Published in final edited form as:

Biochimie. 2015 October ; 117: 3–14. doi:10.1016/j.biochi.2015.03.014.

LncRNAs in Vertebrates: Advances and Challenges

Allison Mallory^{1,2,3,*} and Alena Shkumatava^{1,2,3,*}

¹Institut Curie, 26 Rue d'Ulm, 75248 Paris Cedex 05, France

²CNRS UMR3215, 75248 Paris Cedex 05, France

³INSERM U934, 75248 Paris Cedex 05, France

Abstract

Beyond the handful of classic and well-characterized long noncoding RNAs (lncRNAs), more recently, hundreds of thousands of lncRNAs have been identified in multiple species including bacteria, plants and vertebrates, and the number of newly annotated lncRNAs continues to increase as more transcriptomes are analyzed. In vertebrates, the expression of many lncRNAs is highly regulated, displaying discrete temporal and spatial expression patterns, suggesting roles in a wide range of developmental processes and setting them apart from classic housekeeping ncRNAs. In addition, the deregulation of a subset of these lncRNAs has been linked to the development of several diseases, including cancers, as well as developmental anomalies. However, the majority of vertebrate lncRNA functions remain enigmatic. As such, a major task at hand is to decipher the biological roles of lncRNAs and uncover the regulatory networks upon which they impinge. This review focuses on our emerging understanding of lncRNAs in vertebrate animals, highlighting some recent advances in their functional analyses across several species and emphasizing the current challenges researchers face to characterize lncRNAs and identify their *in vivo* functions.

Keywords

long noncoding RNAs; vertebrate development; genome editing; RNP complexes; ribosome profiling; small ORFs

1 Introduction

The recent advent of low-cost, next generation RNA sequencing (RNA-Seq) technologies has allowed researchers to move away from microarray-based transcriptomic studies, which rely on sequence complementary to fixed oligonucleotide probes, to the high-throughput reading and assembly of the milieu of expressed long RNA molecules. As such, our perspective of what constitutes the functional products of a genome has expanded from the more classical view of a handful of noncoding RNAs and a majority of coding mRNAs, to hundreds of thousands of noncoding RNAs (ncRNAs), including both short (sRNAs) and long (lncRNAs) RNAs. These RNAs were once written off as junk RNA or background transcriptional noise, often produced from what was characterized previously as “intergenic”

*Correspondence: allison.mallory@curie.fr; alena.shkumatava@curie.fr.

regions. However, we now know that ncRNAs have critical regulatory roles in diverse molecular networks. With this thwart of new information, the detailed depiction of numerous eukaryotic genome roadmaps has expanded to include thousands of noncoding genes that produce RNAs rather than proteins as their functional products. This discovery effectively cuts short the classic central dogma of biology, which states that the information in genes flows from DNA to RNA to protein. As such, scientists are now working to unraveling the biological functions of these RNAs.

For the purpose of this review, we will focus on the characteristics, roles and regulatory mechanisms of lncRNAs primarily in vertebrate organisms including human, mouse and zebrafish. Thus, we will not delve into the functions of sRNAs, regulatory RNAs between 20 and 200 nt in length, or classical housekeeping ncRNAs, as they have been reviewed extensively elsewhere [1]. Finally, we will point to some of the challenges faced by the scientific community to tease apart the functions of lncRNAs due both to limitations in technology and to the complex and varying nature of these intriguing molecules.

2 Defining long noncoding RNAs

The most simplified definition of a lncRNA is an RNA longer than 200 nucleotides that are not predicted to be translated to a functional protein product. This definition differentiates lncRNAs from most short housekeeping RNAs, such as small nucleolar (snRNAs) and transfer RNAs (tRNAs), and from short regulatory RNAs such as microRNAs (miRNAs), piwi-interacting RNAs (piRNAs) and short interfering RNAs (siRNAs) based solely on length [2]. LncRNAs, like protein-coding mRNAs, are primarily transcribed by RNA Polymerase II and, thus, have features such as a 5'-cap, polyadenylation and splicing [3, 4]. LncRNAs can be sub-divided into different classes based on the genomic position of the locus from which they are transcribed. For example, a lncRNA can be transcribed from a loci embedded in the intron of a protein-coding gene, in sense or antisense orientation to protein-coding genes (referred to as antisense lncRNAs) and from intergenic regions located between annotated protein-coding or noncoding genes (referred to as lincRNAs). Like protein-coding genes, lncRNA genes display signature histone methylation signals at their promoter and transcribed sequences, which often are indicative of their expression status [3]. Although a handful of lncRNAs displays long stretches or short blocks of sequence conservation as well as syntenic conservation among vertebrates, some lncRNAs share synteny but do not display discernable sequence conservation [5, 6]. Furthermore, numerous of lncRNAs display neither sequence conservation nor shared synteny among vertebrates. This lack of or limited evolutionary constraints often complicates the identification of lncRNA functional motifs or domains. Extensive sets of lncRNAs in several model organisms have been described recently [3, 5–14]. However, identifying a gene as a *bona fide* lncRNA gene is not always an easy task. By definition, lncRNA genes lack many of the features routinely used to annotate protein-coding genes, such as the absence of translational cues and long open reading frames (ORFs). In addition, many lncRNAs have only limited or no sequence conservation among species. As such, current lists of lncRNAs are long and assorted in character, likely including both properly annotated as well as improperly annotated RNAs.

Indeed, recent analyses of a subset of previously annotated lncRNAs expressed in mammalian brain [15, 16] revealed that many of the predicted lncRNAs were actually stable 3'UTR extensions of mRNAs originating from adjacent upstream protein-coding genes rather than true lncRNAs [17–19]. Therefore, to more confidently annotate lncRNA genes and distinguish them from protein-coding genes, it is important to take advantage, when available, of multiple datasets (Figure 1A). For example, chromatin signature maps, histone methylation patterns such as H3K4me3 that marks the promoters of genes actively transcribed by RNA polymerase II [20], CAGE (cap analysis of gene expression) that maps transcription start sites (TSS) at single nucleotide resolution [21], 3P-Seq libraries that reliably define the 3' ends of transcripts [22], ribosome profiling landscapes, and sequence and syntenic conservation can be combined with RNA-Seq data and gene-boundary data sets for annotation, such as in the case of the pipeline GRIT [3, 5, 6, 23–25].

From the current lists of annotated lncRNA genes, their transcripts can be sub-classified based on a myriad of distinguishing features such as nucleotide length, sequence or syntenic conservation (or lack thereof), cellular localization and additional structural features. Moreover, a subset of transcripts classified as lncRNAs contains putative short ORFs, and profiling studies show they are bound by ribosomes suggesting active translation, which, by definition, would make these transcripts coding (see below for a detailed discussion on short ORFs). With ongoing intense efforts by researchers to analyze lncRNA in multiple eukaryotes, it is highly likely that the number of lncRNA genes will continue to increase beyond the thousands of already identified loci. Indeed, a recent update on the LNCipedia database of human lncRNA sequences (<http://www.lncipedia.org>) led to the addition of over 90 000 new lncRNA transcripts, a more than 5-fold expansion [24]. What is clear from current studies is that the more we understand the features of lncRNA genes and the functional lncRNA products, as well as their biological roles, the better able we will be to recognize trends and train algorithms to identify more reliably additional lncRNAs.

3 Distinguishing between peptide-coding and noncoding functions

At present, one of the most highly debated topics in the lncRNA field is how to reliably distinguish lncRNAs from coding RNAs. In the first attempts to computationally identify lncRNAs, algorithms that calculated the presence of a long ORF and its conservation through evolution such as CPC, CPAT, PORTRAIT, PhyloCSF and RNACode [26–30] were used to differentiate coding RNAs from lncRNAs. Recently, a previously annotated lncRNA was shown to harbor a 138nt fragment encoding a conserved 46 amino acid small peptide in the 3rd and last exon of both the mouse and human genes [31]. Indeed, mouse knock-out studies together with behavioral analyses revealed that this skeletal muscle-specific micropeptide myoregulin (MLN) affected muscle performance by regulating the Ca²⁺ pump SERCA. Moreover, several studies have used ribosome profiling, a method to identify RNA fragments bound and protected by ribosomes, together with computational analyses to more precisely annotate lncRNAs and to separate them from transcripts that are engaged by ribosomes and potentially coding for short peptides [32–35]. Indeed, this technique revealed that a portion of the transcripts originally annotated as noncoding actually have the potential to code short proteins or peptides. For example, a study using the Translated ORF Classifier [34] to analyze zebrafish transcript and ribosome profiling data sets for non-annotated ORFs

revealed that several zebrafish genes annotated as noncoding RNAs actually have the capacity to encode secreted peptides [36]. This work showed that one such “previously-annotated lncRNA” gene named *elabela/toddler*, produces a short, conserved peptide that promotes cell motility and is essential for heart development in early zebrafish embryos [36, 37] (Figure 1B). Thus, although a handful of loci originally annotated as producing lncRNAs has been shown to produce short peptides, more work is required to determine whether numerous additional annotated lncRNAs also produce peptides or if these examples are only isolated cases.

