nature structural & molecular biology

Structure and function of long noncoding RNAs in epigenetic regulation

Tim R Mercer[1,2] & John S Mattick[3]

Genomes of complex organisms encode an abundance and diversity of long noncoding RNAs (lncRNAs) that are expressed throughout the cell and fulfill a wide variety of regulatory roles at almost every stage of gene expression. These roles, which encompass sensory, guiding, scaffolding and allosteric capacities, derive from folded modular domains in lncRNAs. In this diverse functional repertoire, we focus on the well-characterized ability for lncRNAs to function as epigenetic modulators. Many lncRNAs bind to chromatin-modifying proteins and recruit their catalytic activity to specific sites in the genome, thereby modulating chromatin states and impacting gene expression. Considering this regulatory potential in combination with the abundance of lncRNAs suggests that lncRNAs may be part of a broad epigenetic regulatory network.

Global transcriptional analyses have revealed that the vast majority of the human genome is dynamically and differentially transcribed to produce a range and complexity of lncRNAs[1]. These observations are supplemented by an increasing number of targeted functional studies showing that lncRNAs fulfill regulatory roles at almost every stage of gene expression, from targeting epigenetic modifications in the nucleus to modulating mRNA stability and translation in the cytoplasm (**Box 1**). As a result of these studies, lncRNAs are being increasingly accepted as a major new gene class.

The abundance of lncRNAs, in conjunction with these emerging functional insights, has fuelled considerable excitement and enthusiasm for research into lncRNA biology. As the assumption of nonfunctionality has been discarded, researchers are starting to appreciate the potential importance of lncRNAs in the ontogeny of complex organisms[2]. Research into lncRNAs has progressed so rapidly that it is becoming increasingly difficult to comprehensively catalog the functionally validated cases[3]. In this wide and diverse emerging functional landscape, we consider features of lncRNA structure, expression, evolution and function with respect to one of the currently best characterized role of lncRNAs—the regulation of epigenetic dynamics.

Defining lncRNAs

The majority of characterized lncRNAs are generated by the same transcriptional machinery as are other mRNAs, as evidenced by RNA polymerase II occupancy and histone modifications associated with transcription initiation and elongation[4]. These lncRNAs have a 5′ terminal methylguanosine cap and are often spliced and polyadenylated. Alternate pathways also contribute to the generation of known lncRNAs, which include a poorly characterized contingent of non-polyadenylated lncRNAs likely expressed from RNA polymerase III promoters[5,6] and lncRNAs that are excised during splicing and small nucleolar RNA production[7].

Because lncRNAs have a biogenesis pathways in common with mRNA and other noncoding RNA classes, no defining biochemical features can be exclusively ascribed to lncRNAs. Rather a lack of defining features, such as the lack of an extended open reading frame (ORF), provides theoretical evidence that many transcripts function intrinsically as an RNA[8]. Conservation of an extended ORF, particularly when nucleotides in the ORF's codons exhibit different rates of selective constraint, can be used to distinguish coding transcripts[9]. Exceptions to these assumptions result from short or noncanonical peptides encoded in transcripts that evade screening attempts for viable ORFs[10]. Empirical support against the categorization of an lncRNA can be provided by matching ribosome footprints or peptide fragments from mass spectrometry that indicate translation[11,12]. However, although the ability to encode a protein does not necessarily preclude a transcript from having a function as an RNA—and indeed there is a growing catalog of bifunctional mRNA that are also lncRNAs[13]—the demonstration of function as an RNA may be ultimately required for annotation as a lncRNA.

The dynamic evolutionary interface between coding and noncoding components of the transcriptome also obscures a clear annotation of lncRNAs. Coding transcripts can lose their ability to encode a protein, and noncoding transcripts can acquire a coding function[14–16]. The complexity of this interface is seen at the X-inactivation center, where the *Xist* gene resulted from 'pseudogenization' of an ancestral protein-coding gene conserved in vertebrates, combined with the integration of flanking repetitive mobile elements[17]. These events remodeled the structure and sequence of the X-inactivation center in the eutherian lineage to generate noncoding transcripts that include not only *Xist*, but also *Tsix*, *Jpx* (*Enox*), *Xite* (*Rr18*) and *Ftx* (*Thcytx*) lncRNAs, each of which have acquired individual roles in X-chromosome inactivation[18].

Alternative splicing can also incorporate exons from multiple coding and noncoding genes, thereby merging gene structures and generating ambiguous transcripts that eschew simple classification[1,19,20]. Such examples of transcriptional complexity have contributed to a revised concept of the transcript being the basic unit of genome

[1]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia. [2]Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Brisbane, Queensland, Australia. [3]Garvan Institute of Medical Research, Sydney, New South Wales, Australia. Correspondence should be addressed to J.S.M. (j.mattick@garvan.org.au).

Received 17 September 2012; accepted 20 November 2012; published online 5 March 2013; doi:10.1038/nsmb.2480

BOX 1 Cytoplasmic lncRNAs

A substantial proportion of lncRNAs reside within, or are dynamically shuttled, to the cytoplasm where they regulate protein localization, mRNA translation and stability. For example, the NFAT transcription factor is trafficked from the cytoplasm to the nucleus to activate target genes in response to calcium-dependent signals. A lncRNA, *NRON*, complexes with importin-β proteins and regulates the trafficking of NFAT[113]. Notably, *NRON* inhibits the trafficking of NFAT to the nucleus specifically, with other proteins also trafficked by importin-β proteins, such as NF-κB, being unaffected.

By virtue of their ability to base pair with mRNAs, cytoplasmic lncRNAs also can regulate translation. The *UCHL1* mRNA is complemented by an antisense lncRNA, which, in response to stress or the mTOR pathway, is shuttled to the cytoplasm where, via an antisense complementary to the *UCHL1* AUG initiation codon and combined inverted SINEB2 domains, increases UCHL1 protein synthesis[54].

Additional repeat elements common to lncRNAs and mRNA create a broad interface for complementary interactions. Alu elements in cytoplasmic lncRNA can form imperfect complementary RNA duplexes with Alu elements in the 3′ UTRs of target mRNAs[83]. Staufen1 subsequently recognizes and binds the resultant dsRNA elements and initiates mRNA decay.

expression, with the concept of a gene encompassing a hierarchy of transcripts that underlie a given phenotype[21]. Similarly, many functional precedents associated with lncRNAs can be feasibly ascribed to other transcriptional elements. For example, untranslated regions (UTRs) of mammalian mRNAs often range over kilobases, sometimes dwarfing the size of the upstream open reading frame[22]. Given their common chemistry, such noncoding regions can feasibly mediate functions that are similar to those of other lncRNAs[23]. Although this review is constrained to the narrow definition of lncRNAs as long transcripts that exclude small regulatory RNAs, such as microRNAs, Piwi-interacting RNAs and small nucleolar RNAs, this arbitrary definition does not capture the nuanced and complex nature of the transcriptome, and efforts should be supported to evolve the concept of noncoding RNAs to a more inclusive definition of functional RNAs.

