Linking RNA biology to lncRNAs

Loyal A. Goff and John L. Rinn

Genome Res. 2015 25: 1456-1465
Access the most recent version at doi:10.1101/gr.191122.115

Perspective

Linking RNA biology to lncRNAs

Loyal A. Goff[1,2] and John L. Rinn[3,4,5]

[1]McKusick-Nathans Institute of Genetic Medicine, [2]Department of Neuroscience, Johns Hopkins University, Baltimore, Maryland 21205, USA; [3]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA; [4]Department of Pathology, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA; [5]The Broad Institute, Cambridge, Massachusetts 02142, USA

The regulatory potential of RNA has never ceased to amaze: from RNA catalysis, to RNA-mediated splicing, to RNA-based silencing of an entire chromosome during dosage compensation. More recently, thousands of long noncoding RNA (lncRNA) transcripts have been identified, the majority with unknown function. Thus, it is tempting to think that these lncRNAs represent a cadre of new factors that function through ribonucleic mechanisms. Some evidence points to several lncRNAs with tantalizing physiological contributions and thought-provoking molecular modalities. However, dissecting the RNA biology of lncRNAs has been difficult, and distinguishing the independent contributions of functional RNAs from underlying DNA elements, or the local act of transcription, is challenging. Here, we aim to survey the existing literature and highlight future approaches that will be needed to link the RNA-based biology and mechanisms of lncRNAs in vitro and in vivo.

Perhaps one of the biggest surprises of the post-genome era is the vast amount of transcription emanating from the noncoding regions of the genome. The human genome sequence has provided a unique opportunity to systematically survey genomic regions for biological activity. As a logical first proxy for activity, many groups began by mapping observed transcriptional events to the genome (Kapranov et al. 2002, 2005; Rinn et al. 2003; Carninci et al. 2005; Cheng et al. 2005; Maeda et al. 2006; Nordström et al. 2009). These early efforts identified thousands of long noncoding RNA (lncRNA) transcripts that did not appear to code for protein coding genes (Okazaki et al. 2002; Carninci et al. 2005; The ENCODE Project Consortium 2007; Kapranov et al. 2007; Guttman et al. 2009; Nordström et al. 2009; Cabili et al. 2011).

This approach quickly expanded the catalog of putative genomic regions in which RNA transcription may serve as a proxy for functional biological activities. Because RNA has a longstanding history of powerful regulatory responsibilities, with numerous classic examples of RNA-mediated roles in splicing, translation, and genomic imprinting, many have speculated that this vast catalog of transcribed noncoding genes represented a wealth of novel regulatory RNA elements. Indeed, lncRNAs are still routinely detected across a variety of cellular- and tissue-specific contexts with many contributing diverse functional roles (Harrow et al. 2012; Morris and Mattick 2014). However, although it is alluring to think that the sheer numbers of reproducibly expressed RNA species must signify an important biological contribution, an equivalent hypothesis is that these RNA molecules are by-products or transcriptional noise. In fact, current evidence suggests that the answer is somewhere in the middle. We define a functional lncRNA locus as being required for a specific cellular activity, but importantly, not constrained by any particular mechanism. In this regard, a functional locus may produce a gene product with a specific activity or serve to harbor regulatory sequences or properties that facilitate proper cellular physiology. Several lines of evidence support the claim that some lncRNAs are functional through RNA-mediated mechanisms, yet the extent and number of important lncRNAs remains highly contested. While this broader question remains unresolved, there is a clear path forward for identifying physiologically important lncRNA loci and their underlying RNA biology.

Identifying functional lncRNAs has been difficult, as it requires multiple lines of evidence that first identify candidate loci that contribute to a physiological function and then confirm that the observed effect is primarily mediated through an RNA-based mechanism. There is currently no universal experimental approach to characterizing lncRNA functional contributions, owing both to the complexities of transcriptional regulation and the diversity of functions that can be attributed to an RNA molecule. In the end, a thorough characterization of a functional lncRNA gene will involve the incorporation of an array of independent experiments, each designed to elucidate either the effect of the gene on cellular physiology or the precise molecular mechanism through which the gene achieves this effect.

An important first step in this process is to determine whether or not a lncRNA gene has any functional consequences within a particular cellular context. As a reasonable first approach to this question, many groups have performed loss- and gain-of-function studies and, as a result, have identified lncRNAs that reproducibly perturb in vitro physiology (Bond et al. 2009; Guttman et al. 2011; Nagano and Fraser 2011; Sun et al. 2013). With functional evidence accumulating, several groups began using mutant mouse models to test the functional roles of lncRNAs at an organismal scale. Taken together, the aggregate of dozens of studies has begun to show that several lncRNAs are critical regulators of cell and developmental biology (Rinn and Chang 2012; Ulitsky and Bartel 2013).

Yet with thousands more uncharacterized lncRNAs remaining, it is critical to take a step back from a serial model of functional lncRNA analysis. In contrast, efforts should be directed toward the identification and dissection of the RNA domains, sequences,

Corresponding authors: loyalgoff@jhmi.edu, john_rinn@harvard.edu
Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.191122.115. Freely available online through the *Genome Research* Open Access option.

25:1456–1465 Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/15; www.genome.org

structures, and characteristics that may be shared across multiple lncRNAs or other classes of RNAs more broadly. Here, we synthesize current lncRNA literature with a focus on the significance of their underlying RNA-mediated biological activities. For brevity, we have selected a small subset of functional lncRNAs (Table 1) to emphasize specific experimental strategies required to evaluate RNA, DNA, or transcriptional functionalities, or combinations thereof. Readers wishing to explore the current breadth of

Table 1. Summary table of lncRNAs highlighted in this review

Gene	Functional	RNA	DNA	Transcription	in vivo	in vitro	*cis*	*trans*	Activity	References
AIRN	+	+	+	+	+	+	+	–	Regulation of IGF2R via imprinting	Sleutels et al. 2002 Nagano et al. 2008 Latos et al. 2012 Santoro et al. 2013
BORG	?	+	?	?	?	?	?	?	Nuclear localization via pentamer RNA domain	Zhang et al. 2014
DALIR	+	+	?	?	?	+	+	+	Proximal regulation of transcription and distal regulation of promoter methylation at select target genes	Chalei et al. 2014
EVF2	+	+	?	+	+	+	+	+	Recruitment of regulatory proteins to intergenic enhancers (*trans*) and antisense competition with *DLX6* (*cis*)	Bond et al. 2009
FENDRR	+	+	+	?	+	+	+	?	Interaction with PRC2 complex, required for viability	Grote et al. 2013 Sauvageau et al. 2013
FIRRE	+	+	+	?	?	+	+	+	Required for adipogenesis. 3D organization of specific genomic sites within the nucleus	Sun et al. 2013 Hacisuleyman et al. 2014 Bergmann et al. 2015
Group I introns	+	+	–	–	+	+	+	–	Catalysis of self-excision from primary transcripts	Cech 1990
H19	+	+	–	+	+	+	+	?	Imprinting of *IGF2* locus	Schoenfelder et al. 2007
HAUNT	+	+	+	+	+	+	+	–	Opposing activities of RNA and DNA on *HOXA* gene regulation	Maamar et al. 2013 Yin et al. 2015
HOTAIR	+	+	+	?	+	+	?	–	Interaction with epigenetic modifying complexes (cancer)	Rinn et al. 2007 Gupta et al. 2010 Yoon et al. 2013
IFNG-AS1	+	+	?	?	+	?	–	+	Activates *TMEVP3* locus	Gomez et al. 2013
LINC00461	–	–	–	?	?	–	–	–	None observed	Oliver et al. 2014
MALAT1	+	+	–	?	+	+	?	?	Regulates metastasis-associated gene expression	Gutschner et al. 2011
NEAT1	+	+	+	+	+	+	+	+	Coordinates assembly of nuclear paraspeckles	Chen and Carmichael 2009 Sunwoo et al. 2009 Mao et al. 2010
PARTICL	+	+	?	?	+	+	?	+	Scaffold for *MAT2A* and triplex-mediated silencing of *MAT2A* promoter	O'Leary et al. 2015
pRNA	+	+	?	?	?	+	+	+	Triplex-mediated methylation of rDNA promoter via DNMT3B recruitment	Schmitz et al. 2010
Ribosomal RNAs	+	+	?	+	+	+	–	+	Catalyzation of peptide-bond formation during protein synthesis and primary structural component of ribosome	Brimacombe and Stiege 1985 Lazdins et al. 1997
RNase P *M1*	+	+	–	–	+	+	–	+	Catalytic cleavage of RNA	Guerrier-Takada et al. 1983
TERC	+	+	–	–	+	+	–	+	RNA template for telomere replication	Zappulla and Cech 2004 Mozdy and Cech 2006
XIST	+	+	+	–	+	+	+	–	X Chromosome inactivation	Penny et al. 1996 Beletskii et al. 2001 Sarma et al. 2010 Leung and Panning 2014

