Epigenetic Regulation by Long Noncoding RNAs

Jeannie T. Lee

Epigenetic Regulation by Long Noncoding RNAs

Jeannie T. Lee

Recent studies show that transcription of the mammalian genome is not only pervasive but also enormously complex. It is estimated that an average of 10 transcription units, the vast majority of which make long noncoding RNAs (lncRNAs), may overlap each traditional coding gene. These lncRNAs include not only antisense, intronic, and intergenic transcripts but also pseudogenes and retrotransposons. Do they universally have function, or are they merely transcriptional by-products of conventional coding genes? A glimpse into the molecular biology of multiple emerging lncRNA systems reveals the "Wild West" landscape of their functions and mechanisms and the key problems to solve in the years ahead toward understanding these intriguing macromolecules.

RNA has become widely suspected as the culprit behind almost every case of epigenetic regulation. There continues to be a shift in how we conceptualize this remarkably versatile macromolecule, once regarded primarily as mere intermediary of the "central dogma" stating that information moves unidirectionally from DNA to RNA to protein. The latest interests center around the genomic "dark matter" that has for years been dismissed as transcriptional noise (1–4). Only 1% of the mammalian genome carries protein-coding potential, yet 70 to 90% is transcribed at some point during development to produce a large transcriptome of long noncoding RNA (lncRNA, defined as RNA > 100 nucleotides in length). Some estimate total membership to exceed 200,000, whereas others suggest fewer than 10,000 (5–8). The ENCODE project has revealed an enormous complexity, with ~10 isoforms overlapping any previously annotated genes, thereby challenging the traditional definition of a gene (8). Although there is now little doubt that pervasive transcription occurs, whether this activity is universally functional is unknown. These transcripts are often poorly conserved, unstable, and/or present in few copies (7, 9, 10). Nonetheless, clear roles have emerged for some lncRNAs, and a survey of some examples illustrates how this class of RNA is helping to establish new paradigms for epigenetic regulation.

Lessons from the X Chromosome

The intriguing story of lncRNAs first debuted in the phenomena of genomic imprinting and X-chromosome inactivation (XCI) (11–14). Nowhere is the abundance of lncRNA more evident than the X-inactivation center (Xic). To balance X-chromosome gene expression between males and females, the Xic on the mammalian X

Howard Hughes Medical Institute (HHMI), Department of Molecular Biology, Massachusetts General Hospital, and Department of Genetics, Harvard Medical School, Boston, MA 02138, USA. E-mail: lee@molbio.mgh.harvard.edu

chromosome controls the initiation steps of XCI through a series of RNA-based switches (13, 14). Today, the Xic serves as a model for understanding epigenetic regulation by lncRNA (Fig. 1).

The X-inactive-specific transcript (XIST/Xist) was one of the first lncRNAs to be discovered in mammals (15). The Xist locus produces a 17- to 20-kb RNA that coats the X chromosome in cis (Fig. 1, A and B) (16), is expressed only from the inactive X chromosome (Xi), and is essential for the silencing processing (14). Its noncoding status immediately implied that RNA itself could be an effector of chromatin and transcriptional change, an idea substantiated years later by the isolation of Xist's first interacting factors, Polycomb repressive complex 2 (PRC2) (17) and YY1 (18). Through a conserved repeat motif (Repeat A, RepA), Xist RNA directly binds PRC2 (17), the epigenetic complex responsible for trimethylation of histone H3 at Lys^{27} (H3K27me3), and targets PRC2 to the Xi. These findings indicated that lncRNAs may be crucial accessory factors for Polycomb function. Like other regulatory factors, epigenetic complexes must be targeted in space (genomic location) and time (during development), but many, such as PRC2, do not possess sequence-specific DNA-binding subunits to guide them. The involvement of RNA, a macromolecule with inherent sequence information, would at once provide targeting specificity and introduce new regulatory capabilities (e.g., action in cis).

Targeting PRC2, however, is biologically separable from loading onto chromatin (17). The 1.6-kb RepA transcript recruits PRC2 to the Xist promoter, but docking of PRC2 is precluded by expression of Xist's antisense transcript, Tsix. Only when Tsix expression is down-regulated during development does the Xist-PRC2 complex load onto the Xi nucleation center within Xist's exon 1. Loading depends on the transcription factor, YY1, bound only to Xi (18) (Fig. 2). By cotranscriptionally tethering Xist RNA to the Xic, YY1 bridges the lncRNA and chromatin, accounting

for the allele-specific binding of Xist RNA to the Xi.

Xist is controlled by two other lncRNAs, one acting negatively (Tsix), the other positively (Jpx). Tsix determines allelic choice by repressing Xist transcription on one allele (19, 20). From numerous genetic manipulations at Tsix, it now appears that Tsix regulates Xist in several ways. Tsix coordinates X-chromosome pairing to generate epigenetic asymmetry within the Xist locus (21); it recruits DNA methyltransferase (Dnmt3a) to silence Xist (22, 23); and it blocks recruitment of PRC2 to Xist by RepA (17). Tsix RNA directly binds PRC2 (17) and also duplexes with Xist-RepA RNA (24), thereby potentially serving as decoy for PRC2 recruitment (by titrating Xist-RepA RNA or PRC2) in its role as repressor of XCI.