It is important to note that evidence of ribosome binding to a given lncRNA is not sufficient to say that a functional protein is produced from that lncRNA. Indeed, careful analyses of ribosome profiling data revealed that many lncRNAs contain small ORFs (smORFs) that are unlikely to encode functional peptides due to their small size (sometimes just 1-2 amino acids), and a lack of conservation at the amino acid level [34, 38]. These lncRNA smORFs are biased to appear close to the 5' end of the transcript and resemble the upstream ORFs (uORFs) of mRNAs. uORFs are not known to produce functional peptides but play regulatory roles affecting translation of downstream main ORFs or mRNA stability [39, 40]. Thus, these smORFs may serve roles that are not directly linked to the production of a particular peptide, but instead could implicate the translational machinery in lncRNA regulation. Several models have emerged to account for the presence of smORFs in lncRNAs. One model posits that over time some smORFs could potentially become functional and thus, smORFs may serve as a repository or birth pool of new peptides [41, 42]. Alternatively, given that nearly nothing is known about how lncRNAs avoid degradation by the cellular RNA surveillance machinery, another appealing model is that smORFs serve to recruit the translational machinery to the lncRNA, effectively protecting the transcripts from the RNA degradation machinery. This model does not exclude the fact that some lncRNA-derived peptides could be functionally important, but it instead evokes the alternative explanation that ribosome binding or the act of translation could be a plausible strategy to adjust the stability of lncRNAs by modulating decay pathways such as nonsense-mediated decay (NMD) or other cellular exo- or endo-ribonuclease-mediated degradation pathways [43–45].

3.1 Techniques to distinguish peptide-coding RNAs from lncRNAs

To distinguish between functional peptide production and ribosome engagement for RNA stability, several experimental approaches can be employed. The previously mentioned computational and experimental methods all allow for a first prediction of lncRNA coding potential. However, in order to establish if the putative noncoding transcript is indeed noncoding, one must introduce point substitutions in the putative ORF region that change the amino acid sequence of the produced peptide and test if the functionality of this modified transcript is maintained. Introducing point substitutions that lead to frame shifts or creating translational fusions between fluorescent tags and the lncRNA peptides are additional techniques that can be used to evaluate the importance of the peptide sequence to lncRNA function, but caution is warranted when employing protein fusion techniques because they create longer peptides that may be more stable than the wild-type peptides. Also, to evaluate if ribosome binding itself rather than the production of a specific peptide sequence is the

functionally relevant act, the putative ORF sequence should be scrambled or replaced by another non-cognate ORF, and the phenotypic output and effect on lncRNA transcription and stability should be evaluated. Indeed, the engagement of lncRNAs with ribosomes represents an interesting model to dissect the interface between truly noncoding vs. coding RNAs and to examine the potential roles of translation beyond protein production.

4 Recently reported examples of vertebrate lncRNA *in vivo* functions

As more and more DNA genomes are sequenced, transcriptomes cataloged and epigenomes profiled, lncRNA annotation continues to improve and expand in numerous organisms. Many lncRNAs are expressed at specific times and in specific tissues or cell types during development, and their expression patterns can change during differentiation. Although the *in vivo* characterization of lncRNA functions has lagged behind lncRNA discovery, dozens of annotated lncRNAs now have known or predicted biological *in vivo* roles. As studies examining the *in vivo* roles of lncRNAs begins to accumulate, it is now becoming clear that their tentacles reach far and wide, affecting numerous physiological and pathological regulatory networks, including the regulation of protein-coding genes, genome architecture and protein activity, and broadly controlling development and pathogenesis. Here, giving priority to those studies that have inferred lncRNA function by analyzing the consequences of lncRNA knock-out in model organisms, we will highlight a selection of newly discovered mammalian and vertebrate *in vivo* lncRNA functions. These examples are not only representative of the current progress in the field but also, in some cases, particularly relevant to development and disease. In addition, as a note of caution, we will discuss a handful of individual lncRNA mutant analyses where the phenotypes observed for knock-out mutant lines do not correlate well with those observed upon knock-down of the corresponding lncRNA.

4.1 lncRNAs, disease and cancer

Even though lncRNA functional discovery is in its infancy, several mammalian lncRNAs already have been directly linked to disease establishment. For example, recently, a cluster of myocardium-specific, primarily nuclear mouse lncRNAs transcribed in antisense orientation to the myosin heavy chain 7 gene (*Myh7*) and referred to as *Myheart* (*Mhrt* for myosin heavy-chain-associated RNA transcripts) was implicated in the protection against cardiomyopathy in mouse [46]. The *Mhrt* lncRNAs were shown to bind to and antagonize the BRG1 helicase, which is part of the BRG1-HDAC-PARP chromatin repressor complex. Under stress conditions, *Mhrt* transcription is repressed, allowing the Brg1 chromatin-remodeling factor to be activated and to direct aberrant gene expression, triggering the formation of cardiac myopathy. The authors went on to show that human *MHRT* is also repressed in myopathic hearts, suggesting that the cardiomyopathy protective function of mouse *Mhrt* is conserved in human.

In addition to playing a role in heart disease, several lncRNA genes have been characterized either as tumor suppressors or as oncogenic, and dozens of lncRNAs are deregulated in multiple cancers. For example, one recent study implicated the lncRNA *PVT1* in the regulation of the *MYC* (myelocytomatosis) oncogene in certain human cancer cell lines. The

MYC and *PVT1* loci, referred to as the 8q24.21 region, are located in adjacent regions of the genome and often are gained in human tumors, such as in breast cancers [47]. It was reported previously that phosphorylation of threonine 58 of the *MYC* protein promotes *MYC* degradation [48]. Here, the authors showed in human cancer cell lines that the *PVT1* lncRNA and *MYC* co-localized in the nucleus and that *PVT1* binds to the *MYC* protein, protecting it from phosphorylation, and consequently stabilizing the *MYC* protein. Indeed, increased *PVT1* lncRNA levels correlated with increased *MYC* protein levels, but did not affect *MYC* mRNA levels. Consistent with this regulation, knocking out the *Pvt1* locus in mouse using Crispr/Cas9 technology led to reduced cell proliferation and impaired tumor formation in colon cancer cells.

Importantly, the finding that the *PVT1* lncRNA regulates *MYC* phosphorylation is reminiscent of the regulatory action of another lncRNA, *lnc-DC*. During human dendritic cell differentiation, *lnc-DC* binds to STAT3 in the cytoplasm and prevents STAT3 dephosphorylation, thus leading to STAT3 activation [49]. The molecular mechanisms employed by *lnc-DC* and *PVT1* suggest that modulation of phosphorylation levels may be a common tactic used by lncRNAs to regulate the levels of their protein targets.

Another recent study reported a role for the *BCAR4* lncRNA in breast cancer metastasis [50]. In this study, chemokine-induced *BCAR4* binding of the transcription factors SNIP1 and PNUTS promoted cell migration through the activation of a non-canonical hedgehog/GLI2 transcription cascade. Because elevated levels of *BCAR4* correlate with increased metastatic potential in breast cancer patients, the authors targeted this lncRNA by intravenous injection of locked nucleic acids (LNAs) antisense oligos in mouse, which led to a marked suppression of breast cancer metastasis, highlighting the utility of lncRNA LNAs as a cancer therapeutic agent.

In addition to these lncRNA examples, a recent study profiling the lncRNA landscape in a human lung cancer transcriptome database containing more than 500 tumor types identified over 3000 novel lncRNAs not previously annotated and more than 100 lung cancer-associated lncRNAs, many of which appear to play roles in cellular proliferation. In addition, analysis of over 300 additional tumor types identified both lung cancer-specific lncRNAs as well as a subset of lncRNAs that are broadly deregulated in many human cancers [51]. Furthermore, Tseng and co-authors surveyed the Cancer Genome Atlas database, representing more than 15,000 tumors, and showed that 18% of these tumors showed a gain in copy-number of the *PVT1*-containing 8q24.21 region and that nearly 100% of this subpopulation also showed a *MYC* copy number increase [47]. Although the correlations in this latter study are certainly strong, additional studies will be necessary to determine if *PVT1* RNA-directed *MYC* deregulation is a general mechanistic feature of multiple cancer types.

These studies as well as many others suggest that lncRNAs are likely important actors in disease and cancer biology. It is important to point out that in most cases it is not known whether these deregulations are the cause or the consequence of cancers. However, with the accumulating evidence that lncRNA expression profiles are dramatically altered in multiple

cancer types, their role in cancer certainly warrants further investigation (see [52, 53] for focused reviews of this subject).

4.2 lncRNAs and development

Because lncRNAs impinge upon both transcriptional and posttranscriptional regulatory activities, they have the potential to affect numerous developmental networks and cellular processes. As such, lncRNAs have been proposed to play major roles in a diverse array of developmental processes. However, there are only a handful of examples where *in vivo* lncRNA function has been examined at the organismal level by direct genetic disruption of lncRNA loci. Recently, two *in vivo* roles of the *Neat1* lncRNA have been described in knock-out mice. The mammalian *Neat1* lncRNA is highly expressed and is required for the formation of paraspeckles [54–56], dynamic nuclear structures that have been proposed to indirectly control gene expression by sequestering certain transcriptional regulators and splicing factors, to play a role in regulating the activity of its protein components by directing their nuclear localization, and to act as storage depots for nuclear-retained RNAs, edited RNAs and dsRNA [57–62]. Although its role as a gene expression regulator has been established in cell culture studies, its *in vivo* role has remained unclear until now. Indeed, the original study of a *Neat1* knock-out mouse line reported that *Neat1* mutant mice developed normally, in a manner indistinguishable from wild-type mice, displaying no obvious growth, viability or behavioral abnormalities. However, two recent studies, further characterization of *Neat1* knock-out mice [63] and analysis of *Neat1* knock-out progeny mice [64] revealed interesting phenotypes.

