Leading Edge
Review

Cell

# Evolution and Functions of Long Noncoding RNAs

Chris P. Ponting,[1,*] Peter L. Oliver,[1] and Wolf Reik[2]
[1]MRC Functional Genomics Unit, University of Oxford, Department of Physiology, Anatomy and Genetics, South Parks Road, Oxford, OX1 3QX, UK
[2]Laboratory of Developmental Genetics and Imprinting, The Babraham Institute, Cambridge CB22 3AT and Centre for Trophoblast Research, University of Cambridge, Cambridge CB2 3EG, UK
*Correspondence: chris.ponting@dpag.ox.ac.uk
DOI 10.1016/j.cell.2009.02.006

**RNA is not only a messenger operating between DNA and protein. Transcription of essentially the entire eukaryotic genome generates a myriad of non-protein-coding RNA species that show complex overlapping patterns of expression and regulation. Although long noncoding RNAs (lncRNAs) are among the least well-understood of these transcript species, they cannot all be dismissed as merely transcriptional "noise." Here, we review the evolution of lncRNAs and their roles in transcriptional regulation, epigenetic gene regulation, and disease.**

Eukaryotic genomes are not the simple, well-ordered substrates of gene transcription that was once believed. We now know them to transcribe a broad spectrum of RNA molecules, ranging from long protein-coding mRNAs to short noncoding transcripts, which frequently overlap or are interleaved on either strand (Figure 1). If RNA types were to have their own color, each eukaryotic genome would continuously be emitting a riot of hues, with some regions radiating across the entire spectrum as, for example, development unfolds. Although untranscribed nucleotides in genomes are rare, alternative combinations of exons are widespread. The large proportion of a eukaryotic genome that is transcribed thus produces a huge array of RNA molecules differing in size, abundance and protein-coding capability.

In stark contrast to this diversity of RNA species, only a small number of non-protein-coding transcripts currently have experimentally-derived functions. Moreover, only rarely have disease-associated mutations been identified outside of protein-coding genes. Might, therefore, this colorful pageant of genomic transcription be a mirage? Might much of a genome's repertoire of non-protein-coding transcripts be inconsequential transcriptional "noise"? Here, we review evidence for whether pervasive transcription is consequential, drawing first upon evolutionary signatures of functionality in genome sequences, and then upon experimental findings about the functions of noncoding transcripts, particularly with respect to transcriptional regulation. We will focus on long noncoding RNAs (lncRNAs, >200 nucleotides) that are, perhaps, the least well-understood products of transcription from genomes.

Defining lncRNAs simply on the basis of size and lack of protein-coding capability is intellectually far from satisfying. However, the ease by which these transcripts can now be sequenced, together with our current imperfect understanding of their functions, explain the need for such a broad categorization of what are likely to be functionally heterogeneous molecules. There is sense, too, in defining noncoding transcripts by their absence of protein-coding capacity given that our ability to identify protein-coding transcripts has improved immeasurably in recent years. (Methods that seek to distinguish coding from noncoding transcripts are reviewed elsewhere; see Dinger et al., 2008.) As the predicted numbers of protein-coding genes present in mammalian genomes have tumbled in recent years, the realization has grown that many long, previously wrongly annotated protein-coding genes, instead represent nonprotein-coding transcripts (Ponting, 2008). It is true that an unknown fraction of these transcripts might, indeed, encode (particularly small) proteins, hence some would prefer them labeled as "transcripts of unknown function" (TUFs) (Cheng et al., 2005). Others describe sets of transcripts that share expressed regions and splicing events, transcription start sites or termination events as "transcriptional frameworks" (Carninci et al., 2005). For the purposes of this Review, we shall describe as noncoding RNAs all transcripts that have neither experimental nor evolutionary evidence for an open reading frame encoding a functional protein.

## Pervasive Transcription

It is expected that at some time, and in at least one cell type, virtually every euchromatic nucleotide in the human euchromatic genome will be transcribed (Birney et al., 2007). Nevertheless, only a small proportion (~5%–10%) of the genome is covered by sequences of mRNAs and spliced noncoding RNAs that are stably transcribed in cell lines (Bertone et al., 2004; Cheng et al., 2005; Kapranov et al., 2007a). Of these sequences, the vast majority, however, do not encode protein. Rather, they represent untranslated sequences within transcripts emanating from protein-coding loci or else from loci without protein-coding capacity. Only ~1% of the human genome encodes proteins, leaving another ~4%–9% that is transcribed but yet whose functions are unknown.

The near ubiquity of transcription across genomes has been demonstrated by diverse methods, including whole genome tiling arrays and transcriptome sequencing (reviewed in Kapranov et al., 2007b). It has also been shown for diverse eukaryotes

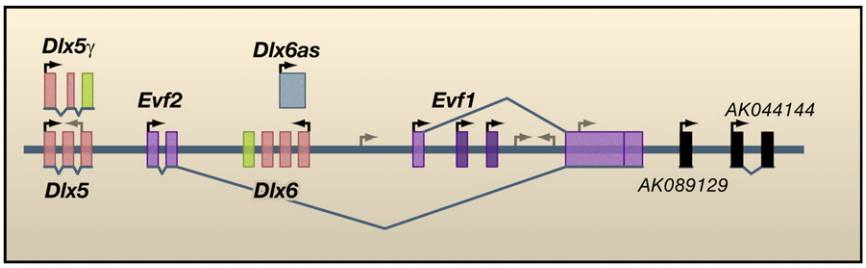


**Figure 1. Pervasive Transcription from the *Dlx5/6* Locus**

This locus resides on mouse chromosome 6 (~6.527–6.628 Mb for version mm5 of the genome assembly). Coding and noncoding exons are shown in pink and blue, respectively, with the orientations of transcription indicated by arrows. Transcriptional start sites indicated by CAGE (cap analysis of gene expression) tags but currently without known transcript sequence are shown in gray. With respect to the five categories of noncoding RNAs: two single exon noncoding RNAs (purple) are intronic and are transcribed in sense to *Evf1/2*; *Evf2* and *Dlx6as* noncoding RNAs are antisense to *Dlx6*; *Evf1* is an intergenic noncoding RNA (this locus is not known to give rise to bidirectional transcripts). Figure modified from http://fantom32p.gsc.riken.jp/gev-f3/gbrowse/mm5/.

ranging from plants to animals and, most recently, to fungi such as the fission yeast *Schizosaccharomyces pombe* (Dutrow et al., 2008; Wilhelm et al., 2008) and the budding yeast *Saccharomyces cerevisiae* (Nagalakshmi et al., 2008). Genome-wide transcription thus appears to have been prevalent among many, if not all, eukaryotes since their last common ancestor over a billion years ago.

In mammals, transcription on a genome-wide scale is achieved using a breathtaking number of transcription events. Approximately 180,000 mouse cDNAs have been identified that, together with a similar number of core promoters, are distributed among 44,000 transcript clusters (Carninci et al., 2005; Carninci et al., 2006). By way of contrast, the mouse genome contains only ~20,210 protein-coding gene loci (L. Goodstadt and C.P.P., unpublished data). A large number of promoters have been inferred from the mapping to the genome of short sequence tags derived from the capped 5′ ends of cDNAs (Carninci et al., 2005). These CAGE (cap analysis of gene expression) tags were located, as expected, at the 5′ ends of well-established protein-coding genes. However, they were also found in more exotic locations, including within exons and, increasingly, toward the ends of the 3′ ends of untranslated regions (UTRs) of coding transcripts (Carninci et al., 2005). For most such CAGE tags, it is unknown whether each tag represents a long or a short RNA, whether this RNA is expressed in few or many tissues, and at few or many developmental stages.

A genome's transcriptional repertoire is made even more complex by different combinations of exons and polyadenylation sites. The majority of protein-coding genes, and ~30% of noncoding RNAs, produce at least one alternative transcript that samples different exon combinations (Kampa et al., 2004; Ravasi et al., 2006). Moreover, about half of human or mouse genes exhibit alternative polyadenylation among their transcripts (Tian et al., 2005). As a result, each defined 3′ end of a transcript has, on average, 1.3 start sites; conversely, each 5′ end has an average of 1.8 3′ ends (Carninci et al., 2005).

Genome-wide transcription thus produces a complex population of transcripts, some that are overlapping, others entirely distinct; some that are same strand, others opposite strand; and, some that encode proteins, whereas others do not. What, then, is a gene (Gerstein et al., 2007; Gingeras, 2007)? One previous definition is that it is a genomic region that, when transcribed, produces a polyadenylated mRNA that encodes a protein. Observations that coding and noncoding transcripts are interwoven in a rich tapestry suggest that this definition needs revision, perhaps by insisting that transcripts, not genes, are the operational unit of a genome (Gingeras, 2007).

Of all these transcripts, it has been those that are protein-coding, rather than noncoding, that have been the principal focus of experimental biology yet, surprisingly, clusters of overlapping noncoding transcripts are more numerous in the mouse genome than coding transcript clusters (Carninci et al., 2005). Protein-coding loci tend to be expressed at high levels, whereas larger numbers of non-protein-coding loci are expressed, sometimes in a regulated manner (see below), at lower levels, suggesting that noncoding RNAs represent cryptic signals that control complex developmental processes (Mattick and Makunin, 2006). If this is so, then proof is required that large numbers of these transcripts are functional in vitro and in vivo, that their expression is regulated, that they have been the subject of selection, and perhaps also, for some, that they contain folded RNA secondary structures.

It is important to emphasize that evidence for functions of lncRNAs remains scant. This may reflect current limitations in experimental tools available to elucidate lncRNA mechanisms, or it may indicate that lncRNAs contribute in only a minor capacity to the functional repertoire of a genome. In the interim, before this issue is finally settled, we should take care not to conclude that pervasive transcription necessarily implies an abundance of functional lncRNAs. Furthermore, we ought not to make sweeping generalizations about molecular mechanisms on the basis of a few, albeit well-established, case studies.