Positive regulation of Xist involves another lncRNA, Jpx (25) (Fig. 1B). Deleting Jpx abolishes Xist activation, its post-transcriptional knockdown recapitulates the deletion, and autosomal expression of Jpx rescues the X-linked deletion. Thus, Jpx is diffusible and acts in trans as an RNA. Combined, these studies demonstrate central functions for lncRNA during XCI: They directly target repressive epigenetic complexes, act as antisense inhibitors, and activate transcription.

Why Long Noncoding RNAs?

Although lncRNAs now dominate the Xic, this region was once coding (26). Evolution of random XCI 150 million years ago in eutherian mammals coincided with a shift from coding to noncoding space, suggesting that lncRNAs offer distinct advantages over proteins for some forms of epigenetic regulation. Allelic and cis control were likely major driving forces behind this "reverse evolution" (27), because XCI treats two X chromosomes in diametrically opposite ways and requires coordinated cis-limited silencing of genes on the Xi. Two properties of mammalian lncRNAs are notably relevant. One feature is lncRNA's tethering capabilities and fast turnover, which enable allelic marking. LncRNAs are naturally tethered to the site of transcription through the RNA-DNA-polymerase (Pol II) ternary complex, thereby enabling function as allele-specific tag (Fig. 2). LncRNAs may have an exposed 5′ business end to capture chromatin complexes and a nascent 3′ tethering end to anchor the RNA-protein complex to a locus. Inclusion of Pol II pausing could transiently stabilize tethering. Tethering could also be enhanced by bridge proteins, such as YY1 (18). Rapid degradation after transcriptional termination would limit the RNA's half-life, thereby preventing diffusion and action at ectopic sites. Tsix's half-life of 30 to 60 min indicates that the 40-kb transcript is degraded as soon as it is created (23), likely explaining its strict cis action, because effective concentrations would only be reached at the site of synthesis. Fast turnover rates and low copy numbers are features of many lncRNAs (7–10). LncRNAs

are thereby distinguished from proteins and small RNAs by the possibility of allele-specific action. Proteins do not retain allelic memory; their transcriptional origin is lost when mRNA is shuttled to the cytoplasm for translation to protein. Small RNAs of the RNA interference (RNAi) pathway would be ineffective tethers because of their size.

Another property of lncRNAs is their ability to specify a unique address through use of a large sequence space. Transcription factors are effective recruiting factors for epigenetic complexes, but an advantage of lncRNAs is the possibility of targeting to a single location. Transcription factors recognize short DNA motifs that typically occur thousands of times in the genome. Transcription factors therefore necessarily act within large networks and affect hundreds of genes at once. By contrast, lncRNAs like Tsix and RepA/Xist occur only once. This singularity enables delivery of epigenetic complexes to a unique address, offering a regulatory specificity not possible with proteins and small RNAs. Because there are no a priori limits on length and composition, the sequence space for lncRNA-mediated targeting greatly exceeds that of binding motifs for the proteome.

Local and Genome-Wide Control

A collaboration between site-specific lncRNAs and network-based transcription factors together with chromatin modifiers would account for spatial and temporal specificity during development. For example, whereas the transcription factor OCT4 responds to leukemia inhibitory factor and bone morphogenetic protein signaling in pluripotent cells to activate a genome-wide transcription program, local effects may be achieved through OCT4-responsive lncRNAs, such as Miat (28), Xite, Tsix, and Xist (21, 29). At the Xic, for instance, OCT4 activates Tsix and Xite and in doing so controls X-chromosome pairing and initiation of the XCI cascade. Tsix RNA in turn interacts with chromatin complexes, such as PRC2 (17) and Dnmt3a (23), breaking the symmetry between the future Xa (active X) and Xi. OCT4's developmental specificity thereby dictates timing of XCI (21, 29). In the same way, its developmental specificity influences other genetic targets, each of which initiates its own local and genome-wide cascades. The net effect of a single transcription factor on multiple downstream lncRNA targets would thus be multiplied by cycles of epigenetic reprogramming on gene-specific and network-wide scales, with each event relying on continuous collaborations among transcription factors, lncRNAs, and the epigenetic complexes recruited by the lncRNAs. The thousands of transcription factors and lncRNAs operating in parallel would then achieve the necessary local and genome-wide changes for development and for discrete responses to environmental signals.

