# lncRNAs in development and differentiation: from sequence motifs to functional characterization

Florian Constanty, Alena Shkumatava

© 2021. Published by The Company of Biologists Ltd | Development (2021) 148, dev182741. doi:10.1242/dev.182741

REVIEW

# lncRNAs in development and differentiation: from sequence motifs to functional characterization

**Florian Constanty and Alena Shkumatava***


## ABSTRACT

The number of long noncoding RNAs (lncRNAs) with characterized developmental and cellular functions continues to increase, but our understanding of the molecular mechanisms underlying lncRNA functions, and how they are dictated by RNA sequences, remains limited. Relatively short, conserved sequence motifs embedded in lncRNA transcripts are often important determinants of lncRNA localization, stability and interactions. Identifying such RNA motifs remains challenging due to the substantial length of lncRNA transcripts and the rapid evolutionary turnover of lncRNA sequences. Nevertheless, the recent discovery of specific RNA elements, together with their experimental interrogation, has enabled the first step in classifying heterogeneous lncRNAs into sub-groups with similar molecular mechanisms and functions. In this Review, we focus on lncRNAs with roles in development, cell differentiation and normal physiology in vertebrates, and we discuss the sequence elements defining their functions. We also summarize progress on the discovery of regulatory RNA sequence elements, as well as their molecular functions and interaction partners.


**KEY WORDS: RNA regulatory motifs, RNA-protein complexes, Cell differentiation, Development, In vivo functions, LncRNAs**

## Introduction

Thousands of genomic loci are transcribed into so-called long noncoding RNAs (lncRNAs) (Rinn and Chang, 2012). A fraction of these has been demonstrated to regulate a range of important biological processes, such as development and cell differentiation (Mallory and Shkumatava, 2015; Perry and Ulitsky, 2016; Ulitsky and Bartel, 2013). LncRNAs have also been linked to human diseases, including nearly every major cancer type as well as various neurological disorders (Bian and Sun, 2011; Esposito et al., 2019; Gutschner and Diederichs, 2012; Huarte, 2015; Li et al., 2018). Although the annotation of lncRNAs and the identification of their biological functions has progressed over the past decade, lncRNAs are still poorly understood at the molecular level, partially due to the diversity of their modes of action and a lack of suitable molecular tools to dissect them.

Three sub-groups of lncRNAs can be distinguished based on their molecular mechanisms of action: (1) lncRNA loci that contain enhancers regulating gene expression; (2) lncRNA loci for which the act of transcription but not the transcript per se is important for regulating neighbouring genes; and (3) lncRNA loci that carry out their cellular functions via lncRNA transcripts that interact with DNA, other RNAs and proteins (Marchese et al., 2017; Wang and Chang, 2011). Notably, some lncRNAs may belong to more than one of these subgroups. The lncRNAs in the final group share biogenesis pathways and post-transcriptional features with mRNAs, i.e. they are transcribed by RNA polymerase II and are capped, spliced and polyadenylated (Cabili et al., 2011; Guttman and Rinn, 2012). For these lncRNAs, the linear transcript sequence often defines function. Comparative analyses of lncRNAs in different vertebrate species has shown that lncRNA sequences undergo rapid evolutionarily turnover, resulting in their overall poor conservation (Hezroni et al., 2015; Kutter et al., 2012; Necsulea et al., 2014; Ulitsky et al., 2011; Washietl et al., 2014). Nonetheless, signatures of primary sequence conservation have been detected in ∼100 lncRNAs conserved from mammals to fish and in ∼1000 lncRNAs conserved among mammals (Hezroni et al., 2015). For the majority of conserved lncRNAs, only short stretches of sequence are retained (Hezroni et al., 2015; Ulitsky et al., 2011). These short sequences may represent functional elements of lncRNA transcripts that facilitate interactions with other RNAs, proteins or genomic loci.

In addition to their linear RNA sequences, the secondary and tertiary structures of lncRNA transcripts have been proposed to be functional determinants (Diederichs, 2014; Guttman and Rinn, 2012; Wutz et al., 2002). Several functionally conserved RNA structure examples support this idea (Holmes et al., 2020; Uroda et al., 2019; Zhang et al., 2010). Moreover, recent development of new technologies to determine RNA structures *in vivo* have demonstrated that RNA structures can be engaged in important long-range interactions (Lu et al., 2016; Ziv et al., 2018). However, it is unclear how prevalent the impact of RNA structure on lncRNA functionality is (Rivas et al., 2017). Furthermore, the unbiased analysis of sequence and context preferences for multiple human RNA-binding proteins (RBPs), which are important drivers of lncRNA function, has demonstrated that RBPs bind short, often low complexity, RNA motifs (Dominguez et al., 2018). RBP-binding specificity is further supported by additional features, such as local RNA secondary structure and flanking nucleotides (Dominguez et al., 2018), suggesting that the functional elements controlling lncRNAs reside in their sequence and adjacent secondary structure motifs, rather than in their global tertiary structure.

In this Review, we discuss a set of vertebrate lncRNAs with characterized developmental, cell biological and physiological functions defined by specific lncRNA sequence elements (Table 1). We discuss how such sequence elements and motifs regulate the biogenesis, localization and interactions of lncRNAs. In addition, we survey recent efforts to experimentally identify functional lncRNA motifs and highlight computational algorithms that have been developed to predict these functions based on lncRNA sequences.

## The regulation of lncRNA biogenesis and processing by RNA motifs

To generate functional noncoding transcripts, tight regulation of lncRNA biogenesis is required. Although the majority of mature

Institut Curie, PSL Research University, CNRS UMR3215, INSERM U934, Paris 75005, France.

*Author for correspondence (alena.shkumatava@curie.fr)

A.S., 0000-0001-9286-4700

Development (2021) 148, dev182741. doi:10.1242/dev.182741

**Table 1. Examples of lncRNA motifs with regulatory functions in development, cell differentiation and animal physiology**

| lncRNA | RNA motif(s) | Function | References |
|---|---|---|---|
| *NEAT1_2* | Two U-rich and one A-rich motifs | 3′ END alternative RNAse P processing | Brown et al. (2012); Sunwoo et al. (2009); Wilusz et al. (2012) |
| *MALAT1* | Two U-rich and one A-rich motifs | 3′ END alternative RNAse P processing | Brown et al. (2012); Wilusz et al. (2008, 2012) |
| *Firre* | Local repeating RNA domains | Nuclear retention signal bound by hnRNPU | Hacisuleyman et al. (2014, 2016) |
| *MALAT1* | RNSP1 binding sites | Paraspeckle localization | Miyagawa et al. (2012) |
| *MEG3* | Nuclear retention element | Nuclear retention signal involving U1 snRNP | Azam et al. (2019) |
| *BORG* | AGCCC motifs | Nuclear retention signal | Zhang et al. (2014) |
| *CTN-RNA* | A-to-I editing site | I-edited RNA as a nuclear retention signal bound by paraspeckle p54$^{nrb}$ protein | Prasanth et al. (2005) |
| *libra*/*NREP* | Near-perfect miR-29 binding site | Regulation of miR-29b levels by TDMD | Bitetti et al. (2018) |
| *Cyrano* | Near-perfect miR-7 binding site | Regulation of miR-7 levels by TDMD | Kleaveland et al. (2018); Ulitsky et al. (2011) |
| *NORAD* | UGURUAUA | Sequestration of PUM proteins | Lee et al. (2016); Tichon et al. (2016) |
| *Megamind* | A conserved region of ∼200bp | RNP complex formation | Lin et al. (2014); Ulitsky et al. (2011) |
| *Braveheart* | G-rich motif (AGIL) | Interaction with the transcription factor CNBP | Xue et al. (2016) |
| *GAS5* | GR mimic-binding site | Decoy of the GRs | Hudson et al. (2014) |
| *PINT* | Conserved PRC2 binding site | Downregulation of pro-invasion gene expression | Marin-Bejar et al. (2013); Marín-Béjar et al. (2017); Sauvageau et al. (2013) |

lncRNA transcripts are generated by the same biogenesis pathways as those regulating mRNAs, two particular lncRNA transcripts – *MALAT1* (*metastasis-associated lung adenocarcinoma transcript 1*) and *NEAT1* (*nuclear enriched abundant transcript 1*) – are processed in a manner that involves conserved motifs (Wilusz et al., 2012) (Fig. 1A).

Both *MALAT1* and *NEAT1_2* (the long isoform of *NEAT1*) are structural components of nuclear bodies, with *MALAT1* localizing to speckles and *NEAT1_2* to paraspeckles (Clemson et al., 2009; Hutchinson et al., 2007; Nakagawa et al., 2011; Sasaki et al., 2009; Sunwoo et al., 2009). Whereas *NEAT1_2* is essential for paraspeckle formation (Clemson et al., 2009; Mao et al., 2011; Naganuma and Hirose, 2013; Naganuma et al., 2012; Sasaki et al., 2009), depletion of *MALAT1* does not noticeably affect speckles (Nakagawa et al., 2012; Zhang et al., 2012). At the organismal level, *Neat1* is required for normal mammary gland development, lactation and corpus luteum formation during the establishment of pregnancy (Nakagawa et al., 2014; Standaert et al., 2014). By contrast, *Malat1* does not appear to be required for normal development, as mouse and zebrafish null alleles of *Malat1* do not exhibit any apparent morphological defects (Eissmann et al., 2012; Lavalou et al., 2019; Nakagawa et al., 2012; Zhang et al., 2012). However, in-depth analysis of one of the mouse null alleles has shown retinal vascularization defects (Michalik et al., 2014). In addition, there is growing evidence that *Malat1* plays an essential role in disease, affecting cancer progression and metastasis (Arun and Spector, 2019).

During their biogenesis, *MALAT1* and *NEAT1* transcripts are processed by an unusual RNase P cleavage step directed by structured tRNA-like motifs (Wilusz et al., 2012). Upon RNase P cleavage, *MALAT1*, which is conserved from fish to human (Ulitsky et al., 2011), is stabilized by a highly conserved triple helical structure located immediately upstream of the RNase P cleavage site; this event prevents exonucleolytic degradation from the 3′ end of the transcript (Brown et al., 2012; Wilusz et al., 2012) (Fig. 1A). The mammalian-specific *NEAT1* locus encodes two lncRNA isoforms generated from the same promoter. Although the short isoform, termed *NEAT1_1* (3.7 kb in humans), is produced by canonical cleavage and polyadenylation using an upstream polyadenylation signal, the long isoform, termed *NEAT1_2* (23 kb in humans), is processed by RNase P cleavage and stabilized by a triple helical structure, similar to that regulating *MALAT1* biogenesis (Fig. 1A). In both cases, the 3′-end stabilizing triple helical structure is generated by conserved motifs of ∼10 nucleotides (nt) in length – two U-rich motifs and an A-rich tract (Brown et al., 2014, 2012; Wilusz et al., 2008, 2012) (Fig. 1A). This biogenesis mechanism remains a unique feature of *MALAT1* and *NEAT1*, as no additional lncRNA transcripts processed by RNase P cleavage and harbouring similar triple-helix forming motifs have been identified so far.