Abundance of lncRNAs encoded in the genome

Initial large-scale sequencing of cDNA libraries have revealed an unexpected abundance of lncRNAs[1]. This was supported by chromosome-wide and genome-wide tiling arrays and RNA sequencing that show the human genome to be prevalently transcribed into lncRNAs[5,24,25]. It is difficult to gauge an exact number of human lncRNAs, with current lncRNA catalogs ranging between 5,000 and 15,000 transcripts[26,27]. However, there is little overlap between these different lncRNA catalogs, and this may merely represent a lower bound, with many lncRNAs yet to be annotated. Whereas the number of known human protein-coding genes has remained stable over recent years, the number of known lncRNAs continues to accumulate, and lncRNAs may eventually rival protein-coding genes in number and diversity.

LncRNAs are expressed in lower amounts generally compared to their protein-coding counterparts, making it difficult to robustly detect and assemble complex transcript structures[26,27]. Indeed, targeted capture and RNA sequencing of intergenic regions affords the detection and assembly of many additional lncRNAs that are expressed in amounts too low to be otherwise detected by conventional high-throughput RNA sequencing (RNA-seq)[28]. Extrapolated across the genome, this abundance of lncRNAs represents a sizable expansion of the transcriptome. Though consistent with the many regulatory functions assigned to lncRNAs, the low expression may restrict these lncRNAs to subtle or redundant roles, or reflect incomplete repression in nonspecific cells. By comparison to protein-coding genes, lncRNA expression is considerably more cell type–specific; thus, RNA sequencing for more developmental stages and tissue types will be required to achieve comprehensive annotations[26,27].

Organization of lncRNA loci in the genome

The genome has a modular architecture with any single sequence incorporated into a range of sense and antisense, interwoven coding and noncoding transcripts[1,29]. The combinatorial application of alternative splicing, transcription initiation and termination exploits this modular architecture to drive diversification of transcription[30]. This transcriptional complexity is notably apparent in the myriad isoforms of lncRNA genes. LncRNAs are also often organized in close association with protein-coding genes (**Fig. 1a**). More than half of mammalian coding genes have complementary noncoding antisense transcription[31], which is also accompanied by overlapping, intronic and bidirectional noncoding transcription[30]. The iterative combination of these organizational modes generates complex transcriptional loci that include both coding and noncoding transcripts[1] (**Fig. 1b**).

Although complex loci containing lncRNAs and protein-coding gene may evolve to have a common local regulatory architecture, this clustered organization may also reflect the *cis*-acting functions of many lncRNAs in mediating changes to the local chromatin and the expression of neighboring genes. The X-inactivation center illustrates how the architecture of a complex locus regulates expression of the central *Xist* transcript (**Fig. 1c**). The lncRNA *Tsix* is transcribed antisense to, and initiates silencing of, *Xist* expression from the active X chromosome[32]. *Tsix* transcription through the 5′ end of *Xist* establishes a repressive chromatin domain by interaction with Polycomb repressive complex 2 (PRC2) and enhancing hypermethylation via DNA methyltransferase 3A (DNMT3A), thereby mediating long-term *Xist* silencing[33,34]. However, neither DNA methylation nor PRC2 is required for *Xist* repression[35], indicating that *Xist* regulation by *Tsix* is complex, with additional and alternative pathways. Indeed, the lncRNAs Tsix and Xist can form long RNA duplexes that are subsequently processed by the RNA interferences pathway into small regulatory RNAs that can contribute to downstream epigenetic changes[36]. Notably, *Tsix* has an additional antisense partner, *Xite* (*Rr18*), that also functions as a *cis*-acting activator of *Tsix*, appending additional layers to this regulatory circuit[37].

RNA structure and chemistry dictate lncRNA function

Given that the defining characteristic of lncRNAs is an intrinsic ability to function as an RNA molecule, it is important to understand the features of RNA chemistry from which this functionality derives. In addition to the four core nucleotides, the RNA sequence can include more than 100 chemically distinct modified nucleotides[38]. To date, most modifications have been detected in tRNAs, rRNAs and small nucleolar RNAs, where they often modulate and stabilize RNA structures[39]. However, the conversion of modified nucleotides for detection with high-throughput sequencing is revealing widespread nucleotide modifications throughout the transcriptome[40], where the modifications may similarly modulate lncRNA function[41,42]. Many of these post-transcriptional modifications are reversible and, given the range of modifications and targets, may comprise an additional

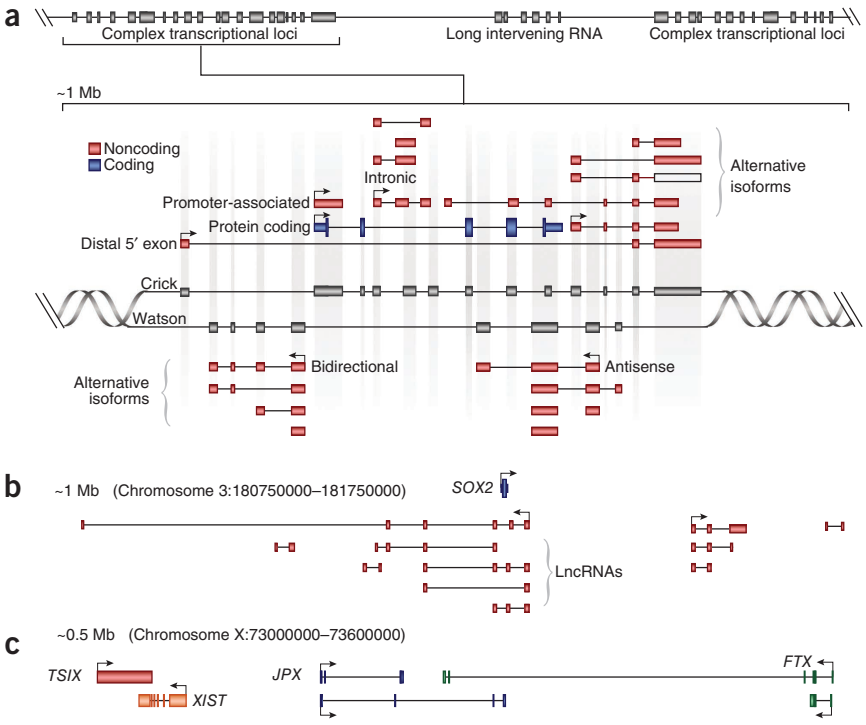

Figure 1 The human genome encodes an abundance and diversity of lncRNAs. (**a**) LncRNAs can be found harbored in intergenic regions or often clustered with protein-coding genes in complex transcriptional loci (top). Schematic of transcriptional networks (bottom) shows examples of lncRNAs (red) organized bidirectionally, or antisense to protein-coding genes or in introns of protein-coding genes (blue). Alternative splicing generates many lncRNAs isoforms and can merge gene structures by incorporating both coding and noncoding exons into a single transcript. (**b**) The X-inactivation center illustrates a complex lncRNA locus with numerous overlapping lncRNAs[17], including *Tsix*, *Ftx* and *Jpx*, that together regulate the expression of *Xist*, which inactivates the female X chromosome. (**c**) An example of the complex organization of lncRNAs associated with developmental genes. The *Sox2* gene involved in pluripotency and development is surrounded by many overlapping lncRNAs[112].

layer of post-transcriptional regulation analogous to the epigenetic landscape[43].