Classes of LncRNAs

Genomic imprinting. Genomic imprinting is an epigenetic phenomenon in which genes are
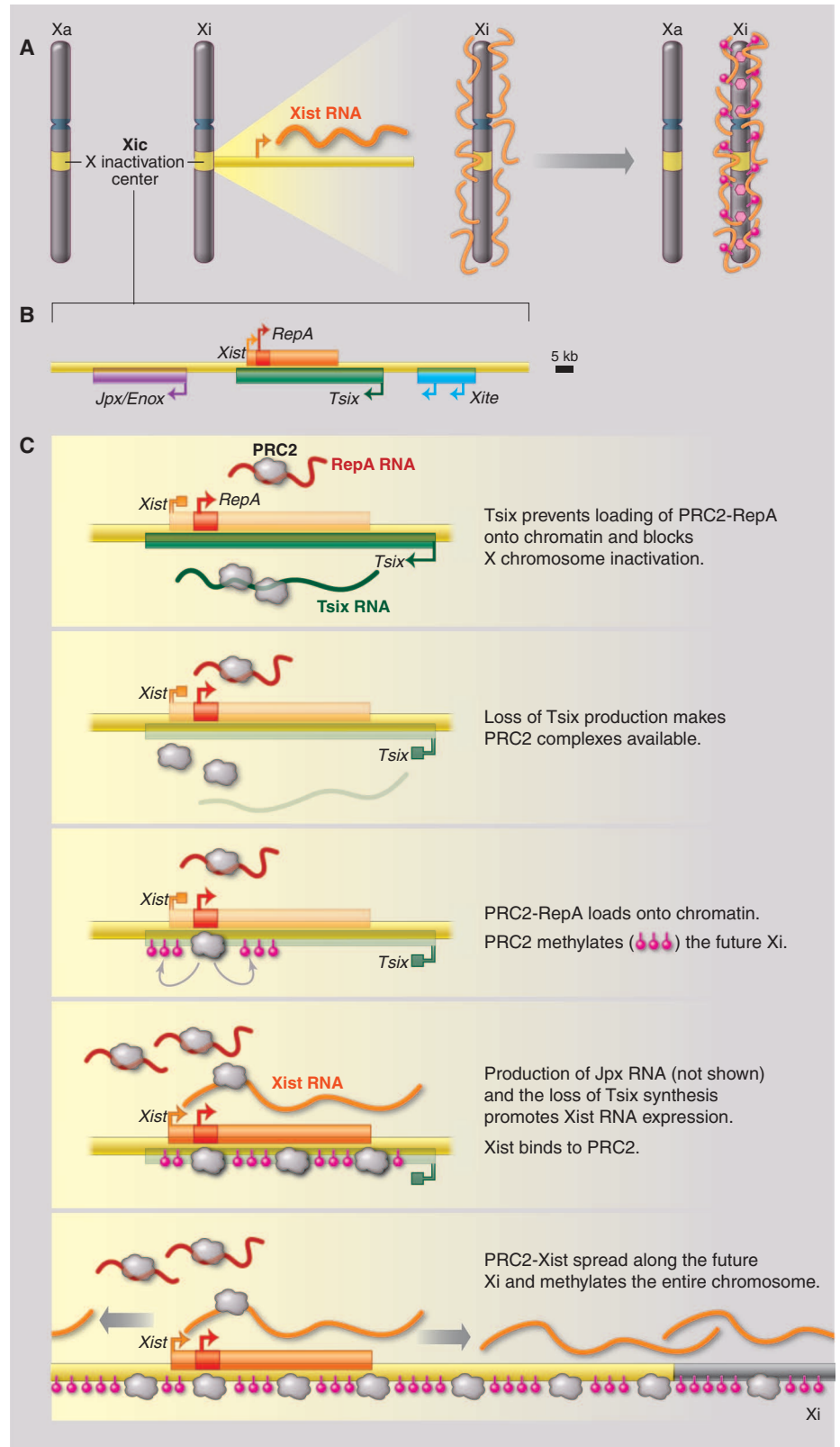
Fig. 1. LncRNAs in X-chromosome inactivation. (**A**) The lncRNA Xist is transcribed from the *Xic* of the inactive X chromosome (Xi). Xist RNA covers the entire chromosome and silences gene expression through epigenetic modification of histones and DNA. (**B**) The core region of the *Xic* and its lncRNAs. (**C**) LncRNA-protein interactions at the initiation of XCI.

expressed from the allele of only one parent (*11*, *12*). The process bears a striking resemblance to XCI in that imprinting occurs within clusters and requires coordinated allelic regulation within the region, and its control regions are graced by lncRNAs. Whether these lncRNAs function as regulatory transcripts is currently debated. One of the first lncRNAs to be identified, H19 (*30*), is reciprocally imprinted with *insulin-like growth factor 2* (*Igf2*) and is highly expressed, but its