This limited set of lncRNAs exhibit a functional RNA product with a ribonucleic base-mediated mechanism (RNA), with one exception (*LINC00461*). For each gene, we denote whether there is any preliminary evidence for additional functionalities from DNA or transcription at these loci. (+) indicates that a particular function or activity has been experimentally confirmed; (–) indicates a lack of functional contribution based on existing experimental data; and fields with (?) have not been evaluated or remain unanswered. It is important to note that these are approximate categorizations for a select set of examples that serve the general purpose of this review (e.g., DNA is considered positive if a phenotype is observed after gene deletion, *cis* in some cases is not necessarily proven as stringently as the same allele) (for more, see Guttman and Rinn 2012). Note the *BORG alias AB010885*, *HAUNT* is also known as *HALR1*, and *EVF2* is also known as *DLX6-AS1 in human and Dlx6os1 in mouse*.

functional lncRNA genes are directed to http://lncRNAdb.org for an up-to-date catalog (Amaral et al. 2011).

Transient methods for lncRNA functional determination

Early studies into lncRNA function primarily used cell-based assays as a means to dissect the contribution of a lncRNA to a particular cellular context (Guttman et al. 2011; Nagano and Fraser 2011; Wapinski and Chang 2011; Flynn et al. 2015). Numerous studies have found functional lncRNAs across a broad spectrum of cellular biology (Mattick 2009). Despite the diverse biological contexts, these studies used a common strategy of using RNAi-mediated LoF to screen for phenotypes (Fig. 1A). One advantage of such an approach is the ability to directly test for RNA-mediated function, as presumably, the RNA itself is targeted for degradation. Although there are many caveats with this experimental strategy that require independent confirmation (see below), this has been a powerful and scalable approach to screen for potentially functional lncRNAs. It is important to note that RNAi-mediated targeting is an acute exposure to an LoF context, often only measured for 72 h. In contrast, genetic approaches, as described in detail below, allow for longer-term compensatory mechanisms to arise and potentially recover acute phenotypes.

One of the first large-scale lncRNA screens used lentiviral shRNA-mediated LoF screening to survey the impact of 147 lincRNAs (Guttman et al. 2011) on the global transcriptional network of mouse embryonic stem cells. Microarray analysis of each knockdown indicated that the repression of the majority of tested lncRNAs resulted in a significant perturbation effect on the transcriptome. Interestingly, 26 (~15%) of the lncRNA knockdown assays demonstrated significant reduction in expression of a panel of pluripotency marker genes, indicative of loss of the pluripotent state. Moreover, almost all of these 26 exhibited cellular phenotypes, such as morphology differences, indicative of a loss of pluripotency. Overall, this study demonstrates that lncRNAs can have a range of no-to-strong phenotypic effects in a given cellular context, somewhat on par with what would be expected for protein coding genes. Similarly, a study using RNAi-mediated LoF function to identify lncRNAs required for adipogenesis found 8/38 (~20%) tested lncRNAs to have a cell based phenotype (Sun et al. 2013).

Transient exposure to targeting oligonucleotides can also be used to evaluate particular RNA-based mechanisms for lncRNAs. For example, Sarma et al. adopted transient exposure of "locked nucleic acid" (LNA) and Protein Nucleic Acids (PNA) oligomers, chemically modified to bind but not degrade the RNA transcript (Beletskii et al. 2001; Sarma et al. 2010). In these studies, LNAs and PNAs were designed against key regions of the *XIST* RNA thought to be responsible for the localization of *XIST* RNA at the future inactive X (Xi) Chromosome during dosage compensation. In this manner, *XIST* is displaced from Xi without the destruction of the RNA itself (Fig. 1B). Both studies found that loss of localization of otherwise intact *XIST* RNA resulted in loss of X-inactivation. Sarma and colleagues also systematically evaluated which regions of the XIST RNA were required for localization and the kinetics therein across a variety of regions of the *XIST* RNA to determine whether other "domains" also contributed to *XIST* localization (Sarma et al. 2010). Together, these and other studies have demonstrated the power of using transient oligonucleotides to target or block lncRNA functionality.

New technologies for the transient manipulation of RNA expression levels are also emerging, the most successful of which utilize the CRISPR/Cas9 system. In one particularly useful iteration, this system can provide a flexible control over RNA transcription rates. CRISPR-inhibition and activation (CRISPRi and CRISPRa, respectively) are two examples of Cas9-mediated activities to directly modulate gene transcription levels (Gilbert et al. 2013, 2014; Qi et al. 2013; Zalatan et al. 2015). These applications involve the coupling of a catalytically inert Cas9 variant with either a strong transcriptional activator (e.g., VP64, a synthetic concatemer of the Herpes Simplex Viral Protein 16, as in CRISPRa) (Maeder et al. 2013) or a transcriptional repressor (e.g., KRAB as in CRISPRi) (Gilbert et al. 2013) to directly modulate the expression levels of a target gene (Fig. 1C).

One major advantage of this approach is that higher abundance of targets is not a factor, as all targeted sites are present at two copies (relative to thousands of fold differences when targeting a lncRNA versus an mRNA directly). However, similarly to RNAi approaches, CRISPR-targeting of genomic

