

 NON-CODING RNA

# Unique features of long non-coding RNA biogenesis and function

Jeffrey J. Quinn<sup>1,2</sup> and Howard Y. Chang<sup>1</sup>

**Abstract** | Long non-coding RNAs (lncRNAs) are a diverse class of RNAs that engage in numerous biological processes across every branch of life. Although initially discovered as mRNA-like transcripts that do not encode proteins, recent studies have revealed features of lncRNAs that further distinguish them from mRNAs. In this Review, we describe special events in the lifetimes of lncRNAs — before, during and after transcription — and discuss how these events ultimately shape the unique characteristics and functional roles of lncRNAs.

**Long non-coding RNAs** (lncRNAs). RNA transcripts with a length of > 200 nt that do not encode proteins.

#### Imprinting

An epigenetic mechanism of regulating gene expression in an allele-specific manner.

#### Cis or trans

Long non-coding RNAs can act on their neighbouring genomic environment (in *cis*) or diffuse to distant sites of action (in *trans*).

#### Transcriptional noise

A hypothesis explaining pervasive transcription by which RNA polymerase II randomly initiates transcription throughout the genome.

The transcriptional landscape of all organisms is far more complex than was originally imagined, as the vast majority of genomic sequence is pervasively transcribed into a diverse range of protein-coding RNAs and non-coding RNAs (ncRNAs)<sup>1,2</sup>. In this expanded view of both the genome and the transcriptome, our catalogue of genetic elements is now brimming with long non-coding RNAs (lncRNAs), a loosely classified group of long RNA transcripts with no apparent protein-coding role<sup>3,4</sup>. lncRNAs are found in every branch of life, and organismal complexity is better correlated with the diversity and size of non-coding RNA expression repertoires than with that of protein-coding genes<sup>5</sup>. lncRNAs are diverse and numerous; by most estimates, the number of human lncRNAs outstrips the number of protein-coding genes<sup>1</sup>. The total number of lncRNAs continues to climb, catalysed by deeper and more sensitive RNA sequencing, improved epigenomic technologies and computational prediction techniques<sup>6,7</sup>. Their growing ranks have motivated an increased focus on understanding the roles of lncRNAs in biology.

lncRNAs are involved in numerous important biological phenomena such as imprinting genomic loci, shaping chromosome conformation and allosterically regulating enzymatic activity (reviewed in REFS 8,9). Specific patterns of lncRNA expression coordinate cell state, differentiation, development and disease (reviewed in REFS 10,11). The overexpression, deficiency or mutation of lncRNA genes has been implicated in numerous human diseases (reviewed in REF. 12). The functions of the majority of lncRNAs are unknown, and many lncRNAs may not have appreciable functions, but the functional roles and mechanisms of action of some classically defined lncRNAs are well understood, such as X inactive specific transcript (XIST; in X chromosome inactivation), HOX transcript antisense RNA (HOTAIR; in positional identity) and telomerase RNA component

(TERC; in telomere elongation) — and this list of characterized lncRNAs continues to grow. Their functional mechanisms are diverse, including lncRNAs that act as scaffolds, decoys or signals and can act through genomic targeting, regulation in *cis* or *trans*, and antisense interference (reviewed in REFS 13,14). lncRNAs can be generally divided into three broad categories on the basis of their roles: non-functional lncRNAs that are likely to be the result of transcriptional noise; lncRNAs for which the act of transcription alone is sufficient for their function but the transcript itself is not necessary; and functional lncRNAs that can act in *cis* and/or in *trans* (reviewed in REFS 13,15).