## Motifs and sequences that control the subcellular localization of lncRNA transcripts

The subcellular localization of lncRNAs is also tightly regulated and is an important determinant of lncRNA function (Chen, 2016; Guo et al., 2020). In the nucleus, lncRNAs regulate chromatin organization, transcription, RNA maturation and nuclear protein activity (Sun et al., 2018; Wang and Chang, 2011). In the cytoplasm, by contrast, lncRNAs interact with microRNAs (miRNAs) and regulate translation efficiency and cytoplasmic protein activity (Noh et al., 2018; Paraskevopoulou and Hatzigeorgiou, 2016). Notably, lncRNAs tend to be less enriched in the cytoplasm compared with mRNAs (Derrien et al., 2012; Mukherjee et al., 2017). Specifically, for lncRNAs whose functions are linked to activities in the nucleus, repression of their export to the cytoplasm is required. Many nuclear lncRNAs have been studied in detail and specific RNA elements that facilitate their binding to proteins defining nuclear localization have been determined. Below, we highlight examples that illustrate the variety of different RNA elements and molecular mechanisms directing the nuclear enrichment of lncRNAs.

The nuclear localization of *MALAT1* can be explained in part by the single-exon structure of its transcript, which does not engage with exon junction complexes promoting nuclear export (Hutchinson et al., 2007). In addition, two distinct sequence elements have been found to promote the specific enrichment of *MALAT1* in nuclear speckles (Miyagawa et al., 2012). These nuclear retention sequences of ∼600 nt are bound *in vitro* by the nuclear speckle protein RNSP1. Depletion of RNSP1 and two other key nuclear speckle proteins, SRM160 and IBP160, leads to diffusion of *MALAT1* into the nucleoplasm, suggesting that these proteins may drive the nuclear retention of *MALAT1*. However, sequence and structure conservation analyses did not detect any common motifs within the two *MALAT1* nuclear localization elements (Miyagawa et al., 2012).

In contrast to *MALAT1*, the lncRNA *Firre* (*functional intergenic repeating RNA element*) is a spliced and polyadenylated nuclear transcript, the knockout of which leads to cell-specific defects in

Development (2021) 148, dev182741. doi:10.1242/dev.182741

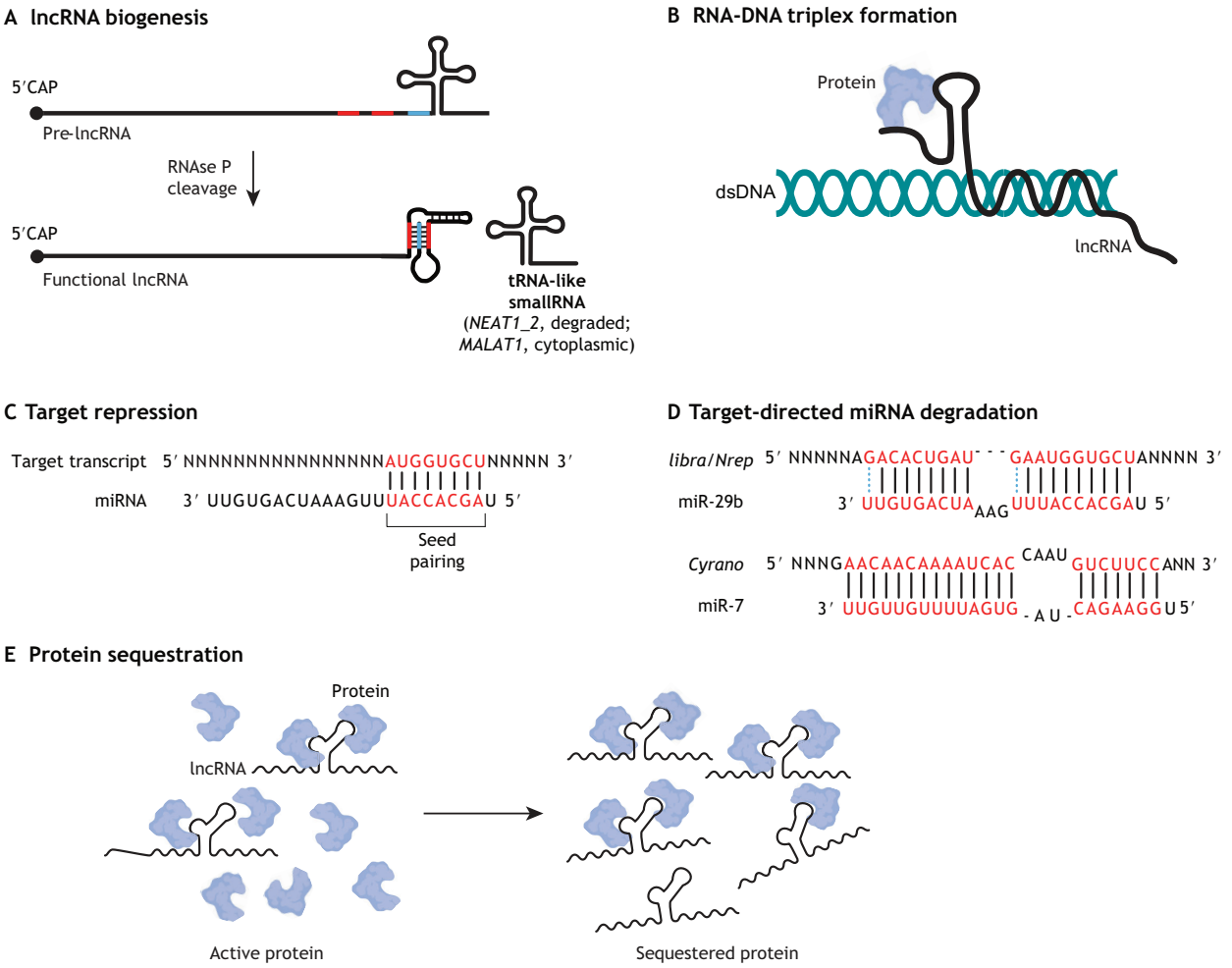


**Fig. 1. Functions of conserved lncRNA motifs.** (A) Two U-rich motifs (red) and one A-rich motif (blue) are essential for the RNAse P-mediated cleavage and processing of the lncRNAs *NEAT1_2* and *MALAT1* (black). This event is also directed by structured tRNA-like motifs. (B) lncRNAs (black) can also act as guides during the formation of RNA-DNA triplex structures that interact with proteins (blue). (C) Canonical animal miRNA-binding sites (with partial miRNA-target RNA pairing complementarity) can regulate post-transcriptional target repression. (D) Nearly perfect miRNA-binding sites regulate miRNA decay through target-directed miRNA degradation (TDMD). The vertical dotted lines represent wobble pairing. (E) lncRNAs (black) can also sequester proteins (blue) to inhibit their functionality.

hematopoietic populations in mice (Lewandowski et al., 2019). Although *Firre* is syntenically conserved in almost all eutherians (Hezroni et al., 2015), its sequence comprising multiple repetitive elements of ∼150 nt, termed repeating RNA domains (RRDs), has diverged between primate and rodent lineages (Hacisuleyman et al., 2016). Despite this divergence in sequence evolution resulting in limited sequence similarity between primate and rodent RRD elements, orthologous RRDs share the same function and act as strong nuclear retention signals (Hacisuleyman et al., 2014, 2016). RRDs drive nuclear retention of the *Firre* transcript and are sufficient to localize an otherwise cytoplasmic mRNA to the nucleus when added to the 3′ end of this mRNA (Hacisuleyman et al., 2016). Mechanistically, it has been proposed that RRD elements interact with the nuclear matrix protein HNRNPU, which insures correct localization of RNAs to the nucleus (Hacisuleyman et al., 2014, 2016).

The nuclear retention of the lncRNA *MEG3* is also regulated by a linear sequence motif of 356 nt (Azam et al., 2019). Deletion of this nuclear retention element (NRE), which is bound by several components of U1 snRNP, effectively relocates *MEG3* reporter transcripts from the nucleus to the cytoplasm (Azam et al., 2019). By contrast, the function of *MEG3*, which acts as a tumour suppressor that stimulates the p53 pathway and controls the cell cycle, is mediated by a tertiary structure that is conserved in mammals (Uroda et al., 2019).

The spliced and polyadenylated lncRNA *BORG* regulates BMP-induced differentiation of C2C12 cells into osteoblastic cells and is exclusively localized to the nucleus (Takeda et al., 1998). The nuclear localization of *BORG* is mediated by a short AGCCC RNA motif. Although the mechanism of nuclear retention controlled by this RNA motif remains unknown, the motif was found to be present in multiple other nuclear lncRNAs as well as in protein-coding RNAs, indicating that nuclear localization motifs are shared between noncoding and coding transcripts (Zhang et al., 2014).

The 3′UTR of *CTN-RNA*, a polyadenylated transcript that localizes to paraspeckles under regular cellular conditions,

Development (2021) 148, dev182741. doi:10.1242/dev.182741

harbours another type of a NRE (Prasanth et al., 2005). This NRE consists of a ∼100 nt forward repeat and three inverted repeats originating from SINE retro-elements. Each of the inverted repeats can form a stem loop with the forward repeat. The stem loop serves as a target site for ADAR enzymes to catalyse adenosine-to-inosine (A-to-I) RNA editing (Prasanth et al., 2005). It has been proposed that the paraspeckle protein p54$^{nrb}$ interacts with *CTN-RNA* at I-edited residues, triggering retention of the transcript in the nucleus. Under cellular stress conditions, *CTN-RNA* is released from the nucleus by cleavage of its edited 3′UTR, generating an mRNA that contains a 5′UTR and a coding region. This cleaved transcript is then transported to the cytoplasm, where it is translated into the mCAT2 protein that modulates cellular uptake of L-arginine (Prasanth et al., 2005), demonstrating that tight regulation of *CTN-RNA* subcellular localization is essential during homeostatic and stress conditions.

From the few examples presented here, it is evident that the repertoire of RNA nuclear retention motifs and their associated interactors far exceeds the number of known protein nuclear localization signal motifs, presenting a major challenge in their identification. It remains to be seen whether individual transcripts adapted various nuclear retention mechanisms or whether common sets of lncRNA sequence elements execute their retention in the nucleus. High-throughput approaches (discussed below) have the potential to find common sets of motifs with similar functions and group them into mechanistically defined sub-classes.

## Unbiased identification of lncRNA subcellular localization motifs

Identification of the aforementioned subcellular localization elements has been restricted to individual lncRNA and mRNA transcripts, and has been achieved using reporter constructs and classical genetics. However, with recent advances in comparative genomics, data analysis and systematic experimental approaches, several computational and high-throughput experimental strategies have been applied to unbiasedly decipher sequence elements dictating RNA localization.