## Long Noncoding RNAs

Classes of noncoding transcripts can be divided between housekeeping noncoding RNAs and regulatory noncoding RNAs. Housekeeping noncoding RNAs include ribosomal, transfer, small nuclear and small nucleolar RNAs and are usually expressed constitutively. Among short regulatory noncoding RNAs are microRNAs, small interfering RNAs and Piwi-associated RNAs (see Reviews by R.W. Carthew and E.J. Sontheimer on page 642 and C.D. Malone and G.J. Hannon on page 656 of this issue). Most transcribed, yet not protein-coding, sequence, however, is associated with lncRNAs (Mattick and Makunin, 2006). These may be located within nuclear or cytosolic fractions, may or may not be polyadenylated, and are often transcribed from either strand within a protein-coding locus (Birney et al., 2007; Carninci et al., 2005). A database providing expression and other information on mammalian lncRNAs has recently become available (Dinger et al., 2009).

Before considering their potential for functionality, it is important to consider whether a large proportion of proposed lncRNAs are, instead, artifacts of either experiment or computation. Among mouse transcripts assigned as noncoding RNAs in the FANTOM2 set, very few (<2%) have both long open reading frames and suppression of nucleotide substitutions at putative nonsynonymous sites, attributes that would have otherwise indicated coding capability (Ravasi et al., 2006). These noncoding RNAs also tend to be shorter and to have fewer introns than protein-coding transcripts (Ravasi et al., 2006), again arguing for an absence of coding capacity. Although few of the FANTOM2 noncoding RNAs appear to represent contamination from genomic DNA, up to 30% may represent fragments of unprocessed pre-mRNAs (Ravasi et al., 2006). The majority of this set thus appear to represent bona fide transcripts, a view that is corroborated by frequent regulation of their expression levels (Cawley et al., 2004; Ravasi et al., 2006).

An lncRNA can be placed into one or more of five broad categories: (1) sense, or (2) antisense, when overlapping one or more exons of another transcript on the same, or opposite, strand, respectively; (3) bidirectional, when the expression of it and a neighboring coding transcript on the opposite strand is initiated in close genomic proximity, (4) intronic, when it is derived wholly from within an intron of a second transcript (although these, as noted above, sometimes may represent pre-mRNA sequences), or (5) intergenic, when it lies within the genomic interval between two genes (Figure 1).

Between 5 and 30% of transcriptional units in diverse eukaryotes have been found to harbor *cis*-natural antisense transcripts (*cis*-NATs); the exact proportion is strongly dependent on the quantity of transcriptome sequence considered (Lapidot and Pilpel, 2006). When, for example, the large human CAGE tag dataset is analyzed, each protein-coding locus is associated with an average of nearly 6 *cis*-NATs (Conley et al., 2008). A *cis*-NAT may function by forming a double-stranded RNA with its complementary sense RNA to subsequently regulate transcription levels. However, several other antisense mechanisms are known, principally those involving transcriptional interference or the regulation of monoallelic expression (see below). These mechanisms imply that sense-antisense transcript pairs should be more coexpressed and more inversely expressed than expected by chance. Although such weak tendencies have been observed (Chen et al., 2005), individual sense-antisense pairs often exhibit more complex and irregular patterns of expression (Mercer et al., 2008).

Of the large numbers of intronic lncRNAs that have been proposed (Louro et al., 2008) many may, instead, be pre-mRNA fragments (see above). In one study, ~80% of such noncoding RNAs appear not to be expressed, at least in the three tissues studied (prostate, kidney and liver) (Louro et al., 2008). Those that are transcribed may yet be found as alternative, and perhaps coding, exons within rarely-expressed transcripts of protein-coding genes. Nevertheless, some intronic noncoding RNAs whose expression profiles contrast with those of their host protein-coding gene have been reported (Mercer et al., 2008). Such instances, particularly the 27% of those noncoding RNAs whose nucleotide sequences are conserved (Mercer et al., 2008), deserve further experimental investigation.

Even when compared to other noncoding RNAs, molecules for which we know next-to-nothing, intergenic noncoding RNAs remain a complete mystery. Those that are transcribed well away from protein-coding loci appear to have little opportunity to *cis*-regulate transcription within such loci. It appears unlikely that many of these act in *trans* by forming triplexes with double-stranded DNA, as they rarely show strong complementarity to sequence elsewhere in the genome. They may, however, often act in *trans* within large ribonucleoprotein complexes. Many intergenic noncoding RNAs are transcribed in close proximity to protein-coding genes (Bertone et al., 2004) and these are more likely to act in *cis*, perhaps through transcriptional interference (see below). About half of the intergenic noncoding RNAs in one study are transcribed near (<10kb) to protein-coding genes (Ponjavic et al., 2007). These, perhaps, represent the best candidates for investigating the transcriptional regulation of neighboring genes.

## Transcriptional Noise?

How many of these vast numbers of lncRNAs are functional, and how many represent "noisy" inconsequential transcription, such as from selectively unconstrained promoters that have arisen serendipitously in genomic sequence? This question cannot be addressed experimentally until we have accumulated diverse instances of the molecular mechanisms of noncoding RNAs (see below). In the meantime, it can be addressed theoretically by considering how sequence that is not functional is expected to evolve over long periods of time. If most lncRNAs result from transcriptional noise then it is expected that: (a) their expression would not vary from one tissue to another, or between developmental time-points; and, (b), their rates of sequence change would not differ from those of other sequences, such as transposable elements, that almost always have evolved neutrally.

In studies that detect transcripts whose expression differs among tissues, or over time, or in response to retinoic acid, comparable proportions of noncoding RNAs and mRNAs have been found (Cawley et al., 2004; Ravasi et al., 2006). Although this implies that, like mRNAs, the expression of such noncoding RNAs may be regulated, formally it remains possible that this regulation is fortuitous and thus not necessarily under selective constraint. To consider this, Khaitovich et al. (2006) compared the evolution of expression levels between and among humans and chimpanzees. They considered that if intergenic noncoding RNAs were not under selection then their expression diversity and divergence patterns would be the same among three tissues (brain, heart, and testis). Their observation, instead, that these patterns differed between these tissues was interpreted as indicating that intergenic transcripts possess functional roles. Intergenic noncoding RNAs represented half of all sequences showing differential expression between these two primate species. Although genetic drift may also contribute greatly to this expression level diversity, Khaitovich et al. (2006) propose that intergenic lncRNAs may have contributed as many functional changes to these two primate lineages as have protein-coding genes.

Evolutionary studies initially concluded that noncoding RNA sequences bear no evidence of functionality because they

appear as poorly conserved as other intergenic sequences (Pang et al., 2006; Wang et al., 2004). More recently, Ponjavic et al. (2007) described how a set of 3122 full-length noncoding RNAs exhibits signatures of functionality that are more usually associated with protein-coding genes. These noncoding RNAs show reduced nucleotide substitutions, insertions and deletions, both within their promoters and within their sequences; moreover, their dinucleotide splice sites are more frequently conserved than expected by chance. By contrast, transposable element sequences within these noncoding RNAs show no evidence for reduced nucleotide substitution. Thus, these noncoding RNAs seem to include many whose nonrepetitive sequences have been under constraint to preserve transcription, sequence and splicing over many tens of millions of years, the period of time necessary for reductions in evolutionary rates to become apparent.

Compared with protein-coding sequence, however, noncoding RNA sequence tends to be only weakly constrained. Traditionally, constraint is estimated from the nucleotide substitution rate in functional sequence as a proportion of the rate in neutral, unconstrained, sequence. For proteins, on average, this ratio is ~10%, whereas for the set of 3122 full-length, and mainly intergenic, noncoding RNAs it is ~90%–95% (Ponjavic et al., 2007). One interpretation of this is that considerably more nucleotide substitutions are deleterious in protein-coding sequence compared with noncoding sequence. This would not be too great a surprise given the stringent thermodynamic, structural and functional constraints on protein sequences.

A new study recapitulates these earlier findings that lncRNAs are frequently under constraint (Guttman et al., 2009). Another interpretation of the frequently low sequence conservation of noncoding RNAs is that they may be frequently acted upon by positive selection (Hyashizaki, 2004; Pang et al., 2006). This, however, appears unlikely as only one example of positive selection on lncRNAs, resulting from Darwinian evolution, has been proposed. *HAR1F*, a noncoding RNA, was identified as the single region of the human genome that has undergone the most rapid sequence change in the human lineage since our last common ancestor with chimpanzees (Pollard et al., 2006). *HAR1F* is expressed specifically in Cajal–Retzius neurons in the developing human neocortex, thus providing support for claims that its rapid evolution has contributed to human-specific alterations in brain size and function. However, more recent scrutiny has indicated that it is not frequent episodes of adaptive evolution that has driven the evolution of *HAR1F*, and other similar regions of the human genome. Instead, rapid change has resulted from recent and local increases in the underlying mutation rates of these regions via the recombination-driven process of biased gene conversion (Dreszer et al., 2007; Galtier and Duret, 2007). This mutational process is distinguished from directional selection in that substituting nucleotides are more frequently G or C nucleotides than they are A or T nucleotides. In the biased gene conversion model, a high substitution rate results in frequent deleterious changes in *HAR1F*; these mutations can be compensated by subsequent nucleotide substitutions that tend to restore RNA secondary structure. In short, it appears more likely that human *HAR1F* has evolved rapidly not because these substitutions, when considered together, have been advantageous along the *Homo sapiens* terminal lineage. Rather, *HAR1F*'s rapid evolution results from frequently introduced deleterious changes that are often then "repaired" by subsequent mutations.

The evolutionary arguments, advanced above, suggest that a large, but as yet unknown, number of noncoding RNAs are not solely the product of transcriptional noise. Nevertheless, transcriptional noise has indeed been observed, in the form of the expression of one gene, coupled to the transcription of another lying within a radius of ~100 kb along the mouse genome (Ebisuya et al., 2008). This "rippling" of transcription occurs unevenly along adjacent sequence: it appears to be highly specific to loci with previous evidence of transcription, rather than inducing transcriptional events uniformly across chromosomal sequence. Rippling is induced irrespective of whether transcription is initiated at a coding or noncoding locus (Ebisuya et al., 2008). This effect may thus explain, at least in part, similarities in coexpression patterns between noncoding RNAs and mRNAs that are transcribed from adjacent genomic loci.