In many instances, mRNAs and lncRNAs are more alike than they are different in terms of their biogenesis and form. Despite these categorical similarities, there are many features of individual lncRNAs or classes of lncRNAs that distinguish them from mRNAs. Indeed, features of numerous lncRNAs are not found in mRNAs (for example, *cis*-regulatory capacity, lack of robustly translated open reading frames (ORFs), special 3'-terminal processing, templating of nucleic acid polymerization or assembly, and others). In this Review, we discuss only those features of lncRNAs that are unique relative to protein-coding genes, primarily focusing on human and mammalian lncRNAs, although lncRNAs in yeast, viruses and other species are also discussed. Throughout this Review, we guide readers to additional reviews that describe in greater detail certain aspects of lncRNA biology or features that are not unique to lncRNAs. We first define lncRNAs and then describe noteworthy processes that many lncRNAs experience during their lifetime: from their transcriptional regulation, post-transcriptional processing and localization, to their eventual degradation (FIG. 1). Finally, we highlight unique characteristics of their functional roles in diverse biological phenomena.

<sup>1</sup>Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, California 94305, USA.

<sup>2</sup>Department of Bioengineering, Stanford University School of Medicine and School of Engineering, Stanford, California 94305, USA.

Correspondence to H.Y.C. [howchang@stanford.edu](mailto:howchang@stanford.edu)

doi:10.1038/nrg.2015.10

### The operational definition of lncRNAs

lncRNAs have a particular penchant for defying absolute classification, as any comprehensive definition of 'lncRNA' invariably engenders countless exceptions. Perhaps the most agreeable definition of lncRNAs is the tautological one: lncRNAs are long RNA transcripts that do not encode proteins. But even this seemingly simple definition is contentious and complex. First, lncRNAs are commonly defined as being longer than 200 nucleotides. However, this classification of 'long' is not a universal definition that is rooted in first principles, but is rather an arbitrary *ad hoc* cutoff, which primarily serves to partition well-known short ncRNAs (for example, tRNAs, microRNAs (miRNAs), small nucleolar RNAs (snoRNAs) and so on) from longer, more mRNA-like transcripts. Second, the distinction between protein-coding and non-coding transcripts has begun to blur upon closer inspection of the transcriptome and proteome, and this has been aided by deeper and more sensitive enumeration techniques (for example, mass spectrometry proteomics, deep RNA sequencing and ribosome profiling) showing that some lncRNAs contain cryptic ORFs<sup>16,17</sup>.

However, this operational definition of lncRNAs does have a purpose, insofar as comparing lncRNAs and protein-coding mRNAs unveils distinct characteristics. Indeed, many lncRNAs are very much like mRNAs: they are transcribed by RNA polymerase II (Pol II) from genomic loci with similar chromatin states to mRNAs<sup>6</sup>; they are often 5'-capped, spliced and polyadenylated; in most instances, they lack any biochemical distinction from mRNAs besides the absence of a translated ORF. But there are also general trends that discriminate lncRNAs from mRNAs: lncRNAs tend to be shorter than mRNAs, have fewer but longer exons, be expressed at relatively low levels and exhibit poorer primary sequence conservation<sup>3,18</sup>. Of course, exceptions to all of these trends abound. As a class, lncRNAs run the gamut from mRNA-like to truly exotic, such as chemically circular RNAs<sup>19</sup>, lncRNAs spanning 100 kb<sup>20</sup> and abundant lncRNAs with a restricted subnuclear localization<sup>21</sup>. Perhaps the sooner we dispense with categorical definitions of lncRNAs and recognize that they exist on multidimensional spectra of biogenesis, form and function, the sooner we can appreciate the enormous diversity of these genes.

### Regulation of lncRNA transcription

**The exceptional cell type and cell state specificity of lncRNA expression.** Pioneering studies of lncRNA expression in mouse brain revealed their precise expression patterns in specific tissues, cell types and subcellular compartments<sup>22</sup>. Later, transcriptome-wide studies showed that lncRNAs in general exhibit more specific expression profiles than mRNAs<sup>3,18</sup>; that is, they are expressed in a cell type-, tissue-, developmental stage- or disease state-specific manner (reviewed in REFS 10, 11). This trend is true even after correcting for the markedly lower expression levels of lncRNAs<sup>3,18</sup>. Furthermore, lncRNA expression patterns are often correlated with mRNA expression patterns both in *cis* and in *trans*, suggesting that certain lncRNAs may be co-regulated in expression networks<sup>6</sup>.