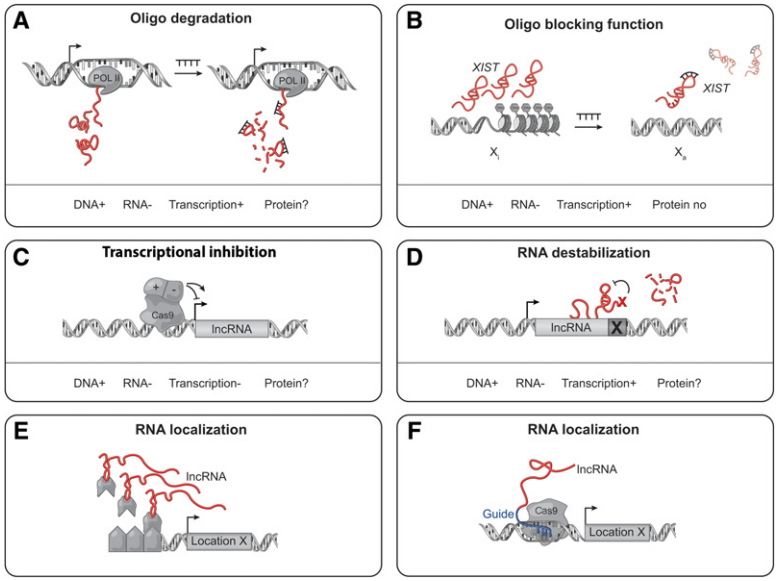

Figure 1. Experimental approaches to manipulate the expression, perturb the activity, or evaluate the functions of long noncoding RNAs. Most commonly used are transient expression of exogenous oligonucleotides designed to exploit the endogenous RNAi machinery or RNase H activity to degrade an RNA (*A*) or occlude putative functional regions (*B*). Recently, strategies have used the flexible CRISPR/Cas9 system to positively or negatively affect the transcription of a lncRNA gene (*C*) or the incorporation of an auto-catalytic regulatory RNA element to destabilize the nascent lncRNA transcript (*D*). An alternative to manipulating lncRNA expression levels involves the recruitment of the RNA of interest to a particular genomic locus or reporter gene through fusion of DNA-binding proteins with RNA-binding elements such as MS2 stem–loops (*E*) or covalent tethering of a lncRNA transcript to the CRISPR/Cas9 guide RNA (*F*).

loci is also prone to off-target effects. Another confounding property is that any local transcription may be affected, and therefore one may not be able to disambiguate the act of transcription from the function of a mature transcript.

One unique example of a successful alternative approach involved the engineering of a destabilization element at the end of a lncRNA to permit normal transcription while enabling the degradation of the mature transcript. This approach was applied to *MALAT1* by adding a sequence into the lncRNA transcript that would be recognized, cleaved, and thus degraded by RNase P (Fig. 1D; Gutschner et al. 2011). As these studies indicate, there is no shortage of effective or creative ways to directly target and manipulate the expression level of a given lncRNA. Few approaches, however, come without significant caveats that can serve to confound the identification of a functional RNA molecule.

Considerations for oligo-based dissection of lncRNA function

The aforementioned studies illustrate the utility of transient LoF approaches to identify and dissect RNA-based mechanisms, and these experiments were key to identifying functional candidates and/or functional RNA domains. Generally speaking, these approaches have been hit-or-miss in their efficacy in targeting lncRNAs. More importantly however, a significant drawback to these approaches is the large number of off-target effects that may confound interpretations (Jackson et al. 2003; Adamson et al. 2012; Sigoillot et al. 2012). It is most important to consider several universally relevant stoichiometric caveats to any oligo-based methodology. One of the biggest concerns with any RNAi or RNase H-mediated depletion of lncRNAs is the significant differences in abundance that exist between mRNAs and lncRNAs; even if the most judicious controls are used (Jackson et al. 2003; Adamson et al. 2012; Sigoillot et al. 2012). For example, the lncRNA *DALIR* was recently detected at two molecules per cell (Chalei et al. 2014), and shRNA-mediated depletion of this RNA molecule resulted in the significant differential regulation of more than 200 genes in N2A cells. Given the low abundance of *DALIR* transcripts in this context and the excess of targeting shRNA molecules, the likelihood of off-target affects rises by orders of magnitude as has been previously suggested (Bassett et al. 2014).

Indeed, if targets are dramatically different in stoichiometric abundance with potential subcomplementary region, significant artifacts can arise. This was illustrated using single-molecule RNA FISH approaches. A single 8-nucleotide match in one of dozens of probes (25 bp in length) targeting one particular lncRNA also matched *NEAT1* (at ~1000-fold higher abundance) (Cabili et al. 2015). Strikingly, this resulted in a staining pattern similar to that of *NEAT1*. Thus, a single 8-mer off-target match of one oligo was able to out-compete all the other probe signals. More generally, oligo-based LoF approaches present even greater challenges in determining "seed"-based off-target effects (Gutschner et al. 2011; Cabili et al. 2015), especially when considering the relatively low abundance of their presumptive lncRNA targets.

With these caveats in mind, it is still possible that low abundance RNA molecules, such as the RNA component of telomerase (*TERC*), which can perform multiple turnover reactions, could accomplish super-stoichometric functionalities (Zappulla and Cech 2004; Mozdy and Cech 2006). In these instances, however, a brute-force LoF approach, such as oligo-based targeting, may not have the desired sensitivity or specificity to ascribe or dissect a precision function.

Ectopic RNA localization: a direct test for RNA function

One way to directly test the role of an RNA is to ectopically localize the RNA via RNA:Protein cognates—protein domains that recognize DNA and RNA domains that recognize fusions to the protein (Fig. 1E). An elegant study using a similar conceptual strategy was applied to the lncRNA *NEAT1* (Mao et al. 2010) that plays an important functional role in paraspeckle formation (Chen and Carmichael 2009; Sunwoo et al. 2009). This is a nuclear "compartment" or "domain" associated with RNA editing and other functions. In this landmark study, the authors ectopically expressed inducible and visible *NEAT1* RNA at a specific genomic location defined by an array of *LacO* DNA binding motifs. This approach allowed imaging of the DNA locus, paraspeckle protein localization and RNA accumulation relative to the site of transcription in living cells. The authors observed that the mature *NEAT1* RNA was required to maintain the paraspeckle but only the act of transcription could establish it.

These studies have been informative but are often laborious, using genetic alterations and in ectopic locations. A recent advance has made localizing RNA to specific genomic locations much easier to test the sufficiency of RNA-based mechanisms. A technology termed CRISPR-Display allows lncRNAs to be covalently connected to short guide RNA (sgRNA) sequences to enable Cas9-mediated localization systems (Shechner et al. 2015). In this manner, a noncoding RNA can be localized to a specific genomic location, via sgRNA complementarity, to evaluate what, if any, biological activities might be attributed to the mature RNA itself (Fig. 1F). As a proof of principle, CRISPR-Display was developed to localize lncRNAs to reporter genes both endogenously and on transient reporters. This study demonstrated that when lncRNAs are directly localized to a particular genomic locus, some have modest but significant influences on activation and repression of targeted genes as predicted from previous studies (Shechner et al. 2015).

Although ectopic RNA localization can directly test an RNA's functional role, independent of the act of transcription, several caveats should still be considered, such as (1) the local chromatin environment of the targeted gene locus at which the RNA is presented (e.g., heterochromatin may prevent access to targeted site resulting in false negatives); (2) the stoichiometric consideration of the number of RNA molecules required at a locus to perform its function; and (3) the position of the presented RNA relative to the target.

It should be increasingly clear that it is important to consider multiple approaches to disentangle the precise mechanism of action of a putative lncRNA gene. We next explore this issue in more detail with respect to testing the functionality of a given lncRNA locus.

Assessing the biology of lncRNAs in vivo

After accumulating evidence for RNA-based mechanisms for lncRNAs using approaches described above, the next major challenge is determining the physiological relevance of these genes in vivo. Historically, this has been achieved through the generation of mutant animal models and investigation of organismal-scale phenotypes. Many of the first discovered lncRNAs (e.g., *H19*, *XIST*, *AIRN*) were investigated using a battery of genetically modified gain- and loss-of-function mouse models. Numerous additional lncRNA loci from more modern catalogs are currently

being tested for in vivo functional roles as well. These studies have found that, much like mRNA genes, many lncRNAs appear dispensable, whereas others can have profound contributions to biology (Nakagawa et al. 2012; Sauvageau et al. 2013; Oliver et al. 2014). Despite their strength, genetic approaches can also be confounding, even in combination, and care must be taken in the interpretation of both positive and negative results.