A major feature of lncRNAs is a propensity to fold into thermodynamically stable secondary and higher-order structures. RNA has the capacity to form hydrogen bonds on the Watson-Crick face but also the Hoogsteen and ribose face. These collective interactions contribute to RNA secondary structures that include double helices, hairpin loops, bulges and pseudoknots, and that are connected in higher-order tertiary interactions primarily mediated by non–Watson-Crick base-pairing[44]. This results in an RNA architecture dominated by coaxial stacks of helices that are organized in parallel or orthogonal to one another. This architecture is also modular, with recurrent structural motifs, including the sarcin-ricin loop, the K-turn and the C-loop, that exhibit only minor dependencies on neighboring sequences[45]. Furthermore, given that Watson-Crick base pairing and base stacking provide the greatest contribution to the energetic stability of RNA structures, these modular secondary structures generally fold initially and independently, before subsequent tertiary interactions occur, resulting in the hierarchal assembly of RNA structure[46].

In many cases, the secondary structure of lncRNAs dictates their function. For example, conservation of the secondary structure maintains the tumor suppressor function of lncRNA MEG3, rather than its primary sequence[47]. Although most RNA structural motifs originally had been described from rRNA and tRNA genes[48], increasing attention has since focused on empirical determination of lncRNA structure. Recently, the human lncRNA SRA1, a coactivator for several hormone receptors, was subjected to detailed chemical and enzymatic probing to determine four broad domains that encompass a suite of secondary structures[49]. Combining enzymatic and chemical probing with sequencing enables high-throughput characterization of RNA secondary structures required to delineate structure-function relationships[50,51]. These initial studies have confirmed a complex structural landscape in lncRNAs that can be distinguished from mRNAs on the basis of their high folding energy[51].

Range of functional domains in lncRNAs

To dissect the functional structures and sequences in lncRNAs, we can borrow insight and terminology developed in the field of synthetic biology, where RNA is commonly used as a regulatory device in genetic circuits[52]. RNA is a preferred substrate for such devices because it can rapidly shift between multiple stable structural conformations and undergo allosteric transitions, thereby acting as a responsive switch. Owing to the omission of translational processes, synthetic or noncoding RNAs are processed faster than other molecules dependent on transcription- or translation-dependent processes, a compact genetic footprint and a reduced energetic and resource load on the host cell[53]. These advantages of synthetic RNA regulatory devices similarly apply to lncRNAs.

RNA-binding domains. By virtue of their ability to base pair with other RNAs, lncRNAs can act as highly specific sensors of mRNA, microRNA and other lncRNA expression (**Fig. 2**), and RNA aptamers are often designed to induce a programmable response in riboswitches[52]. Antisense lncRNAs can regulate the stability and translation of complementary mRNAs. For example, translation of the *UCHL1* mRNA is regulated under stress by action of an lncRNA with antisense sequence that is complementary to, and encompasses, the *UCHL1* start codon[54].

Although antisense lncRNAs are prevalent in the genome, it is notable that small and often imperfect regions of nucleotide complementarity are sufficient for specific interactions, as demonstrated by the potent ability of microRNAs to target a select suite of mRNAs via short, imperfect seed sequences[55]. This imperfect complementarity has the advantage of allowing multiple RNA agonists, each with a range of dissociation constants, to compete for binding, thereby permitting lncRNAs to sense disparate RNA expression signals in the cell. For example, imperfect base pairing between Alu elements in lncRNAs and the 3′ UTR of translationally active mRNAs results in a dsRNA structure recognized and bound by Staufen1 and subsequently targeted for degradation.

The cleavage of lncRNAs can also generate small RNAs that serve as an output signal[56]. A small tRNA-like sequence is cleaved from the 3′ end of the lncRNA MALAT1 and trafficked from the nucleus to the cytoplasm[57]. Similarly, the formation of extended RNA duplexes or

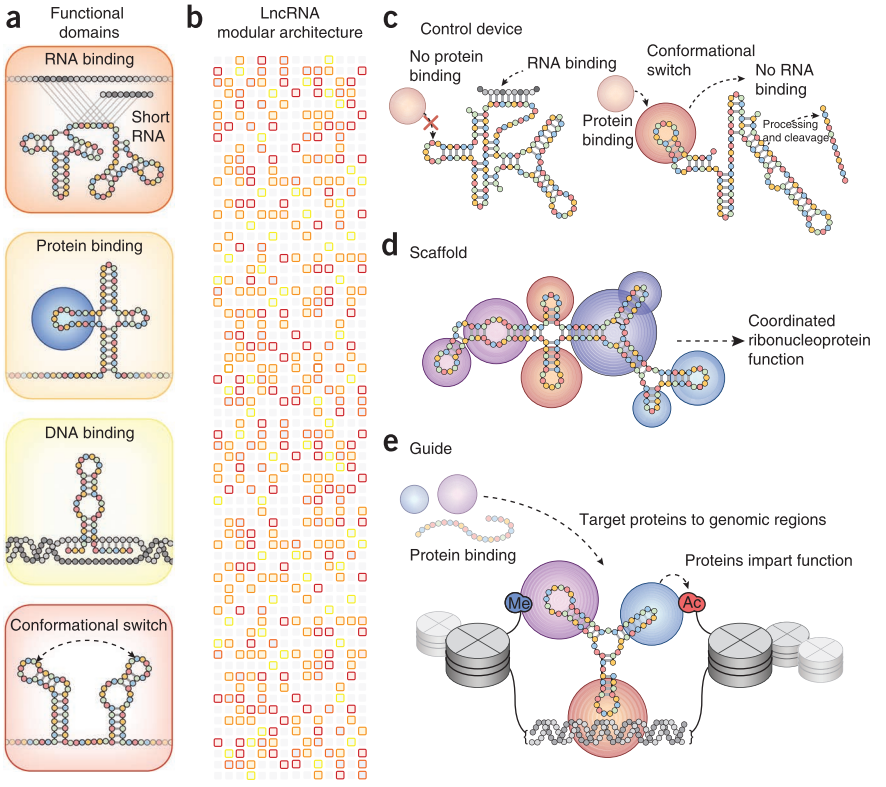

Figure 2 Domain architecture of lncRNAs. (**a**) LncRNAs contain structural domains that can sense or bind other RNAs via complementary base pair interactions, proteins and possibly DNA that can induce allosteric conformational changes to other structures in the lncRNA. (**b**) Alternative splicing can combine these structural domains into the modular architecture of lncRNAs. Each row represents an individual lncRNA; colors correspond to those in **a**. (**c**) Coupling sensory and actuator domains permits lncRNAs to act as a control device. In the example on the left, the binding of RNA (gray) induces a conformational change that prevents protein binding. Alternatively, as illustrated on the right, the protein can bind in the absence of the RNA, inducing the formation of a stem-loop secondary structure that can be processed and cleaved to generate an RNA output. (**d**) LncRNAs such as HOTAIR can act as molecular scaffold by binding multiple proteins to form complex ribonucleoprotein structures[78]. (**e**) LncRNAs, such as Xist, can target the catalytic function of proteins to specific sites in the genome[79]. LncRNAs can recruit chromatin-modifying proteins (purple and blue) to target sites by association with a DNA-binding protein such as YY1 (red)[71]. The chromatin modifiers then modify local histones to influence the expression of adjacent genes.