In the first of these two studies, analysis of pregnancy frequencies showed that adult female homozygous *Neat1* knock-out mice stochastically displayed reduced pregnancy rates compared to wild-type mice [63]. In wild-type pregnant mice, *Neat1* expression as well as paraspeckle number increased during the formation of the corpus luteum, a temporary structure in female mammals that secretes high levels of progesterone and is essential for the establishment and maintenance of pregnancy. By contrast, paraspeckle formation was not observed in a fraction of *Neat1* knock-out mice. Further phenotypic analysis revealed that nearly 50% of *Neat1* knock-out mice displayed reduced serum progesterone levels and impaired formation of the corpus luteum. The reduced fertility of *Neat1* knock-out mice was restored by subcutaneous progesterone administration, suggesting that the fertility defects were primarily due to impaired progesterone synthesis. Although this study clearly points to the involvement of *Neat1* in corpus luteum development, it remains unclear why only a fraction of the *Neat1* knock-out mice display this phenotype.

In the second study, a phenotypic analysis of mouse pups deriving from homozygous *Neat1* knock-out females revealed a reduced survival rate as well as reduced size and weight compared to pups deriving from wild-type mothers [64]. Through a series of elegant phenotypic analyses of mammary glands, as well as genetic and molecular analyses, the authors attributed the defects observed in the progeny of the *Neat1* knock-out mothers to aberrant mammary gland morphogenesis and faulty lactation. Further characterization of these phenotypes revealed impaired paraspeckle formation in the mammary gland of *Neat1* knock-out mice compared to wild-type mice and pointed to the inability of *Neat1*-mutant

cells to maintain proliferation rates during lobular-alveolar development. Although the exact mechanism of *Neat1* action remains elusive, the authors suggested that the formation or overall number of paraspeckles could be linked to proper mammary gland development and lactation. These studies add to the growing number of lncRNA mutant mice that have been generated to date that display obvious developmental anomalies.

Indeed, contrary to the handful of knock-out studies that have revealed phenotypes for lncRNA mutants, several studies suggest that the developmental contributions of lncRNAs are overestimated and are likely much lower than that of protein-coding genes. For example, although molecular phenotypes have been observed in cell culture for the disruption of several lncRNAs including *Malat1*, no obvious physiological or pre- or post-natal developmental anomalies were observed in the corresponding knock-out mice [65–67]. Moreover, no changes in global gene expression or splicing were observed in *Malat1* knock-out mice. Indeed, only a small number of genes were deregulated in *Malat1* mutants, including several genes that are genomically adjacent to the *Malat1* gene, suggesting that *Malat1* has cis-regulatory roles in gene expression. This limited impact on gene expression was surprising for the highly and widely expressed lncRNA *Malat1*, particularly given the multitude of molecular and developmental phenotypes previously observed upon *Malat1* disruption in cell lines [66, 68–71].

In another study examining the *in vivo* role of the highly-expressed and deeply sequence-conserved *Visc-2* vertebrate lncRNA, whose brain-specific expression is both spatially and temporally controlled, no obvious defects or anomalies were observed despite extensive anatomical and behavioral analyses of *Visc-2* mouse knock-out lines [72]. In addition to *Visc-2* knock-out mice, only five out of 18 recently reported lncRNA deletion mouse lines displayed detectable phenotypes at the organismal level [73]. Among these five mutants, two (*linc-Brn1b* and *linc-Pint*) displayed growth defects and three (*Fendrr*, *Peril*, *Mdgt*) showed lethal phenotypes. However, *Fendrr* was the only knock-out mutant displaying a fully penetrant perinatal lethal phenotype due to defects in multiple organs [73]. A previous study also generated *Fendrr* loss-of-function mouse mutants by using a strategy alternative to deleting the *Fendrr* locus [74]. In this study, *Fendrr* loss-of-function mouse mutants were made by inserting a strong transcriptional stop signal just after the *Fendrr* transcription start site. Like the *Fendrr* mouse deletion mutant, these mutants also displayed heart formation defects, leading ultimately to early death [74, 75]. However, unlike the deletion mutant, the organ defects in these *Fendrr* loss-of-function mutants were restricted to the heart, possibly reflecting differences in the nature of the *Fendrr* knock-out strategies used in these two studies. Importantly, the defects in heart development observed in the *Fendrr* loss-of-function mutants were rescued by expressing *in trans* *Fendrr* under the control of its endogenous promoter [74]. This rescue experiment sets this *in vivo* study apart from other knock-out studies because it directly implicates the absence of the *Fendrr* lncRNA as the cause of the developmental defects and excludes the contribution of other genomic perturbations, such as alterations in chromatin structure, that could have been inadvertently introduced by manipulating the *Fendrr* locus.

4.3 lncRNA knock-out versus knock-down studies

As with protein-coding gene mutants, the absence of a lncRNA mutant phenotype under a restricted set of laboratory conditions is not sufficient to label the lncRNA as nonfunctional. Indeed, important roles for lncRNAs could be revealed under certain physiologically relevant stress conditions or in subpopulations of mutant progeny under particular conditions. However, researchers must proceed with caution when inferring the roles of lncRNAs at the organismal level from results in cell culture or from knock-down rather than knock-out studies, particularly in the absence of phenotypic rescue experiments. Moreover, a recent report highlighted the complexities of lncRNA functional analysis at the organismal level [76]. Previously, splice site-targeted morpholino-directed knock-down of the zebrafish lncRNA *megamind* revealed roles in both brain and eye development [5]. Importantly, these phenotypes were directly linked to the down-regulation of *megamind* because the defects were rescued by injecting either the mature *megamind* RNA from zebrafish or its human or mouse ortholog. However, a zebrafish line that has a 3195bp deletion spanning the conserved region of the *megamind* locus and that does not express detectable *megamind* transcripts did not recapitulate the knock-down developmental defects seen in the morpholino knock-down zebrafish. In fact, the *megamind* knock-out mutant was indistinguishable from wild-type zebrafish [76]. Curiously, this “absence of developmental defects in knock-out mutants” was not limited to *megamind* mutants but extended to additional gene knock-out lines including protein-coding genes that were previously shown to display defects upon morpholino-induced or other transient knock-down methods. Together, these studies emphasize the need for the systematic production, screening and thorough characterization of numerous lncRNA knock-out mutants both to obtain a more global view of the *in vivo* contributions of lncRNAs and to confirm, at the organismal level, the cellular roles and molecular mechanisms that have been observed in specific cell types and cell culture lines or in whole organism knock-down mutants.

5 Molecular functions of vertebrate lncRNAs

As with lncRNA organismal-level functions, our understanding of their molecular mechanisms of action is progressing at a fast pace. Although the regulatory mechanisms of most lncRNA still remain enigmatic, several studies have shown that vertebrate lncRNAs localize in both the nucleus and the cytoplasm and can act at the transcriptional, posttranscriptional or posttranslational level [77]. However, because only a few lncRNAs have been studied in detail using techniques such as classical or single molecule RNA FISH (fluorescence in situ hybridization) [78], general functional principles have not emerged yet, particular for cytoplasmically-localized lncRNAs. Indeed, even though the majority of lncRNAs appears to be cytoplasmic [23, 79], currently more is known about the molecular mechanisms of action of those lncRNAs that preferentially localize in the nucleus.

5.1 Nuclear lncRNAs

5.1.1 lncRNAs and genome architecture—Nuclear lncRNAs can act in *cis* by regulating the expression of their neighboring genes and in *trans* by regulating spatially distant target genes [80]. Indeed, several lncRNAs have been shown to regulate or act in concert with their genomically adjacent transcription factor genes, suggesting that

modulation of transcriptional programs is a common functional theme of lncRNAs [18, 67, 81, 82]. However, one of the most commonly reported functions of nuclear lncRNAs is their association with chromatin modifying complexes [4, 10, 83, 84]. Results from these studies and others have led to the proposal that nuclear lncRNAs guide these ribonucleoprotein (RNP) complexes to target genes to activate or repress their expression. This model is particularly appealing in cases where the protein complexes display no obvious DNA binding sequence specificity, such as in the case of PRC2. Although several nuclear lncRNAs have been shown to localize to numerous genomic sites [85] and associate with chromatin modifying proteins such as PRC1 and PRC2 [4, 10, 75, 86–91], the H3K4 histone methyltransferase TRXG/MLL [75], MEDIATOR [92] and the DNA methyltransferase DNMT1 [80, 93, 94], the mechanistic details of how lncRNAs specifically recognize their targets and recruit chromatin modifying complexes are still not understood. Moreover, how lncRNAs that often are present at low cellular levels [5, 80, 82, 95, 96] can recruit protein complexes to numerous (hundreds or thousands) genomic sites is not clear. Recently, a model alternative to specific interactions between chromatin modification complexes and lncRNAs has been proposed. In this model, PRC2, a histone methyltransferase localized to thousands of genomic loci and required for maintaining epigenetic silencing [97, 98], promiscuously binds to nascent lncRNAs to scan for target genes that have escaped repression [99]. A complementary study showed that EZH2, a subunit of PRC2, has high affinity for RNA and generally binds with low specificity [100]. Due to contrasting reports regarding the role of lncRNAs in chromatin regulation and because the role of lncRNAs in chromatin regulation has been reviewed extensively elsewhere [98, 101–103], we refer readers to these citations for further reading on this subject.

In addition to directing broadly chromatin regulation, several studies suggest that specific lncRNAs can function as nuclear organizing factors that establish discrete nuclear domains. For example, and as discussed above, the *Neat1* lncRNA is necessary for the assembly of paraspeckles [60], whereas a more classic example of a lncRNA that has a role in nuclear positioning is *Xist*, which directs silencing and repositioning of the X chromosome within the nucleus [104]. The role of lncRNAs in modulating nuclear architecture appears to be highly complex. Indeed, although X chromosome inactivation in mammals was first reported over 30 years ago and the *Xist* ncRNA gene was discovered in the early 1990s, the mechanism by which it mediates X inactivation remains poorly understood [105, 106].