## Origins of lncRNAs

As most noncoding RNAs are subjected to a low degree of evolutionary constraint (that is, purifying selection) only a small minority exhibit sequence conservation between species as diverse as mammals and fish, for example. Nevertheless, a limited phyletic range of noncoding RNAs might also be explained if they emerge and decline rapidly within particular lineages (Hyashizaki, 2004). If so, then noncoding RNA genes would evolve very differently from protein-coding genes that arise, virtually in every case, by a process of partial or wholesale duplication and subsequent sequence divergence. In Figure 2, we outline various evolutionary scenarios for the emergence of functional noncoding RNAs.

The first such scenario is the metamorphosis of protein-coding into noncoding RNA gene sequence. The *Xist* gene encodes an lncRNA that is critical for the inactivation of the X chromosome in eutherian mammals. It has been recently established that several *Xist* exons and its promoter derive from the "debris" of a protein-coding gene *Lnx3* that had acquired frame-disrupting mutations early in the evolution of placental mammals (Duret et al., 2006; Elisaphenko et al., 2008). It is unknown whether this was a two-step metamorphosis, involving an initial degeneration of an erstwhile protein-coding gene, followed by these exons then being subsumed within an emerging *Xist* gene, or whether these steps occurred concurrently (Figure 2A). However, given that protein-coding and noncoding transcripts are frequently interwoven perhaps *Xist* emerged gradually, rather than stepwise, and coexisted within a locus that also was transcribed into protein-coding sequence.

Instead of deriving from pre-existing protein-coding sequence, can noncoding RNAs emerge from genomic sequence that was previously devoid of exonic sequence? Predicting noncoding RNA genes that arose from such sequence remains a considerable challenge. It is difficult to build a compelling case that sequence across a broad phyletic range has always lacked exons until after the emergence, in one restricted lineage, of a noncoding RNA locus. Nevertheless, one relatively-recent chromosomal rearrangement provides clues that a dog testis-derived noncoding

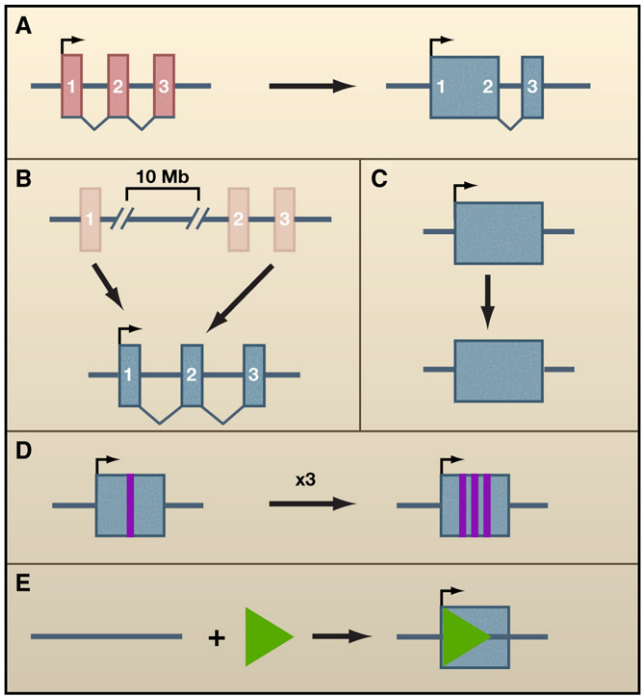


**Figure 2. Possible Origins of lncRNAs**

(A) A protein-coding gene (left, pink) acquires frame disruptions and is transformed into a functional noncoding RNA (right, blue) that incorporates some previous coding sequence. The *Xist* lncRNA originated by undergoing a metamorphosis from a previous protein-coding gene while incorporating transposable element sequence.

(B) Following a chromosome's rearrangement, two untranscribed and previously well-separated sequence regions are juxtaposed and give rise to a multi-exon noncoding RNA. A dog noncoding RNA (supported by ESTs BM537447, C0597044, and DN744681) appears to have arisen following such a lineage-specific change.

(C) Duplication of a noncoding gene by retrotransposition generates either a functional noncoding retrogene or a nonfunctional noncoding retropseudogene.

(D) Neighboring repeats within a noncoding RNA have their origins in two tandem duplication events.

(E) Insertion of a transposable element (green triangle) gives rise to a functional noncoding RNA.

RNA has arisen only recently in the canid lineage (Figure 2B). This noncoding RNA locus spans two regions that, in other eutherian mammals retaining ancestral chromosomal structures, are separated by tens of megabases. It appears reasonable to assume that it arose following the intrachromosomal rearrangement after the last common ancestor of canids and bovids ~80 million years ago.

Unlike protein-coding genes, noncoding RNA genes do not appear to form large homologous families. There is scant evidence for large numbers of noncoding RNAs whose sequences are sufficiently similar, outside of transposable elements, to allow common ancestry to be inferred. One study (Ravasi et al., 2006) proposed a family of mouse lncRNAs, typified by transcript AK014924. However, transposable elements span much of this transcript, and homologous sequences elsewhere in the genome are nonexonic and thus appear to be the relics of transposable element insertions. Rare examples of duplicated lncRNA loci include those for mouse nuclear enriched abundant transcript 2 (Neat2) (Hutchinson et al., 2007) and a mouse testis-derived lncRNA (AK019616) that are separately paralogous to nonexonic sequences elsewhere in the genome (Figure 2C). These may reflect past retrotransposition events and thus represent nonfunctional noncoding RNA pseudogenes. Local, tandem, duplications may also generate repeats, such as those observed in 5′ regions of *Kcnq1ot1* (see below) and *Xist* transcripts (Figure 2D).

LncRNAs may also emerge following insertions of transposable element sequences. BC1 (brain cytoplasmic RNA 1) and BC200 (brain cytoplasmic RNA 200-nucleotide) noncoding RNAs arose from separate episodes of transposable element exaptation, in the rodent and anthropoid lineages, respectively (Figure 2E). Despite their lack of a common origin, these noncoding RNAs appear to possess similar roles in translational regulation (reviewed in Cao et al., 2006) (see below). As transposable element sequences often coincide with transcription start sites they are also likely to contribute frequently to changes in gene transcript repertoires (Conley et al., 2008).

## RNA Localization, Processing, and Secondary Structure

LncRNAs are observed in a wide range of different tissues and, as also seen for mRNAs, most are expressed in the brain (Mercer et al., 2008). The spatiotemporal expression profiles of noncoding RNAs in the brain can be exquisite. The *Evf2* noncoding RNA (Figure 1) shows highly specific expression in the developing mouse brain, for example (Figure 3). The cellular localization of lncRNAs is also as varied as that seen for protein-coding genes, and a full range of subcellular patterns of expression have been described. By analyzing over 800 noncoding RNAs from the Allen mouse brain atlas, Mercer et al. (2008) observed transcripts in the nucleus, cell body or at one or more foci of adult cerebellar Purkinje cells. Some noncoding RNAs show unusual or unique localization patterns, thereby classifying entirely new subcellular compartments. For example, the subcellular localization of one nuclear, yet polyadenylated, lncRNA occurs exclusively in nuclear speckles, and thus has been given the name Gomafu, meaning "spotted pattern" in Japanese (Sone et al., 2007). Gomafu expression reveals a new domain that does not colocalize with any known nuclear compartment marker.

Most lncRNAs, as well as shorter (<200 nucleotide) noncoding RNAs, tend to be transcribed away from the 5′ or 3′ ends of genes. Nevertheless, transcription of lncRNAs has been found to be more concentrated near the promoters and initial exons and introns of genes (Kapranov et al., 2007a). A similar clustering of shorter noncoding RNAs is also observed (Kapranov et al., 2007a). Although these noncoding RNAs may be by-products of abortive initiation or transcriptional pausing, they may also reflect an origin of shorter noncoding RNAs by post-transcriptional processing of lncRNAs (Kapranov et al., 2007a). Nevertheless, as 95% of transcribed sequence lies either in lncRNAs or in short (<200 nucleotide) noncoding RNAs, but not in both, the widespread processing of lncRNAs has yet to be proved (Ponjavic and Ponting, 2007).

Functional RNA sequence thus cannot typically be identified from the processing of longer noncoding RNAs, whose functional

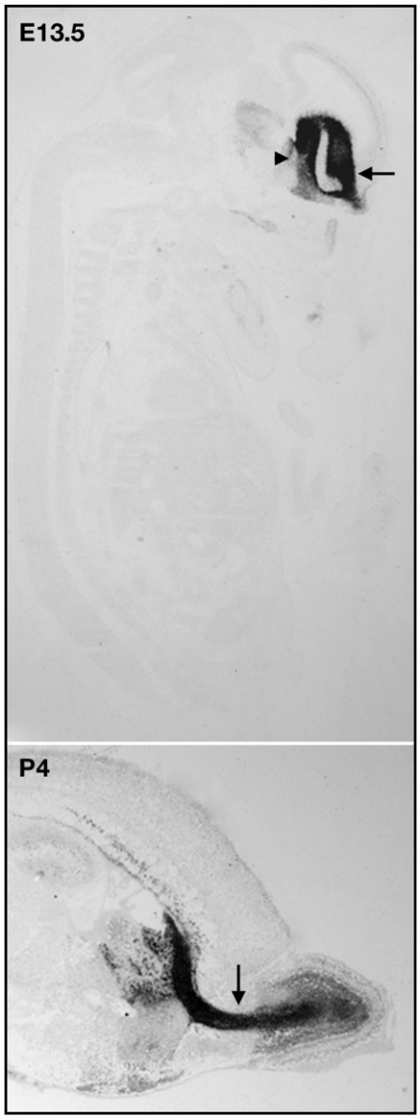


**Figure 3. Highly Specific Expression of *Evf2* in the Developing Mouse Brain**

(Top) The in situ hybridization image shows that at embryonic day E13.5 two major populations of migrating interneurons express *Evf2*. These are the latero-caudal migration from the basal telencephalon to the striatum (arrowhead) and the medio-rostral migration from the subpallial basal telencephalon to the olfactory bulb (arrow), which later becomes the rostral migratory stream as shown at P4 (bottom). Images courtesy of P. Oliver.

sequence is relatively diffuse, to shorter noncoding RNAs containing high densities of functional sequence. Instead, perhaps it can be pinpointed by *in silico* predictions of short stem-loop secondary structures. Indeed, we will highlight below examples of proteins that bind to these structured RNA regions. Approaches to this problem all focus on scoring apparent compensatory mutations for paired bases within conserved sequence. However, approaches that predict stem-loop structures with high precision and sensitivity have yet to be developed (Babak et al., 2007). In contrast to protein-coding genes, therefore, the numbers and repertoires of structured noncoding RNA elements encoded in animal genomes remain to be determined. What is known, however, is that such structures, as expected, are strongly depleted in coding sequence, but are enriched in intergenic, intronic and UTR sequence (Babak et al., 2007).