In vivo lncRNA loss of function

One reasonable genetic approach to lncRNA characterization is whole-gene ablation (Fig. 2A). Indeed, this approach has been adopted to evaluate thousands of mRNAs (Austin et al. 2004; The International Mouse Knockout Consortium 2007) and in turn, has been the predominant means by which phenotypes have been attributed to a given gene. When combined with the introduction of a reporter gene at the target locus (Fig. 2A; Valenzuela et al. 2003), this approach serves to control for the act of transcription at that genomic locus and enables the identification of the spatial-temporal expression of the lncRNA locus in vivo, further reducing the space for phenotypic investigation (Sauvageau et al. 2013; Lai et al. 2015). Moreover, absence of a phenotype in whole gene ablation strategies is a strong indicator that the lncRNA locus is nonfunctional or dispensable (or highly context specific in function). As an example, for the well-conserved lncRNA *C130071C03Rik* (the mouse ortholog of human *LINC00461*), whole-gene ablation was found not to have a function (Oliver et al. 2014). Yet it is important to consider that the absence of evidence for an overt phenotype should not necessarily serve as evidence of absence.

With ablation of any large piece of DNA (e.g., mRNA, lncRNA, enhancer) it is important to consider the possible overlapping molecular contribution(s) of DNA, RNA, and transcription on the observed phenotype (Bassett et al. 2014). Thus, to properly disentangle these attributes requires multiple genetic models to discern between the individual contributions of each mechanism at a given lncRNA locus.

Another in vivo LoF strategy involves generating mutants with transcriptional terminators inserted into lncRNA gene bodies (Sleutels et al. 2002; Bond et al. 2009; Grote et al. 2013). Here, multiple copies of a polyadenylation signal are introduced at the lncRNA locus toward the 5′ end of the gene (Fig. 2B). In this manner, transcription through the gene is halted, and an abortive, presumably nonfunctional RNA product is produced. Due consideration must be given to both of these consequences. For example, Grote et al. (2013) used premature polyA termination to determine that the *Foxf1a*-adjacent lncRNA *Fendrr* is required for viability in mice. The authors further demonstrated a transgene rescue of the *Fendrr* RNA, strongly suggesting the RNA product is the functional element. Interestingly, different approaches targeting the same lncRNA locus may result in identifying similar phenotypic contributions. By way of example, the *Fendrr* locus has been subjected to both polyA termination (Grote et al. 2013) as well as whole gene ablation (Sauvageau et al. 2013). In both instances, a similar phenotype of decreased viability was observed.

This and other studies demonstrate that transcriptional termination is a viable approach to functional characterization of a noncoding RNA locus, but cannot be used in isolation to definitively determine the precise mechanisms of action. A positive result from a transcriptional termination assay may indicate a functional RNA product. However, an alternative hypothesis could be that transcription simply must occur at the locus for its activity. In either case, additional validations are needed to confirm the mechanism.

The *AIRN* locus is a primary example of how multiple approaches were required to determine the functional aspect of the locus. It has recently been revealed that the imprinting activity regulated by the lncRNA *AIRN* is a function of the local act of transcription across the locus and not mediated by the transcript itself (Latos et al. 2012; Santoro et al. 2013). Prior to this, it was thought to have an activity that was attributed to a functional RNA molecule. Sleutels et al. (2002) disrupted transcription through the *AIRN* locus and observed a loss of imprinting phenotype. Similarly, Nagano et al. (2008) conducted both a genetic ablation and a truncation assay and noted a significant reduction in G9A recruitment and also a loss of imprinting. These observations, combined with an earlier study demonstrating that transgene expression of the *AIRN* RNA was insufficient to induce imprinting (Sleutels and Barlow 2001) were initially used as evidence of a functional *AIRN* RNA acting in *cis*. More recent studies have demonstrated that antisense transcription of *AIRN* across the *IGF2R* promoter, regardless of the RNA product, is the specific functional activity that results in imprinting of this locus (Latos et al. 2012).

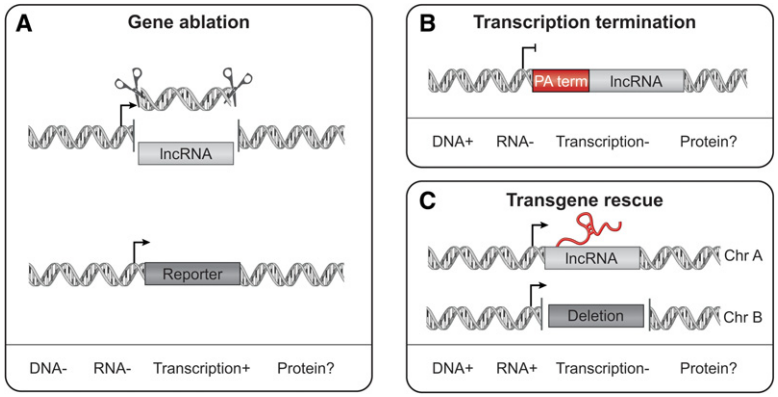

Figure 2. Genetic approaches to evaluate lncRNA functional contributions. (*A*) Genetic ablation of a lncRNA gene or a lncRNA promoter is a powerful technique that can be used to confirm any regulatory activity arising from a particular locus. The optional incorporation of a reporter gene at the deleted locus can be used to visualize the biological contexts in which the lncRNA gene is expressed and evaluate any contribution of the local act of transcription to an observed phenotype. (*B*) An alternative strategy is to introduce a premature transcriptional terminator sequence, which will prevent transcription of the full-length lncRNA transcript. With this approach, however, it is not possible to assess the contribution of the act of transcription. (*C*) With either approach, the rescue of any observed phenotype by transgene expression is currently still considered the gold standard to confirm a functional lncRNA molecule; however, a rescue may not be possible if the RNA works exclusively in *cis*.

In vivo gain of function: transgene expression

lncRNA rescue or gain-of-function assay by transgene expression has in many ways been considered the "gold standard" to prove that a lncRNA molecule

has a direct functional role (Fig. 2C). Examples of lncRNAs with demonstrated activity in *trans* include *Fendrr* (Grote et al. 2013), *Evf2* (Bond et al. 2009), and *Tmevpg1* (Gomez et al. 2013). In this last example, transgenic expression of *Tmevpg1* RNA in both $CD4^+$ and $CD8^+$ T-cells was sufficient to confer a protective phenotype against *Salmonella* infection through activation of the *TMEVP3* locus (Gomez et al. 2013). Yet again, there remain many possible interpretations. A lncRNA gene whose primary function is exerted locally in *cis* will not be able to rescue via *trans* expression, resulting in a false negative.

Conversely, a positive result in a *trans* rescue assay may also be falsely attributed to a functional RNA and rather due to novel small protein coding genes. Emerging evidence suggests that several lncRNAs may harbor cryptic micropeptides (Slavoff et al. 2013; Sun et al. 2013; Bazzini et al. 2014; Anderson et al. 2015). A recent study identified a conserved 46 amino acid micropeptide myoregulin (*MLN*) that was initially found within an annotated muscle-specific lncRNA gene and is an important regulator of Ca^{2+}-mediated contractile machinery in skeletal muscle (Anderson et al. 2015). As in this particular case, the expression and translation of a cryptic, functional peptide within a lncRNA gene may in fact be the functional element resulting in an observed phenotype. With this in mind, many lncRNA catalogs have made concerted efforts to filter out lncRNAs with small peptides that are easily detectable by codon substitution frequency analysis (Lin et al. 2011), and genes potentially harboring such peptides are referred to as transcripts of unknown coding potential (TUCP) (Cabili et al. 2011; Guttman and Rinn 2012; Molyneaux et al. 2015).