deletion has no phenotype (*31*). H19 now appears to function as microRNA (miRNA) precursor (*32*). In mice, the Igf2 receptor gene (*Igf2r*) is regulated by the antisense *Airn* gene (*33*). Some genetic studies suggest that *Airn* regulates *Igf2r* through its act of transcription rather than through its lncRNA (*34*, *35*). However, a biochemical study proposes that Airn RNA recruits the histone methyltransferase, G9a, to silence *Igf2r* (*36*). By using formaldehyde–cross-linked chromatin, the study left open the possibility that the contacts occur indirectly through a chromatin-bound intermediary rather than directly between lncRNA and G9a. Within the Beckwith-Wiedemann syndrome locus, a long antisense transcript (Kncq1ot1) may likewise associate with G9a and PRC2 (*37*). A genome-wide analysis using RNA immunoprecipitation-sequencing (RIP-seq) suggests that thousands of lncRNAs associate with PRC2 (*38*): At the *Dlk1-Gtl2* locus, Gtl2 RNA may target PRC2 to the reciprocally imprinted *Dlk1* locus in cis; in the *Nesp/Gnas* cluster, the antisense transcript of *Nesp* (Nespas) may recruit PRC2 to control *Nesp* (*38*). Because the mechanisms of action remain to be investigated in detail, it is too soon to state whether these noncoding elements will have universal roles as lncRNAs in genomic imprinting.

Beyond allelic phenomena. LncRNA's function is not limited to the control of allelic expression (Table 1). In addition to the RIP-seq analysis identifying thousands of RNAs in the PRC2 transcriptome (*38*), analysis of RIP products on a microarray (RIP-chip) suggests that PRC2 and the LSD1/REST/coREST complex associate with hundreds of RNA (*39–41*) and that promoters of Polycomb target genes often make short transcripts that also associate with PRC2 (*42*). These PRC2 targets include hundreds that map to nonallelically regulated loci, including those involved in cancer and stem cell differentiation (*38*). The biochemically distinct Polycomb complex, PRC1, also interacts with RNA. The *INK4b/ARF/INK4a* tumor suppressor locus is controlled by PRC1 in a manner dependent on ANRIL lncRNA (*43*). Interestingly, interactions between a PRC1 subunit (PC2) and MALAT1 and TUG lncRNAs occur during movement of genes between nuclear compartments and during gene activation (*44*).

Noncoding RNAs frequently localize to gene promoters. These

promoter-associated short RNAs (PASRs) are typically short (50 to 200 nt) and have been considered abortive transcripts made by stalled or paused polymerases (*42*, *45–48*). Pausing or stalling could prolong tethering and facilitate recruitment of factors (Fig. 2). For example, short lncRNAs made from *CCND1* (cyclin D1) may tether a transcription repressor, TLS, to the *CCND1* promoter and allosterically modify the repressor to turn off transcription (*49*). LncRNAs also occur in exons, introns, and other unusual spaces (*8*). In the plant *Arabidopsis thaliana*, the 1.1-kb intronic COLDAIR transcript controls flowering time by targeting PRC2 to silence *FLC* (*Flowering Locus C*) (*50*). In mice, RepA (within *Xist*'s first exon) targets PRC2 to the *Xist* promoter (*17*). Antisense transcripts may originate anywhere within or downstream of the genes they regulate [e.g., Tsix and Bdnf-as (*19*, *51*)]. LncRNAs also originate within retrotransposons, the ubiquitous repetitive elements once regarded as junk. Transcription from short interspersed element (SINE) B2, for example, may form a chromatin boundary for the growth hormone locus (*52*).

LncRNAs as activators. LncRNAs also serve as activators of gene expression. The brain-specific Evf2 originates within an enhancer between *Dlx5* and *Dlx6* and is proposed to aid Dlx2-mediated activation of *Dlx5/Dlx6* (*53*). In XCI, Jpx RNA is required to induce *Xist* expression (*25*). From the *HoxA* cluster, two lncRNAs could recruit H3K4 trimethylases: Interaction between MLL and Hottip RNA is proposed to control activation of proximal *HoxA* genes (*54*), and interaction between MLL and Mistral RNA is thought to activate neighboring *HoxA6* and *HoxA7* genes (*55*). Recent studies showed abundant transcription through neuronal enhancers (*56*) and enhancer-like regions (*6*). Principles governing repressive lncRNA could also apply to activating lncRNAs, but much work remains to be done in this area because mechanistic details are currently lacking. In some cases, activation may depend on the act of transcription through associated chromatin rather than on the transcript.

Pseudogenes as regulators. Long regarded as a genomic graveyard, pseudogene space may turn out to be a vast repository of regulatory lncRNAs. For example, a pseudogene of *Makorin1* makes a truncated noncoding transcript that could stabilize the mRNA of the parent gene (*57*). The *asOct4-ps5* pseudogene could regulate *Oct4* by hybridizing with the Oct4 mRNA, thereby targeting silencing complexes to the *Oct4* promoter (*58*). Pseudogenes of the tumor suppressor, *Pten*, have been proposed as decoys for miRNAs that bind to and down-regulate Pten mRNA (*59*). The proposed mechanisms remain controversial and require further investigation.