stem loops provides a ready substrate for Dicer enzyme to generate multiple small regulatory RNAs that have cascading ability to mediate downstream epigenetic changes[58]. Ribozymes comprise RNA secondary structures capable of the phosphodiester bond cleavage within themselves or in other RNAs[59]. Comparison between long and short RNA populations in human cells suggests widespread evidence of post-transcriptional cleavage, with lncRNAs being a preferred substrate for the generation of small RNAs[21]. The use of RNA as both output and input signals promotes RNA as a standard medium for transferring information within and between regulatory pathways, thereby assembling complex, multilayered and modular regulatory networks in the cell.

Protein-binding domains. Proteins are a major partner of lncRNAs (**Fig. 2**), with complexed ribonucleoprotein (RNP) particles acting as chaperones, transport aids or effectors[60]. Although RNA-binding proteins are one of the most abundant human protein classes, they are assembled from relatively few RNA-binding modules, and these domains are deployed in modular combinations with intervening disordered linker regions to accommodate the large diversity of RNA structures[61,62]. Additionally, many noncanonical domains supplement these RNP catalogs, with kinases, DNA-binding proteins and metabolic enzymes also found bound to RNA[63], and analysis of RNP fractions estimated that at least ~15% of expressed proteins are associated with polyadenylated RNA[64].

Approaches such as photoactivatable ribonucleoside–enhanced crosslinking and immunoprecipitation (PAR-CLIP) provide the complementary footprint to protein interactions across the transcriptome[65], and details of such protein-RNA interfaces can be resolved by X-ray crystallography or NMR spectroscopy. Proteins tend to interact with RNA where it forms complex secondary structures, positioning protein structures into the groove of an RNA stem-loop helix or providing a binding pocket in β-sheets for unpaired RNA nucleotides[66]. Almost all such interactions characterized to date involve conformational changes to the protein, the RNA or both. The structural diversity of RNA in combination with the commensurate abundance of RNA-binding proteins provide a broad interface for communication between the proteome and transcriptome.

DNA-binding domains? There is currently little evidence for direct interaction between lncRNAs and DNA. RNA:DNA hybrids or triplex structures can allow single strands of RNA to interact with DNA duplexes by base-pair interactions. These direct RNA-DNA interactions could efficiently and selectively target RNA signals to genomic loci (**Fig. 2**). However, such interactions may also expose the genome to deamination and damage[67,68]. A promoter-associated lncRNA, pRNA, can occlude binding of the transcription termination factor 1 (TTF1), while simultaneously recruiting DNMT3b to repress rRNA gene expression[69]. Notably, this lncRNA can form a triplex with the TTF1-binding site *in vitro*, supporting a direct interaction with the genomic locus. Rather than being involved in direct complementary interactions, RNA folds may create a DNA-binding pocket in a manner analogous to the DNA-binding domains of a protein transcription factor. Similar to protein transcription factors, an enriched DNA sequence motif has been identified in the binding sites of an lncRNA, HOTAIR[70]. Alternatively, rather than directly interacting with DNA, Xist harnesses the sequence-specific YY1 transcription factor to tether Xist to sites in the X chromosome[71].

Modular architecture of lncRNAs

LncRNAs can act as regulatory devices by allosterically coupling binding domains with the switching of structural conformations and thereby activating or suppressing linked functional domains[52]. Incorporating multiple sensors into the architecture of a single lncRNA permits the integration and processing of multiple inputs through logic gates to produce a single output[53]. An instructive example

of such a switch embedded in an RNA structure involves a protein-dependent riboswitch in the VEGF 3′ UTR that is bound by either GAIT or heterogeneous nuclear ribonucleoprotein L (hnRNP L) protein complexes[72]. The binding of one protein, in response to inflammatory and hypoxic cues, induces a conformational change that precludes binding of the other protein, thereby processing different protein inputs to regulate mRNA translation and gene expression.

When each of these modular domains specifies alternate protein interactions, an lncRNA can scaffold many disparate proteins into a single cogent ribonucleoprotein[73]. The lncRNA CDKN2B-AS1 (ANRIL) binds the Suz12 subunit of the PRC2 complex and the chromodomain of CBX7 in the PRC1 complex to trimethylate H3K27, which then mediates repression of the *CDKN2A/CDKN2B* (*IFNK4a/IFNK4b*) tumor suppressor locus[74]. With at least 12 chromatin-modifying proteins having been associated with lncRNAs to date[75], the composition of possible chromatin-modifying proteins in a single ribonucleoprotein can be varied by shuffling the modular components within an lncRNA, in concordance with the combinatorial histone code[76]. For example, in mice the lncRNA Kcnq1ot1 can recruit both PRC2 and G9a, which impart H3K4 trimethylation and H3K9 methylation, respectively[77]. Similarly, HOTAIR binds and targets the catalytic methyltransferase subunit EZH2 of PRC2 via a structural domain at the 5′ end of HOTAIR. The 3′ end binds and recruits lysine-specific demethylase 1 (LSD1), which catalyzes demethylation of H3K4, a modification similarly associated with gene repression[78]. HOTAIR can therefore act as a scaffold that, via its modular architecture, can combine PRC2 and LSD1 into a single RNP complex.

Many functional domains in lncRNAs derive from repeat elements. *Xist* exhibits such a modular functional architecture, containing an A-region comprised of up to nine repetitive elements that fold into two stem-loop structures that recruit PRC1 for H3K27 trimethylation, and a C-repeat region that binds to both hnRNP U and YY1 and is responsible for localizing and tethering Xist to the X chromosome[71,79–81]. Each of these functional domains, the A- and C-elements, derives from the insertion and remodeling of repetitive elements. This repetitive architecture is also present in the functionally analogous lncRNA Rsx, which mediates X-chromosome inactivation in marsupials. Although Rsx likely evolved independently and has no homology with Xist, Rsx is similarly enriched for tandem repeats in the 5′ region and conserved stem-loop motifs[82]. Similar repetitive sequences identified across many lncRNAs may also reflect underlying functional domains in lncRNAs, analogous to protein domains[26]. Transposon-derived repeats, such as short interspersed elements (SINEs) and Alu elements, can also be integrated within lncRNAs, where they act as recognition domains that target complementary mRNA for translation or degradation[83,84]. In these cases, the transmission of transposons can result in the insertion of a ready functional domain into target lncRNAs.

