilik IA, Akhtar A, Chang HY. 2016. Rapid evolutionary turnover underlies conserved lncRNA-genome interactions. *Genes Dev*. 30(2):191–207
86. Cabili MN, Dunagin MC, McClanahan PD, Biaesch A, Padovan-Merhar O, et al. 2015. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol*. 16(1):20
87. Zhang B, Gunawardane L, Niazi F, Jahanbani F, Chen X, Valadkhan S. 2014. A novel RNA motif mediates the strict nuclear localization of a long non-coding RNA. *Mol. Cell. Biol.* 34:2318–29
88. Shukla CJ, McCorkindale AL, Gerhardinger C, Korthauer KD, Cabili MN, et al. 2018. High-throughput identification of RNA nuclear enrichment sequences. *EMBO J*. 37(6):e98452
89. Lubelsky Y, Ulitsky I. 2018. Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. *Nature* 555(7694):107–11
90. Chen C-K, Blanco M, Jackson C, Aznauryan E, Ollikainen N, et al. 2016. Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science* 354:468–72
91. Miyagawa R, Tano K, Mizuno R, Nakamura Y, Ijiri K, et al. 2012. Identification of cis- and trans-acting factors involved in the localization of MALAT-1 noncoding RNA to nuclear speckles. *RNA* 18(4):738–51
92. Martin L, Meier M, Lyons SM, Sit RV, Marzluff WF, et al. 2012. Systematic reconstruction of RNA functional motifs with high-throughput microfluidics. *Nat. Methods* 9(12):1192–94
93. Buenrostro JD, Araya CL, Chircus LM, Layton CJ, Chang HY, et al. 2014. Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat. Biotechnol.* 32(6):562–68
94. Dietrich JS. 1996. Edward B. Lewis, Nobel Laureate 1995: A classical geneticist is recognized for his insights into the process by which genes control the development of an organism from egg to adult. *Calif. Inst. Technol. Eng. Sci.* LIX(1):2–7
95. Birnbaum RY, Clowney EJ, Agamy O, Kim MJ, Zhao J, et al. 2012. Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res*. 22(6):1059–68
96. Groff AF, Sanchez-Gomez DB, Soruco MML, Gerhardinger C, Barutcu AR, et al. 2016. In vivo characterization of *Linc-p21* reveals functional *cis*-regulatory DNA elements. *Cell Rep*. 16(8):2178–86
97. Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, et al. 2010. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 142(3):409–19
98. Dimitrova N, Zamudio JR, Jong RM, Soukup D, Resnick R, et al. 2014. *LincRNA-p21* activates *p21* in *cis* to promote Polycomb target gene expression and to enforce the G1/S checkpoint. *Mol. Cell* 54:777–90
99. Lai K-MV, Gong G, Atanasio A, Rojas J, Quispe J, et al. 2015. Diverse phenotypes and specific transcription patterns in twenty mouse lines with ablated LincRNAs. *PLOS ONE* 10(4):e0125522
100. Groff AF, Barutcu AR, Lewandowski JP, Rinn JL. 2018. Enhancers in the *Peril* lincRNA locus regulate distant but not local genes. *Genome Biol*. 19(1):219
101. Yin Y, Yan P, Lu J, Song G, Zhu Y, et al. 2015. Opposing roles for the lncRNA *Haunt* and its genomic locus in regulating *HOXA* gene activation during embryonic stem cell differentiation. *Cell Stem Cell* 16(5):504–16
102. Maamar H, Cabili MN, Rinn J, Raj A. 2013. *linc-HOXA1* is a noncoding RNA that represses *Hoxa1* transcription in *cis*. *Genes Dev*. 27(11):1260–71
103. Melé M, Rinn JL. 2016. "Cat's Cradling" the 3D genome by the act of lncRNA transcription. *Mol. Cell* 62(5):657–64
104. Schmitt S. 2005. Intergenic transcription through a Polycomb group response element counteracts silencing. *Genes Dev*. 19(6):697–708
105. Nozawa R-S, Boteva L, Soares DC, Naughton C, Dun AR, et al. 2017. SAF-A regulates interphase chromosome structure through oligomerization with chromatin-associated RNAs. *Cell* 169(7):1214–18
106. Vance KW, Ponting CP. 2014. Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends Genet*. 30:348–55
107. Paralkar VR, Taborda CC, Huang P, Yao Y, Kossenkov AV, et al. 2016. Unlinking an lncRNA from its associated *cis* element. *Mol. Cell* 62(1):104–10
108. Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, et al. 2016. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* 539(7629):452–55

109. Goff L, Groff AF, Sauvageau M, Trayes-Gibson Z, Sanchez-Gomez DB, et al. 2015. Spatiotemporal expression and transcriptional perturbations by long noncoding RNAs in the mouse brain. *PNAS* 112:6855–62
110. Zakany J, Darbellay F, Mascrez B, Necsulea A, Duboule D. 2017. Control of growth and gut maturation by *HoxD* genes and the associated lncRNA *Haglr*. *PNAS* 114(44):E9290–99
111. Bonora G, Deng X, Fang H, Ramani V, Qiu R, et al. 2018. Orientation-dependent Dxz4 contacts shape the 3D structure of the inactive X chromosome. *Nat. Commun.* 9(1):1445
112. Barutcu AR, Maass PG, Lewandowski JP, Weiner CL, Rinn JL. 2018. A TAD boundary is preserved upon deletion of the CTCF-rich *Firre* locus. *Nat. Commun.* 9(1):1444