Fundamental Differences: Cis Versus Trans

LncRNAs operate not only in cis but also in trans (Table 1). Fundamental differences exist between these two categories. Cis-acting RNAs are restricted to the site of synthesis and directly act
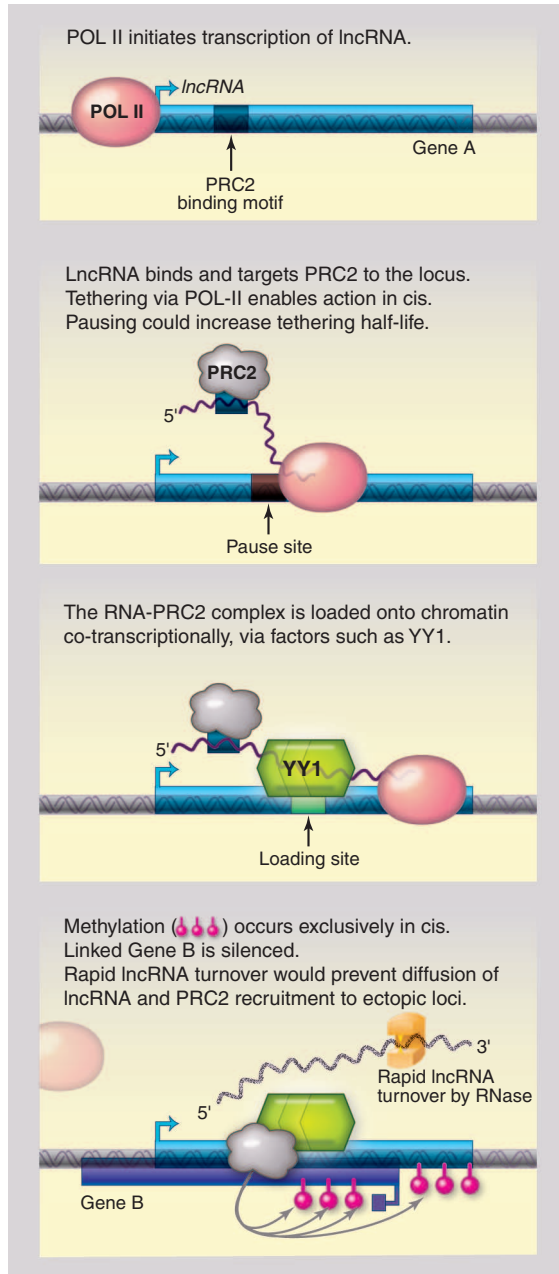
Fig. 2. LncRNAs tether epigenetic complexes to chromatin, enabling allele- and locus-specific regulation. LncRNA that is synthesized binds to an epigenetic complex (such as PRC2) and, together, are loaded onto chromatin cotranscriptionally through DNA-bound factors (such as YY1 for Xist RNA). Epigenetic modifications then silence the gene, and rapid lncRNA turnover prevents its diffusion to other loci.

on one or several linked, generally contiguous, genes on the same chromosome. By contrast, trans-acting RNAs diffuse from the site of synthesis and can act directly on many genes at great distances, including at other chromosomes. In doing so, trans-acting RNAs, like transcription factors and small RNAs, are more likely to converge on and act within large genic networks. Both mechanisms imply direct action on target genes rather than secondary downstream effects. Whereas cis-acting lncRNAs are exemplified by mammalian XCI, trans-acting lncRNAs are exemplified by fruit fly dosage compensation, which occurs by hypertranscribing the single male X chromosome and which therefore does not require allelic discrimination. Indeed, the lncRNAs, roX1 and roX2 (60, 61), scaffold the MSL-MOF protein complex, which in turn binds hundreds of X-linked sites (62, 63).

Whether mammalian lncRNAs generally work in cis or trans has been debated (6, 7, 40, 56, 64, 65). There will likely be many members in each class. Some might even blur the distinction. Xist RNA, for example, acts in cis to repress genes on Xi in the normal context but could diffuse to ectopic sites when *Xist* transgenes are introduced de novo (18). This observation indicates that Xist is actually diffusible but is prevented from acting in trans by developmental programming that masks non-*Xic* YY1-binding sites. Thus, cis-acting transcripts may operate in cis only to the extent that they are properly anchored and ectopic sites masked.

Although mechanistically and consequentially very different, the Xist case illustrates why cis versus trans lncRNAs are not always easily distinguished. Complicating the matter is a lack of knowledge of how the vast majority of lncRNAs work. One key unanswered question is whether trans-acting lncRNAs could also target chromatin complexes. The property of targeting is reserved for cases in which RNA is directly involved in guiding a complex to a specific locus. Targeting is easier to conceive for cis-regulatory RNAs, which are naturally tethered to the site of transcription (Fig. 2). A targeting role for diffusible, trans-acting RNA is possible, however, if they engage in RNA:DNA triplex formation via Hoogsteen base-pairing, as has been proposed for pRNA (promoter-associated)–mediated recruitment of DNMT3b to ribosomal gene (rDNA) promoters (66). Another hypothesized targeting mechanism involves hybridization between complementary RNA strands of the antisense Oct4 pseudogene and the target *Oct4* mRNA (58).