## Noncoding RNAs as Transcriptional Regulators

Eukaryotic gene regulation is traditionally explained by the direct interactions of proteins with other proteins or with DNA to modulate the expression of protein-coding genes. Regulatory networks also have an additional layer of complexity consisting of specific and dynamic interactions between RNA species and DNA or proteins (Goodrich and Kugel, 2006). Transcription of lncRNAs is now known to regulate the expression of genes in close genomic proximity (*cis*-acting regulation) and to target distant transcriptional activators or repressors (*trans*-acting) via a variety of mechanisms (for selected examples see Figure 4).

The genomic location of lncRNA transcription is an important facet of their regulatory potential. *Cis*-NATs, and other exotic RNA species, are frequently derived from within, or near to, protein-coding loci (Carninci et al., 2005; Conley et al., 2008; Katayama et al., 2005; Ponjavic et al., 2007). If promoters lie in close genomic proximity then transcriptional events initiated from them may be coregulated. Coregulation may occur via chromatin remodelling of chromosomal domains (Gribnau et al., 2000), but it also may occur because of ''collisions'' between transcriptional machineries that are processing along adjacent sequences (Osato et al., 2007). When elongation of one transcriptional event proceeds through a promoter sequence it can suppress this sequence's ability to initiate a second transcriptional event (Osato et al., 2007). An example of this ''transcriptional interference'' effect is the perturbed expression of the yeast *SER3* gene when the upstream lncRNA *SRG1* is actively transcribed or overexpressed (Martens et al., 2004) (Figure 4A). The 3′ end of *SRG1* contains regulatory elements within the *SER3* promoter, and premature termination of *SRG1* transcription prevents repression of *SER3* (Martens et al., 2004). Transcriptional interference, regulating expression levels in *cis*, may represent a widespread function for lncRNAs. This is because strong conservation in promoter sequences of lncRNAs (Carninci et al., 2005), and weaker conservation in the sequences of their transcripts (Ponjavic et al., 2007), are consistent with the act of transcription itself having a greater biological consequence than the transcript sequence.

Transcription of an lncRNA may promote the accessibility of protein-coding genes to RNA polymerases. In yeast, glucose starvation results in the induction of the *fbp1* gene and also of several lncRNAs transcribed 5′ upstream (Hirota et al., 2008). Transcription of these noncoding RNAs disrupts chromatin structure, which allows further passage of the transcriptional machinery through the promoter region, thereby promoting expression of *fbp1* (Figure 4B). Similar episodes of transcription have been observed to regulate expression within the human *β-globin* locus (Gribnau et al., 2000).

In other contexts, lncRNA sequences themselves convey functions through binding to DNA or protein. One such sequence, transcribed from a minor promoter upstream of the human dihydrofolate reductase *DHFR* gene, acts to repress

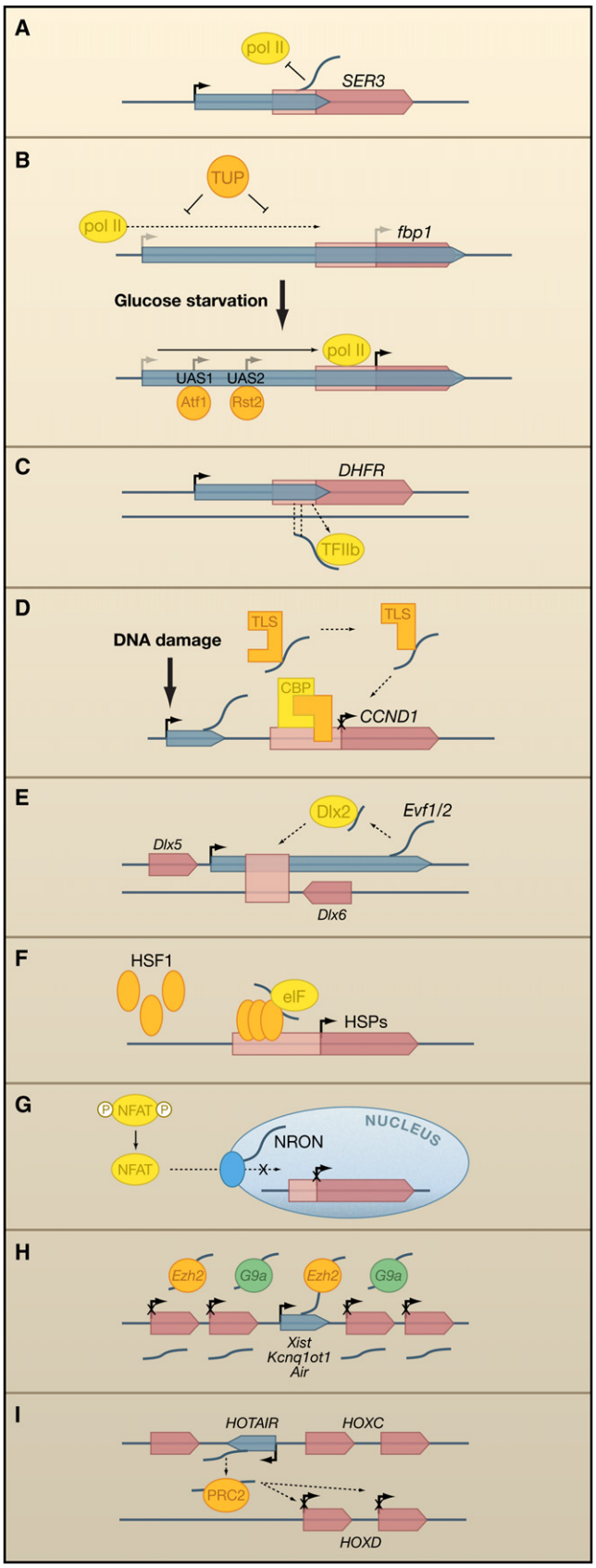


**Figure 4. Mechanisms of lncRNA Function in Transcriptional Regulation**

LncRNAs are blue and protein-coding genes are pink, with pale pink regions indicating promoter/enhancer elements.

(A) Transcriptional interference. Transcription of the lncRNA *SRG1* through the promoter of the adjacent *SER3* gene.

(B) Initiation of chromatin remodeling. RNA pol II processivity upstream of *fbp1* is normally repressed by Tup proteins, however, rare lncRNAs are transcribed. Upon glucose starvation, the Atf1 activator binds to the UAS1 element, facilitating chromatin remodeling by RNA pol II and the subsequent binding of Rst2 to a second UAS2 element. As further lncRNAs are transcribed, the chromatin structure around the *fbp1* initiation site is then accessible to the transcriptional machinery allowing induction of the gene to occur.

(C) Promoter inactivation by binding to basal transcription factors. Formation of a complex between an lncRNA and both the *DHFR* promoter and TFIIB prevents normal preinitiation of transcription.

(D) Activation of an accessory protein. In response to stress, lncRNAs upstream of *CCND1* form a complex with an RNA-binding protein TLS (translocated in liposarcoma) in which the inactive conformation of the protein is altered, facilitating repression of *CCND1* via chromatin-binding protein (CBP).

(E) Activation of transcription factors. The lncRNA *Evf2* cooperates with the *Dlx2* homeodomain protein to activate the *Dlx5/6* enhancer.

(F) Oligomerization of an activator protein. In response to heat shock, an lncRNA assists the trimerization of the HSF1 protein, which in turn forms a complex with the translation factor EIF to facilitate HSP expression.

(G) Transport of transcription factors. Dephosphorylated NFAT is prevented from translocating to the nucleus and activating its targets due to interactions between the lncRNA NRON and importin proteins.

(H) Epigenetic silencing of gene clusters by lncRNAs. The *Xist*, *Kcnq1ot1*, and *Air* RNAs establish a nuclear domain (or "coating") for gene silencing of genes in *cis*. The lncRNAs may directly or indirectly attract epigenetic modifiers such as histone methyltransferases (G9a or Ezh2) to bring about repressive epigenetic marks in the cluster.

(I) Epigenetic repression of genes by an intergenic lncRNA in *trans*. *HOTAIR* RNA, transcribed within the *HOXC* cluster, interacts with the Polycomb repressor complex 2 (PRC2) resulting in the methylation and silencing of several genes in the *HOXD* locus.

transcription from the major *DHFR* promoter (Martianov et al., 2007). Overexpression of the full-length noncoding RNA represses the major *DHFR* promoter in an RNA-dependent manner. The lncRNA binds to both the major *DHFR* promoter and the general transcription factor IIB, leading to dissociation of the preinitiation complex from the major promoter (Figure 4C). Triplex formation between the single-stranded lncRNA and double-stranded *DHFR* promoter is proposed to form a stable purine-purine-pyrimidine triplex structure. Such structures are predicted to be most concentrated around human promoters (Goni et al., 2004), but it remains to be seen whether binding to promoters forming triplex structures is a common lncRNA mechanism.