To identify RNA motifs that drive nuclear localization, several independent studies have used a tiling-based strategy. In this approach, pools of thousands of tiled oligonucleotides covering human lncRNAs and selected 3′UTRs are evaluated for their ability to retain an otherwise cytoplasmic reporter transcript in the nucleus. The Ulitsky laboratory constructed a library of ∼5500 109-mers that tiled exons of 37 human lncRNAs and 13 3′UTRs of mRNAs enriched in the nucleus. The library was cloned into the 5′ and 3′ UTRs of a cytoplasmic GFP reporter and transfected into human cells; RNA from cytoplasmic and nuclear fractions was then sequenced and analysed for enrichment of sequence motifs. A 42 nt C-rich sequence element derived from an Alu-repeat named SIRLOIN was found to drive nuclear localization of a set of lncRNAs and mRNAs by interacting with the protein HNRNPK (Lubelsky and Ulitsky, 2018). In parallel, the Rinn laboratory used a similar tiling array-reporter method to identify *cis* elements driving nuclear retention (Shukla et al., 2018). The authors generated a pool of ∼12,000 110-mers tightly tiled across 38 lncRNAs with different subcellular localization patterns. Similarly, the oligonucleotides were fused to a cytoplasmic reporter transcript and the reporter-oligonucleotide pool was transfected into human cells, which were subjected to subcellular fractionation and RNA-seq (Shukla et al., 2018). This study identified 109 RNA elements, including a 15 nt C-rich motif, that enriched the cytoplasmic reporter in the nucleus. Interestingly, the identified C-rich motif shows sequence similarities to the SIRLOIN motif (Lubelsky and Ulitsky, 2018; Shukla et al., 2018). However, single-molecule *in situ* hybridization analyses showed that this short C-rich region alone is not sufficient to drive nuclear localization and requires a larger RNA element to retain the reporter transcript in the nucleus (Shukla et al., 2018). Furthermore, the Shen team used a tiling-based approach to search for lncRNA sequence elements that can mediate chromatin enrichment and identified a short 7 nt motif that is essential for reporter RNA localization to chromatin. It has been proposed that this motif is recognized by U1 snRNP to tether lncRNAs to chromatin (Yin et al., 2020).

In addition to the systematic experimental approaches described above, several computational approaches have been developed to evaluate the subcellular localization of lncRNAs based on their sequences. One approach generated a map of transposable element (TE) insertions in lncRNAs and found that a set of evolutionarily conserved TEs functionalized to nuclear retention elements (Carlevaro-Fita et al., 2019). Specifically, this study identified four TEs that significantly correlate with lncRNA nuclear enrichment, as well as a set of GC-rich TEs that correlates with lncRNA cytoplasmic localization (Carlevaro-Fita et al., 2019). In addition, efficient splicing has been found to be one of the main predictive factors for lncRNA cytoplasmic localization, whereas inefficient splicing strongly correlates with nuclear enrichment (Zuckerman and Ulitsky, 2019). Moreover, by training computational algorithms to compare sequences of lncRNAs with known localization and detecting a link between specific sequence motifs and subcellular localization, one can predict the subcellular localization of lncRNA transcripts (Cao et al., 2018; Su et al., 2018). Another approach used nuclear and cytoplasmic fractionated RNA-seq data to develop a deep learning algorithm that predicts lncRNA subcellular localization based on lncRNA sequences (Gudenas and Wang, 2018). Taken together, the ability to computationally predict and experimentally identify functional sequence elements that drive lncRNA subcellular localization is an important step towards understanding lncRNA functions.

## Sequences that control the formation of lncRNA-DNA triplexes

Nuclear pyrimidine (T and C)-rich RNAs can interact with purine (A and G)-rich DNA via Hoogsteen-type hydrogen bonding to form RNA:DNA triple helices that may have regulatory functions (Fig. 1B) (Felsenfeld and Rich, 1957). As triplex formation requires pairing between double-stranded DNA and single-stranded RNA, the formation of the triplexes depends on the lncRNA primary sequence (Li et al., 2016). *Fendrr* is one example of a nuclear lncRNA that forms RNA:DNA triple helices and plays a regulatory role during development. Loss of function of *Fendrr*, which is exclusively expressed in the lateral plate mesoderm during mouse embryogenesis, leads to embryonic lethality due to heart and body wall developmental defects (Grote et al., 2013; Sauvageau et al., 2013). *Fendrr* interacts with both PRC2 (polycomb repressive complex 2) and TrxG/MLL complexes *in vivo*, and its loss results in changes in epigenetic modifications and gene expression of mesoderm differentiation factors (Grote et al., 2013). A 40 nt sequence element of *Fendrr* RNA was identified *in silico* and validated experimentally to form triplexes with promoter regions of two targets of *Fendrr*, *Foxf1* and *Pitx2*, resulting in increased PRC2 occupancy at these promoters and a subsequent inhibition of expression (Grote et al., 2013).

Although reports of DNA:lncRNA triplexes and their functional relevance have been restricted to specific examples, computational

Development (2021) 148, dev182741. doi:10.1242/dev.182741

methods have been developed to predict the formation of lncRNA-DNA triplex structures based on lncRNA sequence *in silico* (Buske et al., 2012; Kuo et al., 2019). One algorithm, called TDF (triplex domain finder), enables retrieval and statistical ranking of sequence elements already known to form triplex structures, including characterized domains of the lncRNAs *Fendrr*, *HOTAIR* and *MEG3* (Kuo et al., 2019). Importantly, the TDF algorithm has also identified and ranked novel sequence elements with triplex-forming potential and detected previously unknown DNA-binding domains in *MEG3* and in lncRNAs relevant to human cardiomyocyte differentiation (Kuo et al., 2019).

### RNA motifs that mediate miRNA degradation

While RNA sequence elements of nuclear lncRNA transcripts have received a lot of attention, RNA motifs in cytoplasmic lncRNAs that have defined molecular mechanisms linked to cellular or developmental functions are as yet under-represented. One of the most studied interaction partners of cytoplasmic lncRNAs are miRNAs, which are small RNAs of ∼22 nt that regulate virtually every aspect of development and normal physiology by pairing with their targets (Alberti and Cochella, 2017; Bartel, 2018). Because one can easily predict miRNA interactions based on the presence of miRNA-binding sites in their target transcripts, miRNA-lncRNA interactions have been extensively investigated (Ulitsky, 2018). The majority of miRNA binding sites within vertebrate lncRNA transcripts are canonical miRNA sites with partial miRNA-lncRNA pairing complementarity (Fig. 1C), which can regulate lncRNA expression similar to miRNA-mediated regulation of mRNA transcripts. It has also been suggested that canonical miRNA target sites within lncRNAs can regulate miRNA activity via competition for binding sites (Ulitsky, 2018).

In addition to the canonical miRNA-binding sites with partial complementarity, several independent studies have found lncRNAs that harbour miRNA-binding sites with extensive, near-perfect complementarity (Fig. 1D), which is rather unusual for animal miRNAs and has not been reported in endogenous transcripts before (Bitetti et al., 2018; Kleaveland et al., 2018; Ulitsky et al., 2011). By using these high-complementarity miRNA sites in artificially engineered targets and natural viral RNAs, it has been demonstrated that the extensive target-miRNA pairing triggers efficient degradation of miRNAs by shortening (‘trimming’) or untemplated lengthening ‘tailing’) the 3′ end of miRNAs, a phenomenon known as target-directed miRNA degradation, or TDMD (Ameres et al., 2010; de la Mata et al., 2015; Marcinowski et al., 2012).

More recently, an endogenous transcript that harbours a RNA sequence element conserved among all vertebrates that binds and specifically degrades miR-29b was discovered (Bitetti et al., 2018). The conserved RNA sequence element is part of the cytoplasmic lncRNA *libra* in zebrafish and the 3′ UTR of the protein-coding gene *Nrep* in mammals. Genetic scrambling of the miR-29 site resulted in ectopic accumulation of miR-29b in the brain, leading to abnormal animal behaviour, thus demonstrating that the miRNA sites embedded in genome-encoded lncRNAs have crucial *in vivo* relevance (Bitetti et al., 2018). A similar activity of the endogenous high complementarity miR-7 site (Ulitsky et al., 2011) was reported for the lncRNA *Cyrano* (Kleaveland et al., 2018). The miR-7 site embedded within the *Cyrano* transcript is conserved among all vertebrates and triggers efficient miR-7 degradation *in vivo* in mouse tissues and in human cells (Kleaveland et al., 2018). The biological relevance of miR-7 degradation by TDMD remains to be addressed, as genetic perturbations of *Cyrano* do not lead to obvious morphological defects in mice or zebrafish (Goudarzi et al., 2019; Kleaveland et al., 2018; Lavalou et al., 2019), in contrast to previous morpholino-based knockdown zebrafish studies (Sarangdhar et al., 2018; Ulitsky et al., 2011). Nonetheless, it is possible that additional lncRNAs contain high complementarity miRNA sites that control miRNA turnover and potentially have key biological functions. This is further supported by another finding of a high-complementarity miRNA site for miR-30 located in the 3′UTR of the *Serpine1* (Ghini et al., 2018). Similar to *libra*/*NREP* and *Cyrano*, the endogenously encoded miR-30 site reduces levels of miR-30b-p5 and 30c-p5 by TDMD, regulating mitotic rate and apoptosis in mouse fibroblasts (Ghini et al., 2018).

### LncRNA motifs that sequester cellular proteins and regulate normal physiology

Although multiple lncRNAs have been shown to have regulatory functions, detailed structure-function analyses have been performed only for a few of them. One of the striking examples of a lncRNA for which specific RNA sequences have been assigned a function is *NORAD* (*noncoding RNA activated by DNA damage*). *NORAD* is an abundant cytoplasmic transcript present at 80-1400 copies per cell that shows sequence conservation across mammals and is upregulated upon DNA damage (Lee et al., 2016; Tichon et al., 2016). Remarkably, *NORAD* contains at least 17 consensus binding sites for PUMILIO proteins (PUM1 and PUM2), which are highly conserved RBPs that bind to consensus motifs typically located in the 3′ untranslated regions (UTRs) of mRNAs, resulting in reduced translation and turnover of mRNA targets (Miller and Olivas, 2011). *NORAD*, with its multiple PUMILIO binding sites and high copy number, acts as a negative regulator of PUMILIO proteins by limiting their amount in the cell (Fig. 1E). NORAD-mediated limitation of PUMILIO modulates the abundance of PUMILIO targets, which are enriched for mRNAs with mitotic, DNA repair and DNA replication functions (Lee et al., 2016; Tichon et al., 2016). As a result, *NORAD* loss of function leads to excess repression of PUMILIO targets, resulting in chromosomal instability, aberrant mitosis and aneuploidy in human cells (Lee et al., 2016; Tichon et al., 2016). In mice, *Norad* deletion leads to genomic instability and mitochondrial dysfunction, with animals showing signs of premature aging; *Norad* null allele mice were found to have grey and thin fur, to show abnormal spine curvature characteristic of aging and to die earlier than wild-type mice (Kopp et al., 2019). While *NORAD* interacts with additional proteins (Munschauer et al., 2018; Tichon et al., 2018), a series of genetic experiments, including PUMILIO 2 overexpression in mice and *NORAD* expression in human cells deficient for *NORAD*, demonstrated that the aforementioned cellular and organismal phenotypes are linked to PUMILIO hyperactivity (Elguindy et al., 2019; Kopp et al., 2019; Lee et al., 2016; Tichon et al., 2016). This example demonstrates the importance of clustered RNA motifs/elements in establishing functional lncRNA-protein interactions that play key roles in cell biology and normal physiology.