Site-specific targeting seems unlikely to be a property of most trans-acting lncRNAs. Existing examples suggest that many have functions akin to those of protein factors. Steroid receptor RNA (SRA), a "subunit" of nuclear hormone receptors for estrogen, androgen, glucocorticoid, and progesterone (67), functions as a coactivator in the same way as traditional protein-based coactivators such as p300 and the transcription factor cyclic adenosine monophosphate (cAMP) response element binding protein (CREB). SINE B2 RNAs, a product of retrotransposons, directly bind Pol II and down-regulate Pol II activity at mammalian heat-shock genes, thereby functioning like protein-based co-repressors (68). LncRNAs also scaffold protein complexes. HOTAIR from the *HOX-C* locus (39) scaffolds PRC2 and LSD1 (a histone H3K4 demethylase) (41, 69), and genome-wide analyses show that it localizes to thousands of sites in disease states (62, 70) as well as to the *HOX-D* cluster in trans (39). Other functions of trans-acting lncRNAs are only beginning to emerge, and many remain incompletely defined. The expression of an lncRNA called LincRNA-p21 and the p21-associated ncRNA DNA damage–activated (PANDA) lncRNA (71, 72) are induced during the p53 response and correlate with gene expression changes. The abundant MALAT1 and TUG1 lncRNAs differentially interact with methylated and unmethylated forms of PC2 (PRC1 subunit) to relocate growth-control genes between nuclear substructures for transcription activation (44). Thus, unlike the site-specific cis-acting lncRNAs, trans-acting transcripts operate within large genic networks.

Prospects and Conclusions

The lncRNA field, once a boutique field for imprinting and dosage compensation, has now grown to include mainstream biochemists, genomicists, and computational biologists. The accelerated discovery of thousands of lncRNAs over the past 5 years is outpacing our ability to vet function and mechanism. Indeed, few lncRNA knockouts have yielded robust phenotypes. H19 knockout mice have a normal phenotype (31). A *HoxC* deletion including *Hotair* does not abolish PRC2 targeting (73). Knockouts of Malat1 and Neat1, two of the most abundant nuclear lncRNAs, yielded normal, viable mice (74–76). Although these findings lead some to question the importance other lncRNAs, they do not exclude the possibility of developmental compensation, as often happens in knockout models, nor do they exclude the possibility that subtle effects would emerge in tissue-specific studies.

Related issues are the inherent differences in comparing knockout to knockdown phenotypes. Before RNAi, gene deletions were the staple of proving function in vivo. Reduction of gene expression by small RNAs may not recapitulate knockouts, especially because nuclear lncRNAs are often difficult to knock down and residual levels of lncRNAs after a knockdown could yield phenotypes different from knockouts that completely eliminate lncRNA production. Analysis of lncRNA function would ideally also include

Table 1. Emerging themes in lncRNA regulation. Potential groupings of lncRNA based on proposed interactions, functions, and mechanisms. Representative lncRNAs of each group are shown.

Functions	Examples	Action	Hypothesized mechanisms
Cis-tether cis-targeting	RepA, Xist, Tsix, Gtl2, Kcnq1ot1, Airn, Hottip, ANRIL, Oct4-ps5, COLDAIR, Evf2, BDNF-AS	Cis	Co-transcriptional targeting of chromatin factors; allelic and locus-specific action; repressive or activating
Trans-targeting	pRNA, asOct4-ps5	Trans	RNA:DNA triplex via Hoogsteen base pairing or sense:antisense Watson:Crick base pairing
Enhancer	Xite, ncRNA-a7, eRNAs	Cis	Mediated by RNA, transcriptional force, or chromatin intermediaries
Decoy	PTEN-ps Tsix	Trans Cis	Competitive inhibition of target protein or RNA
Scaffold	MALAT1, TUG1, NEAT1, HOTAIR, roX1, roX2	Trans	RNA subunit of effector complex
Allosteric modification	TLS	Cis (trans also possible)	Alters conformation and activity of protein partners
Coactivator or co-repressor	SRA, SINE B2, Jpx, pRNA	Trans (cis also possible)	Accessory factor for transcriptional activation or repression

gain-of-function experiments, because exaggerated phenotypes would bolster arguments for proposed effects.

Equally important considerations include experiments to distinguish among RNA, transcriptional activity, and underlying chromatin as the mechanism of action for a noncoding locus. Because they are inextricably linked, distinguishing among them may not be trivial. The movement of Pol II unwinds the DNA duplex, alters chromatin structure, and could effect epigenetic change independently of the nascent transcript. For example, transcript truncation experiments have shown that the lncRNA made from *Tsix*'s enhancer, *Xite*, does not equate with deleting *Xite*, suggesting that *Xite* operates through its transcription or chromatin change rather than through Xite RNA (*77*). A similar conclusion was reached regarding transcription of *Airn* (*35*). Thus, transcript truncation experiments might be generally helpful in analysis of lncRNA function. When the RNA per se is shown to be necessary, what they recruit and how they interact with protein partners would need elucidation before general principles could be articulated. To date, few mammalian lncRNAs have been tested in these ways.