Noncoding RNAs also regulate transcription in *cis* indirectly, without binding to DNA. In a recent report, Wang et al. (2008) describe noncoding RNAs whose transcription, upstream of the cyclin D1 (*CCND1*) promoter, is induced by ionizing radiation. These noncoding RNAs bind to an RNA-binding protein TLS (translocated in liposarcoma), thereby permitting, via an allosteric effect, interactions with histone acetyltransferases; the resultant inhibition of these acetyltransferases causes decreased *CCND1* transcription (Wang et al., 2008) (Figure 4D). The authors

speculate that other pairs of noncoding RNA and RNA-binding proteins may yet be found to act in a similar manner.

Another lncRNA whose action depends on forming a ribonucleoprotein complex is *Evf2*. *Evf2* exons lie downstream of *Dlx5* and surround *Dlx6*, both of which are homeodomain genes involved in neuronal differentiation and migration, and limb patterning (Feng et al., 2006) (Figure 1). Single-stranded *Evf2* forms a complex with the product of a third homeodomain gene *Dlx2* whose sequence lies elsewhere in the mouse genome. This complex activates *Dlx5/6* enhancer activity by an as yet unknown mechanism, perhaps by binding directly to an enhancer sequence that is also transcribed within the second exon of *Evf2* (Figure 4E). The complex structure of this feedback loop perhaps provides an indication of the level of detail that will be required to fully appreciate how noncoding RNAs might function as regulators of protein-coding genes. These findings, together with the developmental expression profile of *Evf2* (Figure 3), show how temporal coordination of the expression of protein-coding genes and noncoding RNAs is likely to be critical to fundamental developmental processes.

A third protein-binding lncRNA is *HSR1* (heat shock RNA-1), a 604 nucleotide RNA that, together with eukaryotic translation-elongation factor 1A, stimulates trimerization of heat-shock factor 1 (HSF1) (Shamovsky et al., 2006) (Figure 4F). Trimeric, but not monomeric, HSF1 induces the transcription of heat-shock-induced genes by binding to their promoters. Not only is the formation of this complex induced by heat shock, but knockdown of *HSR1* by RNA interference causes cells to become thermosensitive. This suggests that *HSR1* contributes to a cellular thermosensing device, similar to those in bacteria (Shamovsky et al., 2006).

Other lncRNAs regulate transcription indirectly by controlling the subcellular localization of transcription factors. One such lncRNA affects the localization of the transcription factor NFAT (nuclear factor of activated T cells) perhaps by interactions with nuclear transport factors (Willingham et al., 2005) (Figure 4G). Knockdown of this lncRNA (named NRON, noncoding repressor of NFAT) results in increased NFAT in the nucleus and increased NFAT activity. Further investigation of the binding partners and functions of NRON should, perhaps, focus on its numerous predicted stem-loop secondary structures that are conserved between diverse vertebrates.

The 7SK noncoding RNA (331 nucleotides) represses transcript elongation by Pol II in complex with the elongation factor P-TEFb by downregulating its kinase activity (Nguyen et al., 2001; Yang et al., 2001). Repression also involves a second protein HEXIM1 (hexamethylene bisacetamide-induced protein-1). Dissociation of 7SK and HEXIM1 from P-TEFb promotes P-TEFb kinase activity, and thus enhances Pol II elongation. Stem-loop structures within 7SK are apparent for species as diverse as mammals, molluscs and annelids implying an ancient bilaterian ancestry for this abundant and important transcriptional regulator (Gruber et al., 2008).

Additional noncoding RNAs act as Pol II inhibitors. Mouse *B2* is a ~178 nucleotide RNA that was derived originally from short interspersed repeat elements (SINEs). Upon stress conditions such as heat shock, *B2* expression is dramatically enhanced leading to repression of transcription for genes such as actin and hexokinase II (Allen et al., 2004; Espinoza et al., 2004). Surprisingly, human *Alu* RNA also acts as a transcriptional repressor during heat shock, again by binding to Pol II, despite the substantial divergence between human *Alu* and mouse *B2* SINE sequences (Mariner et al., 2008). This has led to the proposal that many noncoding RNAs, including many from active transposable elements, control transcription of specific genes by targeting Pol II in *trans* (Mariner et al., 2008).

## LncRNAs in Epigenetics

LncRNAs have often been implicated in epigenetic gene regulation, and recent results now suggest a more unified model of how they might work. The initial links between imprinting and X chromosome inactivation, the two major epigenetic gene silencing phenomena in mammals, and lncRNAs, were made through the discovery of the *H19* and *Xist* RNAs, respectively. *H19* is an imprinted and maternally expressed lncRNA that is spliced, polyadenylated, and exported into the cytoplasm where it accumulates to very high levels (Bartolomei et al., 1991). The function of the *H19* RNA is still enigmatic, although it is likely that it plays some role in growth regulation. Recently, it was found that the *H19* RNA is host to an exonic microRNA, *miR-675*, which, as a result, is also imprinted and maternally expressed (Cai and Cullen, 2007). Characteristic features of both the *H19* transcript and of *miR-675* are conserved in all therian mammals (that is, for at least 150 million years), suggesting perhaps that there is purifying selection on both the lncRNA and its embedded miRNA (Smits et al., 2008). Although the (cytoplasmic) role of *H19* needs further investigation, it is interesting to ask how an RNA that is spliced and exported from the nucleus evades destruction by the nonsense mediated decay pathway, which normally surveys RNAs and ensures that only those with extensive open reading frames reach the cytoplasm intact for translation. Indeed, key components of the nuclear mRNA degradation and nonsense mediated decay pathways may regulate the levels of *H19* RNA during embryonic stem cell differentiation (Ciaudo et al., 2006).

The *Xist* RNA, which is crucial for X chromosome inactivation in *cis* in eutherian mammals (Brockdorff et al., 1991; Brown et al., 1991), shares some similarities with *H19* in that it is also spliced and polyadenylated, and its stability is regulated by the same nonsense mediated decay pathway as *H19*. However, this is probably where the similarities end, because *Xist* evades export into the cytoplasm and instead is associated as an RNA domain or compartment with the X chromosome that it inactivates (Clemson et al., 1996). This ''coating'' of the chromatin region that is silenced provided the first model of how lncRNAs might be involved in stable epigenetic gene silencing in *cis*. Indeed, it is now thought that the *Xist* RNA establishes a specialized nuclear compartment devoid of Pol II, into which most of the chromatin of the future inactive X chromosome becomes localized during inactivation (Chaumeil et al., 2006). A particular region of the *Xist* RNA is necessary for the formation of this specialized nuclear domain, whereas another region is required for translocation into the domain of X-linked genes and their consequent silencing (Chaumeil et al., 2006). Interestingly, ''coating'' is stably associated in *cis* with the inactive X chromosome even in metaphase, thus potentially providing one layer of an epigenetic memory for the inactive X to remain silent over many cell divisions (Jonkers et al., 2008).

Subsequent to locating into the *Xist* RNA domain, X-linked genes become silenced and lose activating histone modifications (such as acetylation) and gain repressive ones (particularly H3K27me3 and H2A K119ub1 induced by the Polycomb repressive complex PRC2 and PRC1, respectively). In addition, they become marked by the incorporation of the histone variant macroH2A and localized close to the nucleolus (reviewed in Wutz and Gribnau, 2007). Recently, it was shown that a segment of the *Xist* RNA (called RepA) is important for the targeting of Ezh2 (a component of PRC2) and hence H3K27me3 to the X chromosome (Zhao et al., 2008). Finally, many CpG islands of gene promoters on the inactive X become methylated; RNA coating together with histone modifications and DNA methylation together probably constitute the epigenetic memory by which the inactive state of the chromosome is mitotically heritable (Wutz and Gribnau, 2007). The *Xist* RNA is indeed required during a critical window in early development, overlapping with the commitment to differentiation of pluripotent stem cells, for the establishment of gene silencing and epigenetic memory (Wutz and Jaenisch, 2000). In somatic cells, however, the *Xist* gene can be deleted without significant loss of gene silencing, suggesting that once a memory based on DNA and perhaps histone methylation is established, the silencing domain originally created by the lncRNA is no longer required (Brown and Willard, 1994).

Imprinted gene clusters in the eutherian genome contain many other lncRNA genes aside from *H19*, most of which are imprinted themselves (Peters and Robson, 2008). A subclass of these noncoding RNAs is of primary importance in epigenetic regulation throughout the clusters, whereas others may have more local functions. Hence, the lncRNAs *Kcnq1ot1*, *Air*, and *Nespas* are all paternally expressed, and repressed on the maternal allele by promoter DNA methylation originating in the oocyte (Peters and Robson, 2008). There is very little splicing of these RNAs, which are therefore largely colinear with DNA, and possibly as a consequence, very long (Pandey et al., 2008; Redrup et al., 2009; Seidl et al., 2006). Perhaps during their evolution they have lost the capacity to be spliced in order to evade the nonsense mediated decay pathway. By deletion of the promoters of *Air* and *Kcnq1ot1* (Fitzpatrick et al., 2002; Wutz et al., 1997), or by truncation of the RNA through insertion of premature polyA signals (Mancini-Dinardo et al., 2006; Shin et al., 2008; Sleutels et al., 2002), it was found that the lncRNAs or the act of their transcription is necessary for epigenetic gene silencing of imprinted genes (on the paternal chromosomes) in the *Igf2r*/*Air* and *Kcnq1ot1* clusters, respectively. In addition to gene silencing, the lncRNAs (or their transcription) are required for the acquisition of repressive histone marks (H3K27me3, H3K9me2) and DNA methylation (in some genes) throughout the imprinting clusters, which are up to 800kb in size (Lewis et al., 2004; Nagano et al., 2008; Pandey et al., 2008; Regha et al., 2007; Umlauf et al., 2004). Further striking similarities with the process of X chromosome inactivation include the fact that both *Xist* and *Kcnq1ot1* (Lewis et al., 2006) are paternally expressed from the two-cell stage of embryonic development (this initial phase of X inactivation occurs on the paternal X, which hence is imprinted). In addition, *Kcnq1ot1* apparently induces epigenetic gene silencing during a critical window of opportunity in early development (Green et al., 2007), just as *Xist* does (Wutz and Jaenisch, 2000).