### LncRNA elements that mediate lncRNA-protein interactions

Interactions between lncRNAs and RBPs can lead to different molecular, cellular and developmental outcomes. Indeed, lncRNAs can interact with proteins to inhibit or enhance transcription (Feng, 2006; Wutz et al., 2002), stabilize proteins (Zhao et al., 2018) or sequester them (Kino et al., 2010; Lee et al., 2016; Tichon et al., 2016), regulate translation efficiency (Gong et al., 2015), and/or regulate scaffold protein complexes (Rinn et al., 2007). Because RBPs tend to bind low-complexity sequences (Dominguez et al., 2018), it is often difficult to identify protein consensus binding

Development (2021) 148, dev182741. doi:10.1242/dev.182741

motifs within RNA transcripts; however, it is possible to attribute protein binding to specific lncRNA regions. Indeed, it has been shown that, while lncRNA transcripts can interact with multiple RBPs, often only a short RNA region, embedded in the long transcript and interacting with a specific RBP or with a specific set of RBPs, is sufficient and required to drive the biological function of a number of lncRNAs.

One of the best characterized lncRNAs with a crucial developmental function is *Xist* (*X-inactive specific transcript*), which orchestrates transcriptional silencing of one of the female X chromosomes in placental mammals (Borsani et al., 1991; Brown, 1991; Moindrot and Brockdorff, 2016), a process known as X-chromosome inactivation (XCI) (Gendrel and Heard, 2014; Marahrens et al., 1997). *Xist* knockout leads to loss of X-inactivation and female-specific lethality in mice (Marahrens et al., 1997). One of the conserved regions of *Xist*, named the A-repeat, is composed of repetitive sequence elements that are believed to be derived from TEs (Elisaphenko et al., 2008) and is required for the initiation of XCI (Nesterova et al., 2001; Wutz et al., 2002). The A-repeat interacts with multiple RBPs (Chu et al., 2015b; Graindorge et al., 2019; Lu et al., 2016; McHugh et al., 2015; Minajigi et al., 2015; Moindrot et al., 2015; Monfort et al., 2015), among which SPEN (also known as SHARP in human and MINT in mouse) is essential for initiating XCI in mouse embryonic stem cells (ESCs) and preimplantation embryos (Carter et al., 2020; Dossin et al., 2020). By specifically binding the *Xist* A-repeat region, SPEN recruits additional silencing factors to initiate XCI, bridging *Xist* RNA, transcription machinery and chromatin remodellers (Chu et al., 2015b; Dossin et al., 2020; McHugh et al., 2015; Minajigi et al., 2015). Interestingly, in mouse ESCs, Spen binds and silences endogenous retroviral element (ERV) RNAs that resemble the A-repeat of *Xist* (Carter et al., 2020). Insertion of an ERV into an A-repeat-deficient *Xist* transcript rescues *Xist*-mediated gene silencing, suggesting that *Xist* might have adapted a functional TE RNA-protein interaction for dose compensation (Carter et al., 2020).

*Braveheart* is a mouse ∼590 nt lncRNA that carries out a regulatory function in cardiovascular lineage commitment (Klattenhoff et al., 2013). It contains a short G-rich motif termed AGIL that is necessary for the differentiation of mouse ESCs to cardiomyocytes (Xue et al., 2016). This motif interacts with the zinc-finger transcription factor CNBP/ZNF9 (Xue et al., 2016), a heart-specific transcription factor known to bind single-stranded G-rich DNA and RNA motifs (Calcaterra et al., 2010; Chen, 2003). It has been proposed that *Braveheart* and CNBP function together to regulate cardiac gene expression (Xue et al., 2016). Thus, *Braveheart* directs cell fate via a short RNA motif that interacts with a tissue-specific transcription factor (Xue et al., 2016).

The lncRNA *growth arrest-specific 5* (*GAS5*) is one of the first studied lncRNAs (Schneider et al., 1988). *GAS5* regulates the survival of female germline stem cells *in vitro* (Wang et al., 2018) and the self-renewal of human ESCs (Xu et al., 2016). In certain cases, *GAS5* acts as a glucocorticoid receptor (GR) decoy (Kino et al., 2010), participating in the negative-feedback loop of activated GRs, directing cell fate and regulating transcription. The region implicated in this interaction is located in the last exon of the 12-exon *GAS5* transcript and includes a motif that 'mimics' genomic GR-binding sites (Hudson et al., 2014). Interestingly, a single nucleotide substitution is sufficient to abolish the GR-*GAS5* interaction. Sequence alignment showed that this recognition site is conserved among the haplorhine lineage (a suborder of primates), suggesting that other RNAs containing this motif may also interact with GRs (Hudson et al., 2014).

*Megamind* (also known as *Tunar*) harbours a region of ∼200 nt that is deeply conserved among vertebrates (Lin et al., 2014; Ulitsky et al., 2011). Knockdown of *Tunar* in mouse ESCs inhibits their differentiation into neural lineages (Lin et al., 2014). It has been proposed that *Tunar* carries out its regulatory function in pluripotency through an interaction with a protein complex composed of PTBP1, HNRNPK and NCL. Specifically, RNA pulldown experiments have shown that the conserved sequence binds to PTBP1, HNRNPK and NCL proteins with an affinity comparable with that of the full-length *Tunar* transcript, whereas binding affinity of the proteins to a synthetic transcript without the conserved sequence is decreased. The conserved region of *Tunar* therefore appears to act as an RBP-binding platform driving cellular differentiation (Lin et al., 2014).

lncRNA-protein interactions that play regulatory functions in development and organ growth can also contribute to human diseases. For example, the ubiquitously expressed lncRNA *Pint* (p53-induced noncoding transcript), which is under the direct control of p53 (Marín-Béjar et al., 2013), has been implicated in cancer. In mouse cells, *Pint* controls cell proliferation and survival by interacting with PRC2 and regulating the expression of hundreds of genes (Marín-Béjar et al., 2013). In line with this, it has been demonstrated that $Pint^{-/-}$ mice are noticeably smaller and have reduced body weight compared with wild-type mice (Sauvageau et al., 2013). Similar to mouse *Pint*, human *PINT* is regulated by p53 and interacts with PRC2. Of note, *PINT* inhibits the migration and invasion of cancer cells, and this function is carried out by a short sequence motif conserved in mammals that is required for the interaction between *PINT* and PRC2, and the downregulation of pro-invasion genes (Marín-Béjar et al., 2017).

## Identifying RNA-binding proteins associated with lncRNA elements

As highlighted above, a number of lncRNA molecular functions are mediated by interactions with proteins. As such, reliable identification of lncRNA-associated proteins is key for delineating the molecular mechanisms of lncRNA action. Unbiased identification of lncRNA associated proteins is typically carried out by applying so-called RNA-centric approaches, which identify RBPs that interact with a specific RNA of interest. The most commonly used strategy relies on cross-linking and probe-based affinity capture of a test RNA followed by mass spectrometry-based identification of the co-purified proteins (Fig. 2A) (Chu et al., 2015a). Despite the high potential of this methodology, the identification of lncRNA-protein interactions via this approach remains technically challenging due to the general inefficiency of RNA pull-downs. Recently, an alternative approach termed incPRINT, which enables in-cell identification of protein interactions with any RNA of interest, was developed (Graindorge et al., 2019). This technique is based on screening a library of tagged proteins (including the majority of all known human RBPs, transcription factors and chromatin modifiers) with a test RNA that is tethered to a luciferase detector. Instead of pulling down the test RNA, high-throughput immunoprecipitation of thousands of tagged RBPs is performed followed by luciferase detection of their interactions with the test RNA (Graindorge et al., 2019; Sabaté-Cadenas and Shkumatava, 2020). incPRINT is suitable for the identification of in-cell RNA-interacting proteomes of full-length transcripts of various endogenous abundancies, and for mapping proteins that bind to different regions of longer RNA transcripts (Graindorge et al., 2019).

An alternative method to identify proteins that bind to specific sequence motifs harnesses existing data from crosslinking and

Development (2021) 148, dev182741. doi:10.1242/dev.182741

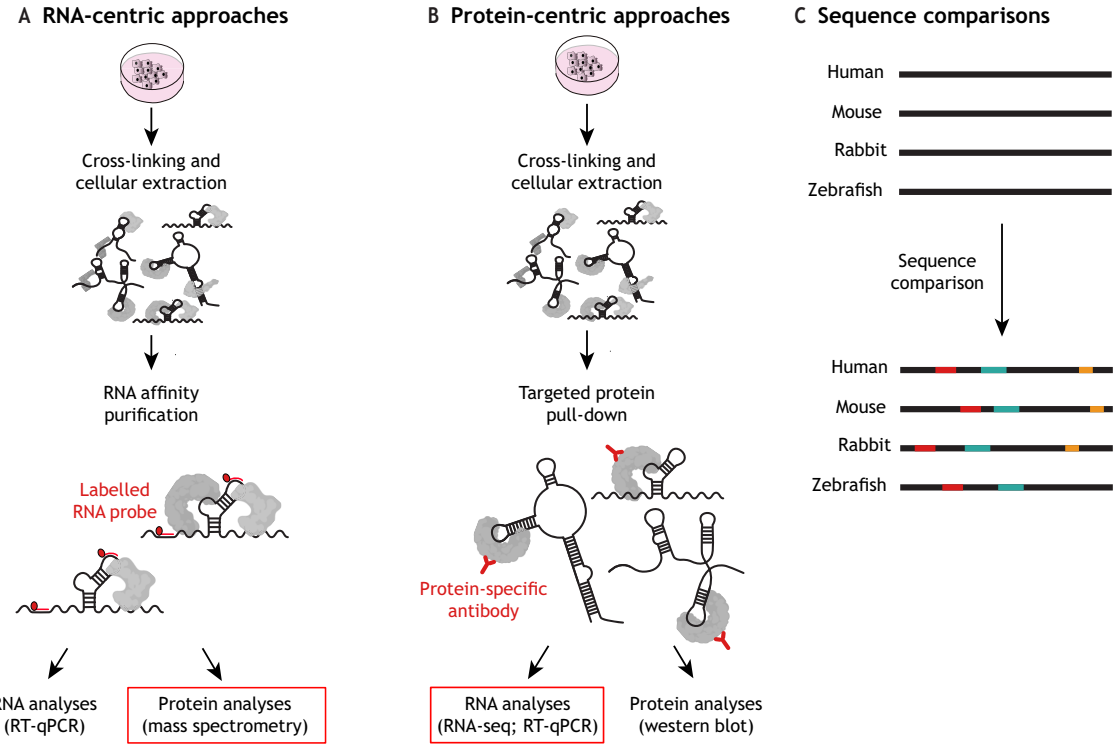


**Fig. 2. Identifying lncRNA-binding proteins and conserved sequence elements.** (A) RNA-centric approaches allow the identification of proteins that can bind to specific RNAs. In these approaches, probe-based affinity capture of a test RNA is performed after UV or chemical crosslinking and subsequent cell lysis. Following biochemical purification of RNA-protein complexes, RNA-binding proteins are identified by mass-spectrometry. (B) Protein-centric approaches enable the identification of RNAs bound by specific proteins. CLIP (crosslinking and immunoprecipitation) approaches use UV or chemical crosslinking to covalently bind protein-RNA complexes, allowing stringent purification of RNA-binding proteins with specific antibodies along with small fragments of RNA, which are amplified and sequenced. (C) Sequence comparisons of homologous lncRNAs enable discoveries of conserved sequence elements.