Cis and *trans* mechanisms of epigenetic regulation

The regulatory potential of lncRNAs has been best characterized in the context of their role as epigenetic modulators. During development, DNA and histone modifications comprise a detailed and dynamic template that is overlaid across the genome to guide the regulation and expression of the underlying DNA sequences. It has become apparent that lncRNA can guide many of the proteins that catalyze these modifications[75,76,85,86].

The lncRNA HOTAIR guides chromatin proteins and their catalytic action in *trans* to multiple sites spread across the genome. *HOTAIR* is expressed from the end of the *HOXC* (*HOXC@*) cluster in cells with distal and posterior identities[87]. HOTAIR binds and targets PRC2 to the *HOXD* (*HOXD@*) cluster as well as hundreds of additional sites around the genome to impart repressive histone modifications[70]. These focal interactions of HOTAIR with target genome sites are likely pioneering events that subsequently nucleate broad regions of Polycomb occupancy and H3K27 trimethylation. By the expression of *HOTAIR*, distal developmental states can initiate an epigenetic regulatory cascade that maintains the cells' positional identity and continually refines a progressive developmental trajectory.

Many lncRNAs also mediate local functions in *cis*, recruiting chromatin-modifying proteins to modify their surrounding epigenetic neighborhood. For example, loci with parent-of-origin–specific expression patterns, termed imprinted regions, often encompass many lncRNAs[86,88]. These lncRNAs can silence expression of neighboring genes on the parental allele from which they are expressed. Air (Airn), an 108-kb lncRNA, is transcribed from the paternal allele to which it recruits G9a to methylate H3K9 residues over an adjacent 300-kb region of the genome, thereby silencing the expression of neighboring paternal *Igf2r*, *Slc22a2* and *Slc22a3* genes[89]. The ability of lncRNAs to act in *cis* to remodel the local epigenetic landscape and thereby regulate neighboring gene expression may, in part, explain the concentration of lncRNAs around many critical developmental genes whose expression is subject to strict control[90] (**Fig. 1c**).

Local epigenetic changes initiated by lncRNAs can seed broader epigenetic effects, with the silencing of an entire female X chromosome by Xist providing a dramatic example. After transcription, Xist accumulates at many sites across the X chromosome that subsequently seed the progression of facultative heterochromatin formation[91,92]. LncRNAs similarly seed the spread of heterochromatin at pericentric satellites[93]. Pericentric satellite repeats, localized around the centromere, are required for genome integrity and replication. The peripheral regions of these satellite repeats undergo bidirectional transcription during early development, and the resulting forward-strand lncRNAs can recruit heterochromatin protein 1 (HP1) proteins to the satellite repeats to impart a distinct modified chromatin profile[94]. The continued local accumulation of HP1 results in the spreading of facultative heterochromatin to demarcate broad repressive chromosomal domains.

Role for lncRNAs in nuclear organization

Considering complex transcriptional loci and their surrounding epigenetic landscape in the three-dimensional context of the nucleus can obscure the distinctions between *trans*- and *cis*-acting lncRNA functions. The HOTTIP lncRNA binds WD repeat-containing protein 5 (WDR5), a cofactor to the histone methyltransferase MLL1 that mediates H3K4 trimethylation associated with gene activation[95]. Although HOTTIP remains bound at the site of transcription, chromatin looping positions the HOTTIP–WDR5–MLL1 complex into close spatial proximity to multiple distal sites to coordinate active chromatin domains and gene expression throughout the *HOXC* loci. Chromosome looping thereby extends the range to which HOTTIP can impact epigenetic changes, similar to the role of looping in facilitating enhancer action over large genomic distances.

Immunofluorescence microscopy has revealed a complex internal structure to the nucleus, with modified chromatin often localized in cogent subnuclear structures. LncRNAs can scaffold and nucleate the assembly of such structures, recruiting the accumulation of proteins required for their assembly. For example, the lncRNA MALAT1 binds to and sequesters serine/arginine–rich splicing factors to nuclear speckles[96] and the lncRNA NEAT1 is a requisite structural component of interchromatin paraspeckles[97]. Additional subnuclear structures, including histone locus bodies and stress

BOX 2 The 'evolvability' of RNA?

The close relationship between nucleotide sequence and RNA and protein secondary structure has proved an ideal model system for the analysis of genotype-phenotype relationships in the context of adaptive evolution[114]. The structure (or phenotype) of RNA or proteins can be generated by a number of alternative sequences, as different RNA or protein sequences may fold into similar structures. As a group, these different sequences are called a genotype network. Generally, many sequences fold into similar RNA secondary structures, and therefore RNA structures have very large and diverse genotype networks[111]. Indeed, entirely dissimilar sequences can often fold into similar RNA structures, and RNA structures can therefore traverse the nucleotide space completely. For example, RNA molecules that adopt the characteristic tRNA cloverleaf secondary structure can differ in up to >90% of their nucleotides[115], and it is typical for naturally occurring RNAs to have astronomically large genotype networks[116]. Protein structures have much smaller genotype networks that are closely restrained in their ability to traverse the nucleotide space[111], and have a much lower affinity to explore other viable secondary structures. Random protein libraries rarely fold into soluble and compact protein structures[117]. By contrast, random RNAs collapse with high probability into compact and ordered structures[118].

The large genotype network size of RNA structures means they are robust, capable of preserving secondary structure while accumulating mutations[119]. However, these accumulated mutations also serve to diversify the genotype network, spreading the network into more neighborhoods and making a broader spectrum of a novel phenotype in immediate neighborhood directly accessible by only a small number of additional mutations[119]. This ultimately increases the likelihood of encountering a beneficial phenotype and therefore achieving evolutionary adaptation and innovation. This advantage is demonstrated in ribozymes that harbor greater levels of latent cryptic variation being better suited to evolutionary adaption in response to changing environment conditions[120]. Collectively, this suggests that RNAs, in comparison to proteins, have a higher affinity for adaptive evolution as a result of their greater genotype networks.

The complexity of the transcriptome, driven in large part by a massive expansion of lncRNA transcripts, is a hallmark of the eukaryotic genome. It is tempting to speculate that this expansion of lncRNAs has been driven in part by selection for their rapid evolvibility, which reciprocally has been a primary factor in driving the radiation and evolution of eukaryotic lineages[121].

bodies also require RNA for nucleation and assembly, suggesting a broad role for lncRNAs in nuclear organization[98].