Moreover, the spatial location of a given lncRNA transcription unit may guide how it shapes local nuclear genomic architecture. As such, the protein complexes, particularly in the case of chromatin modifying complexes that interact with lncRNAs likely also play an important role in determining how lncRNAs influence chromatin structure. For example, the lncRNA *CCAT1-L*, which is positioned within a strong enhancer sequence upstream of *MYC*, was recently shown to be specifically expressed in human colorectal cancer and to play a role in *MYC* transcriptional regulation. *CCAT1-L* lncRNA promotes chromatin looping, and its knock-down led to a reduction in the interactions between the *MYC* promoter and its enhancer sequences [107].

In addition, a recent study aimed at identifying RNA-RNA interactions used a method based on RNA antisense purification (RAP) to systemically examine the RNA contacts of the RNA

splicing component *U1* and the lncRNA *Malat1*, which localizes to nuclear speckles [108]. Both RNAs were shown to interact with nascent RNA transcripts, and *Malat1* was shown to interact with pre-mRNAs via their protein partners. Due to their contacts with nascent pre-mRNAs, both *U1* and *Malat1* localized to chromatin sites of active genes, showing that RNA-RNA interactions can have a role in targeting lncRNAs to specific genomic regions [108]. Although the mechanistic details of this process are not fully understood, it is evident that lncRNAs are important players in determining nuclear organization.

5.1.2 Techniques to dissect the role of nuclear lncRNAs in chromatin

organization—Because lncRNAs have been proposed to play key roles in modulating chromatin architecture by mediating DNA looping and bringing distant chromosomal regions in contact, techniques that allow the 3D visualization of chromatin structure [109–111], such as the chromatin conformation capture (3C) technique and its many derivatives (4C, 5C, 6C, Hi-C, CHIA-PET and Capture-C), will be useful for dissecting the contribution of individual lncRNA to chromatin structure [112]. Indeed, studies have revealed that one major way that lncRNAs influence gene expression is by modulating chromatin topology, thus understanding how individual lncRNAs influence local or global chromatin structure will help to identify their potential gene targets and to dissect the mechanism underlying how a given lncRNA regulates gene expression.

Already, several techniques have been developed to identify the genomic regions bound by lncRNAs. Capture hybridization analysis of RNA targets (CHART), which uses biotinylated oligonucleotides complementary to a lncRNA of interest to isolate the DNA sites bound by specific lncRNA [113]. As such, CHART can be used to determine the *trans* genomic target sites as well as protein targets of endogenous human and fly lncRNAs. In this study, multiple genomic binding sites of the roX2 lncRNA, a ncRNA involved in dosage compensation in *Drosophila*, were determined. ChiRP-Seq (Chromatin Isolation by RNA purification), a technique similar to CHART, uses tiling oligonucleotides to retrieve specific lncRNAs together with their bound proteins and DNA sequences combined with massively parallel sequencing to identify genomic regions bound by lncRNAs or with a ribonucleoprotein containing the lncRNA of interest [114]. In addition, a variation on the ChiRP technique called dChiRP (domain-specific chromatin isolation by RNA purification) has been used both to identify intra- and inter-molecular RNA interactions with DNA and protein as well as other RNAs, and also to precisely map the binding sites. Indeed, in the case of the roX1 lncRNA, dChiRP was used to tease apart its domain architecture, pinpoint its protein- and chromatin-interacting domains, and determine its localization on the X chromosome [115]. Importantly, these techniques, when combined with 3C-based techniques, result in an extremely powerful and broad toolbox of strategies that can be used to reveal, on a genome-wide scale, the genomic sites, RNA and proteins that interact with a given lncRNA and to identify how chromatin structure, and thus global gene expression, is influenced by RNAs. However, one caveat of these techniques is that so far they have only been tested on abundant lncRNAs or in lines over-expression lncRNAs. Thus, it remains to be shown if these techniques will also be successful in the study of low abundance lncRNAs, which represent the majority of lncRNAs.

5.1.3 lncRNA and protein regulation—In addition to having roles in the establishment of chromatin structure and influencing gene expression, several lncRNAs have been implicated in the regulation of protein stability or activity. As discussed in the previous section, it was recently reported that the *PVT1* lncRNA and the MYC protein colocalize in the nucleus of the SK-BR-3 human breast-cancer cell line, and that the *PVT1* lncRNA stabilizes the MYC protein through interactions that protect it from threonine 58 phosphorylation, a modification that normally promotes MYC degradation [47]. A reduction in *PVT1* lncRNA levels led to decreased MYC protein accumulation, whereas gain of *PVT1* expression led to high MYC protein levels, thus promoting cancer in human cancer cells. This nuclear modulation of protein stability is reminiscent of the regulatory action of the cytoplasmic lncRNA, *lnc-DC*. During human dendritic cell differentiation, *lnc-DC* binds to the STAT3 transcription factor in the cytoplasm and prevents STAT3 dephosphorylation, thus leading to STAT3 activation [49]. The molecular mechanisms employed by *lnc-DC* and *PVT1* suggest that modulation of phosphorylation levels may be a common tactic used by both nuclear and cytoplasmic lncRNAs to regulate the levels of their protein targets.

5.1.4 Enhancer RNAs—Another group of ncRNAs that differs from classic lncRNAs are the enhancer RNAs (eRNAs) that are transcribed from the majority of active enhancer elements. In contrast to conventional lncRNAs, eRNAs are unspliced and rapidly degraded by the exosome. It has been proposed that eRNAs can function locally to promote/facilitate enhancer-promoter interactions. A recent report implicated eRNAs in the expression of immediate-early genes in neurons [116]. Through knock-down and RNA interaction assays, eRNAs were shown to bind negative elongation factor (NELF) complex, facilitating its transient disassociation from promoter sequences and leading to the release of paused RNA polymerase II and more productive elongation. Although, it is unclear how eRNAs target particular promoter sequences, the authors propose that chromatin looping might play a role in bringing enhancer sequences, and thus eRNAs, in proximity to specific promoter sequences.

5.1.5 Viral-derived lncRNAs—In addition to endogenously-derived nuclear lncRNAs, lncRNAs have also been shown to be transcribed from exogenous sources such as the case of the abundant Epstein-Barr virus nuclear lncRNAs EBER1 and EBER2 [117]. Recently, CHART analysis revealed that the Epstein-Barr virus (EBV)-derived lncRNA EBER2 binds the terminal repeat (TR) region of EBV, a region that also binds the B cell transcription factor PAX5 [118]. Indeed, EBER2 was shown to interact both with the PAX5 protein and with transcripts deriving from the LMP2 gene of the TR region. This RNA-RNA base pairing, which appears to be evolutionarily conserved to the primate herpesvirus CeHV15, was shown to be important for recruiting host PAX5 to the virus TR target site, and knock-down of EBER2 phenocopied PAX5 knockdown and led to decreased viral lytic replication.

5.2 Cytoplasmic lncRNAs

5.2.1 lncRNAs as small RNA competitors or scaffolds—Although the modes of action of some cytoplasmic lncRNAs have been reported, far less is known about cytoplasmic lncRNAs compared to nuclear lncRNAs. One popular model is that they serve as decoys or molecular sponges for small regulatory RNAs and their associated Argonaute

complexes [119–122]. In addition, several studies proposed that cytoplasmic lncRNAs could compete with miRNAs for target mRNA binding leading to derepression of the targeted mRNAs. Recent examples of lncRNAs acting as competing endogenous RNAs (ceRNAs) include *lincMD1*, which controls muscle differentiation by sponging away myogenic *miR-133* and *miR-135* to regulate expression of transcription factors that activate muscle-specific gene expression [119], lncRNA *H19*, which contains four *let-7* binding sites and appears to regulate muscle differentiation by acting as a molecular sponge modulating *let-7* availability in the cell [120], and *lncRNA-RoR*, which was reported to act as a ceRNA and to regulate human embryonic stem cell self-renewal by sponging miRNAs away from the core pluripotency transcription factors NANOG, OCT4 and SOX2 [122].

Several recent studies have reported that lncRNAs act as miRNA sponges or decoys, but the relevance of this activity remains unclear due to the small number of miRNA binding sites in most lncRNA transcripts and the low expression levels of lncRNAs relative to the miRNAs they are proposed to regulate [121, 123, 124]. A recent study tested the ceRNA model in adult mice and quantified the stoichiometric relationship between miR-122 and its validated targets [125]. The authors calculated the number of RNA molecules needed to effectively sponge away miR-122 from its mRNA and de-repress the endogenous mRNAs normally repressed by miR-122 [125]. The study concluded that, in general, lncRNAs are unlikely to be effective endogenous regulators because the majority of known lncRNAs do not reach the level needed to sponge miRNAs expressed at relevant levels and because most lncRNAs do not contain multiple miRNA binding sites, a characteristic necessary for RNAs to be effectively targeted by miRNAs [125]. However, at least one exception to this conclusion is the lncRNA *CDRIas*, a highly expressed, circular RNA that contains more than 70 conserved *miR-7* sites [126, 127] and reaches a high copy number per cell.