immunoprecipitation (CLIP) assays. CLIP methods are based on chemical or UV crosslinking followed by immunoprecipitation of a test protein to identify all RNA fragments that crosslink to the protein of interest, which in turn is followed by sequencing (Darnell, 2012; Jensen and Darnell, 2008) (Fig. 2B). Currently, the eCLIP data from the ENCODE project includes ∼150 RBPs (Van Nostrand et al., 2020a, 2020b) the binding sites of which can be mapped and analyzed in specific transcripts of interest. Although the application of CLIP data to an RNA of interest is limited by the number of proteins and cell lines for which the data are available (currently, predominantly K562 or HepG2 cells), it is powerful as an orthogonal approach for validating RNA-centric methods (Graindorge et al., 2019) and for mapping RBP-binding sites identified by eCLIP (Van Nostrand et al., 2017) to RNA sequence elements (Kirk et al., 2018; Lee et al., 2016; Tichon et al., 2018). The identification of RNA motifs and their associated proteins can also enable the prediction of RNA-protein interactions based on lncRNA sequences, although this will require the development of algorithms that are able to detect short sequence homologies. These predictions can then be tested by the aforementioned RNA-protein interaction technologies.

### RNA motif discovery as a starting point for lncRNA functional classification

One of the major bottlenecks in identifying physiologically important lncRNAs is the difficulty in establishing a connection between lncRNA sequence and function. Moreover, because even lncRNAs with well-characterized cellular and organismal functions represent a highly heterogeneous group, it is becoming increasingly important to classify them into functional subgroups. The identification of sequence motifs linked to specific molecular events, such as localization and protein binding, and the subsequent discovery of these motifs in other lncRNAs and/or mRNAs, might help in assigning specific functions to individual lncRNAs, as experimental interrogations are often tedious. Indeed, the comparative analysis of lncRNA sequences across different species (Fig. 2C) (Chen et al., 2016; Hezroni et al., 2015; Necsulea et al., 2014) can be a powerful approach for studying the functions and molecular mechanisms of action of lncRNAs, especially when followed by experimental validations. Such alignment-based comparative approaches have led to the identification of conserved sequence motifs within lncRNA transcripts such as *Megamind*, *Cyrano*, *NEAT1*, *MALAT1*, *NORAD*, *PINT* and many more (Cornelis et al., 2016; Lee et al., 2016; Marín-Béjar et al., 2017; Tichon et al., 2016; Ulitsky et al., 2011; Wilusz et al., 2012). Moreover, uncovering RNA elements with known functions within as-yet uncharacterized lncRNA sequences can facilitate the discovery of novel lncRNA functions. For example, a *SRA1*-like (steroid receptor RNA activator 1) sequence element that mediates androgen receptor (AR) binding was found in the lncRNA *SLNCR1* (Schmidt et al., 2016) and, based on the known function of this motif as a co-activator of hormonal pathways (Lanz et al., 1999; Novikova et al., 2012), it was possible to predict the function of *SLNCR1*. As predicted, targeting this specific binding site within *SLNCR1* prevented its interaction with AR, leading to inhibition of melanogenesis (Schmidt et al., 2020).

Although powerful, alignment-based approaches require relatively long stretches of sequence conservation between homologs. However, most lncRNAs lack long stretches with high sequence conservation. Detection of rather short, conserved sequence motifs within long and rapidly evolving transcripts would be an important step towards assigning sequences to biological and molecular functions. Application of the SEEKR algorithm, developed to detect the presence of similar short sequence motifs of three to eight nucleotides (termed *k*-mers), recently allowed the functional classification of lncRNAs with related functions (Kirk et al., 2018). Reasoning that lncRNAs with shared functions harbour a shared set of similar *k*-mers while lacking linear homology, SEEKR groups lncRNAs based on their *k*-mer profiles. Subgroups of lncRNAs with similar *k*-mer contents share biological properties, such as protein-binding and subcellular localization, even between groups with no apparent evolutionary relationships. Based on both computational data and experimental validation, it has been proposed that no linear sequence homology is required for lncRNAs to support conserved functions. Instead, lncRNA functions are driven by similar sets of short *k*-mer motifs that represent RNA-binding protein motifs (Kirk et al., 2018). The classification of lncRNAs into functional groups based on short motifs is very promising but there is a need for improved algorithms for unbiased identification of short conserved motifs within lncRNA transcripts across more distant species.

## Conclusions and perspectives

The regulatory functions of the lncRNA sequence motifs discussed here demonstrate the importance of primary RNA sequences and the necessity of careful analyses of lncRNA sequence elements, as many of them are shared with protein-coding RNAs. RNA motifs regulate a range of different functions, including transcript biogenesis, subcellular localization and transcript interactions with proteins, DNA and other RNAs, such as miRNAs and protein-coding mRNAs. The identification and further characterization of regulatory RNA motifs will facilitate the prediction of lncRNA interactions and mechanisms of action. Establishing a database of functional RNA elements identified experimentally as well as RNA motifs determined by computational approaches would enable the classification of lncRNAs into functional subgroups. Because several thousand of lncRNAs have been annotated and it is likely that only a minor fraction of them are functional, the identification of conserved RNA motifs would help to prioritize lncRNAs for experimental interrogation to identify transcripts with important cellular functions.

The unbiased identification of regulatory sequence elements within lncRNA transcripts has demonstrated that some coding mRNAs employ the same regulatory RNA motifs. As such, the characterization of functional RNA sequence elements might also help us to understand regulatory sequences that are embedded in mRNA transcripts. Indeed, studying regulatory RNA motifs within mRNA transcripts is often challenging, because the impact of genetic manipulation of sequences located in untranslated regions of mRNAs cannot always be easily uncoupled from their impact on protein production and function. Hence, the characterization of RNA sequence elements in noncoding RNAs could provide a more accessible route and would allow predictions of mRNA regulation by these elements. Furthermore, regulatory RNA sequence motifs represent an untapped source of potential therapeutic targets. As the realization that RNA molecules are druggable targets (Warner et al., 2018), there has been an effort in the field to find small molecule inhibitors that target specific RNA transcripts. However, it is challenging to identify selective drugs that bind specific RNAs without broad effects. Targeting regulatory RNA motifs may provide this specificity. Furthermore, many proteins have been found to be undruggable (Warner et al., 2018); thus, targeting RNA motifs that regulate the transcription, stability and/or translation of mRNAs impacting protein production could open the door to novel treatment strategies.

#### Acknowledgements

We thank all members of the Shkumatava laboratory and Dina Zielinsky for useful discussions.

#### Competing interests

The authors declare no competing or financial interests.

#### Funding

The authors' research is supported by LabEx DEEP, France (ANR-11-LABX-0044 and ANR-10-IDEX-0001-02) and by a SDSV n°577, université PSL doctoral fellowship to F.C.

#### References

**Alberti, C. and Cochella, L.** (2017). A framework for understanding the roles of miRNAs in animal development. *Development* **144**, 2548-2559. doi:10.1242/dev.146613

**Ameres, S. L., Horwich, M. D., Hung, J.-H., Xu, J., Ghildiyal, M., Weng, Z. and Zamore, P. D.** (2010). Target RNA-directed trimming and tailing of small silencing RNAs. *Science* **328**, 1534-1539. doi:10.1126/science.1187058

**Arun, G. and Spector, D. L.** (2019). MALAT1 long non-coding RNA and breast cancer. *RNA Biol.* **16**, 860-863. doi:10.1080/15476286.2019.1592072

**Azam, S., Hou, S., Zhu, B., Wang, W., Hao, T., Bu, X., Khan, M. and Lei, H.** (2019). Nuclear retention element recruits U1 snRNP components to restrain spliced lncRNAs in the nucleus. *RNA Biol.* **16**, 1001-1009. doi:10.1080/15476286.2019.1620061

**Bartel, D. P.** (2018). Metazoan MicroRNAs. *Cell* **173**, 20-51. doi:10.1016/j.cell.2018.03.006

**Bian, S. and Sun, T.** (2011). Functions of noncoding RNAs in neural development and neurological diseases. *Mol. Neurobiol.* **44**, 359-373. doi:10.1007/s12035-011-8211-3

**Bitetti, A., Mallory, A. C., Golini, E., Carrieri, C., Carreño Gutiérrez, H., Perlas, E., Pérez-Rico, Y. A., Tocchini-Valentini, G. P., Enright, A. J., Norton, W. H. J. et al.** (2018). MicroRNA degradation by a conserved target RNA regulates animal behavior. *Nat. Struct. Mol. Biol.* **25**, 244-251. doi:10.1038/s41594-018-0032-x

**Borsani, G., Tonlorenzi, R., Simmler, M. C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D., Lawrence, C. et al.** (1991). Characterization of a murine gene expressed from the inactive X chromosome. *Nature* **351**, 325-329. doi:10.1038/351325a0

**Brown, S. D. M.** (1991). XIST and the mapping of the X chromosome inactivation centre. *BioEssays* **13**, 607-612. doi:10.1002/bies.950131112

**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 MENbeta noncoding RNAs. *Proc. Natl. Acad. Sci. USA* **109**, 19202-19207. doi:10.1073/pnas.1217338109

**Brown, J. A., Bulkley, D., Wang, J., Valenstein, M. L., Yario, T. A., Steitz, T. A. and Steitz, J. A.** (2014). Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nat. Struct. Mol. Biol.* **21**, 633-640. doi:10.1038/nsmb.2844

**Buske, F. A., Bauer, D. C., Mattick, J. S. and Bailey, T. L.** (2012). Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res.* **22**, 1372-1381. doi:10.1101/gr.130237.111

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

**Calcaterra, N. B., Armas, P., Weiner, A. M. J. and Borgognone, M.** (2010). CNBP: A multifunctional nucleic acid chaperone involved in cell death and proliferation control. *IUBMB Life* **62**, 707-714. doi:10.1002/iub.379

**Cao, Z., Pan, X., Yang, Y., Huang, Y. and Shen, H.-B.** (2018). The lncLocator: a subcellular localization predictor for long non-coding RNAs based on a stacked ensemble classifier. *Bioinformatics* **34**, 2185-2194. doi:10.1093/bioinformatics/bty085

**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. doi:10.1101/gr.229922.117

**Carter, A. C., Xu, J., Nakamoto, M. Y., Wei, Y., Zarnegar, B. J., Shi, Q., Broughton, J. P., Ransom, R. C., Salhotra, A., Nagaraja, S. D. et al.** (2020).