The preceding examples demonstrate that no one approach to functionally characterize a lncRNA locus can uniquely satisfy all possible models. Rather, this effort should involve the coordinated use of multiple approaches to identify the often multiple overlapping functional modalities of a gene locus. One excellent example of a detailed, systematic approach is illustrated in a recent study that used multiple genetic models to dissect the *HAUNT* lncRNA locus. Intriguingly, they found that DNA, RNA, and transcription all play an important role in regulating *HOX* gene expression (Yin et al. 2015). Deletion of the *HAUNT* genomic locus resulted in a significant reduction in the expression and activation of *HOXA* genes. In striking contrast, however, reduction of *HAUNT* RNA by either RNAi or TSS inclusion resulted in an increase in *HOXA* gene expression. Furthermore, overexpression of *HAUNT* RNA *in cis*, but not *in trans*, was sufficient to down-regulate many of the *HOXA* genes as well. Using this suite of approaches, the authors are able to conclude that there are independent, and opposing, functional activities conferred by both the *HAUNT* genomic DNA, and the *HAUNT* RNA. Intriguingly, another study identified a functional role for the lncRNA in repression of *HOXA1 in cis* (Maamar et al. 2013). These studies underscore the potential complexity of any given genomic locus and highlight the multifaceted approach that should be used to thoroughly evaluate the functional activity of a lncRNA genomic locus.

A classic example bringing all these principles and approaches together is the *XIST* locus that controls X Chromosome dosage compensation. From this locus, a 17 kb transcript is sufficient to silence the majority of an entire X Chromosome. In many ways, *XIST* can be considered a model example of lncRNA biology, and the dissection of its mechanism of action can serve as a guide for future lncRNA studies. Principally, ablation of the DNA and promoter of *XIST* results in loss of X Chromosome inactivation (Penny et al. 1996), yet these models cannot rule out the contribution of the local act of transcription. It took additional experiments (Beletskii et al. 2001; Sarma et al. 2010) that blocked the localization region of *XIST* RNA (REPC) with a complimentary modified oligo that maintained transcription, preserved the *XIST* genomic DNA, and kept the mature transcript intact, but resulted in reactivation of the X Chromosome, thus demonstrating a mechanism that is dependent on the mature *XIST* RNA localizing to the inactive X Chromosome *in cis*. Thus, both genetic and molecular biology experiments were required to demonstrate, and specifically ascribe, *XIST* function to the activity of a functional RNA molecule.

A path forward toward common modalities of ncRNA biology

RNA has a long history of breaking the rules of biology through diverse and fascinating molecular modalities (Wapinski and Chang 2011; Cech and Steitz 2014; Rinn and Guttman 2014). Yet, RNA does almost always obey one rule, i.e., form ribonucleic protein complexes to carry out functional roles. Despite this property, we know very little about the common RNA sequences, structures, and domains of lncRNAs that facilitate these associations. In sharp contrast, primary sequences of many proteins are predictive of their functionality. For example, if a protein harbors a bHLH domain (Fig. 3), it is likely to be involved in DNA binding. Similarly, other protein motifs might suggest functional activities such as specific catalytic domains. Yet, almost no such common features have been identified for lncRNAs.

One complication in identifying such domains may be the inherent flexibility of RNA and its unique ability (as compared to DNA) to form complex secondary structures or interact with a diversity of protein partners. Much of our understanding of motifs has arisen from analyses of primary sequence, and many of the tools used to define primary sequence motifs do not translate to more complex two- and three-dimensional structures. However, many current efforts are underway to develop techniques to resolve RNA 2D structures en masse (Wilkinson et al. 2008; Kertesz et al. 2010; Mortimer et al. 2012; Siegfried et al. 2014; Spitale et al. 2014; Wan et al. 2014). Additionally, classic biochemistry tools, such as RNA footprinting, are being adapted to genome-scale approaches (Silverman et al. 2014) to identify RNase-protected regions of the transcriptome that may represent sites of RNA:protein interactions. The data generated by these and other efforts will provide important insight into potential common structural and functional motifs for RNA.

Here, we summarize by discussing a few of the types of RNA domains that have been identified that have similar functions to proteins. By discussing just a few of these examples, we hope to highlight a logical framework toward unraveling the molecular grammar of lncRNAs.

Subcellular localization signals

Both RNA and proteins harbor endogenous sequences and structures that provide direction to their appropriate subcellular location. These localization queues (Fig. 3, first row), such as the signal peptide or nuclear localization signal for proteins, have recently been described in different noncoding RNA sequences. The lncRNA, *BORG*, harbors a pentamer RNA motif that is required for its proper nuclear retention (Zhang et al. 2014). A similar

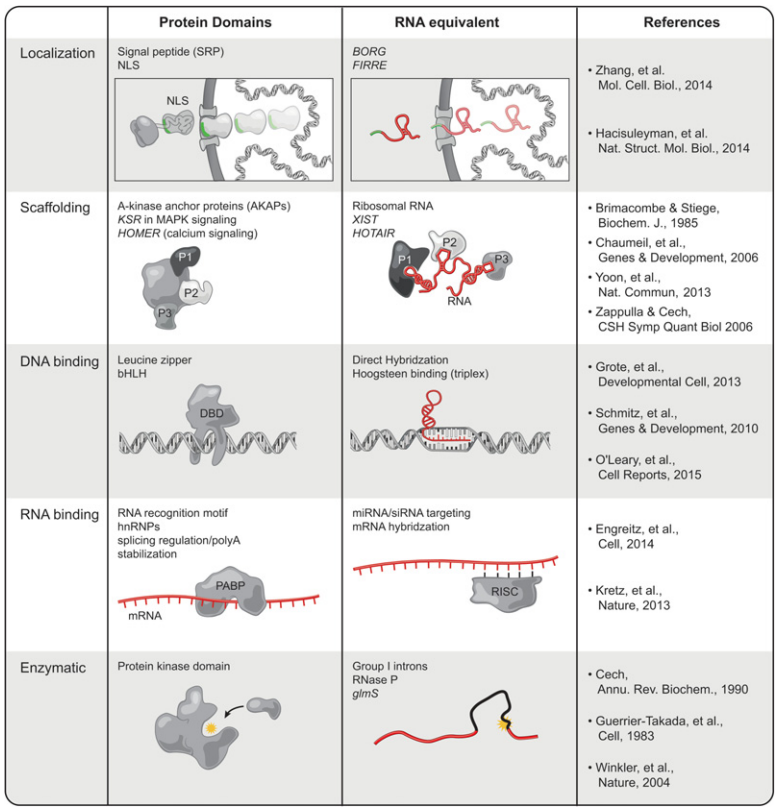

Figure 3. Functionally equivalent RNA and DNA domains. Despite distinct mechanisms by which they achieve their equivalence, RNA domains are capable of affecting many biological activities that have traditionally been ascribed exclusively to proteins. Much work remains, however, in identifying common RNA primary, secondary, and tertiary structures that comprise orthologous functional domains and can mediate these and other subcellular activities.

functional motif can be observed within the RRD repeat domain of the lncRNA, *FIRRE* (Hacisuleyman et al. 2014), which is also responsible for the nuclear localization of this lncRNA. In the future, it will be important to catalog the RNA localization sequences and structures within lncRNAs toward understanding shared modalities.