Examples discussed herein caution against treating all lncRNAs as a single class of molecules. Some classification schemes have been based on geography, resulting in the coinage of catchy names like PASR for promoter-associated short RNAs (*45*), NAT for natural antisense transcripts (*78*), lincRNA for large intervening noncoding RNA (*2*), and eRNA for enhancer-associated RNAs (*6*, *56*). Yet a transcript's location may not instruct function or mechanism. Jpx resides in cis to *Xist*, but Jpx acts in trans (*25*), and an antisense orientation does not necessarily imply repressive effects (*6*). Distinction by cis versus trans action might be a logical first step. Eventually, it will be useful to think in terms of transcriptomes. Although a transcriptome has so far only been identified for PRC2 (*17*, *38–40*, *42*), those for other epigenetic complexes would not be difficult to envision.

The mechanisms outlined in Table 1 likely provide only a glimpse of the full range of lncRNA functions as others await discovery. In addition to the more traditional roles in scaffolding, coactivation, and co-repression of trans-acting RNAs, cis-acting lncRNAs are now known to perform crucial targeting functions in a site-specific manner. Until recently, the proteome was thought to possess its own targeting specificity. LncRNAs are now known to confer a degree of temporal and spatial specificity not possible with proteins and small RNAs. Allelic and locus-

specific control might have been a major driving force behind the genome-wide fixation of lncRNAs. Eventually, lncRNAs may be seen to play roles as far-reaching as small RNAs and proteins, given that disease-associated mutations occur more often in coding than noncoding space (*79*) and that species-specific differences among organisms now appear to originate more in noncoding space than in the few thousand coding genes we share. With the site-specific action of cis-acting lncRNAs, drugs designed against lncRNAs could circumvent pleiotropic effects that plague many current treatment modalities that target enzymatic activities within epigenetic complexes. Indeed, the Wild West is a rich landscape waiting to unfold.