A number of models have been suggested to explain how lncRNAs in imprinting clusters might regulate epigenetic gene silencing in *cis*. Imprinting clusters also contain small RNAs, some of which are processed from lncRNAs (Peters and Robson, 2008). This may suggest the possibility that the RNA interference pathway is involved in targeting gene silencing by local formation of heterochromatin. This idea has been tested genetically using a conditional knockout of Dicer. Imprinting in the *Kcnq1ot1* cluster was unaffected by the loss of Dicer (Redrup et al., 2009), and so was X inactivation in one study (Nesterova et al., 2008), whereas in another study some aspects of X inactivation were apparently affected by the small RNA pathway (Ogawa et al., 2008). It is also possible that the act of transcription through the gene regions of the imprinted lncRNA is important. Indeed, a specific proposal has been made that transcription through an enhancer element for *Igf2r* by the *Air* RNA might be critical for the silencing of this gene on the paternal chromosome (Pauler et al., 2007). On the other hand, the many similarities with X inactivation raise the question of whether "coating" by autosomal lncRNAs might be involved in epigenetic gene silencing.

A number of recent studies suggest that this is indeed the case (Pandey et al., 2008; Redrup et al., 2009; Terranova et al., 2008; Nagano et al., 2008). For both *Air* and *Kcnq1ot1*, it was found that the RNA appears to establish a nuclear domain, which is closely associated with the genes that are inactivated in *cis*, whereas genes outside the cluster, which are not regulated by the lncRNA, are found outside the nuclear RNA domain (Figures 4H and 5). The *Kcnq1ot1* RNA domain excludes Pol II and is enriched with Prc1 and Prc2 components (Terranova et al., 2008). The *Kcnq1ot1* RNA also binds to the histone methyltransferase G9a (Pandey et al., 2008); the *Air* RNA also binds to G9a and appears to target this enzyme to a silenced gene in the cluster (Nagano et al., 2008). In combination, therefore, these new findings show that critical mechanistic aspects of epigenetic silencing in mammalian gene clusters are probably shared between X inactivation and autosomal imprinting. An intriguing possibility is that mechanisms based on nuclear RNA domains not only operate in epigenetic silencing in *cis*, but also perhaps in *trans*, where for example the *HOTAIR* lncRNA is involved in targeting gene silencing and histone marks to particular genes in the *Hox* cluster (Figure 4I) (Rinn et al., 2007). The idea that lncRNAs themselves target epigenetic modifiers (such as Ezh2 or G9a) to the regions they inactivate is enticing, but this possibility is difficult to disentangle from that of the RNAs establishing nuclear compartments for silencing, which inevitably brings them close to repressive chromatin and the epigenetic modifiers that regulate repressive chromatin. Another important link between lncRNAs and DNA methylation in imprinted genes in germ cells has recently been discovered; it seems that transcription by lncRNAs through imprinting control elements attracts DNA methylation to these elements (Chotalia et al., 2009). Overall, an important role of lncRNAs might be the targeting of epigenetic gene silencing (or activation) across cell divisions, thus contributing to the maintenance of cell identity in multicellular organisms.

It is worth noting that in contrast to the conservation of the cytoplasmic lncRNA H19 in therian mammals, lncRNAs that

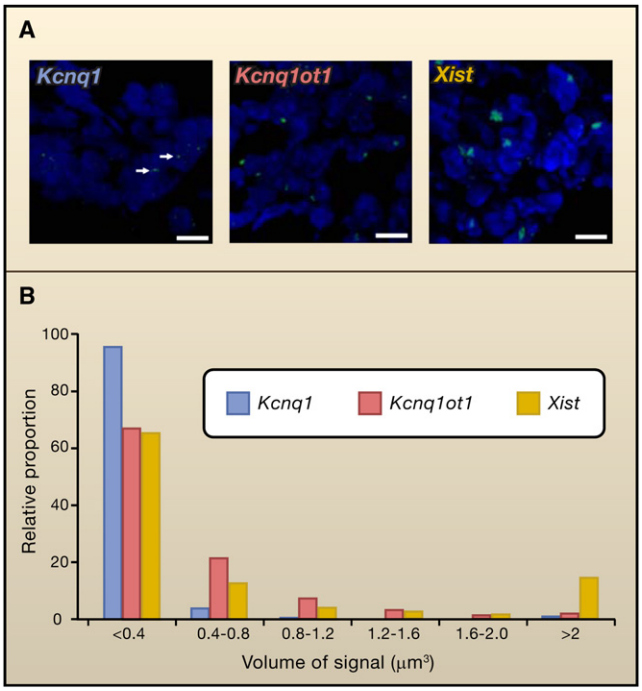


**Figure 5. Long-Distance Silencing by *Kcnq1ot1***

The paternally expressed long noncoding RNA (lncRNA) *Kcnq1ot1* regulates epigenetic gene silencing in an imprinted gene cluster in *cis* over a distance of 780 kb. Gene silencing by the *Kcnq1ot1* RNA involves repressive histone modifications including H3K9me2 and H3K27me3, which are brought about by G9a and Ezh2 histone methyltransferases. *Kcnq1ot1* is transcribed by RNA polymerase II, is unspliced, relatively stable, and localized in the nucleus. RNA/DNA fluorescence in situ hybridization (FISH) experiments show that the *Kcnq1ot1* RNA establishes a nuclear domain within which the genes that are epigenetically inactivated in *cis* are frequently found, whereas nearby genes that are not regulated by *Kcnq1ot1* are localized outside of the domain.

(A) RNA FISH signals of *Kcnq1* (protein coding), *Kcnq1ot1* (lncRNA in imprinting cluster), and *Xist* (lncRNA critical for X chromosome inactivation) in mouse placental sections. Note the small signal of *Kcnq1* (arrows), which is typical of the primary transcript of a protein-coding gene, the large coating signal of *Xist*, and the intermediate size signal of *Kcnq1ot1*.

(B) Volumes of RNA FISH signals were measured following confocal microscopy and the frequencies of different volume classes were plotted. Volume distributions of the three transcripts are different, with *Kcnq1* occupying the smallest volume, *Kcnq1ot1* intermediate, and *Xist* the largest volume. Hence, autosomal lncRNAs may be able to establish nuclear domains, which might create a repressive environment for epigenetic silencing of adjacent genes. LncRNAs in imprinting clusters and the *Xist* RNA on the inactive X chromosome may thus regulate epigenetic gene silencing by similar mechanisms. (Images in A, courtesy of L. Redrup; B, modified from Redrup et al., 2009). The scale bar represents 10 μm.

are involved in epigenetic gene inactivation in imprinting clusters or the X chromosome seem to have arisen more recently and are only present in eutherians (Peters and Robson, 2008; Smits et al., 2008). Hence, *Xist* is a protein-coding gene in marsupials that is unlikely to be involved in X inactivation (Duret et al., 2006), and neither Air nor Kcnq1ot1 have been identified in marsupials (Killian et al., 2000). Therefore lncRNAs involved in *cis* inactivation of larger genomic regions by epigenetic mechanisms may have particularly arisen in response to selective pressures that intensified the evolution of genomic imprinting in eutherian mammals (Smits et al., 2008).

## LncRNAs and Disease

LncRNAs have been linked to disease. In most instances, however, evidence has relied on differences in transcript expression levels between disease- and nondisease-associated states (reviewed in Szymanski et al., 2005). For example, increased expression of either *BC200* or an antisense transcript of the β-secretase-1 (*BACE1*) gene has been implicated in the progression of Alzheimer's disease (Faghihi et al., 2008; Mus et al., 2007). Nevertheless, it is important to stress that altered levels of lncRNA expression need not necessarily be relevant to disease etiology.

More definitive evidence would require results from genome-wide approaches. Nevertheless, whole genome mutagenesis experiments have yet to locate causative lesions in noncoding sequence. Moreover, genome-wide association studies have only rarely pinpointed high-risk alleles within noncoding RNA genes (e.g., Broadbent et al., 2008) and these may yet be found to reflect changes in long-range control of protein-coding gene expression, rather than altered noncoding RNA function. Evidence for disease-association might also accrue from directed experiments in model organisms, such as mouse. One such example concerns *BC1*, a rodent-specific noncoding RNA. *BC1*-deficient mice exhibit reduced exploration and increased anxiety and they show increased mortality, relative to wild-type, when kept within a semi-natural environment (Lewejohann et al., 2004). The neural phenotype of *BC1* is consistent with its known interaction with both FMRP, the product of the fragile X mental retardation gene (*FMR1*), and with mRNAs regulated by FMRP (Zalfa et al., 2003). It is the 5′ stem-loop structure of *BC1* that is required for the FMRP interaction, forming a stable FMRP-*BC1*-mRNA complex that represses translation (Zalfa et al., 2003).

Demonstration that mutations in noncoding RNAs are associated with disease has been provided by a mouse model of human spinocerebellar ataxia type 8 (SCA8) (Moseley et al., 2006). Patients show a trinucleotide (CUG) expansion in a noncoding RNA termed ataxin 8 opposite strand (*ATXN8OS*), an antisense transcript to the *KLHL1* gene (Koob et al., 1999). Transgenic mice containing this pathogenic CTG expansion in their DNA show a progressive neurological phenotype, with mice containing the highest CTG copy number, as with human patients, being worst affected. From these studies it appears that SCA8 proceeds via gain-of-function mechanisms involving both a noncoding RNA (*ATXN8OS*) and a polyglutamine protein mRNA transcribed from the opposite strand (Moseley et al., 2006). Repeat expansions appear also to be pathogenic at the RNA level in myotonic dystrophies (reviewed in Ranum and Cooper, 2006; see Review by T. A. Cooper, L. Wan, and G. Dreyfuss on page 777 of this issue). Tri- or tetra-nucleotide repeat expansions, when transcribed into noncoding RNA, are proposed to form hairpin structures that sequester splicing regulators away from their normal pre-mRNA targets. Symptoms of myotonic dystrophies in multiple organ systems are suggested to arise from the systemic misregulation of pre-mRNA splicing.