Spen links RNA-mediated endogenous retrovirus silencing and X chromosome inactivation. *eLife* **9**, e54508. doi:10.7554/eLife.54508
**Chen, W.** (2003). The zinc-finger protein CNBP is required for forebrain formation in the mouse. *Development* **130**, 1367-1379. doi:10.1242/dev.00349
**Chen, L.-L.** (2016). Linking long noncoding RNA localization and function. *Trends Biochem. Sci.* **41**, 761-772. doi:10.1016/j.tibs.2016.07.003
**Chen, J., Shishkin, A. A., Zhu, X., Kadri, S., Maza, I., Guttman, M., Hanna, J. H., Regev, A. and Garber, M.** (2016). Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. *Genome Biol.* **17**, 19. doi:10.1186/s13059-016-0880-9
**Chu, C., Spitale, R. C. and Chang, H. Y.** (2015a). Technologies to probe functions and mechanisms of long noncoding RNAs. *Nat. Struct. Mol. Biol.* **22**, 29-35. doi:10.1038/nsmb.2921
**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.** (2015b). Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404-416. doi:10.1016/j.cell.2015.03.025
**Clemson, C. M., Hutchinson, J. N., Sara, S. A., Ensminger, A. W., Fox, A. H., Chess, A. and Lawrence, J. B.** (2009). An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of Paraspeckles. *Mol. Cell* **33**, 717-726. doi:10.1016/j.molcel.2009.01.026
**Cornelis, G., Souquere, S., Vernochet, C., Heidmann, T. and Pierron, G.** (2016). Functional conservation of the lncRNA NEAT1 in the ancestrally diverged marsupial lineage: Evidence for NEAT1 expression and associated paraspeckle assembly during late gestation in the opossum *Monodelphis domestica*. *RNA Biol.* **13**, 826-836. doi:10.1080/15476286.2016.1197482
**Darnell, R.** (2012). CLIP (cross-linking and immunoprecipitation) identification of RNAs bound by a specific protein. *Cold Spring Harb. Protoc.* **2012**, 1146-1160. doi:10.1101/pdb.prot072132
**de la Mata, M., Gaidatzis, D., Vitanescu, M., Stadler, M. B., Wentzel, C., Scheiffele, P., Filipowicz, W. and Grosshans, H.** (2015). Potent degradation of neuronal miRNAs induced by highly complementary targets. *EMBO Rep.* **16**, 500-511. doi:10.15252/embr.201540078
**Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D. G. et al.** (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775-1789. doi:10.1101/gr.132159.111
**Diederichs, S.** (2014). The four dimensions of noncoding RNA conservation. *Trends Genet.* **30**, 121-123. doi:10.1016/j.tig.2014.01.004
**Dominguez, D., Freese, P., Alexis, M. S., Su, A., Hochman, M., Palden, T., Bazile, C., Lambert, N. J., Van Nostrand, E. L., Pratt, G. A. et al.** (2018). Sequence, structure, and context preferences of human RNA binding proteins. *Mol. Cell* **70**, 854-867.e859. doi:10.1016/j.molcel.2018.05.001
**Dossin, F., Pinheiro, I., Żylicz, J. J., Roensch, J., Collombet, S., Le Saux, A., Chelmicki, T., Attia, M., Kapoor, V., Zhan, Y. et al.** (2020). SPEN integrates transcriptional and epigenetic control of X-inactivation. *Nature* **578**, 455-460. doi:10.1038/s41586-020-1974-9
**Eissmann, M., Gutschner, T., Hämmerle, M., Günther, S., Caudron-Herger, M., Gross, M., Schirmacher, P., Rippe, K., Braun, T., Zörnig, M. et al.** (2012). Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. *RNA Biol.* **9**, 1076-1087. doi:10.4161/rna.21089
**Elguindy, M. M., Kopp, F., Goodarzi, M., Rehfeld, F., Thomas, A., Chang, T.-C. and Mendell, J. T.** (2019). PUMILIO, but not RBMX, binding is required for regulation of genomic stability by noncoding RNA NORAD. *eLife* **8**, e48625. doi:10.7554/eLife.48625
**Elisaphenko, E. A., Kolesnikov, N. N., Shevchenko, A. I., Rogozin, I. B., Nesterova, T. B., Brockdorff, N. and Zakian, S. M.** (2008). A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements. *PLoS ONE* **3**, e2521. doi:10.1371/journal.pone.0002521
**Esposito, R., Bosch, N., Lanzós, A., Polidori, T., Pulido-Quetglas, C. and Johnson, R.** (2019). Hacking the Cancer genome: profiling therapeutically actionable long non-coding RNAs using CRISPR-Cas9 screening. *Cancer Cell* **35**, 545-557. doi:10.1016/j.ccell.2019.01.019
**Felsenfeld, G. and Rich, A.** (1957). Studies on the formation of two- and three-stranded polyribonucleotides. *Biochim. Biophys. Acta* **26**, 457-468. doi:10.1016/0006-3002(57)90091-4
**Feng, J.** (2006). The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev.* **20**, 1470-1484. doi:10.1101/gad.1416106
**Gendrel, A.-V. and Heard, E.** (2014). Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. *Annu. Rev. Cell Dev. Biol.* **30**, 561-580. doi:10.1146/annurev-cellbio-101512-122415
**Ghini, F., Rubolino, C., Climent, M., Simeone, I., Marzi, M. J. and Nicassio, F.** (2018). Endogenous transcripts control miRNA levels and activity in mammalian cells by target-directed miRNA degradation. *Nat. Commun.* **9**, 3119. doi:10.1038/s41467-018-05182-9
**Gong, C., Li, Z., Ramanujan, K., Clay, I., Zhang, Y., Lemire-Brachat, S. and Glass, D. J.** (2015). A long non-coding RNA, LncMyoD, regulates skeletal muscle differentiation by blocking IMP2-mediated mRNA translation. *Dev. Cell* **34**, 181-191. doi:10.1016/j.devcel.2015.05.009
**Goudarzi, M., Berg, K., Pieper, L. M. and Schier, A. F.** (2019). Individual long non-coding RNAs have no overt functions in zebrafish embryogenesis, viability and fertility. *eLife* **8**, e40815. doi:10.7554/eLife.40815
**Graindorge, A., Pinheiro, I., Nawrocka, A., Mallory, A. C., Tsvetkov, P., Gil, N., Carolis, C., Buchholz, F., Ulitsky, I., Heard, E. et al.** (2019). In-cell identification and measurement of RNA-protein interactions. *Nat. Commun.* **10**, 5317. doi:10.1038/s41467-019-13235-w
**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. doi:10.1016/j.devcel.2012.12.012
**Gudenas, B. L. and Wang, L.** (2018). Prediction of LncRNA subcellular localization with deep learning from sequence features. *Sci. Rep.* **8**, 16385. doi:10.1038/s41598-018-34708-w
**Guo, C.-J., Ma, X.-K., Xing, Y.-H., Zheng, C.-C., Xu, Y.-F., Shan, L., Zhang, J., Wang, S., Wang, Y., Carmichael, G. G. et al.** (2020). Distinct processing of lncRNAs contributes to non-conserved functions in stem cells. *Cell* **181**, 621-636.e622. doi:10.1016/j.cell.2020.03.006
**Gutschner, T. and Diederichs, S.** (2012). The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol.* **9**, 703-719. doi:10.4161/rna.20481
**Guttman, M. and Rinn, J. L.** (2012). Modular regulatory principles of large non-coding RNAs. *Nature* **482**, 339-346. doi:10.1038/nature10887
**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. doi:10.1038/nsmb.2764
**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. doi:10.1038/ncomms11021
**Hezroni, H., Koppstein, D., Schwartz, M. G., Avrutin, A., Bartel, D. P. and Ulitsky, I.** (2015). Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep.* **11**, 1110-1122. doi:10.1016/j.celrep.2015.04.023
**Holmes, Z. E., Hamilton, D. J., Hwang, T., Parsonnet, N. V., Rinn, J. L., Wuttke, D. S. and Batey, R. T.** (2020). The Sox2 transcription factor binds RNA. *Nat. Commun.* **11**, 1805. doi:10.1038/s41467-020-15571-8
**Huarte, M.** (2015). The emerging role of lncRNAs in cancer. *Nat. Med.* **21**, 1253-1261. doi:10.1038/nm.3981
**Hudson, W. H., Pickard, M. R., de Vera, I. M. S., Kuiper, E. G., Mourtada-Maarabouni, M., Conn, G. L., Kojetin, D. J., Williams, G. T. and Ortlund, E. A.** (2014). Conserved sequence-specific lincRNA–steroid receptor interactions drive transcriptional repression and direct cell fate. *Nat. Commun.* **5**, 5395. doi:10.1038/ncomms6395
**Hutchinson, J. N., Ensminger, A. W., Clemson, C. M., Lynch, C. R., Lawrence, J. B. and Chess, A.** (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* **8**, 39. doi:10.1186/1471-2164-8-39
**Jensen, K. B. and Darnell, R. B.** (2008). CLIP: crosslinking and immunoprecipitation of in vivo RNA targets of RNA-binding proteins. *Methods Mol. Biol.* **488**, 85-98. doi:10.1007/978-1-60327-475-3_6
**Kino, T., Hurt, D. E., Ichijo, T., Nader, N. and Chrousos, G. P.** (2010). Noncoding RNA Gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal.* **3**, ra8. doi:10.1126/scisignal.2000568
**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. doi:10.1038/s41588-018-0207-8
**Klattenhoff, C. A., Scheuermann, J. C., Surface, L. E., Bradley, R. K., Fields, P. A., Steinhauser, M. L., Ding, H., Butty, V. L., Torrey, L., Haas, S. et al.** (2013). Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* **152**, 570-583. doi:10.1016/j.cell.2013.01.003
**Kleaveland, B., Shi, C. Y., Stefano, J. and Bartel, D. P.** (2018). A network of noncoding regulatory RNAs acts in the mammalian brain. *Cell* **174**, 350-362.e317. doi:10.1016/j.cell.2018.05.022
**Kopp, F., Elguindy, M. M., Yalvac, M. E., Zhang, H., Chen, B., Gillett, F. A., Lee, S., Sivakumar, S., Yu, H., Xie, Y. et al.** (2019). PUMILIO hyperactivity drives premature aging of Norad-deficient mice. *eLife* **8**, e42650. doi:10.7554/eLife.42650
**Kuo, C.-C., Hänzelmann, S., Sentürk Cetin, N., Frank, S., Zajzon, B., Derks, J.-P., Akhade, V. S., Ahuja, G., Kanduri, C., Grummt, I. et al.** (2019). Detection of RNA–DNA binding sites in long noncoding RNAs. *Nucleic Acids Res.* **47**, e32-e32. doi:10.1093/nar/gkz037
**Kutter, C., Watt, S., Stefflova, K., Wilson, M. D., Goncalves, A., Ponting, C. P., Odom, D. T. and Marques, A. C.** (2012). Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet.* **8**, e1002841. doi:10.1371/journal.pgen.1002841
**Lanz, R. B., McKenna, N. J., Onate, S. A., Albrecht, U., Wong, J., Tsai, S. Y., Tsai, M.-J. and O'Malley, B. W.** (1999). A steroid receptor coactivator, SRA, functions