References and Notes

1. P. Kapranov, A. T. Willingham, T. R. Gingeras, *Nat. Rev. Genet.* **8**, 413 (2007).
2. M. Guttman et al., *Nature* **458**, 223 (2009).
3. C. P. Ponting, P. L. Oliver, W. Reik, *Cell* **136**, 629 (2009).
4. J. E. Wilusz, H. Sunwoo, D. L. Spector, *Genes Dev.* **23**, 1494 (2009).
5. H. Jia et al., *RNA* **16**, 1478 (2010).
6. U. A. Ørom et al., *Cell* **143**, 46 (2010).
7. M. N. Cabili et al., *Genes Dev.* **25**, 1915 (2011).
8. S. Djebali et al., *Nature* **489**, 101 (2012).
9. M. B. Clark et al., *Genome Res.* **22**, 885 (2012).
10. H. Tani et al., *Genome Res.* **22**, 947 (2012).
11. C. A. Edwards, A. C. Ferguson-Smith, *Curr. Opin. Cell Biol.* **19**, 281 (2007).
12. L. B. Wan, M. S. Bartolomei, *Adv. Genet.* **61**, 207 (2008).
13. J. T. Lee, *Nat. Rev. Mol. Cell Biol.* **12**, 815 (2011).
14. A. Wutz, *Nat. Rev. Genet.* **12**, 542 (2011).
15. C. J. Brown et al., *Cell* **71**, 527 (1992).
16. C. M. Clemson, J. A. McNeil, H. F. Willard, J. B. Lawrence, *J. Cell Biol.* **132**, 259 (1996).
17. J. Zhao, B. K. Sun, J. A. Erwin, J. J. Song, J. T. Lee, *Science* **322**, 750 (2008).
18. Y. Jeon, J. T. Lee, *Cell* **146**, 119 (2011).
19. J. T. Lee, L. S. Davidow, D. Warshawsky, *Nat. Genet.* **21**, 400 (1999).
20. J. T. Lee, N. Lu, *Cell* **99**, 47 (1999).
21. M. E. Donohoe, S. S. Silva, S. F. Pinter, N. Xu, J. T. Lee, *Nature* **460**, 128 (2009).
22. T. Sado, Y. Hoki, H. Sasaki, *Dev. Cell* **9**, 159 (2005).
23. B. K. Sun, A. M. Deaton, J. T. Lee, *Mol. Cell* **21**, 617 (2006).
24. Y. Ogawa, B. K. Sun, J. T. Lee, *Science* **320**, 1336 (2008).
25. D. Tian, S. Sun, J. T. Lee, *Cell* **143**, 390 (2010).
26. L. Duret, C. Chureau, S. Samain, J. Weissenbach, P. Avner, *Science* **312**, 1653 (2006).
27. J. T. Lee, *Genes Dev.* **23**, 1831 (2009).
28. J. Sheik Mohamed, P. M. Gaughwin, B. Lim, P. Robson, L. Lipovich, *RNA* **16**, 324 (2010).
29. P. Navarro et al., *Science* **321**, 1693 (2008).
30. M. S. Bartolomei, S. Zemel, S. M. Tilghman, *Nature* **351**, 153 (1991).
31. B. K. Jones, J. M. Levorse, S. M. Tilghman, *Genes Dev.* **12**, 2200 (1998).
32. X. Cai, B. R. Cullen, *RNA* **13**, 313 (2007).
33. A. Wutz et al., *Nature* **389**, 745 (1997).
34. F. M. Pauler, M. V. Koerner, D. P. Barlow, *Trends Genet.* **23**, 284 (2007).
35. P. A. Latos et al., *Science* **338**, 1469 (2012).
36. T. Nagano et al., *Science* **322**, 1717 (2008); 10.1125/science.1163802.
37. R. R. Pandey et al., *Mol. Cell* **32**, 232 (2008).
38. J. Zhao et al., *Mol. Cell* **40**, 939 (2010).
39. J. L. Rinn et al., *Cell* **129**, 1311 (2007).
40. A. M. Khalil et al., *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11667 (2009).
41. M. C. Tsai et al., *Science* **329**, 689 (2010); 10.1126/science.1192002.
42. A. Kanhere et al., *Mol. Cell* **38**, 675 (2010).
43. K. L. Yap et al., *Mol. Cell* **38**, 662 (2010).
44. L. Yang et al., *Cell* **147**, 773 (2011).
45. P. Kapranov et al., *Science* **316**, 1484 (2007); 10.1126/science.1138341.
46. L. J. Core, J. T. Lis, *Science* **319**, 1791 (2008).
47. A. C. Seila et al., *Science* **322**, 1849 (2008); 10.1126/science.1162253.
48. R. J. Taft et al., *Nat. Genet.* **41**, 572 (2009).
49. X. Wang et al., *Nature* **454**, 126 (2008).
50. J. B. Heo, S. Sung, *Science* **331**, 76 (2011); 10.1126/science.1197349.
51. F. Modarresi et al., *Nat. Biotechnol.* **30**, 453 (2012).
52. V. V. Lunyak et al., *Science* **317**, 248 (2007).
53. J. Feng et al., *Genes Dev.* **20**, 1470 (2006).
54. K. C. Wang et al., *Nature* **472**, 120 (2011).
55. S. Bertani, S. Sauer, E. Bolotin, F. Sauer, *Mol. Cell* **43**, 1040 (2011).
56. T. K. Kim et al., *Nature* **465**, 182 (2010).
57. S. Hirotsune et al., *Nature* **423**, 91 (2003).
58. P. G. Hawkins, K. V. Morris, *Transcription* **1**, 165 (2010).
59. L. Salmena, L. Poliseno, Y. Tay, L. Kats, P. P. Pandolfi, *Cell* **146**, 353 (2011).
60. H. Amrein, R. Axel, *Cell* **88**, 459 (1997).
61. V. H. Meller, K. H. Wu, G. Roman, M. I. Kuroda, R. L. Davis, *Cell* **88**, 445 (1997).
62. C. Chu, K. Qu, F. L. Zhong, S. E. Artandi, H. Y. Chang, *Mol. Cell* **44**, 667 (2011).
63. M. D. Simon et al., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20497 (2011).
64. M. Guttman et al., *Nature* **477**, 295 (2011).
65. M. Guttman, J. L. Rinn, *Nature* **482**, 339 (2012).
66. K. M. Schmitz, C. Mayer, A. Postepska, I. Grummt, *Genes Dev.* **24**, 2264 (2010).
67. R. B. Lanz et al., *Cell* **97**, 17 (1999).
68. P. D. Mariner et al., *Mol. Cell* **29**, 499 (2008).
69. S. Kaneko et al., *Genes Dev.* **24**, 2615 (2010).
70. R. A. Gupta et al., *Nature* **464**, 1071 (2010).
71. M. Huarte et al., *Cell* **142**, 409 (2010).
72. J. Hung, H. Y. Chang, *RNA Biol.* **7**, 582 (2010).
73. P. Schorderet, D. Duboule, *PLoS Genet.* **7**, e1002071 (2011).
74. S. Nakagawa, T. Naganuma, G. Shioi, T. Hirose, *J. Cell Biol.* **193**, 31 (2011).
75. M. Eißmann et al., *RNA Biol.* **9**, 1076 (2012).
76. B. Zhang et al., *Cell Rep* **2**, 111 (2012).
77. Y. Ogawa, J. T. Lee, *Mol. Cell* **11**, 731 (2003).
78. M. Magistri, M. A. Faghihi, G. St Laurent 3rd, C. Wahlestedt, *Trends Genet.* **28**, 389 (2012).
79. T. A. Manolio, *N. Engl. J. Med.* **363**, 166 (2010).