This study provides substantial evidence that argues against the widespread function of lncRNAs as microRNA sponges, suggesting mechanisms alternative to the titration of miRNA/Argonaute complexes when lncRNAs and small RNAs interact. Some of these interactions show striking sequence conservation and complementarity, with pairing between the miRNA and lncRNA extending well beyond the typical 5' seed pairing that typifies most miRNA targeted mRNAs, suggesting that these interactions are functionally relevant. We recently found a nearly perfect *miR-7* site in the most conserved region of *cyrano* [5], a lncRNA that is expressed at approximately 20–40 molecules per cell (A.S. unpublished data). The 26-nucleotide *miR-7* site in *cyrano* transcript is perfectly conserved from mammals to basal vertebrates and blocking the *miR-7* site in zebrafish with an antisense morpholino leads to developmental defects similar to the loss of *cyrano* function [5].

So far, the most established model for cytoplasmic lncRNA function is their cross-talk with small RNAs, one can imagine that cytoplasmic lncRNAs, like nuclear lncRNAs, may act as molecular scaffolds by assembling various proteins into functional complexes, regulating protein localization or directing protein degradation. As many cytoplasmic lncRNAs are engaged with ribosomes [32–35, 38, 79], it is possible that some lncRNAs produce short functional peptides, or that the act of ribosome binding may serve to regulate localized protein synthesis or RNA stability (see the section “Distinguishing between protein-coding and noncoding functions” above for a detailed discussion).

5.2.2 Identification of functional lncRNA-protein complexes—Whether cytoplasmic or nuclear, in all of the proposed mechanistic models, lncRNAs form functional ribonucleoprotein (RNP) complexes with proteins. As such, the identification of proteins interacting with lncRNAs is key to understanding the molecular functions of lncRNAs. To date several RNA-centric methods have been described for identifying both *in vitro* and *in vivo* RNA-protein interactions [128] and more recently, several studies have reported protein binding at the transcriptome-wide scale [129, 130]. These types of global studies, as well as more direct analyses profiling lncRNA-binding proteins will not only help to expand our knowledge of the molecular functions of lncRNAs but will also help to identify functional lncRNA motifs that have eluded efforts thus far due to lack of sequence conservation.

One of the most commonly used methods for the identification of lncRNA interacting proteins is conventional RNA pull down assays followed by mass spectrometry. Briefly, an exogenously expressed and typically biotin-labeled lncRNA of interest is captured with streptavidin beads and subsequently incubated with a cell lysate to allow the formation of RNPs. The bound proteins are then eluted and identified by mass spectrometry or Western blot detection. Using these types of RNA pull down assays, multiple studies have reported several lncRNA interacting proteins [80, 82, 88, 89, 131–133]. However, this technique has some major limitations: first, the repertoire of identified interaction partners is usually limited to a few abundantly expressed, “sticky” RNA-binding proteins, as such, this technique has a high false positive rate. Thus, defining RNA controls for exogenous expression of lncRNAs are critical in order to limit the number of false positives, but the identity of these controls is not always evident and should be customized to each lncRNA. Secondly, this technique does not provide information about the function of the bound proteins, but rather simply that they are associated with a given lncRNA and a given time-point. Indeed, not all proteins that are bound to lncRNAs are part of the functionally relevant RNP complex.

An alternative high-throughput strategy to RNA pull down assays has recently been applied to the human terminal differentiation-induced ncRNA (*TINCR*). In this study, evaluation of the binding capacity of *TINCR* to more than 9,000 human proteins identified direct *TINCR* binding proteins, with STAUFEN 1 (STAU1) displaying the strongest *TINCR* RNA binding [134]. Although this technique has the ability to identify global interactions, at present, this type of protein array binding method has not been applied to other lncRNAs.

The CHART technique discussed above [113] has been adapted to identify the proteins associated with a few strongly expressed lncRNAs [85, 135]. Indeed, using the original CHART technique in combination with mass spectrometry (CHART-MS), proteins associated with lncRNAs of interest can be identified. In one of these studies, the proteins PSPC1, PSF and SRSF1, which are known to interact with the lncRNAs *NEAT1* and *MALAT1*, were confirmed by CHART-MS, but also several additional proteins not previously known to bind to *NEAT1* or *MALAT1* were identified [85], indicating that this assay might be useful as an alternative approach to the conventional RNA pull-down assays. However, because these lncRNAs are among the most highly abundant known lncRNAs, it is likely that more sensitive techniques will need to be developed for the efficient isolation of

proteins interacting with much less abundant lncRNAs, which represent the majority of lncRNAs.

Today, the discrepancy between the variety of cellular processes reported to involve lncRNAs and the limited number of identified lncRNA interacting proteins suggests that the protein-binding assays used currently have several technical limitations. As the identification of proteins interacting with lncRNAs is paramount to the characterization of lncRNA functional domains and their biological functions, it is clear that the lncRNA field would greatly benefit from the development of new robust high-throughput methods for RNA interacting protein identification.

6 Strategies for the identification of novel *in vivo* lncRNA functions

While forward genetic screens have been used successfully for the functional dissection of many protein-coding genes, these approaches are more limited in identifying the functions of lncRNAs for several potential reasons. It is possible that the majority of lncRNAs is not functional or has only limited effects on development under typical laboratory conditions. It is also possible that, given the absence of the restrictions placed by codon sequences, lncRNA sequences exhibit higher nucleotide flexibility and, as such, are inherently more resistant to point substitutions than coding RNAs. Indeed, in our recent study, introduction of 5-9 nucleotide point mutations in the most conserved regions of the lncRNAs *cyrano* and *megamind* led to loss of lncRNA *in vivo* functions and developmental abnormalities, while more limited substitutions (1-3 nucleotides) within these regions did not have the same negative consequences [5]. Furthermore, as with protein-coding genes, redundancies among lncRNA loci could mask the appearance of loss-of-function phenotypic consequences due to genetic compensation. For example, there are three zebrafish and two human *megamind* loci [5] that could potentially have redundant or partially redundant functions based on their sequence similarities. Lastly, it is not clear for some lncRNAs, if it is the RNA transcript itself that performs the important biological function or if it is the act of transcription of the lncRNA locus that is the functionally relevant activity. In such cases, point mutations are likely to have little or no effect on the functional output of the lncRNA. Therefore, as an alternative approach to forward genetic screens, several reverse genetic strategies involving either transient knock-down of lncRNAs or the creation of stable mutant lines have been employed to investigate their *in vivo* functions in either cell culture or at the organismal level (reviewed in [136]).

6.1 lncRNA knock-down

To date, most functional *in vivo* lncRNA studies have used knock-down assays that rely on antisense technologies, such as RNAi and morpholinos, which lead to transient knock-down. These strategies were employed primarily because they were relatively fast, cost-effective and could be engineered to target a precise RNA sequence while minimizing off-target effects. The main caveats of the knock-down experiments are their transient and incomplete nature. In addition, many of the knock-down systems do not permit the establishment of rescue experiments, which are crucial to test the specificity of the antisense technology. We recently pioneered a lncRNA knock-down strategy in zebrafish embryos that is conducive to

rescue experiments by injecting antisense morpholino oligos targeting either splice sites or conserved regions of the lncRNAs *cyrano* and *megamind* [5]. Injection of splice-site or conserved-site morpholinos both led to similar developmental defects in zebrafish that were rescued efficiently by co-injection of the splice-site morpholinos with mature, spliced lncRNAs that are resistant to morpholinos [5]. Given the relatively straightforward implementation of antisense oligo technology coupled with appropriate sequence specificity controls, knock-down strategies remain a valuable tool for the initial assessment of lncRNA functionality.

6.2 lncRNA knock-out

More recently, the availability and applicability of the Crispr/Cas9 genome editing system has revolutionized the creation of directed genomic knock-out lines (citation). Indeed, various Crispr/Cas9 plasmids have been adapted to multiple organisms, making it possible to quickly establish robust genetic knock-out lines in several model organisms as well as in numerous cell lines. As such, the Crispr/Cas9 system is an invaluable tool for the *in vivo* study of vertebrate lncRNA functions. In addition, an emerging alternative to lncRNA depletion with antisense oligos is the application of modified a Crispr/Cas9 system leading either to CRISPR interference (CRISPRi)-mediated transcriptional repression [137] or endonucleolytic cleavage of the targeted single-stranded lncRNAs [138]. Because conventional RNAi is mediated by cytoplasmic Argonaute proteins and because knock-down through this approach is best suited to cytoplasmic lncRNAs, these Crispr/Cas9 technologies are an attractive alternative to deplete nuclear lncRNAs. Moreover, application of CRISPRi is suitable for setting up genome-wide screens [139, 140], which could be used for unbiased determination of lncRNA functionality.

Because lncRNAs are not translated, generating lncRNA genetic loss-of-function mutants is more technologically challenging than for protein coding genes where a single point mutation can lead to a frame shift or a stop codon and result in the loss of function of the gene of interest. Because this strategy is likely not applicable to the majority of lncRNAs, several alternative strategies to perturb *in vivo* lncRNA expression are employed routinely including the deletion of whole lncRNA loci [66, 76, 141], the deletion of promoter regions [142–145], the insertion of a premature polyadenylation (polyA) termination signal [74, 146, 147], the replacement of a lncRNA locus with a reporter gene [65, 73, 148], and the deletion of putative functional domains in cases where lncRNA function has been characterized. Recently, using these different strategies to achieve loss of lncRNA function, several genetic lncRNA mouse mutants, as well as numerous lncRNA mutant cell lines have been generated and analyzed (reviewed in [136, 149]).