as an RNA and is present in an SRC-1 complex. *Cell* **97**, 17-27. doi:10.1016/S0092-8674(00)80711-4
**Lavalou, P., Eckert, H., Damy, L., Constanty, F., Majello, S., Bitetti, A., Graindorge, A. and Shkumatava, A.** (2019). Strategies for genetic inactivation of long noncoding RNAs in zebrafish. *RNA* **25**, 897-904. doi:10.1261/rna.069484.118
**Lee, S., Kopp, F., Chang, T.-C., Sataluri, A., Chen, B., Sivakumar, S., Yu, H., Xie, Y. and Mendell, J. T.** (2016). Noncoding RNA NORAD regulates genomic stability by sequestering PUMILIO proteins. *Cell* **164**, 69-80. doi:10.1016/j.cell.2015.12.017
**Lewandowski, J. P., Lee, J. C., Hwang, T., Sunwoo, H., Goldstein, J. M., Groff, A. F., Chang, N. P., Mallard, W., Williams, A., Henao-Meija, J. et al.** (2019). The Firre locus produces a trans-acting RNA molecule that functions in hematopoiesis. *Nat. Commun.* **10**, 5137. doi:10.1038/s41467-019-12970-4
**Li, Y., Syed, J. and Sugiyama, H.** (2016). RNA-DNA triplex formation by long noncoding RNAs. *Cell Chem. Biol.* **23**, 1325-1333. doi:10.1016/j.chembiol.2016.09.011
**Li, L., Zhuang, Y., Zhao, X. and Li, X.** (2018). Long non-coding RNA in neuronal development and neurological disorders. *Front. Genet.* **9**, 744. doi:10.3389/fgene.2018.00744
**Lin, N., Chang, K.-Y., Li, Z., Gates, K., Rana, Z. A., Dang, J., Zhang, D., Han, T., Yang, C.-S., Cunningham, T. J. et al.** (2014). An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. *Mol. Cell* **53**, 1005-1019. doi:10.1016/j.molcel.2014.01.021
**Lu, Z., Zhang, Q. C., Lee, B., Flynn, R. A., Smith, M. A., Robinson, J. T., Davidovich, C., Gooding, A. R., Goodrich, K. J., Mattick, J. S. et al.** (2016). RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* **165**, 1267-1279. doi:10.1016/j.cell.2016.04.028
**Lubelsky, Y. and Ulitsky, I.** (2018). Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. *Nature* **555**, 107-111. doi:10.1038/nature25757
**Mallory, A. C. and Shkumatava, A.** (2015). LncRNAs in vertebrates: advances and challenges. *Biochimie* **117**, 3-14. doi:10.1016/j.biochi.2015.03.014
**Mao, Y. S., Sunwoo, H., Zhang, B. and Spector, D. L.** (2011). Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat. Cell Biol.* **13**, 95-101. doi:10.1038/ncb2140
**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. doi:10.1101/gad.11.2.156
**Marchese, F. P., Raimondi, I. and Huarte, M.** (2017). The multidimensional mechanisms of long noncoding RNA function. *Genome Biol.* **18**, 206. doi:10.1186/s13059-017-1348-2
**Marcinowski, L., Tanguy, M., Krmpotic, A., Rädle, B., Lisnić, V. J., Tuddenham, L., Chane-Woon-Ming, B., Ruzsics, Z., Erhard, F., Benkartek, C. et al.** (2012). Degradation of cellular mir-27 by a novel, highly abundant viral transcript is important for efficient virus replication in vivo. *PLoS Pathog.* **8**, e1002510. doi:10.1371/journal.ppat.1002510
**Marín-Béjar, O., Marchese, F. P., Athie, A., Sánchez, Y., González, J., Segura, V., Huang, L., Moreno, I., Navarro, A., Monzó, M. et al.** (2013). Pint lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2. *Genome Biol.* **14**, R104. doi:10.1186/gb-2013-14-9-r104
**Marín-Béjar, O., Mas, A. M., González, J., Martinez, D., Athie, A., Morales, X., Galduroz, M., Raimondi, I., Grossi, E., Guo, S. et al.** (2017). The human lncRNA LINC-PINT inhibits tumor cell invasion through a highly conserved sequence element. *Genome Biol.* **18**, 202. doi:10.1186/s13059-017-1331-y
**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. doi:10.1038/nature14443
**Michalik, K. M., You, X., Manavski, Y., Doddaballapur, A., Zörnig, M., Braun, T., John, D., Ponomareva, Y., Chen, W., Uchida, S. et al.** (2014). Long noncoding RNA MALAT1 regulates endothelial cell function and vessel growth. *Circ. Res.* **114**, 1389-1397. doi:10.1161/CIRCRESAHA.114.303265
**Miller, M. A. and Olivas, W. M.** (2011). Roles of Puf proteins in mRNA degradation and translation. *Wiley Interdiscip. Rev. RNA* **2**, 471-492. doi:10.1002/wrna.69
**Minajigi, A., Froberg, J., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W. et al.** (2015). Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**, aab2276. doi:10.1126/science.aab2276
**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. doi:10.1261/rna.028639.111
**Moindrot, B. and Brockdorff, N.** (2016). RNA binding proteins implicated in Xist-mediated chromosome silencing. *Semin. Cell Dev. Biol.* **56**, 58-70. doi:10.1016/j.semcdb.2016.01.029
**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. doi:10.1016/j.celrep.2015.06.053
**Monfort, A., Di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S. and Wutz, A.** (2015). Identification of Spen as a crucial factor for xist function through forward genetic screening in haploid embryonic stem cells. *Cell Rep.* **12**, 554-561. doi:10.1016/j.celrep.2015.06.067
**Mukherjee, N., Calviello, L., Hirsekorn, A., de Pretis, S., Pelizzola, M. and Ohler, U.** (2017). Integrative classification of human coding and noncoding genes through RNA metabolism profiles. *Nat. Struct. Mol. Biol.* **24**, 86-96. doi:10.1038/nsmb.3325
**Munschauer, M., Nguyen, C. T., Sirokman, K., Hartigan, C. R., Hogstrom, L., Engreitz, J. M., Ulirsch, J. C., Fulco, C. P., Subramanian, V., Chen, J. et al.** (2018). The NORAD lncRNA assembles a topoisomerase complex critical for genome stability. *Nature* **561**, 132-136. doi:10.1038/s41586-018-0453-z
**Naganuma, T. and Hirose, T.** (2013). Paraspeckle formation during the biogenesis of long non-coding RNAs. *RNA Biol.* **10**, 456-461. doi:10.4161/rna.23547
**Naganuma, T., Nakagawa, S., Tanigawa, A., Sasaki, Y. F., Goshima, N. and Hirose, T.** (2012). Alternative 3′-end processing of long noncoding RNA initiates construction of nuclear paraspeckles: LncRNA processing for nuclear body architecture. *EMBO J.* **31**, 4020-4034. doi:10.1038/emboj.2012.251
**Nakagawa, S., Naganuma, T., Shioi, G. and Hirose, T.** (2011). Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. *J. Cell Biol.* **193**, 31-39. doi:10.1083/jcb.201011110
**Nakagawa, S., Ip, J. Y., Shioi, G., Tripathi, V., Zong, X., Hirose, T. and Prasanth, K. V.** (2012). Malat1 is not an essential component of nuclear speckles in mice. *RNA* **18**, 1487-1499. doi:10.1261/rna.033217.112
**Nakagawa, S., Shimada, M., Yanaka, K., Mito, M., Arai, T., Takahashi, E., Fujita, Y., Fujimori, T., Standaert, L., Marine, J.-C. et al.** (2014). The lncRNA Neat1 is required for corpus luteum formation and the establishment of pregnancy in a subpopulation of mice. *Development* **141**, 4618-4627. doi:10.1242/dev.110544
**Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., Baker, J. C., Grützner, F. and Kaessmann, H.** (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**, 635-640. doi:10.1038/nature12943
**Nesterova, T. B., Slobodyanyuk, S. Y., Elisaphenko, E. A., Shevchenko, A. I., Johnston, C., Pavlova, M. E., Rogozin, I. B., Kolesnikov, N. N., Brockdorff, N. and Zakian, S. M.** (2001). Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Res.* **11**, 833-849. doi:10.1101/gr.174901
**Noh, J. H., Kim, K. M., McClusky, W. G., Abdelmohsen, K. and Gorospe, M.** (2018). Cytoplasmic functions of long noncoding RNAs. *Wiley Interdiscip. Rev. RNA* **9**, e1471. doi:10.1002/wrna.1471
**Novikova, I. V., Hennelly, S. P. and Sanbonmatsu, K. Y.** (2012). Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res.* **40**, 5034-5051. doi:10.1093/nar/gks071
**Paraskevopoulou, M. D. and Hatzigeorgiou, A. G.** (2016). Analyzing MiRNA–LncRNA interactions. In *Long Non-Coding RNAs* (ed. Y. Feng and L. Zhang), pp. 271-286. New York, NY: Springer New York.
**Perry, R. B.-T. and Ulitsky, I.** (2016). The functions of long noncoding RNAs in development and stem cells. *Development* **143**, 3882-3894. doi:10.1242/dev.140962
**Prasanth, K. V., Prasanth, S. G., Xuan, Z., Hearn, S., Freier, S. M., Bennett, C. F., Zhang, M. Q. and Spector, D. L.** (2005). Regulating gene expression through RNA nuclear retention. *Cell* **123**, 249-263. doi:10.1016/j.cell.2005.08.033
**Rinn, J. L. and Chang, H. Y.** (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145-166. doi:10.1146/annurev-biochem-051410-092902
**Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E. et al.** (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311-1323. doi:10.1016/j.cell.2007.05.022
**Rivas, E., Clements, J. and Eddy, S. R.** (2017). A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat. Methods* **14**, 45-48. doi:10.1038/nmeth.4066
**Sabaté-Cadenas, X. and Shkumatava, A.** (2020). In-cell discovery of RNA-protein interactions. *Trends Biochem. Sci.* **45**, 272-273. doi:10.1016/j.tibs.2019.12.003
**Sarangdhar, M. A., Chaubey, D., Srikakulam, N. and Pillai, B.** (2018). Parentally inherited long non-coding RNA Cyrano is involved in zebrafish neurodevelopment. *Nucleic Acids Res.* **46**, 9726-9735. doi:10.1093/nar/gky628
**Sasaki, Y. T. F., Ideue, T., Sano, M., Mituyama, T. and Hirose, T.** (2009). MENε/β noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc. Natl. Acad. Sci. USA* **106**, 2525-2530. doi:10.1073/pnas.0807899106
**Sauvageau, M., Goff, L. A., Lodato, S., Bonev, B., Groff, A. F., Gerhardinger, C., Sanchez-Gomez, D. B., Hacisuleyman, E., Li, E., Spence, M. et al.** (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife* **2**, e01749. doi:10.7554/eLife.01749
**Schmidt, K., Joyce, C. E., Buquicchio, F., Brown, A., Ritz, J., Distel, R. J., Yoon, C. H. and Novina, C. D.** (2016). The lncRNA SLNCR1 mediates melanoma invasion through a conserved SRA1-like region. *Cell Rep.* **15**, 2025-2037. doi:10.1016/j.celrep.2016.04.018