Molecular scaffold

Both proteins and RNAs exhibit the ability to serve as a scaffold for larger complexes (Fig. 3, second row). Protein scaffolding complexes such as the A-kinase anchor proteins (AKAP) (Diviani et al. 2011) use a specific EVH1 domain to tether and complex key neuronal proteins (Luo et al. 2012). Similarly, RNA is commonly used as a platform to scaffold many proteins for a related function. The ribosome is a dramatic example of RNA scaffolding structural organization of RNA-protein interactions, as well as catalytic activity to enable translation (Brimacombe and Stiege 1985). This and many other classic examples demonstrate the powerful roles of RNA scaffolds (Zappulla and Cech 2006). More recently, several long noncoding RNAs such as *HOTAIR*, *XIST*, *FIRRE*, and many others (Chaumeil et al. 2006; Tattermusch and Brockdorff 2011; Wang and Chang 2011; Yoon et al. 2013; Hacisuleyman et al. 2014; Quinn et al. 2014; Chu et al. 2015a,b; McHugh et al. 2015) have also been identified to function as RNA scaffolds that interface with and influence epigenetic regulatory complexes.

Nucleic acid interaction and targeting

Some of the earliest defined protein domains to bind directly to DNA were the basic helix-loop-helix (bHLH) domain (Murre et al. 1989) and the leucine zippers (Fig. 3, third row; Landschulz et al. 1988). Likewise, RNA possesses an apt property in that it can readily hybridize to nascent DNA via sequence complementarity, or DNA:DNA:RNA triplexes, to target specific genomic loci. Indeed, such examples have been observed for lncRNAs such as *Fendrr* (Grote et al. 2013), *PRNA* (Schmitz et al. 2010), and the recently described lncRNA PARTICLE (*PARTICL*) (O'Leary et al. 2015). In much the same manner, RNA–RNA interactions can occur through direct sequence complementarity and regulatory roles (Fig. 3, fourth row; Kretz et al. 2013; Engreitz et al. 2014; Tay et al. 2014). To this end, lncRNAs can also serve as host transcripts for various classes of small RNAs, including microRNAs (da Rocha et al. 2008; Royo and Cavaillé 2008; Augoff et al. 2012; Sabin et al. 2013).

RNA catalysis

Perhaps the most exciting yet elusive property RNA shares with protein is catalysis (Fig. 3, fifth row). Although the vast majority of catalytic activity appears to be mediated by proteins, there is increasing evidence for a number of mammalian catalytic RNAs. Most notably, this list includes the self-cleaving group I introns (Cech 1990); the RNase P *M1* RNA, which catalyzes tRNA biosynthesis (Guerrier-Takada et al. 1983); the ribosome (Brimacombe and Stiege 1985); and others involved in RNA processing (Teixeira et al. 2004; Winkler et al. 2004; Strobel and Cochrane 2007).

These examples, and the methods used to discern their functional activities, serve as a conceptual framework for how to identify functional RNA sequences and domains. We propose that efforts within the lncRNA community should proceed with similar strategies (as have been applied extensively for protein-coding genes and the original noncoding RNAs classes) and in parallel, adopt and embrace new experimental designs that appropriately reflect the complexities of RNA biology. An important next step forward will be identifying the additional RNA-mediated localization signals, scaffolding motifs, protein-guidance cues, and catalytic domains that we already know RNA molecules are capable of harboring. It will take creative, diverse, and collaborative multifaceted approaches to determine the RNA-based functional importance of any *single lncRNA*. Here, we have summarized emerging approaches and their considerations toward the goal of unlinking the underlying ribonucleic logic of lncRNAs.

Acknowledgments

The authors thank all of the members of the Goff and Rinn labs for their input.

References

Adamson B, Smogorzewska A, Sigoillot FD, King RW, Elledge SJ. 2012. A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response. *Nat Cell Biol* **14:** 318–328.

Amaral PP, Clark MB, Gascoigne DK, Dinger ME, Mattick JS. 2011. lncRNAdb: a reference database for long noncoding RNAs. *Nucleic Acids Res* **39:** D146–D151.

Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, Kasaragod P, Shelton JM, Liou J, Bassel-Duby R, et al. 2015. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* **160:** 595–606.

Augoff K, McCue B, Plow EF, Sossey-Alaoui K. 2012. miR-31 and its host gene lncRNA LOC554202 are regulated by promoter hypermethylation in triple-negative breast cancer. *Mol Cancer* **11:** 5.

Austin CP, Battey JF, Bradley A, Bucan M, Capecchi M, Collins FS, Dove WF, Duyk G, Dymecki S, Eppig JT, et al. 2004. The knockout mouse project. *Nat Genet* **36:** 921–924.

Bassett AR, Akhtar A, Barlow DP, Bird AP, Brockdorff N, Duboule D, Ephrussi A, Ferguson-Smith AC, Gingeras TR, Haerty W, et al. 2014. Considerations when investigating lncRNA function in vivo. *Elife* **3:** e03058.

Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, Vejnar CE, Lee MT, Rajewsky N, Walther TC, et al. 2014. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J* **33:** 981–993.

Beletskii A, Hong YK, Pehrson J, Egholm M, Strauss WM. 2001. PNA interference mapping demonstrates functional domains in the noncoding RNA *Xist*. *Proc Natl Acad Sci* **98:** 9215–9220.

Bergmann JH, Li J, Eckersley-Maslin MA, Rigo F, Freier SM, Spector DL. 2015. Regulation of the ESC transcriptome by nuclear long noncoding RNAs. *Genome Res* **25:** 1336–1346.

Bond AM, VanGompel MJW, Sametsky EA, Clark MF, Savage JC, Disterhoft JF, Kohtz JD. 2009. Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nat Neurosci* **12:** 1020–1027.

Brimacombe R, Stiege W. 1985. Structure and function of ribosomal RNA. *Biochem J* **229:** 1–17.

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **25:** 1915–1927.

Cabili MN, Dunagin MC, McClanahan PD, Biaesch A, Padovan-Merhar O, Regev A, Rinn JL, Raj A. 2015. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol* **16:** 20.

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells CA, et al. 2005. The transcriptional landscape of the mammalian genome. *Science* **309:** 1559–1563.

Cech TR. 1990. Self-splicing of group I introns. *Annu Rev Biochem* **59:** 543–568.

Cech TR, Steitz JA. 2014. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157:** 77–94.

Chalei V, Sansom SN, Kong L, Lee S, Montiel JF, Vance KW, Ponting CP. 2014. The long non-coding RNA *Dali* is an epigenetic regulator of neural differentiation. *Elife* **3:** e04530.

Chaumeil J, Le Baccon P, Wutz A, Heard E. 2006. A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev* **20:** 2223–2237.

Chen LL, Carmichael GG. 2009. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Mol Cell* **35:** 467–478.

Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G, et al. 2005. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308:** 1149–1154.

Chu C, Spitale RC, Chang HY. 2015a. Technologies to probe functions and mechanisms of long noncoding RNAs. *Nat Struct Mol Biol* **22:** 29–35.

Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, Magnuson T, Heard E, Chang HY. 2015b. Systematic discovery of Xist RNA binding proteins. *Cell* **161:** 404–416.

da Rocha ST, Edwards CA, Ito M, Ogata T, Ferguson-Smith AC. 2008. Genomic imprinting at the mammalian *Dlk1-Dio3* domain. *Trends Genet* **24:** 306–316.