However, these various genomic disruption strategies have both strengths and weaknesses. Indeed, one of the caveats of removing a large genomic sequence to achieve lncRNA loss of function is that these types of deletions may inadvertently remove additional regulatory DNA sequences, such as enhancer regions or regions important for maintaining local chromatin architecture, that can affect the observed phenotypic readout. Thus, an essential step to distinguish between an RNA- and a DNA-dependent effect is rescue experiments. These types of experiments entail either expressing in the stable deletion line the lncRNA

from its endogenous regulatory elements or directly injecting lncRNA molecules, as is done commonly in zebrafish embryos. In addition, by performing rescue experiments with cognate RNAs from other species, as we have reported for the zebrafish *megamind* and *cyrano* lncRNAs [5], these types of rescue experiments can determine if lncRNA functions are conserved among different species.

Complementary to deleting the full lncRNA locus, deleting only the promoter region can be used to distinguish between the role of the lncRNA transcript and the importance of transcription. This distinction is particularly important because at least for some lncRNA loci, such as *Aim* [150], the act of transcription and not the lncRNA product *per se* is an important regulatory element for at least a subset of its functions. However, caution is warranted when implementing this promoter deletion strategy as some lncRNAs have alternative promoter usage, and, thus, incomplete deletion of promoter sequences may lead to only partial loss of lncRNA function due to the retention of limited transcription, as has recently been shown for the lncRNA *Kcnq1ot1* [144]. In addition to alternative promoter usage, there are some examples where the promoters of lncRNA genes overlap with the promoters or regulatory elements of their adjacent genes. Thus, in these cases, promoter deletions risk not only to affect the expression of the lncRNA but also to impinge upon the expression of the neighboring genes. As such, the utility of this strategy must be evaluated on an individual locus basis and requires a prior minimal characterization of the lncRNA promoter region of interest. An alternative strategy to sequence deletion is the insertion of a premature polyA termination signal. One advantage of this method is that it is less likely to remove regulatory DNA sequences, but it is not a caveat-free strategy mainly because RNA POL II can read-through certain polyA termination signals, resulting in leaky transcription and incomplete loss of lncRNA function.

An additional strategy to investigate lncRNA *in vivo* function and to tease apart the contributions of local sequences contained within a given lncRNA is the removal and/or replacement of individual lncRNA functional domains. Although, currently, the identity of the RNA sequence or structural domains of most lncRNAs remains elusive, this method is an attractive and technologically feasible alternative to full lncRNA locus deletion due to the development of CRISPR/Cas9 technology.

6.3 Model systems for studying biological functions of lncRNAs

With the identification of hundreds of thousands of lncRNAs in species ranging from humans and mice to fish, flies, nematodes, plants, and protozoans [23], the door is now open for functional studies using a large assortment of powerful molecular and genomic *in vivo* techniques. Experimental systems such as zebrafish and chicken, where molecular manipulations are more affordable, efficient and technically feasible than in mammalian systems, and where some degree of evolutionary conservation has been documented among lncRNAs, have already proven valuable for deciphering mammalian lncRNA function [5, 82, 133]. Indeed, functional genetic and molecular analyses of sequence and syntenically conserved lncRNAs in a variety of species facilitates the identification of lncRNA roles that can then directly be tested in mammalian models to determine if lncRNA functions are evolutionarily conserved.

In the case of the recently identified lincRNA *megamind*, where three and two loci exist in zebrafish and human, respectively, identification of these homologous loci was possible based on the relatively long stretches of nucleotide conservation [5]. However, for those lncRNAs with sparse or no detectable conserved nucleotide sequences, finding putative homologs is computationally challenging, thus recognizing functional domains based on features other than sequence or syntenic conservation is a priority. This being said, it is important to mention that conserved lncRNA function is not necessarily restricted to sequence or syntenic conservation. RNA structure also likely plays a direct role in determining lncRNA function. However, current techniques to identify and understand lncRNA 3-dimensional structure are not well developed, are labor intensive and often rely on stretches of sequence conservation or identified functional domains to be reliably predicted. Thus, again, the more we understand about the biological mechanisms of lncRNAs, the more we will broaden our ability to understand the functional contribution of RNA structure in addition to sequence conservation and shared synteny.

7 Conclusions and Perspectives

The field of lncRNAs is rapidly evolving with exciting new functional discoveries of novel noncoding RNA molecules covering a variety of biological and cellular processes. However, we are still at the beginning of our understanding of lncRNA functions as only a handful of vertebrate lncRNAs have been tested in robust genetic models. Lack of unbiased genetic screens makes it particularly difficult to openly search for the potential biological functions of lncRNAs. Given the multitude of functions that have been described for lncRNAs in cell lines, and the surprise that cognate genetic studies in model organisms do not always lead to the phenotypes predicted by cellular analyses, it is clear that the establishment of more genetic models for lncRNAs is paramount.

One way to help to determine which out of the thousands of identified lncRNAs are functional is to divide them into different classes based on their molecular signatures and characteristics shared with already characterized functional lncRNAs. However, at the moment, it is too premature to use this type of categorization to attempt to divide lncRNAs in distinct classes because there is still very little understanding of the molecular mechanisms of action and the *in vivo* functions of most of the lncRNAs. In fact, it is unclear how many unique mechanistic classes of lncRNAs exist or if the majority of lncRNAs use common molecular mechanisms to carry out their functions. In this respect, investigating lncRNA evolution and the selective pressures acting on lncRNA sequences and genomic positions, and determining the subcellular localization of lncRNAs will help to advance our understanding of ncRNA biology and will reveal new principles of lncRNA modularity and functionality.

Our lack of knowledge of the molecular mechanisms of lncRNA action and our limited view of the RNP complexes containing lncRNAs makes it particularly difficult to identify the biological functions of lncRNAs, as we do not know what to look for when evaluating the phenotypic consequences of lncRNA perturbation. Indeed, the defects could be limited in intensity, space or time, restricted to particular cell types, or unique developmental stages due to temporal regulation, or appearing in a stochastic or partially penetrant fashion and,

thus, not easily visible in normal laboratory conditions. Also, the roles of certain lncRNAs may only be revealed under particular stress conditions. Finally, potential functional redundancy is difficult to predict or evaluate, making it hard to “eliminate” potential genetic compensation in lncRNA mutants. As *in vivo* analytical techniques continue to improve and the number of lncRNA knock-out mutants in a range of vertebrate model organisms increases, our knowledge of lncRNA *in vivo* functions and the molecular mechanisms employed by lncRNAs will continue to expand.

Acknowledgements

We apologize to colleagues who have produced primary research on the topic but could not be cited or discussed owing to space limitations. We thank Raphael Margueron and Igor Ulitsky for helpful comments on this manuscript. This work is supported by grants from ERC (FLAME-337440), ATIP-Avenir, Fondation Bettencourt Schueller. We also thank Eric Anderson for his generous donation.

References

1. Ipsaro JJ, Joshua-Tor L. From guide to target: molecular insights into eukaryotic RNA-interference machinery. *Nat Struct Mol Biol.* 2015; 22(1):20–28. [PubMed: 25565029]
2. Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell.* 2014; 157(1):77–94. [PubMed: 24679528]
3. Guttman M, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature.* 2009
4. Khalil AM, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A.* 2009; 106(28):11667–72. [PubMed: 19571010]
5. Ulitsky I, et al. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell.* 2011; 147(7):1537–50. [PubMed: 22196729]
6. Necseulea A, et al. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature.* 2014; 505(7485):635–40. [PubMed: 24463510]
7. Nam JW, Bartel DP. Long noncoding RNAs in *C. elegans*. *Genome Res.* 2012; 22(12):2529–40. [PubMed: 22707570]
8. Pauli A, et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome research.* 2012; 22(3):577–91. [PubMed: 22110045]
9. Ponjavic J, Ponting CP, Lunter G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome research.* 2007; 17(5):556–65. [PubMed: 17387145]
10. Guttman M, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature.* 2011; 477(7364):295–300. [PubMed: 21874018]
11. Sigova AA, et al. Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc Natl Acad Sci U S A.* 2013; 110(8):2876–81. [PubMed: 23382218]
12. Tan MH, et al. RNA sequencing reveals a diverse and dynamic repertoire of the *Xenopus tropicalis* transcriptome over development. *Genome research.* 2013; 23(1):201–16. [PubMed: 22960373]
13. Washietl S, Kellis M, Garber M. Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res.* 2014; 24(4):616–28. [PubMed: 24429298]
14. Jin J, et al. PLncDB: plant long non-coding RNA database. *Bioinformatics.* 2013; 29(8):1068–71. [PubMed: 23476021]
15. Carninci P, et al. The transcriptional landscape of the mammalian genome. *Science.* 2005; 309(5740):1559–63. [PubMed: 16141072]
16. Ravasi T, et al. Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res.* 2006; 16(1):11–9. [PubMed: 16344565]
17. Chodroff RA, et al. Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. *Genome Biol.* 2010; 11(7):R72. [PubMed: 20624288]