Nuclear organization also forms an overarching three-dimensional context under which lncRNAs mediate their regulatory roles. For example, repressive regions of H3K27 trimethylation and PRC2 occupancy form a network of close physical interactions that appear as discrete Polycomb bodies in the nucleus[99,100]. Given the close and promiscuous association of PRC2 with lncRNAs[85], the *cis*-acting action of many lncRNAs may be similarly involved in mediating the repressive functions of Polycomb bodies. Indeed, two lncRNAs, TUG1 and MALAT1, traffic gene loci between Polycomb bodies to the activating context of interchromatin granules[101]. TUG1 lncRNA specifically associates with methylated PRC2 at E2F1 target gene promoters within Polycomb bodies. However, in response to growth signals, the promoter-localized PRC2 is demethylated and associates with MALAT1, resulting in the relocation of these gene loci to interchromatin granules. This interaction with MALAT1 also regulates E2F1 sumoylation, which permits recruitment of CDCA7L, a histone H2B monoubiquitinase, to switch the preference of the PRC2 chromodomain from repressive to activation-associated histone modifications[101].

LncRNAs in a broader epigenetic regulatory network

Many of the examples discussed above ascribe a function of lncRNAs in guiding, whether in *cis* or *trans*, the catalytic function of chromatin-modifying proteins to specific genomic sites. Thousands of lncRNAs are found in association with chromatin modifiers, associations that are as highly dynamic as the tissue- and development-specific expression of lncRNAs[75,85]. Considering the abundance of lncRNAs along with these functional precedents raises the potential for lncRNAs and chromatin-modifying enzymes to collectively comprise a regulatory network with the requisite complexity to delineate a dynamic epigenetic landscape. A recurrent property of biological regulatory networks is a scale-free topology, in which the majority of nodes have few links whereas a minority of nodes, termed hubs, are furnished with many links that bind the network together[102]. Chromatin-modifying enzymes as ubiquitous and highly connected proteins are archetypal hubs, whereas many lowly expressed, and tissue-specific lncRNAs likely populate sparsely connected, lower-level nodes.

A major advantage of scale-free networks is that they are robust. The network is tolerant to the inactivation of even a large number of sparsely connected nodes without disrupting the integrity of the network, consistent with the few phenotypic effects observed in knockout screens targeting individual lncRNAs[103–105]. However, scale-free networks are vulnerable to the inactivation of highly connected hubs that fatally splinter off the network into many isolated nodes. Indeed, chromatin-modifying proteins are highly conserved, with few genes being lost between worms and humans[106], and mutations in chromatin-modifying proteins causing multiple cancers and developmental disorders[107].

Selective constraint would also be relieved on nonessential lncRNAs nodes. LncRNAs are among the fastest evolving elements in the genome, with a rate of lncRNA gain and loss that is much higher than that of protein-coding genes[108–110], and one-third of human lncRNAs are thought to have arisen solely in the primate lineage[26]. This specificity extends to function, with even the iconic HOTAIR exhibiting divergent roles in human and mouse[27,104]. Therefore, combining chromatin modifiers and lncRNAs in a cogent regulatory network could confer both robust and adaptive capacities[111] (**Box 2**).

Future directions

Since the abundance of lncRNAs was revealed in early sequencing efforts[24], lncRNAs have been the focus of intense research, and already a wide range of functional roles have been ascribed to individual lncRNAs. However, the sheer abundance and diversity of lncRNAs pose a challenge for their characterization.

A greater understanding of structure-to-function relationships—that is, how and which modular elements dictate a specific function—will be required to characterize such an abundance of transcripts. The development of high-throughput approaches to determine secondary structure, protein-binding motifs and other features in the primary sequence could realize a detailed and global landscape of elements in lncRNAs. The functional characterization of such elements, as opposed to individual transcripts, would provide a powerful predictive platform to extrapolate functions across related classes lncRNAs that have similar features. This could ultimately permit the functional assignation and validation of many lncRNAs on the basis of sequence and structure, analogous to

large-scale structural proteogenomic efforts, and be hugely informative in the hypothesis of individual lncRNA function.

LncRNAs have, in a relatively short period of time, become recognized as a legitimate and major new class of genes. LncRNAs may potentially comprise a major component of the genome's information content, complementary and comparable in abundance and complexity to the proteome. Given such huge potential, they have begun to generate considerable excitement in the molecular biology community. It is with time that we will realize whether lncRNAs will live up to such potential.