## Future Perspectives

On the one hand, the low degree of sequence constraint and the current absence of associations to disease might be argued to imply that lncRNAs contribute little to a species' biology. On the other hand, because large numbers of lncRNAs, when considered together, exhibit signatures of evolutionary constraint it is apparent that past mutations in functional sequence have been deleterious and have thus been preferentially purged from populations. Taken together these observations imply that each lncRNA contributes, albeit only slightly, to an organism's fitness, yet large numbers of lncRNAs contribute substantially when they are considered in aggregate. If so, then only rarely would obvious phenotypes arise when the transcription of a single lncRNA is disrupted, and thus only rarely will the mechanisms of individual lncRNAs be determined from simple experiments. Instead, elucidating the more subtle *cis*- and *trans*- regulatory roles of lncRNAs may require technological developments in both in vivo imaging of RNAs at high-resolution, and high-throughput identification of protein, RNA and DNA-binding partners of lncRNAs. Moreover, only when such experiments have been performed across a range of diverse species will the rates of functional lncRNA gain and loss be apparent. Lineage-specific lncRNAs may contribute substantially to innovative biological traits, or they may do so only rarely, leaving such traits to be derived from changes in DNA regulatory elements or in mRNAs. It is to be hoped that the recent burst of interest in lncRNAs will foreshadow resolution of all of these issues from wide-ranging experimental discoveries about the evolution and functional mechanisms of lncRNAs.


### ACKNOWLEDGMENTS

W.R. thanks L. Redrup and M. Branco for discussions and advice and BBSRC, MRC, and EU NoE The Epigenome for funding. C.P.P. and P.L.O. are funded by the MRC.


### REFERENCES


Allen, T.A., Von Kaenel, S., Goodrich, J.A., and Kugel, J.F. (2004). The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. Nat. Struct. Mol. Biol. *11*, 816–821.

Babak, T., Blencowe, B.J., and Hughes, T.R. (2007). Considerations in the identification of functional RNA structural elements in genomic alignments. BMC Bioinformatics *8*, 33.

Bartolomei, M.S., Zemel, S., and Tilghman, S.M. (1991). Parental imprinting of the mouse H19 gene. Nature *351*, 153–155.

Bertone, P., Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., Weissman, S., et al. (2004). Global identification of human transcribed sequences with genome tiling arrays. Science *306*, 2242–2246.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799–816.

Broadbent, H.M., Peden, J.F., Lorkowski, S., Goel, A., Ongen, H., Green, F., Clarke, R., Collins, R., Franzosi, M.G., Tognoni, G., et al. (2008). Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. Hum. Mol. Genet. *17*, 806–814.

Brockdorff, N., Ashworth, A., Kay, G.F., Cooper, P., Smith, S., McCabe, V.M., Norris, D.P., Penny, G.D., Patel, D., and Rastan, S. (1991). Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. Nature *351*, 329–331.

Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature *349*, 38–44.

Brown, C.J., and Willard, H.F. (1994). The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. Nature *368*, 154–156.

Cai, X., and Cullen, B.R. (2007). The imprinted H19 noncoding RNA is a primary microRNA precursor. RNA *13*, 313–316.

Cao, X., Yeo, G., Muotri, A.R., Kuwabara, T., and Gage, F.H. (2006). Noncoding RNAs in the mammalian central nervous system. Annu. Rev. Neurosci. *29*, 77–103.

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

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C., et al. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat. Genet. *38*, 626–635.

Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., et al. (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell *116*, 499–509.

Chaumeil, J., Le Baccon, P., Wutz, A., and 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, J., Sun, M., Hurst, L.D., Carmichael, G.G., and Rowley, J.D. (2005). Genome-wide analysis of coordinate expression and evolution of human cis-encoded sense-antisense transcripts. Trends Genet. *21*, 326–329.

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.

Chotalia, M., Smallwood, S.A., Ruf, N., Dawson, C., Lucifero, D., Frontera, M., James, K., Dean, W., and Kelsey, G. (2009). Transcription is required for establishment of germline methylation marks at imprinted genes. Genes Dev. *23*, 105–117.

Ciaudo, C., Bourdet, A., Cohen-Tannoudji, M., Dietz, H.C., Rougeulle, C., and Avner, P. (2006). Nuclear mRNA degradation pathway(s) are implicated in Xist regulation and X chromosome inactivation. PLoS Genet. *2*, e94. 10.1371/journal.pgen.0020094.

Clemson, C.M., McNeil, J.A., Willard, H.F., and Lawrence, J.B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J. Cell Biol. *132*, 259–275.

Conley, A.B., Miller, W.J., and Jordan, I.K. (2008). Human cis natural antisense transcripts initiated by transposable elements. Trends Genet. *24*, 53–56.

Dinger, M.E., Pang, K.C., Mercer, T.R., Crowe, M.L., Grimmond, S.M., and Mattick, J.S. (2009). NRED: a database of long noncoding RNA expression. Nucleic Acids Res. *37*, D122–D126.

Dinger, M.E., Pang, K.C., Mercer, T.R., and Mattick, J.S. (2008). Differentiating protein-coding and noncoding RNA: challenges and ambiguities. PLoS Comput. Biol. *4*, e1000176. 10.1371/journal.pcbi.1000176.

Dreszer, T.R., Wall, G.D., Haussler, D., and Pollard, K.S. (2007). Biased clustered substitutions in the human genome: the footprints of male-driven biased gene conversion. Genome Res. *17*, 1420–1430.

Duret, L., Chureau, C., Samain, S., Weissenbach, J., and Avner, P. (2006). The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. Science *312*, 1653–1655.

Dutrow, N., Nix, D.A., Holt, D., Milash, B., Dalley, B., Westbroek, E., Parnell, T.J., and Cairns, B.R. (2008). Dynamic transcriptome of Schizosaccharomyces pombe shown by RNA-DNA hybrid mapping. Nat. Genet. *40*, 977–986.

Ebisuya, M., Yamamoto, T., Nakajima, M., and Nishida, E. (2008). Ripples from neighbouring transcription. Nat. Cell Biol. *10*, 1106–1113.

Elisaphenko, E.A., Kolesnikov, N.N., Shevchenko, A.I., Rogozin, I.B., Nesterova, T.B., Brockdorff, N., and Zakian, S.M. (2008). A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements. PLoS ONE *3*, e2521. 10.1371/journal.pone.0002521.

Espinoza, C.A., Allen, T.A., Hieb, A.R., Kugel, J.F., and Goodrich, J.A. (2004). B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. Nat. Struct. Mol. Biol. *11*, 822–829.

Faghihi, M.A., Modarresi, F., Khalil, A.M., Wood, D.E., Sahagan, B.G., Morgan, T.E., Finch, C.E., St Laurent, G., 3rd, Kenny, P.J., and Wahlestedt, C. (2008). Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. Nat. Med. *14*, 723–730.

Feng, J., Bi, C., Clark, B.S., Mady, R., Shah, P., and Kohtz, J.D. (2006). The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. Genes Dev. *20*, 1470–1484.

Fitzpatrick, G.V., Soloway, P.D., and Higgins, M.J. (2002). Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. Nat. Genet. *32*, 426–431.

Galtier, N., and Duret, L. (2007). Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. Trends Genet. *23*, 273–277.

Gerstein, M.B., Bruce, C., Rozowsky, J.S., Zheng, D., Du, J., Korbel, J.O., Emanuelsson, O., Zhang, Z.D., Weissman, S., and Snyder, M. (2007). What is a gene, post-ENCODE? History and updated definition. Genome Res. *17*, 669–681.

Gingeras, T.R. (2007). Origin of phenotypes: genes and transcripts. Genome Res. *17*, 682–690.

Goni, J.R., de la Cruz, X., and Orozco, M. (2004). Triplex-forming oligonucleotide target sequences in the human genome. Nucleic Acids Res. *32*, 354–360.

Goodrich, J.A., and Kugel, J.F. (2006). Non-coding-RNA regulators of RNA polymerase II transcription. Nat. Rev. Mol. Cell Biol. *7*, 612–616.

Green, K., Lewis, A., Dawson, C., Dean, W., Reinhart, B., Chaillet, J.R., and Reik, W. (2007). A developmental window of opportunity for imprinted gene silencing mediated by DNA methylation and the Kcnq1ot1 noncoding RNA. Mamm. Genome *18*, 32–42.

Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R., and Fraser, P. (2000). Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. Mol. Cell *5*, 377–386.

Gruber, A.R., Koper-Emde, D., Marz, M., Tafer, H., Bernhart, S., Obernosterer, G., Mosig, A., Hofacker, I.L., Stadler, P.F., and Benecke, B.J. (2008). Invertebrate 7SK snRNAs. J. Mol. Evol. *66*, 107–115.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature, in press. Published online February 1, 2009. 10.1038/nature07672.

Hirota, K., Miyoshi, T., Kugou, K., Hoffman, C.S., Shibata, T., and Ohta, K. (2008). Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. Nature *456*, 130–134.

Hutchinson, J.N., Ensminger, A.W., Clemson, C.M., Lynch, C.R., Lawrence, J.B., and Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. BMC Genomics *8*, 39.

Hyashizaki, Y. (2004). Mouse transcriptome: Neutral evolution of 'non-coding' complementary DNAs. Nature *431*, 757.

Jonkers, I., Monkhorst, K., Rentmeester, E., Grootegoed, J.A., Grosveld, F., and Gribnau, J. (2008). Xist RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle. Mol. Cell. Biol. *28*, 5583–5594.

Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., Cawley, S., Drenkow, J., Piccolboni, A., Bekiranov, S., Helt, G., et al. (2004). Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. Genome Res. *14*, 331–342.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermuller, J., Hofacker, I.L., et al. (2007a). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science *316*, 1484–1488.

Kapranov, P., Willingham, A.T., and Gingeras, T.R. (2007b). Genome-wide transcription and the implications for genomic organization. Nat. Rev. Genet. *8*, 413–423.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., et al. (2005). Antisense transcription in the mammalian transcriptome. Science *309*, 1564–1566.