18. Ponjavic J, et al. Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. *PLoS Genet.* 2009; 5(8):e1000617. [PubMed: 19696892]
19. Miura P, et al. Widespread and extensive lengthening of 3' UTRs in the mammalian brain. *Genome research.* 2013; 23(5):812–25. [PubMed: 23520388]
20. Zhou VW, Goren A, Bernstein BE. Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet.* 2011; 12(1):7–18. [PubMed: 21116306]
21. Carninci P, et al. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet.* 2006; 38(6):626–35. [PubMed: 16645617]
22. Jan CH, et al. Formation, regulation and evolution of *Caenorhabditis elegans* 3'UTRs. *Nature.* 2011; 469(7328):97–101. [PubMed: 21085120]
23. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell.* 2013; 154(1):26–46. [PubMed: 23827673]
24. Volders PJ, et al. An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res.* 2015; 43(Database issue):D174–80. [PubMed: 25378313]
25. Boley N, et al. Genome-guided transcript assembly by integrative analysis of RNA sequence data. *Nat Biotechnol.* 2014; 32(4):341–6. [PubMed: 24633242]
26. Kong L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 2007; 35(Web Server issue):W345–9. [PubMed: 17631615]
27. Wang L, et al. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* 2013; 41(6):e74. [PubMed: 23335781]
28. Arrial RT, Togawa RC, Brigido Mde M. Screening non-coding RNAs in transcriptomes from neglected species using PORTRAIT: case study of the pathogenic fungus *Paracoccidioides brasiliensis*. *BMC Bioinformatics.* 2009; 10:239. [PubMed: 19653905]
29. Lin MF, Jungreis I, Kellis M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics.* 2011; 27(13):i275–82. [PubMed: 21685081]
30. Washietl S, et al. RNAcode: robust discrimination of coding and noncoding regions in comparative sequence data. *RNA.* 2011; 17(4):578–94. [PubMed: 21357752]
31. Anderson DM, et al. A Micropeptide Encoded by a Putative Long Noncoding RNA Regulates Muscle Performance. *Cell.* 2015
32. Ingolia NT, et al. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep.* 2014; 8(5):1365–79. [PubMed: 25159147]
33. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell.* 2011; 147(4):789–802. [PubMed: 22056041]
34. Chew GL, et al. Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs. *Development.* 2013; 140(13):2828–34. [PubMed: 23698349]
35. Bazzini AA, et al. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J.* 2014; 33(9):981–93. [PubMed: 24705786]
36. Pauli A, et al. Toddler: an embryonic signal that promotes cell movement via Apelin receptors. *Science.* 2014; 343(6172):1248636. [PubMed: 24407481]
37. Chng SC, et al. ELABELA: a hormone essential for heart development signals via the apelin receptor. *Developmental cell.* 2013; 27(6):672–80. [PubMed: 24316148]
38. Guttman M, et al. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell.* 2013; 154(1):240–51. [PubMed: 23810193]
39. Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci U S A.* 2009; 106(18):7507–12. [PubMed: 19372376]
40. Wethmar K, Smink JJ, Leutz A. Upstream open reading frames: molecular switches in (patho)physiology. *Bioessays.* 2010; 32(10):885–93. [PubMed: 20726009]
41. Ruiz-Orera J, et al. Long non-coding RNAs as a source of new peptides. *Elife.* 2014; 3:e03523. [PubMed: 25233276]

42. Andrews SJ, Rothnagel JA. Emerging evidence for functional peptides encoded by short open reading frames. *Nat Rev Genet.* 2014; 15(3):193–204. [PubMed: 24514441]
43. Schweingruber C, et al. Nonsense-mediated mRNA decay - mechanisms of substrate mRNA recognition and degradation in mammalian cells. *Biochim Biophys Acta.* 2013; 1829(6–7):612–23. [PubMed: 23435113]
44. Wolin SL, Sim S, Chen X. Nuclear noncoding RNA surveillance: is the end in sight? *Trends Genet.* 2012; 28(7):306–13. [PubMed: 22475369]
45. Lykke-Andersen S, et al. Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev.* 2014; 28(22):2498–517. [PubMed: 25403180]
46. Han P, et al. A long noncoding RNA protects the heart from pathological hypertrophy. *Nature.* 2014; 514(7520):102–6. [PubMed: 25119045]
47. Tseng YY, et al. PVT1 dependence in cancer with MYC copy-number increase. *Nature.* 2014; 512(7512):82–6. [PubMed: 25043044]
48. Yeh E, et al. A signalling pathway controlling c-Myc degradation that impacts oncogenic transformation of human cells. *Nat Cell Biol.* 2004; 6(4):308–18. [PubMed: 15048125]
49. Wang P, et al. The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation. *Science.* 2014; 344(6181):310–3. [PubMed: 24744378]
50. Xing Z, et al. lncRNA directs cooperative epigenetic regulation downstream of chemokine signals. *Cell.* 2014; 159(5):1110–25. [PubMed: 25416949]
51. White NM, et al. Transcriptome sequencing reveals altered long intergenic non-coding RNAs in lung cancer. *Genome Biol.* 2014; 15(8):429. [PubMed: 25116943]
52. Cheetham SW, et al. Long noncoding RNAs and the genetics of cancer. *Br J Cancer.* 2013; 108(12):2419–25. [PubMed: 23660942]
53. Yang G, Lu X, Yuan L. lncRNA: a link between RNA and cancer. *Biochim Biophys Acta.* 2014; 1839(11):1097–109. [PubMed: 25159663]
54. Clemson CM, et al. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell.* 2009; 33(6):717–26. [PubMed: 19217333]
55. Sasaki YT, et al. MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc Natl Acad Sci U S A.* 2009; 106(8):2525–30. [PubMed: 19188602]
56. Sunwoo H, et al. MEN epsilon/beta nuclear-retained non-coding RNAs are upregulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res.* 2009; 19(3):347–59. [PubMed: 19106332]
57. Yamazaki T, Hirose T. The building process of the functional paraspeckle with long non-coding RNAs. *Front Biosci (Elite Ed).* 2015; 7:1–47. [PubMed: 25553361]
58. Bond CS, Fox AH. Paraspeckles: nuclear bodies built on long noncoding RNA. *J Cell Biol.* 2009; 186(5):637–44. [PubMed: 19720872]
59. Chen LL, Carmichael GG. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Mol Cell.* 2009; 35(4):467–78. [PubMed: 19716791]
60. Mao YS, et al. Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat Cell Biol.* 2011; 13(1):95–101. [PubMed: 21170033]
61. Spector DL. SnapShot: Cellular bodies. *Cell.* 2006; 127(5):1071. [PubMed: 17129789]
62. Fox AH, et al. Paraspeckles: a novel nuclear domain. *Curr Biol.* 2002; 12(1):13–25. [PubMed: 11790299]
63. Nakagawa S, et al. The lncRNA Neat1 is required for corpus luteum formation and the establishment of pregnancy in a subpopulation of mice. *Development.* 2014
64. Standaert L, et al. The long noncoding RNA Neat1 is required for mammary gland development and lactation. *RNA.* 2014
65. Nakagawa S, et al. Malat1 is not an essential component of nuclear speckles in mice. *RNA.* 2012; 18(8):1487–99. [PubMed: 22718948]
66. Eissmann M, et al. Loss of the abundant nuclear non-coding RNA MALAT1 is compatible with life and development. *RNA Biol.* 2012; 9(8):1076–87. [PubMed: 22858678]

67. Zhang B, et al. The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell reports*. 2012; 2(1):111–23. [PubMed: 22840402]
68. Gutschner T, et al. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res*. 2013; 73(3):1180–9. [PubMed: 23243023]
69. Tripathi V, et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell*. 2010; 39(6):925–38. [PubMed: 20797886]
70. Bernard D, et al. A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J*. 2010; 29(18):3082–93. [PubMed: 20729808]
71. Lin R, et al. Control of RNA processing by a large non-coding RNA over-expressed in carcinomas. *FEBS Lett*. 2011; 585(4):671–6. [PubMed: 21266177]
72. Oliver PL, et al. Disruption of Visc-2, a Brain-Expressed Conserved Long Noncoding RNA, Does Not Elicit an Overt Anatomical or Behavioral Phenotype. *Cereb Cortex*. 2014
73. Sauvageau M, et al. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife*. 2013; 2:e01749. [PubMed: 24381249]
74. Grote P, et al. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Developmental cell*. 2013; 24(2):206–14. [PubMed: 23369715]
75. Grote P, Herrmann BG. The long non-coding RNA Fendrr links epigenetic control mechanisms to gene regulatory networks in mammalian embryogenesis. *RNA Biol*. 2013; 10(10):1579–85. [PubMed: 24036695]
76. Kok FO, et al. Reverse Genetic Screening Reveals Poor Correlation between Morpholino-Induced and Mutant Phenotypes in Zebrafish. *Dev Cell*. 2015; 32(1):97–108. [PubMed: 25533206]
77. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annual review of biochemistry*. 2012; 81:145–66.
78. Dunagin M, et al. Visualization of lncRNA by Single-Molecule Fluorescence In Situ Hybridization. *Methods Mol Biol*. 2015; 1262:3–19. [PubMed: 25555572]
79. van Heesch S, et al. Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol*. 2014; 15(1):R6. [PubMed: 24393600]
80. Chalei V, et al. The long non-coding RNA Dali is an epigenetic regulator of neural differentiation. *Elife*. 2014; 3
81. Berghoff EG, et al. Evf2 (Dlx6as) lncRNA regulates ultraconserved enhancer methylation and the differential transcriptional control of adjacent genes. *Development*. 2013; 140(21):4407–16. [PubMed: 24089468]
82. Wang KC, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*. 2011; 472(7341):120–4. [PubMed: 21423168]
83. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. *Nature*. 2012; 482(7385):339–46. [PubMed: 22337053]
84. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Mol Cell*. 2011; 43(6):904–14. [PubMed: 21925379]
85. West JA, et al. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol Cell*. 2014; 55(5):791–802. [PubMed: 25155612]
86. Zhao J, et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Molecular cell*. 2010; 40(6):939–53. [PubMed: 21172659]
87. Schoeftner S, et al. Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J*. 2006; 25(13):3110–22. [PubMed: 16763550]
88. Rinn JL, et al. Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell*. 2007; 129(7):1311–23. [PubMed: 17604720]
89. Tsai MC, et al. Long noncoding RNA as modular scaffold of histone modification complexes. *Science*. 2010; 329(5992):689–93. [PubMed: 20616235]
90. Klattenhoff CA, et al. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell*. 2013; 152(3):570–83. [PubMed: 23352431]
91. Zhao J, et al. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*. 2008; 322(5902):750–6. [PubMed: 18974356]