Diviani D, Dodge-Kafka KL, Li J, Kapiloff MS. 2011. A-kinase anchoring proteins: scaffolding proteins in the heart. *Am J Physiol Heart Circ Physiol* **301:** H1742–H1753.

The ENCODE Project Consortium. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447:** 799–816.

Engreitz JM, Sirokman K, McDonel P, Shishkin AA, Surka C, Russell P, Grossman SR, Chow AY, Guttman M, Lander ES. 2014. RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent pre-mRNAs and chromatin sites. *Cell* **159:** 188–199.

Flynn RA, Martin L, Spitale RC, Do BT, Sagan SM, Zarnegar B, Qu K, Khavari PA, Quake SR, Sarnow P, et al. 2015. Dissecting noncoding and pathogen RNA–protein interactomes. *RNA* **21:** 135–143.

Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, et al. 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154:** 442–451.

Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, et al. 2014. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159:** 647–661.

Gomez JA, Wapinski OL, Yang YW, Bureau JF, Gopinath S, Monack DM, Chang HY, Brahic M, Kirkegaard K. 2013. NeST, a long noncoding RNA, controls microbial susceptibility and epigenetic activation of the interferon-γ locus. *Cell* **152:** 743–754.

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.

Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S. 1983. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35:** 849–857.

Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Hung T, Argani P, Rinn JL, Wang Y, et al. 2010. Long non-coding RNA *HOTAIR* reprograms chromatin state to promote cancer metastasis. *Nature* **464:** 1071–1076.

Gutschner T, Baas M, Diederichs S. 2011. Noncoding RNA gene silencing through genomic integration of RNA destabilizing elements using zinc finger nucleases. *Genome Res* **21:** 1944–1954.

Guttman M, Rinn JL. 2012. Modular regulatory principles of large non-coding RNAs. *Nature* **482:** 339–346.

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, et al. 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458:** 223–227.

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, et al. 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477:** 295–300.

Hacisuleyman E, Goff LA, Trapnell C, Williams A, Henao-Mejia J, Sun L, McClanahan P, Hendrickson DG, Sauvageau M, Kelley DR, et al. 2014. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat Struct Mol Biol* **21:** 198–206.

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, et al. 2012. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22:** 1760–1774.

The International Mouse Knockout Consortium. 2007. A mouse for all reasons. *Cell* **128:** 9–13.

Jackson AL, Bartz SR, Schelter J, Kobayashi SV, Burchard J, Mao M, Li B, Cavet G, Linsley PS. 2003. Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* **21:** 635–637.

Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SPA, Gingeras TR. 2002. Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296:** 916–919.

Kapranov P, Drenkow J, Cheng J, Long J, Helt G, Dike S, Gingeras TR. 2005. Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res* **15:** 987–997.

Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermuller J, Hofacker IL, et al. 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316:** 1484–1488.

Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, Chang HY, Segal E. 2010. Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467:** 103–107.

Kretz M, Siprashvili Z, Chu C, Webster DE, Zehnder A, Qu K, Lee CS, Flockhart RJ, Groff AF, Chow J, et al. 2013. Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* **493:** 231–235.

Lai KM, Gong G, Atanasio A, Rojas J, Quispe J, Posca J, White D, Huang M, Fedorova D, Grant C, et al. 2015. Diverse phenotypes and specific transcription patterns in twenty mouse lines with ablated lincRNAs. *PLoS One* **10:** e0125522.

Landschulz WH, Johnson PF, McKnight SL. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240:** 1759–1764.

Latos PA, Pauler FM, Koerner MV, Şenergin HB, Hudson QJ, Stocsits RR, Allhoff W, Stricker SH, Klement RM, Warczok KE, et al. 2012. *Airn*