Khaitovich, P., Kelso, J., Franz, H., Visagie, J., Giger, T., Joerchel, S., Petzold, E., Green, R.E., Lachmann, M., and Paabo, S. (2006). Functionality of intergenic transcription: an evolutionary comparison. PLoS Genet. *2*, e171. 10.1371/journal.pgen.0020171.

Killian, J.K., Byrd, J.C., Jirtle, J.V., Munday, B.L., Stoskopf, M.K., MacDonald, R.G., and Jirtle, R.L. (2000). M6P/IGF2R imprinting evolution in mammals. Mol. Cell *5*, 707–716.

Koob, M.D., Moseley, M.L., Schut, L.J., Benzow, K.A., Bird, T.D., Day, J.W., and Ranum, L.P. (1999). An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). Nat. Genet. *21*, 379–384.

Lapidot, M., and Pilpel, Y. (2006). Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. EMBO Rep. *7*, 1216–1222.

Lewejohann, L., Skryabin, B.V., Sachser, N., Prehn, C., Heiduschka, P., Thanos, S., Jordan, U., Dell'Omo, G., Vyssotski, A.L., Pleskacheva, M.G., et al. (2004). Role of a neuronal small non-messenger RNA: behavioural alterations in BC1 RNA-deleted mice. Behav. Brain Res. *154*, 273–289.

Lewis, A., Green, K., Dawson, C., Redrup, L., Huynh, K.D., Lee, J.T., Hemberger, M., and Reik, W. (2006). Epigenetic dynamics of the Kcnq1 imprinted domain in the early embryo. Development *133*, 4203–4210.

Lewis, A., Mitsuya, K., Umlauf, D., Smith, P., Dean, W., Walter, J., Higgins, M., Feil, R., and Reik, W. (2004). Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. Nat. Genet. *36*, 1291–1295.

Louro, R., El-Jundi, T., Nakaya, H.I., Reis, E.M., and Verjovski-Almeida, S. (2008). Conserved tissue expression signatures of intronic noncoding RNAs transcribed from human and mouse loci. Genomics *92*, 18–25.

Mancini-Dinardo, D., Steele, S.J., Levorse, J.M., Ingram, R.S., and Tilghman, S.M. (2006). Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. Genes Dev. *20*, 1268–1282.

Mariner, P.D., Walters, R.D., Espinoza, C.A., Drullinger, L.F., Wagner, S.D., Kugel, J.F., and Goodrich, J.A. (2008). Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. Mol. Cell *29*, 499–509.

Martens, J.A., Laprade, L., and Winston, F. (2004). Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. Nature *429*, 571–574.

Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., and Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature *445*, 666–670.

Mattick, J.S., and Makunin, I.V. (2006). Non-coding RNA. Hum. Mol. Genet. *15*, R17–R29.

Mercer, T.R., Dinger, M.E., Sunkin, S.M., Mehler, M.F., and Mattick, J.S. (2008). Specific expression of long noncoding RNAs in the mouse brain. Proc. Natl. Acad. Sci. USA *105*, 716–721.

Moseley, M.L., Zu, T., Ikeda, Y., Gao, W., Mosemiller, A.K., Daughters, R.S., Chen, G., Weatherspoon, M.R., Clark, H.B., Ebner, T.J., et al. (2006). Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. Nat. Genet. *38*, 758–769.

Mus, E., Hof, P.R., and Tiedge, H. (2007). Dendritic BC200 RNA in aging and in Alzheimer's disease. Proc. Natl. Acad. Sci. USA *104*, 10679–10684.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. Science *320*, 1344–1349.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science *322*, 1717–1720.

Nesterova, T.B., Popova, B.C., Cobb, B.S., Norton, S., Senner, C.E., Tang, Y.A., Spruce, T., Rodriguez, T.A., Sado, T., Merkenschlager, M., and Brockdorff, N. (2008). Dicer regulates Xist promoter methylation in ES cells indirectly through transcriptional control of Dnmt3a. Epigenetics Chromatin *1*, 2.

Nguyen, V.T., Kiss, T., Michels, A.A., and Bensaude, O. (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature *414*, 322–325.

Ogawa, Y., Sun, B.K., and Lee, J.T. (2008). Intersection of the RNA interference and X-inactivation pathways. Science *320*, 1336–1341.

Osato, N., Suzuki, Y., Ikeo, K., and Gojobori, T. (2007). Transcriptional interferences in cis natural antisense transcripts of humans and mice. Genetics *176*, 1299–1306.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol. Cell *32*, 232–246.

Pang, K.C., Frith, M.C., and Mattick, J.S. (2006). Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. Trends Genet. *22*, 1–5.

Pauler, F.M., Koerner, M.V., and Barlow, D.P. (2007). Silencing by imprinted noncoding RNAs: is transcription the answer? Trends Genet. *23*, 284–292.

Peters, J., and Robson, J.E. (2008). Imprinted noncoding RNAs. Mamm. Genome. *19*, 493–502.

Pollard, K.S., Salama, S.R., Lambert, N., Lambot, M.A., Coppens, S., Pedersen, J.S., Katzman, S., King, B., Onodera, C., Siepel, A., et al. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. Nature *443*, 167–172.

Ponjavic, J., and Ponting, C.P. (2007). The long and the short of RNA maps. Bioessays *29*, 1077–1080.

Ponjavic, J., Ponting, C.P., and Lunter, G. (2007). Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. Genome Res. *17*, 556–565.

Ponting, C.P. (2008). The functional repertoires of metazoan genomes. Nat. Rev. Genet. *9*, 689–698.

Ranum, L.P., and Cooper, T.A. (2006). RNA-mediated neuromuscular disorders. Annu. Rev. Neurosci. *29*, 259–277.

Ravasi, T., Suzuki, H., Pang, K.C., Katayama, S., Furuno, M., Okunishi, R., Fukuda, S., Ru, K., Frith, M.C., Gongora, M.M., et al. (2006). Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. Genome Res. *16*, 11–19.

Redrup, L., Branco, M.R., Perdeaux, E.R., Krueger, C., Lewis, A., Santos, F., Nagano, T., Cobb, B.S., Fraser, P., and Reik, W. (2009). The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing. Development *136*, 525–530. Published online January 14, 2009. 10.1242/dev.031328.

Regha, K., Sloane, M.A., Huang, R., Pauler, F.M., Warczok, K.E., Melikant, B., Radolf, M., Martens, J.H., Schotta, G., Jenuwein, T., and Barlow, D.P. (2007). Active and repressive chromatin are interspersed without spreading in an imprinted gene cluster in the Mamm. Genome. Mol. Cell *27*, 353–366.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., and Chang, H.Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311–1323.

Seidl, C.I., Stricker, S.H., and Barlow, D.P. (2006). The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. EMBO J. *25*, 3565–3575.

Shamovsky, I., Ivannikov, M., Kandel, E.S., Gershon, D., and Nudler, E. (2006). RNA-mediated response to heat shock in mammalian cells. Nature *440*, 556–560.

Shin, J.Y., Fitzpatrick, G.V., and Higgins, M.J. (2008). Two distinct mechanisms of silencing by the KvDMR1 imprinting control region. EMBO J. *27*, 168–178.

Sleutels, F., Zwart, R., and Barlow, D.P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature *415*, 810–813.

Smits, G., Mungall, A.J., Griffiths-Jones, S., Smith, P., Beury, D., Matthews, L., Rogers, J., Pask, A.J., Shaw, G., VandeBerg, J.L., et al. (2008). Conservation of the H19 noncoding RNA and H19–IGF2 imprinting mechanism in therians. Nat. Genet. *40*, 971–976.

Sone, M., Hayashi, T., Tarui, H., Agata, K., Takeichi, M., and Nakagawa, S. (2007). The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. J. Cell Sci. *120*, 2498–2506.

Szymanski, M., Barciszewska, M.Z., Erdmann, V.A., and Barciszewski, J. (2005). A new frontier for molecular medicine: noncoding RNAs. Biochim. Biophys. Acta *1756*, 65–75.

Terranova, R., Yokobayashi, S., Stadler, M.B., Otte, A.P., van Lohuizen, M., Orkin, S.H., and Peters, A.H. (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. Dev. Cell *15*, 668–679.

Tian, B., Hu, J., Zhang, H., and Lutz, C.S. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res. *33*, 201–212.

Umlauf, D., Goto, Y., Cao, R., Cerqueira, F., Wagschal, A., Zhang, Y., and Feil, R. (2004). Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. Nat. Genet. *36*, 1296–1300.

Wang, J., Zhang, J., Zheng, H., Li, J., Liu, D., Li, H., Samudrala, R., Yu, J., and Wong, G.K. (2004). Mouse transcriptome: neutral evolution of 'non-coding' complementary DNAs. Nature *431*, 757.

Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., Tempst, P., Rosenfeld, M.G., Glass, C.K., and Kurokawa, R. (2008). Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature *454*, 126–130.

Wilhelm, B.T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C.J., Rogers, J., and Bahler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature *453*, 1239–1243.

Willingham, A.T., Orth, A.P., Batalov, S., Peters, E.C., Wen, B.G., Aza-Blanc, P., Hogenesch, J.B., and Schultz, P.G. (2005). A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. Science *309*, 1570–1573.

Wutz, A., and Gribnau, J. (2007). X inactivation Xplained. Curr. Opin. Genet. Dev. *17*, 387–393.

Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol. Cell *5*, 695–705.

Wutz, A., Smrzka, O.W., Schweifer, N., Schellander, K., Wagner, E.F., and Barlow, D.P. (1997). Imprinted expression of the Igf2r gene depends on an intronic CpG island. Nature *389*, 745–749.

Yang, Z., Zhu, Q., Luo, K., and Zhou, Q. (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. Nature *414*, 317–322.

Zalfa, F., Giorgi, M., Primerano, B., Moro, A., Di Penta, A., Reis, S., Oostra, B., and Bagni, C. (2003). The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at synapses. Cell *112*, 317–327.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science *322*, 750–756.