92. Lai F, et al. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature*. 2013; 494(7438):497–501. [PubMed: 23417068]
93. Di Ruscio A, et al. DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature*. 2013
94. Mohammad F, et al. Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. *Development*. 2010; 137(15):2493–9. [PubMed: 20573698]
95. Clark MB, et al. Genome-wide analysis of long noncoding RNA stability. *Genome research*. 2012; 22(5):885–98. [PubMed: 22406755]
96. Derrien T, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. 2012; 22(9):1775–89. [PubMed: 22955988]
97. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. *Nature*. 2011; 469(7330):343–9. [PubMed: 21248841]
98. Brockdorff N. Noncoding RNA and Polycomb recruitment. *RNA*. 2013
99. Davidovich C, et al. A dimeric state for PRC2. *Nucleic Acids Res*. 2014; 42(14):9236–48. [PubMed: 24992961]
100. Cifuentes-Rojas C, et al. Regulatory interactions between RNA and polycomb repressive complex 2. *Mol Cell*. 2014; 55(2):171–85. [PubMed: 24882207]
101. Quinodoz S, Guttman M. Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. *Trends Cell Biol*. 2014; 24(11):651–63. [PubMed: 25441720]
102. Bonasio R, Shiekhattar R. Regulation of transcription by long noncoding RNAs. *Annu Rev Genet*. 2014; 48:433–55. [PubMed: 25251851]
103. Bergmann JH, Spector DL. Long non-coding RNAs: modulators of nuclear structure and function. *Curr Opin Cell Biol*. 2014; 26:10–8. [PubMed: 24529241]
104. Chaumeil J, et al. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev*. 2006; 20(16):2223–37. [PubMed: 16912274]
105. Gendrel AV, Heard E. Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. *Annu Rev Cell Dev Biol*. 2014; 30:561–80. [PubMed: 25000994]
106. Froberg JE, Yang L, Lee JT. Guided by RNAs: X-inactivation as a model for lncRNA function. *J Mol Biol*. 2013; 425(19):3698–706. [PubMed: 23816838]
107. Xiang JF, et al. Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res*. 2014; 24(5):513–31. [PubMed: 24662484]
108. Engreitz JM, et al. RNA-RNA Interactions Enable Specific Targeting of Noncoding RNAs to Nascent Pre-mRNAs and Chromatin Sites. *Cell*. 2014; 159(1):188–99. [PubMed: 25259926]
109. Sajan SA, Hawkins RD. Methods for identifying higher-order chromatin structure. *Annu Rev Genomics Hum Genet*. 2012; 13:59–82. [PubMed: 22703176]
110. Hughes JR, et al. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet*. 2014; 46(2):205–12. [PubMed: 24413732]
111. Dekker J, et al. Capturing chromosome conformation. *Science*. 2002; 295(5558):1306–11. [PubMed: 11847345]
112. Ma W, et al. Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. *Nat Methods*. 2015; 12(1):71–8. [PubMed: 25437436]
113. Simon MD, et al. The genomic binding sites of a noncoding RNA. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108(51):20497–502. [PubMed: 22143764]
114. Chu C, et al. Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. *Mol Cell*. 2011
115. Quinn JJ, et al. Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. *Nat Biotechnol*. 2014; 32(9):933–40. [PubMed: 24997788]
116. Schaukowitz K, et al. Enhancer RNA facilitates NELF release from immediate early genes. *Mol Cell*. 2014; 56(1):29–42. [PubMed: 25263592]

117. Lerner MR, et al. Two small RNAs encoded by Epstein-Barr virus and complexed with protein are precipitated by antibodies from patients with systemic lupus erythematosus. *Proc Natl Acad Sci U S A*. 1981; 78(2):805–9. [PubMed: 6262773]
118. Lee N, et al. EBV Noncoding RNA Binds Nascent RNA to Drive Host PAX5 to Viral DNA. *Cell*. 2015; 160(4):607–18. [PubMed: 25662012]
119. Cesana M, et al. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*. 2011; 147(2):358–69. [PubMed: 22000014]
120. Kallen AN, et al. The imprinted H19 lncRNA antagonizes let-7 microRNAs. *Mol Cell*. 2013; 52(1):101–12. [PubMed: 24055342]
121. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. *Nature*. 2014; 505(7483):344–52. [PubMed: 24429633]
122. Wang Y, et al. Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. *Dev Cell*. 2013; 25(1):69–80. [PubMed: 23541921]
123. Ebert MS, Sharp PA. Emerging roles for natural microRNA sponges. *Curr Biol*. 2010; 20(19):R858–61. [PubMed: 20937476]
124. Bak RO, Mikkelsen JG. miRNA sponges: soaking up miRNAs for regulation of gene expression. *Wiley Interdiscip Rev RNA*. 2014; 5(3):317–33. [PubMed: 24375960]
125. Denzler R, et al. Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. *Mol Cell*. 2014; 54(5):766–76. [PubMed: 24793693]
126. Hansen TB, et al. Natural RNA circles function as efficient microRNA sponges. *Nature*. 2013; 495(7441):384–8. [PubMed: 23446346]
127. Memczak S, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*. 2013; 495(7441):333–8. [PubMed: 23446348]
128. McHugh CA, Russell P, Guttman M. Methods for comprehensive experimental identification of RNA-protein interactions. *Genome Biol*. 2014; 15(1):203. [PubMed: 24467948]
129. Baltz AG, et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Molecular cell*. 2012; 46(5):674–90. [PubMed: 22681889]
130. Castello A, et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*. 2012; 149(6):1393–406. [PubMed: 22658674]
131. Hacisuleyman E, et al. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat Struct Mol Biol*. 2014; 21(2):198–206. [PubMed: 24463464]
132. Huarte M, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*. 2010; 142(3):409–19. [PubMed: 20673990]
133. Lin N, et al. An Evolutionarily Conserved Long Noncoding RNA TUNA Controls Pluripotency and Neural Lineage Commitment. *Molecular cell*. 2014
134. Kretz M, et al. Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature*. 2013; 493(7431):231–5. [PubMed: 23201690]
135. Rapicavoli NA, et al. The long noncoding RNA Six3OS acts in trans to regulate retinal development by modulating Six3 activity. *Neural Dev*. 2011; 6:32. [PubMed: 21936910]
136. Bassett AR, et al. Considerations when investigating lncRNA function in vivo. *Elife*. 2014; 3:e03058. [PubMed: 25124674]
137. Gilbert LA, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*. 2013; 154(2):442–51. [PubMed: 23849981]
138. O'Connell MR, et al. Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature*. 2014
139. Gilbert LA, et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*. 2014; 159(3):647–61. [PubMed: 25307932]
140. Konermann S, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*. 2014
141. Li L, et al. Targeted disruption of hotair leads to homeotic transformation and gene derepression. *Cell reports*. 2013; 5(1):3–12. [PubMed: 24075995]

142. Fitzpatrick GV, Soloway PD, Higgins MJ. Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nat Genet.* 2002; 32(3):426–31. [PubMed: 12410230]
143. Nakagawa S, et al. Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. *J Cell Biol.* 2011; 193(1):31–9. [PubMed: 21444682]
144. Schultz BM, et al. Enhancers compete with a long non-coding RNA for regulation of the Kcnq1 domain. *Nucleic Acids Res.* 2015; 43(2):745–59. [PubMed: 25539921]
145. Anguera MC, et al. Tsx produces a long noncoding RNA and has general functions in the germline, stem cells, and brain. *PLoS Genet.* 2011; 7(9):e1002248. [PubMed: 21912526]
146. Bond AM, et al. Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nat Neurosci.* 2009; 12(8):1020–7. [PubMed: 19620975]
147. Sleutels F, Zwart R, Barlow DP. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature.* 2002; 415(6873):810–3. [PubMed: 11845212]
148. Ripoché MA, et al. Deletion of the H19 transcription unit reveals the existence of a putative imprinting control element. *Genes Dev.* 1997; 11(12):1596–604. [PubMed: 9203585]
149. Li L, Chang HY. Physiological roles of long noncoding RNAs: insight from knockout mice. *Trends Cell Biol.* 2014; 24(10):594–602. [PubMed: 25022466]
150. Latos PA, et al. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science.* 2012; 338(6113):1469–72. [PubMed: 23239737]
151. Bogdanovic O, et al. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome research.* 2012; 22(10):2043–53. [PubMed: 22593555]
152. Nepal C, et al. Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis. *Genome research.* 2013
153. Ulitsky I, et al. Extensive alternative polyadenylation during zebrafish development. *Genome research.* 2012; 22(10):2054–66. [PubMed: 22722342]

Figure 1. Distinguishing between peptide-coding and noncoding transcripts.

(A) Genomic locus of the zebrafish lincRNA *cyrano*. Multiple available datasets, including H3K4me3 ChIP-Seq [151], [152], RNA-Seq [8], 3P-Seq [153], and ribosome profiling data (RFP-Seq; [34]), should be used to annotate lincRNA genes and distinguish them from protein-coding genes. All mentioned datasets were plotted using the UCSC genome browser and correspond to the dome stage of zebrafish development. All mentioned custom genome browser tracks were plotted and generously shared with us by Igor Ulitsky. Spliced zebrafish ESTs from GenBank, RefSeq annotations and sequence conservation among 5 fish species by PhastCons are also shown. (B) Genomic locus of the peptide-coding *elabela/toddler* gene [36, 37]. All data sets and annotations are as in Figure 1A.