transcriptional overlap, but not its lncRNA products, induces imprinted *Igf2r* silencing. *Science* **338:** 1469–1472.
Lazdins IB, Delannoy M, Sollner-Webb B. 1997. Analysis of nucleolar transcription and processing domains and pre-rRNA movements by in situ hybridization. *Chromosoma* **105:** 481–495.
Leung KN, Panning B. 2014. X-inactivation: Xist RNA uses chromosome contacts to coat the X. *Curr Biol* **24:** R80–R82.
Lin MF, Jungreis I, Kellis M. 2011. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **27:** i275–i282.
Luo P, Li X, Fei Z, Poon W. 2012. Scaffold protein Homer 1: implications for neurological diseases. *Neurochem Int* **61:** 731–738.
Maamar H, Cabili MN, Rinn J, Raj A. 2013. *linc-HOXA1* is a noncoding RNA that represses *Hoxa1* transcription in *cis*. *Genes Dev* **27:** 1260–1271.
Maeda N, Kasukawa T, Oyama R, Gough J, Frith M, Engström PG, Lenhard B, Aturaliya RN, Batalov S, Beisel KW, et al. 2006. Transcript annotation in FANTOM3: mouse gene catalog based on physical cDNAs. *PLoS Genet* **2:** e62.
Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. 2013. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods* **10:** 977–979.
Mao YS, Sunwoo H, Zhang B, Spector DL. 2010. Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat Cell Biol* **13:** 95–101.
Mattick JS. 2009. The genetic signatures of noncoding RNAs. *PLoS Genet* **5:** e1000459.
McHugh CA, Chen CK, Chow A, Surka CF, 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.
Molyneaux BJ, Goff LA, Brettler AC, Chen HH, Brown JR, Hrvatin S, Rinn JL, Arlotta P. 2015. DeCoN: genome-wide analysis of in vivo transcriptional dynamics during pyramidal neuron fate selection in neocortex. *Neuron* **85:** 275–288.
Morris KV, Mattick JS. 2014. The rise of regulatory RNA. *Nat Rev Genet* **15:** 423–437.
Mortimer SA, Trapnell C, Aviran S, Pachter L, Lucks JB. 2012. SHAPE-Seq: high-throughput RNA structure analysis. *Curr Protoc Chem Biol* **4:** 275–297.
Mozdy AD, Cech TR. 2006. Low abundance of telomerase in yeast: implications for telomerase haploinsufficiency. *RNA* **12:** 1721–1737.
Murre C, McCaw PS, Baltimore D. 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless*, *MyoD*, and *myc* proteins. *Cell* **56:** 777–783.
Nagano T, Fraser P. 2011. No-nonsense functions for long noncoding RNAs. *Cell* **145:** 178–181.
Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P. 2008. The air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322:** 1717–1720.
Nakagawa S, Ip JY, Shioi G, Tripathi V, Zong X, Hirose T, Prasanth KV. 2012. Malat1 is not an essential component of nuclear speckles in mice. *RNA* **18:** 1487–1499.
Nordström KJV, Mirza MAI, Almén MS, Gloriam DE, Fredriksson R, Schiöth HB. 2009. Critical evaluation of the FANTOM3 non-coding RNA transcripts. *Genomics* **94:** 169–176.
Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, Suzuki H, et al. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420:** 563–573.
O'Leary VB, Ovsepian SV, Carrascosa LG, Buske FA, Radulovic V, Niyazi M, Moertl S, Trau M, Atkinson MJ, Anastasov N. 2015. *PARTICLE*, a triplex-forming long ncRNA, regulates locus-specific methylation in response to low-dose irradiation. *Cell Rep* **11:** 474–485.
Oliver PL, Chodroff RA, Gosal A, Edwards B, Cheung AFP, Gomez-Rodriguez J, Elliot G, Garrett LJ, Lickiss T, Szele F, et al. 2014. Disruption of *Visc-2*, a brain-expressed conserved long noncoding RNA, does not elicit an overt anatomical or behavioral phenotype. *Cereb Cortex* bhu196.
Penny G, Kay G, Sheardown S, Rastan S, Brockdorff N. 1996. Requirement for *Xist* in X chromosome inactivation. *Nature* **379:** 131–137.
Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152:** 1173–1183.
Quinn JJ, Ilik IA, Qu K, Georgiev P, Chu C, Akhtar A, Chang HY. 2014. Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. *Nat Biotechnol* **32:** 933–940.
Rinn JL, Chang HY. 2012. Genome regulation by long noncoding RNAs. *Annu Rev Biochem* **81:** 145–166.
Rinn J, Guttman M. 2014. RNA Function. RNA and dynamic nuclear organization. *Science* **345:** 1240–1241.
Rinn JL, Euskirchen G, Bertone P, Martone R, Luscombe NM, Hartman S, Harrison PM, Nelson FK, Miller P, Gerstein M, et al. 2003. The transcriptional activity of human Chromosome 22. *Genes Dev* **17:** 529–540.
Rinn J, Kertesz M, Wang J, Squazzo S, Xu X, Brugmann S, Goodnough L, Helms J, Farnham P, Segal E, et al. 2007. Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* **129:** 1311–1323.
Royo H, Cavaillé J. 2008. Non-coding RNAs in imprinted gene clusters. *Biol Cell* **100:** 149–166.
Sabin LR, Delás MJ, Hannon GJ. 2013. Dogma derailed: the many influences of RNA on the genome. *Mol Cell* **49:** 783–794.
Santoro F, Mayer D, Klement RM, Warczok KE, Stukalov A, Barlow DP, Pauler FM. 2013. Imprinted *Igf2r* silencing depends on continuous *Airn* lncRNA expression and is not restricted to a developmental window. *Development* **140:** 1184–1195.
Sarma K, Levasseur P, Aristarkhov A, Lee JT. 2010. Locked nucleic acids (LNAs) reveal sequence requirements and kinetics of Xist RNA localization to the X chromosome. *Proc Natl Acad Sci* **107:** 22196–22201.
Sauvageau M, Goff LA, Lodato S, Bonev B, Groff AF, Gerhardinger C, Sanchez-Gomez DB, 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.
Schmitz KM, Mayer C, Postepska A, Grummt I. 2010. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev* **24:** 2264–2269.
Schoenfelder S, Smits G, Fraser P, Reik W, Paro R. 2007. Non-coding transcripts in the *H19* imprinting control region mediate gene silencing in transgenic *Drosophila*. *EMBO Rep* **8:** 1068–1073.
Shechner DM, Hacisuleyman E, Younger ST, Rinn JL. 2015. Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat Methods* **12:** 664–670.
Siegfried NA, Busan S, Rice GM, Nelson JAE, Weeks KM. 2014. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* **11:** 959–965.
Sigoillot FD, Lyman S, Huckins JF, Adamson B, Chung E, Quattrochi B, King RW. 2012. A bioinformatics method identifies prominent off-targeted transcripts in RNAi screens. *Nat Methods* **9:** 363–366.
Silverman IM, Li F, Alexander A, Goff L, Trapnell C, Rinn JL, Gregory BD. 2014. RNase-mediated protein footprint sequencing reveals protein-binding sites throughout the human transcriptome. *Genome Biol* **15:** R3.
Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ, Karger AD, Budnik BA, Rinn JL, Saghatelian A. 2013. Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat Chem Biol* **9:** 59–64.
Sleutels F, Barlow DP. 2001. Investigation of elements sufficient to imprint the mouse air promoter. *Mol Cell Biol* **21:** 5008–5017.
Sleutels F, Zwart R, Barlow DP. 2002. The non-coding *Air* RNA is required for silencing autosomal imprinted genes. *Nature* **415:** 810–813.
Spitale RC, Flynn RA, Torre EA, Kool ET, Chang HY. 2014. RNA structural analysis by evolving SHAPE chemistry. *Wiley Interdiscip Rev RNA* **5:** 867–881.
Strobel SA, Cochrane JC. 2007. RNA catalysis: ribozymes, ribosomes and riboswitches. *Curr Opin Chem Biol* **11:** 636–643.
Sun L, Goff LA, Trapnell C, Alexander R, Lo KA, Hacisuleyman E, Sauvageau M, Tazon-Vega B, Kelley DR, Hendrickson DG, et al. 2013. Long noncoding RNAs regulate adipogenesis. *Proc Natl Acad Sci* **110:** 3387–3392.
Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, Spector DL. 2009. *MEN* ε/β nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res* **19:** 347–359.
Tattermusch A, Brockdorff N. 2011. A scaffold for X chromosome inactivation. *Hum Genet* **130:** 247–253.
Tay Y, Rinn J, Pandolfi PP. 2014. The multilayered complexity of ceRNA crosstalk and competition. *Nature* **505:** 344–352.
Teixeira A, Tahiri-Alaoui A, West S, Thomas B, Ramadass A, Martianov I, Dye M, James W, Proudfoot NJ, Akoulitchev A. 2004. Autocatalytic RNA cleavage in the human β-globin pre-mRNA promotes transcription termination. *Nature* **432:** 526–530.
Ulitsky I, Bartel DP. 2013. lincRNAs: genomics, evolution, and mechanisms. *Cell* **154:** 26–46.
Valenzuela DM, Murphy AJ, Frendewey D, Gale NW, Economides AN, Auerbach W, Poueymirou WT, Adams NC, Rojas J, Yasenchak J, et al. 2003. High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nat Biotechnol* **21:** 652–659.
Wan Y, Qu K, Zhang QC, Flynn RA, Manor O, Ouyang Z, Zhang J, Spitale RC, Snyder MP, Segal E, et al. 2014. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* **505:** 706–709.
Wang KC, Chang HY. 2011. Molecular mechanisms of long noncoding RNAs. *Mol Cell* **43:** 904–914.
Wapinski O, Chang HY. 2011. Long noncoding RNAs and human disease. *Trends Cell Biol* **21:** 354–361.

Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, Mathews DH, Giddings MC, Weeks KM. 2008. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* **6:** e96.

Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR. 2004. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428:** 281–286.

Yin Y, Yan P, Lu J, Song G, Zhu Y, Li Z, Zhao Y, Shen B, Huang X, Zhu H, et al. 2015. Opposing roles for the lncRNA *Haunt* and its genomic locus in regulating *HOXA* gene activation during embryonic stem cell differentiation. *Cell Stem Cell* **16:** 504–516.

Yoon JH, Abdelmohsen K, Kim J, Yang X, Martindale JL, Tominaga-Yamanaka K, White EJ, Orjalo AV, Rinn JL, Kreft SG, et al. 2013. Scaffold function of long non-coding RNA *HOTAIR* in protein ubiquitination. *Nat Commun* **4:** 2939.

Zalatan JG, Lee ME, Almeida R, Gilbert LA, Whitehead EH, La Russa M, Tsai JC, Weissman JS, Dueber JE, Qi LS, et al. 2015. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160:** 339–350.

Zappulla DC, Cech TR. 2004. Yeast telomerase RNA: a flexible scaffold for protein subunits. *Proc Natl Acad Sci* **101:** 10024–10029.

Zappulla DC, Cech TR. 2006. RNA as a flexible scaffold for proteins: yeast telomerase and beyond. *Cold Spring Harb Symp Quant Biol* **71:** 217–224.

Zhang B, Gunawardane L, Niazi F, Jahanbani F, Chen X, Valadkhan S. 2014. A novel RNA motif mediates the strict nuclear localization of a long non-coding RNA. *Mol Cell Biol* **34:** 2318–2329.