ACKNOWLEDGMENTS

We thank the following funding sources: Australian National Health and Medical Research Council Australia Fellowship (631668 to J.S.M. and T.R.M.).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nsmb.2480.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).
One of the earliest analyses of large-scale cDNA sequencing, this work reveals the abundance of lncRNAs and complexity of transcriptional organization in the eukaryotic genome.
2. Mattick, J.S. A new paradigm for developmental biology. *J. Exp. Biol.* **210**, 1526–1547 (2007).
3. Amaral, P.P., Clark, M.B., Gascoigne, D.K., Dinger, M.E. & Mattick, J.S. lncRNAdb: a reference database for long noncoding RNAs. *Nucleic Acids Res.* **39**, D146–D151 (2011).
4. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
5. Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484–1488 (2007).
6. Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M. & Pagano, A. The expanding RNA polymerase III transcriptome. *Trends Genet.* **23**, 614–622 (2007).
7. Yin, Q.F. *et al.* Long noncoding RNAs with snoRNA ends. *Mol. Cell* **48**, 219–230 (2012).
8. Dinger, M.E., Pang, K.C., Mercer, T.R. & Mattick, J.S. Differentiating protein-coding and noncoding RNA: challenges and ambiguities. *PLOS Comput. Biol.* **4**, e1000176 (2008).
9. Lin, M.F., Jungreis, I. & Kellis, M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **27**, i275–i282 (2011).
10. Galindo, M.I., Pueyo, J.I., Fouix, S., Bishop, S.A. & Couso, J.P. Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol.* **5**, e106 (2007).
11. Ingolia, N.T., Lareau, L.F. & Weissman, J.S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011).
12. Banfai, B. *et al.* Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res.* **22**, 1646–1657 (2012).
13. Dinger, M.E., Gascoigne, D.K. & Mattick, J.S. The evolution of RNAs with multiple functions. *Biochimie* **93**, 2013–2018 (2011).
14. Poliseno, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033–1038 (2010).
15. Carvunis, A.R. *et al.* Proto-genes and *de novo* gene birth. *Nature* **487**, 370–374 (2012).
16. Zheng, D. *et al.* Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. *Genome Res.* **17**, 839–851 (2007).
17. Duret, L., Chureau, C., Samain, S., Weissenbach, J. & Avner, P. The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* **312**, 1653–1655 (2006).
18. Lee, J.T. The X as model for RNA's niche in epigenomic regulation. *Cold Spring Harb. Perspect. Biol.* **2**, a003749 (2010).
19. Gerstein, M.B. *et al.* What is a gene, post-ENCODE? History and updated definition. *Genome Res.* **17**, 669–681 (2007).
20. Denoeud, F. *et al.* Prominent use of distal 5′ transcription start sites and discovery of a large number of additional exons in ENCODE regions. *Genome Res.* **17**, 746–759 (2007).
21. Djebali, S., Davis, C.A., LaGarde, J. & Gingeras, T. Landscape of transcription in human cell lines. *Nature* **489**, 101–108 (2012).
22. Mazumder, B., Seshadri, V. & Fox, P.L. Translational control by the 3′-UTR: the ends specify the means. *Trends Biochem. Sci.* **28**, 91–98 (2003).
23. Mercer, T.R. *et al.* Expression of distinct RNAs from 3′ untranslated regions. *Nucleic Acids Res.* **39**, 2393–2403 (2011).
24. Okazaki, Y. *et al.* Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**, 563–573 (2002).
25. Rinn, J.L. *et al.* The transcriptional activity of human chromosome 22. *Genes Dev.* **17**, 529–540 (2003).
26. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
27. Cabili, M.N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).
28. Mercer, T.R. *et al.* Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.* **30**, 99–104 (2012).
29. Kapranov, P., Willingham, A.T. & Gingeras, T.R. Genome-wide transcription and the implications for genomic organization. *Nat. Rev. Genet.* **8**, 413–423 (2007).
30. Kapranov, P. *et al.* Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res.* **15**, 987–997 (2005).
This systematic study reveals the complex organization of gene loci that is a common feature within the genome's modular architecture.
31. Katayama, S. *et al.* Antisense transcription in the mammalian transcriptome. *Science* **309**, 1564–1566 (2005).
32. Ohhata, T., Hoki, Y., Sasaki, H. & Sado, T. Crucial role of antisense transcription across the Xist promoter in Tsix-mediated Xist chromatin modification. *Development* **135**, 227–235 (2008).
33. Sun, B.K., Deaton, A.M. & Lee, J.T. A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. *Mol. Cell* **21**, 617–628 (2006).
34. Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J. & Lee, J.T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756 (2008).
35. Sado, T., Okano, M., Li, E. & Sasaki, H. De novo DNA methylation is dispensable for the initiation and propagation of X chromosome inactivation. *Development* **131**, 975–982 (2004).
36. Ogawa, Y., Sun, B.K. & Lee, J.T. Intersection of the RNA interference and X-inactivation pathways. *Science* **320**, 1336–1341 (2008).
37. Lee, J.T. Regulation of X-chromosome counting by Tsix and Xite sequences. *Science* **309**, 768–771 (2005).
38. Cantara, W.A. *et al.* The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res.* **39**, D195–D201 (2011).
39. Helm, M. Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res.* **34**, 721–733 (2006).
40. Kellner, S., Burhenne, J. & Helm, M. Detection of RNA modifications. *RNA Biol.* **7**, 237–247 (2010).
41. Squires, J.E. *et al.* Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res.* **40**, 5023–5033 (2012).
42. Jia, G. *et al.* N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7**, 885–887 (2011).
43. He, C. Grand challenge commentary: RNA epigenetics? *Nat. Chem. Biol.* **6**, 863–865 (2010).
44. Cruz, J.A. & Westhof, E. The dynamic landscapes of RNA architecture. *Cell* **136**, 604–609 (2009).
45. Lescoute, A. & Westhof, E. The interaction networks of structured RNAs. *Nucleic Acids Res.* **34**, 6587–6604 (2006).
46. Talkington, M.W., Siuzdak, G. & Williamson, J.R. An assembly landscape for the 30S ribosomal subunit. *Nature* **438**, 628–632 (2005).
47. Zhang, X. *et al.* Maternally expressed gene 3 (MEG3) noncoding ribonucleic acid: isoform structure, expression, and functions. *Endocrinology* **151**, 939–947 (2010).
48. Steitz, T.A. A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell Biol.* **9**, 242–253 (2008).
49. Novikova, I.V., Hennelly, S.P. & Sanbonmatsu, K.Y. Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res.* **40**, 5034–5051 (2012).
50. Underwood, J.G. *et al.* FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* **7**, 995–1001 (2010).
51. Kertesz, M. *et al.* Genome-wide measurement of RNA secondary structure in yeast. *Nature* **467**, 103–107 (2010).
52. Liang, J.C., Bloom, R.J. & Smolke, C.D. Engineering biological systems with synthetic RNA molecules. *Mol. Cell* **43**, 915–926 (2011).
53. Win, M.N. & Smolke, C.D. Higher-order cellular information processing with synthetic RNA devices. *Science* **322**, 456–460 (2008).
54. Carrieri, C. *et al.* Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* **491**, 454–457 (2012).
55. Cisse, I.I., Kim, H. & Ha, T. A rule of seven in Watson-Crick base-pairing of mismatched sequences. *Nat. Struct. Mol. Biol.* **19**, 623–627 (2012).
56. Affymetrix ENCODE Transcriptome Project & Cold Spring Harbor Laboratory ENCODE Transcriptome Project. Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs. *Nature* **457**, 1028–1032 (2009).
57. Wilusz, J.E., Freier, S.M. & Spector, D.L. 3′ end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135**, 919–932 (2008).