113. Cho SW, Xu J, Sun R, Mumbach MR, Carter AC, et al. 2018. Promoter of lncRNA gene *PVT1* is a tumor-suppressor DNA boundary element. *Cell* 173(6):1398–412.e22
114. Fulco CP, Munschauer M, Anyoha R, Munson G, Grossman SR, et al. 2016. Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* 354(6313):769–73
115. Tseng Y-Y, Moriarity BS, Gong W, Akiyama R, Tiwari A, et al. 2014. *PVT1* dependence in cancer with *MYC* copy-number increase. *Nature* 512:82–86
116. Lam MTY, Li W, Rosenfeld MG, Glass CK. 2014. Enhancer RNAs and regulated transcriptional programs. *Trends Biochem. Sci.* 39(4):170–82
117. Lai F, Shiekhattar R. 2014. Enhancer RNAs: the new molecules of transcription. *Curr. Opin. Genet. Dev.* 25:38–42
118. Lewandowski JP, Lee JC, Hwang T, Sunwoo H, Goldstein JM, et al. 2019. The *Firre* locus produces a *trans*-acting RNA molecule that functions in hematopoiesis. *Nat. Commun.* 10:5137
119. Fang H, Bonora G, Lewandowski JP, Thakur J, Filippova GN, et al. 2019. Trans- and cis-acting effects of the lncRNA *Firre* on epigenetic and structural features of the inactive X chromosome. bioRxiv 687236. **https://doi.org/10.1101/687236**
120. Froberg JE, Pinter SF, Kriz AJ, Jégu T, Lee JT. 2018. Megadomains and superloops form dynamically but are dispensable for X-chromosome inactivation and gene escape. *Nat. Commun.* 9(1):5004
121. Froberg JE, Yang L, Lee JT. 2013. Guided by RNAs: X-inactivation as a model for lncRNA function. *J. Mol. Biol.* 425(19):3698–706
122. Andergassen D, Smith ZD, Lewandowski JP, Gerhardinger C, Meissner A, Rinn JL. 2019. In vivo *Firre* and *Dxz4* deletion elucidates roles for autosomal gene regulation. *eLife* 8:e47214
123. Grote P, Herrmann BG. 2013. The long non-coding RNA *Fendrr* links epigenetic control mechanisms to gene regulatory networks in mammalian embryogenesis. *RNA Biol.* 10(10):1579–85
124. Szafranski P, Dharmadhikari AV, Brosens E, Gurha P, Kolodziejska KE, et al. 2013. Small noncoding differentially methylated copy-number variants, including lncRNA genes, cause a lethal lung developmental disorder. *Genome Res.* 23(1):23–33
125. Atianand MK, Hu W, Satpathy AT, Shen Y, Ricci EP, et al. 2016. A long noncoding RNA lincRNA-EPS acts as a transcriptional brake to restrain inflammation. *Cell* 165(7):1672–85
126. Elling R, Robinson EK, Shapleigh B, Liapis SC, Covarrubias S, et al. 2018. Genetic models reveal *cis* and *trans* immune-regulatory activities for lincRNA-Cox2. *Cell Rep.* 25(6):1511–16
127. Wang P, Xu J, Wang Y, Cao X. 2017. An interferon-independent lncRNA promotes viral replication by modulating cellular metabolism. *Science* 358(6366):1051–55
128. Kopp F, Elguindy MM, Yalvac ME, Zhang H, Chen B, et al. 2019. PUMILIO hyperactivity drives premature aging of Norad-deficient mice. *eLife* 8:e42650
129. Elguindy MM, Kopp F, Goodarzi M, Rehfeld F, Thomas A, et al. 2019. PUMILIO, but not RBMX, binding is required for regulation of genomic stability by noncoding RNA *NORAD*. *eLife* 8:e48625
130. Ang CE, Ma Q, Wapinski OL, Fan S, Flynn RA, et al. 2019. The novel lncRNA *lnc-NR2F1* is pro-neurogenic and mutated in human neurodevelopmental disorders. *eLife* 8:e41770
131. Bassett AR, Akhtar A, Barlow DP, Bird AP, Brockdorff N, et al. 2014. Considerations when investigating lncRNA function in vivo. *eLife* 3:e03058
132. Goff LA, Rinn JL. 2015. Linking RNA biology to lncRNAs. *Genome Res.* 25(10):1456–65
133. Meng L, Ward AJ, Chun S, Bennett CF, Beaudet AL, Rigo F. 2015. Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. *Nature* 518(7539):409–12

134. Kim J, Piao H-L, Kim B-J, Yao F, Han Z, et al. 2018. Long noncoding RNA MALAT1 suppresses breast cancer metastasis. *Nat. Genet.* 50(12):1705–15
135. Arun G, Spector DL. 2019. MALAT1 long non-coding RNA and breast cancer. *RNA Biol.* 16(6):860–63
136. Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, et al. 2015. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517(7536):583–88
137. Zhu S, Li W, Liu J, Chen C-H, Liao Q, et al. 2016. Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. *Nat. Biotechnol.* 34(12):1279–86
138. Yeo NC, Chavez A, Lance-Byrne A, Chan Y, Menn D, et al. 2018. An enhanced CRISPR repressor for targeted mammalian gene regulation. *Nat. Methods* 15:611–16
139. Cox DBT, Gootenberg JS, Abudayyeh OO, Franklin B, Kellner MJ, et al. 2017. RNA editing with CRISPR-Cas13. *Science* 358(6366):1019–27
140. Konermann S, Lotfy P, Brideau NJ, Oki J, Shokhirev MN, Hsu PD. 2018. Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. *Cell* 173(3):665–76.e14
141. Niederer RO, Hass EP, Zappulla DC. 2017. Long noncoding RNAs in the yeast *S. cerevisiae*. *Adv. Exp. Med. Biol.* 1008(5705):119–32