**Schmidt, K., Weidmann, C. A., Hilimire, T. A., Yee, E., Hatfield, B. M., Schneekloth, J. S., Weeks, K. M. and Novina, C. D.** (2020). Targeting the oncogenic long non-coding RNA SLNCR1 by blocking its sequence-specific binding to the androgen receptor. *Cell Rep.* **30**, 541-554.e545. doi:10.1016/j.celrep.2019.12.011

**Schneider, C., King, R. M. and Philipson, L.** (1988). Genes specifically expressed at growth arrest of mammalian cells. *Cell* **54**, 787-793. doi:10.1016/S0092-8674(88)91065-3

**Shukla, C. J., McCorkindale, A. L., Gerhardinger, C., Korthauer, K. D., Cabili, M. N., Shechner, D. M., Irizarry, R. A., Maass, P. G. and Rinn, J. L.** (2018). High-throughput identification of RNA nuclear enrichment sequences. *EMBO J.* **37**, e98452. doi:10.15252/embj.201798452

**Standaert, L., Adriaens, C., Radaelli, E., Van Keymeulen, A., Blanpain, C., Hirose, T., Nakagawa, S. and Marine, J.-C.** (2014). The long noncoding RNA Neat1 is required for mammary gland development and lactation. *RNA* **20**, 1844-1849. doi:10.1261/rna.047332.114

**Su, Z.-D., Huang, Y., Zhang, Z.-Y., Zhao, Y.-W., Wang, D., Chen, W., Chou, K.-C. and Lin, H.** (2018). iLoc-lncRNA: predict the subcellular location of lncRNAs by incorporating octamer composition into general PseKNC. *Bioinformatics* **34**, 4196-4204. doi:10.1093/bioinformatics/bty508

**Sun, Q., Hao, Q. and Prasanth, K. V.** (2018). nuclear long noncoding RNAs: key regulators of gene expression. *Trends Genet.* **34**, 142-157. doi:10.1016/j.tig.2017.11.005

**Sunwoo, H., Dinger, M. E., Wilusz, J. E., Amaral, P. P., Mattick, J. S. and Spector, D. L.** (2009). MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res.* **19**, 347-359. doi:10.1101/gr.087775.108

**Takeda, K., Ichijo, H., Fujii, M., Mochida, Y., Saitoh, M., Nishitoh, H., Sampath, T. K. and Miyazono, K.** (1998). Identification of a novel bone morphogenetic protein-responsive gene that may function as a noncoding RNA. *J. Biol. Chem.* **273**, 17079-17085. doi:10.1074/jbc.273.27.17079

**Tichon, A., Gil, N., Lubelsky, Y., Havkin Solomon, T., Lemze, D., Itzkovitz, S., Stern-Ginossar, N. and Ulitsky, I.** (2016). A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nat. Commun.* **7**, 12209. doi:10.1038/ncomms12209

**Tichon, A., Perry, R. B.-T., Stojic, L. and Ulitsky, I.** (2018). SAM68 is required for regulation of Pumilio by the NORAD long noncoding RNA. *Genes Dev.* **32**, 70-78. doi:10.1101/gad.309138.117

**Ulitsky, I.** (2018). Interactions between short and long noncoding RNAs. *FEBS Lett.* **592**, 2874-2883. doi:10.1002/1873-3468.13085

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

**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. doi:10.1016/j.cell.2011.11.055

**Uroda, T., Anastasakou, E., Rossi, A., Teulon, J.-M., Pellequer, J.-L., Annibale, P., Pessey, O., Inga, A., Chillón, I. and Marcia, M.** (2019). Conserved pseudoknots in lncRNA MEG3 are essential for stimulation of the p53 pathway. *Mol. Cell* **75**, 982-995.e989. doi:10.1016/j.molcel.2019.07.025

**Van Nostrand, E. L., Nguyen, T. B., Gelboin-Burkhart, C., Wang, R., Blue, S. M., Pratt, G. A., Louie, A. L. and Yeo, G. W.** (2017). Robust, cost-effective profiling of RNA binding protein targets with single-end enhanced crosslinking and immunoprecipitation (seCLIP). *Methods Mol. Biol.* **1648**, 177-200. doi:10.1007/978-1-4939-7204-3_14

**Van Nostrand, E. L., Freese, P., Pratt, G. A., Wang, X., Wei, X., Xiao, R., Blue, S. M., Chen, J. Y., Cody, N. A. L., Dominguez, D. et al**. (2020a). A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711-719. doi:10.1038/s41586-020-2077-3

**Van Nostrand, E. L., Pratt, G. A., Yee, B. A., Wheeler, E. C., Blue, S. M., Mueller, J., Park, S. S., Garcia, K. E., Gelboin-Burkhart, C., Nguyen, T. B. et al.** (2020b). Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. *Genome Biol.* **21**, 90. doi:10.1186/s13059-020-01982-9

**Wang, K. C. and Chang, H. Y.** (2011). Molecular mechanisms of long noncoding RNAs. *Mol. Cell* **43**, 904-914. doi:10.1016/j.molcel.2011.08.018

**Wang, J., Gong, X., Tian, G. G., Hou, C., Zhu, X., Pei, X., Wang, Y. and Wu, J.** (2018). Long noncoding RNA growth arrest-specific 5 promotes proliferation and survival of female germline stem cells in vitro. *Gene* **653**, 14-21. doi:10.1016/j.gene.2018.02.021

**Warner, K. D., Hajdin, C. E. and Weeks, K. M.** (2018). Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov.* **17**, 547-558. doi:10.1038/nrd.2018.93

**Washietl, S., Kellis, M. and Garber, M.** (2014). Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res.* **24**, 616-628. doi:10.1101/gr.165035.113

**Wilusz, J. E., Freier, S. M. and Spector, D. L.** (2008). 3′ end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135**, 919-932. doi:10.1016/j.cell.2008.10.012

**Wilusz, J. E., JnBaptiste, C. K., Lu, L. Y., Kuhn, C.-D., Joshua-Tor, L. and Sharp, P. A.** (2012). A triple helix stabilizes the 3′ ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev.* **26**, 2392-2407. doi:10.1101/gad.204438.112

**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. doi:10.1038/ng820

**Xu, C., Zhang, Y., Wang, Q., Xu, Z., Jiang, J., Gao, Y., Gao, M., Kang, J., Wu, M., Xiong, J. et al.** (2016). Long non-coding RNA GAS5 controls human embryonic stem cell self-renewal by maintaining NODAL signalling. *Nat. Commun.* **7**, 13287. doi:10.1038/ncomms13287

**Xue, Z., Hennelly, S., Doyle, B., Gulati, A. A., Novikova, I. V., Sanbonmatsu, K. Y. and Boyer, L. A.** (2016). A G-rich motif in the lncRNA braveheart interacts with a zinc-finger transcription factor to specify the cardiovascular lineage. *Mol. Cell* **64**, 37-50. doi:10.1016/j.molcel.2016.08.010

**Yin, Y., Lu, J. Y., Zhang, X., Shao, W., Xu, Y., Li, P., Hong, Y., Cui, L., Shan, G., Tian, B. et al.** (2020). U1 snRNP regulates chromatin retention of noncoding RNAs. *Nature* **580**, 147-150. doi:10.1038/s41586-020-2105-3

**Zhang, X., Rice, K., Wang, Y., Chen, W., Zhong, Y., Nakayama, Y., Zhou, Y. and Klibanski, A.** (2010). Maternally Expressed Gene 3 (MEG3) noncoding ribonucleic acid: isoform structure, expression, and functions. *Endocrinology* **151**, 939-947. doi:10.1210/en.2009-0657

**Zhang, B., Arun, G., Mao, Y. S., Lazar, Z., Hung, G., Bhattacharjee, G., Xiao, X., Booth, C. J., Wu, J., Zhang, C. et al.** (2012). The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep.* **2**, 111-123. doi:10.1016/j.celrep.2012.06.003

**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. doi:10.1128/MCB.01673-13

**Zhao, J., Du, P., Cui, P., Qin, Y., Hu, C., Wu, J., Zhou, Z., Zhang, W., Qin, L. and Huang, G.** (2018). LncRNA PVT1 promotes angiogenesis via activating the STAT3/VEGFA axis in gastric cancer. *Oncogene* **37**, 4094-4109. doi:10.1038/s41388-018-0250-z

**Ziv, O., Gabryelska, M. M., Lun, A. T. L., Gebert, L. F. R., Sheu-Gruttadauria, J., Meredith, L. W., Liu, Z.-Y., Kwok, C. K., Qin, C.-F., MacRae, I. J. et al.** (2018). COMRADES determines in vivo RNA structures and interactions. *Nat. Methods* **15**, 785-788. doi:10.1038/s41592-018-0121-0

**Zuckerman, B. and Ulitsky, I.** (2019). Predictive models of subcellular localization of long RNAs. *RNA* **25**, 557-572. doi:10.1261/rna.068288.118