58. Djupedal, I. & Ekwall, K. Epigenetics: heterochromatin meets RNAi. *Cell Res.* **19**, 282–295 (2009).
59. Talini, G., Branciamore, S. & Gallori, E. Ribozymes: flexible molecular devices at work. *Biochimie* **93**, 1998–2005 (2011).
60. Hogan, D.J., Riordan, D.P., Gerber, A.P., Herschlag, D. & Brown, P.O. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biol.* **6**, e255 (2008).
61. Lunde, B.M., Moore, C. & Varani, G. RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.* **8**, 479–490 (2007).
62. Castello, A. *et al.* Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406 (2012).
63. Ciesla, J. Metabolic enzymes that bind RNA: yet another level of cellular regulatory network? *Acta Biochim. Pol.* **53**, 11–32 (2006).
64. Baltz, A.G. *et al.* The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **46**, 674–690 (2012).
65. Scheibe, M., Butter, F., Hafner, M., Tuschl, T. & Mann, M. Quantitative mass spectrometry and PAR-CLIP to identify RNA-protein interactions. *Nucleic Acids Res.* **40**, 9897–9902 (2012).
66. Reiter, N.J., Chan, C.W. & Mondragon, A. Emerging structural themes in large RNA molecules. *Curr. Opin. Struct. Biol.* **21**, 319–326 (2011).
67. Buske, F.A., Mattick, J.S. & Bailey, T.L. Potential *in vivo* roles of nucleic acid triple-helices. *RNA Biol.* **8**, 427–439 (2011).
68. Aguilera, A. & Garcia-Muse, T. R loops: from transcription byproducts to threats to genome stability. *Mol. Cell* **46**, 115–124 (2012).
69. Schmitz, K.M., Mayer, C., Postepska, A. & Grummt, I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev.* **24**, 2264–2269 (2010).
70. Chu, C., Qu, K., Zhong, F.L., Artandi, S.E. & Chang, H.Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* **44**, 667–678 (2011).
71. Jeon, Y. & Lee, J.T. YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* **146**, 119–133 (2011).
72. Ray, P.S. *et al.* A stress-responsive RNA switch regulates VEGFA expression. *Nature* **457**, 915–919 (2009).
73. Delebecque, C.J., Lindner, A.B., Silver, P.A. & Aldaye, F.A. Organization of intracellular reactions with rationally designed RNA assemblies. *Science* **333**, 470–474 (2011).
74. Yap, K.L. *et al.* Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol. Cell* **38**, 662–674 (2010).
75. Guttman, M. *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295–300 (2011).
76. Koziol, M.J. & Rinn, J.L. RNA traffic control of chromatin complexes. *Curr. Opin. Genet. Dev.* **20**, 142–148 (2010).
77. Pandey, R.R. *et al.* Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* **32**, 232–246 (2008).
78. Tsai, M.C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
79. Plath, K. *et al.* Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131–135 (2003).
This work shows the recruitment of the Eed–Ezh2 complex to mediate H3K27 methylation at the inactive X chromosome by Xist.
80. Wutz, A., Rasmussen, T.P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nat. Genet.* **30**, 167–174 (2002).
81. Hasegawa, Y., Brockdorff, N., Kawano, S., Tsutui, K. & Nakagawa, S. The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev. Cell* **19**, 469–476 (2010).
82. Grant, J. *et al.* Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation. *Nature* **487**, 254–258 (2012).
83. Gong, C. & Maquat, L.E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature* **470**, 284–288 (2011).
84. Mariner, P.D. *et al.* Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol. Cell* **29**, 499–509 (2008).
85. Khalil, A.M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. USA* **106**, 11667–11672 (2009).
86. Sanchez-Elsner, T., Gou, D., Kremmer, E. & Sauer, F. Noncoding RNAs of Trithorax response elements recruit *Drosophila* Ash1 to Ultrabithorax. *Science* **311**, 1118–1123 (2006).
A model for the local recruitment of the epigenetic regulator Ash1 by lncRNAs is proposed.
87. Rinn, J.L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
This work is a demonstration of a lncRNA, HOTAIR, that targets Polycomb proteins in *trans* to the *HOXD* loci sites in the human genome.
88. Mohammad, F., Mondal, T. & Kanduri, C. Epigenetics of imprinted long noncoding RNAs. *Epigenetics* **4**, 277–286 (2009).
89. Nagano, T. *et al.* The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717–1720 (2008).
90. Engstrom, P.G. *et al.* Complex Loci in human and mouse genomes. *PLoS Genet.* **2**, e47 (2006).
91. Pinter, S.F. *et al.* Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. *Genome Res.* **22**, 1864–1876 (2012).
92. Wutz, A. Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nat. Rev. Genet.* **12**, 542–553 (2011).
93. Almouzni, G. & Probst, A.V. Heterochromatin maintenance and establishment: lessons from the mouse pericentromere. *Nucleus* **2**, 332–338 (2011).
94. Maison, C. *et al.* SUMOylation promotes de novo targeting of HP1alpha to pericentric heterochromatin. *Nat. Genet.* **43**, 220–227 (2011).
95. Wang, K.C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
96. Tripathi, V. *et al.* The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* **39**, 925–938 (2010).
97. Mao, Y.S., Sunwoo, H., Zhang, B. & Spector, D.L. Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat. Cell Biol.* **13**, 95–101 (2011).
98. Shevtsov, S.P. & Dundr, M. Nucleation of nuclear bodies by RNA. *Nat. Cell Biol.* **13**, 167–173 (2011).
99. Bantignies, F. *et al.* Polycomb-dependent regulatory contacts between distant Hox loci in *Drosophila*. *Cell* **144**, 214–226 (2011).
100. Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F. & Orlando, V. Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat. Cell Biol.* **9**, 1167–1174 (2007).
101. Yang, L. *et al.* ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell* **147**, 773–788 (2011).
102. Barabasi, A.L. & Oltvai, Z.N. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).
103. Mattick, J.S. The genetic signatures of noncoding RNAs. *PLoS Genet.* **5**, e1000459 (2009).
104. Schorderet, P. & Duboule, D. Structural and functional differences in the long non-coding RNA hotair in mouse and human. *PLoS Genet.* **7**, e1002071 (2011).
105. Zhang, B. *et al.* The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep.* **2**, 111–123 (2012).
106. On, T. *et al.* The evolutionary landscape of the chromatin modification machinery reveals lineage specific gains, expansions, and losses. *Proteins* **78**, 2075–2089 (2010).
107. Fog, C.K., Jensen, K.T. & Lund, A.H. Chromatin-modifying proteins in cancer. *APMIS* **115**, 1060–1089 (2007).
108. Kutter, C. *et al.* Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet.* **8**, e1002841 (2012).
109. Pollard, K.S. *et al.* An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**, 167–172 (2006).
110. Pollard, K.S. *et al.* Forces shaping the fastest evolving regions in the human genome. *PLoS Genet.* **2**, e168 (2006).
111. Ferrada, E. & Wagner, A. A comparison of genotype-phenotype maps for RNA and proteins. *Biophys. J.* **102**, 1916–1925 (2012).
112. Amaral, P.P. *et al.* Complex architecture and regulated expression of the Sox2ot locus during vertebrate development. *RNA* **15**, 2013–2027 (2009).
113. Willingham, A.T. *et al.* A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science* **309**, 1570–1573 (2005).
114. Schuster, P., Fontana, W., Stadler, P.F. & Hofacker, I.L. From sequences to shapes and back: a case study in RNA secondary structures. *Proc. Biol. Sci.* **255**, 279–284 (1994).
115. Saks, M.E., Sampson, J.R. & Abelson, J. Evolution of a transfer RNA gene through a point mutation in the anticodon. *Science* **279**, 1665–1670 (1998).
116. Jorg, T., Martin, O.C. & Wagner, A. Neutral network sizes of biological RNA molecules can be computed and are not atypically small. *BMC Bioinformatics* **9**, 464 (2008).
117. Yomo, T. *et al.* Properties of artificial proteins with random sequences. *Ann. NY Acad. Sci.* **864**, 131–135 (1998).
118. Schultes, E.A., Spasic, A., Mohanty, U. & Bartel, D.P. Compact and ordered collapse of randomly generated RNA sequences. *Nat. Struct. Mol. Biol.* **12**, 1130–1136 (2005).
119. Wagner, A. The role of robustness in phenotypic adaptation and innovation. *Proc. Biol. Sci.* **279**, 1249–1258 (2012).
120. Hayden, E.J., Ferrada, E. & Wagner, A. Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature* **474**, 92–95 (2011).
This work is an empirical demonstration that variation in an RNA genotype network can drive rapid phenotypic adaptation.
121. Pigliucci, M. Is evolvability evolvable? *Nat. Rev. Genet.* **9**, 75–82 (2008).