**Figure 1 | The busy lifetimes of certain lncRNAs differ from those of mRNAs — in birth, life and death.** Some long non-coding RNAs (lncRNAs) or classes of lncRNAs are regulated differentially at different points of their biogenesis, maturation and degradation. **a** | At the level of the chromatin state, lncRNAs and mRNAs exhibit similar properties, such as an enrichment of H3K4me3 at promoters; however, lncRNA genes have a higher enrichment of H3K27ac and are more strongly repressed by certain chromatin remodelling complexes, such as Swr1, Isw2, Rsc and Ino80. **b** | Transcriptional initiation from divergent promoters differs for the sense (mRNA) and the antisense (lncRNA) directions; divergent antisense transcription is enriched for H3K56ac and phosphorylation of RNA polymerase II (Pol II) Tyr1. Transcription in the divergent direction is further enhanced by the SWI/SNF proteins and repressed by CAF-1. **c** | Transcriptional elongation is more strongly regulated by DICER1 and MYC for lncRNAs than for mRNAs. **d** | The occurrence of U1 and polyadenylation signals differs on either side of bidirectional promoters (along the U1-PAS axis), favouring the splicing of mRNAs in the sense direction and the cleavage and polyadenylation in the divergent, antisense direction. **e** | Whereas mRNAs localize very specifically to ribosomes in the cytoplasm, lncRNA localization is much more varied, as certain lncRNAs can occupy the chromatin, subnuclear domains, the nucleoplasm or the cytoplasm. **f** | Finally, whereas mRNAs are primarily degraded in the cytoplasm by decapping and 5'-to-3' exonuclease digestion, many unstable lncRNA transcripts are subject to the nuclear exosome or to cytosolic nonsense-mediated decay (NMD). TSS, transcription start site.

This specificity has been used as evidence that lncRNA expression is even more tightly regulated than that of protein-coding genes, thereby arguing for the essential role of lncRNAs in determining cell state<sup>6</sup>. lncRNA promoters are approximately as evolutionarily conserved as mRNA promoters in humans<sup>3</sup> and mice<sup>4,6</sup>, further implying the importance of lncRNA expression programmes. But does this necessarily mean that they are under stricter regulation than protein-coding genes, or could the exceptionally specific expression patterns of lncRNAs be more or less a function of the cell state — the incidental by-product of the very specific chromatin states of particular physiological contexts? That is, cell state may establish a characteristic chromatin landscape, and therefore transcription occurs in regions that are accessible, thus resulting in a more-or-less distinct transcriptome. Alternatively, lncRNAs may cause or reinforce these specific cell states. Undoubtedly, both of these rules are likely to be true on a lncRNA-by-lncRNA basis. Nevertheless, lncRNA expression profiles are important markers for disease or developmental state<sup>23</sup>, such as diverse human cancers<sup>24–26</sup>, T cell differentiation<sup>27</sup> and development<sup>28</sup>.

**Chromatin effects on lncRNA expression.** From the perspective of chromatin state, lncRNAs seemingly follow the same rules as protein-coding genes<sup>3</sup>. That is to say, expressed lncRNA promoters are enriched for active histone modifications (for example, H3K4me3, H3K9ac and H3K27ac) similar to their protein-coding counterparts, and histone modification patterns can be used to

#### Biogenesis

The production by organisms of new biological material, such as RNAs, proteins or organelles.

#### Open reading frames

(ORFs). Continuous stretches of codons that have the potential to encode a protein.

#### MicroRNAs

(miRNAs). Short (~22 nt) non-coding RNAs that post-transcriptionally silence target RNAs by base pairing.

#### Small nucleolar RNAs

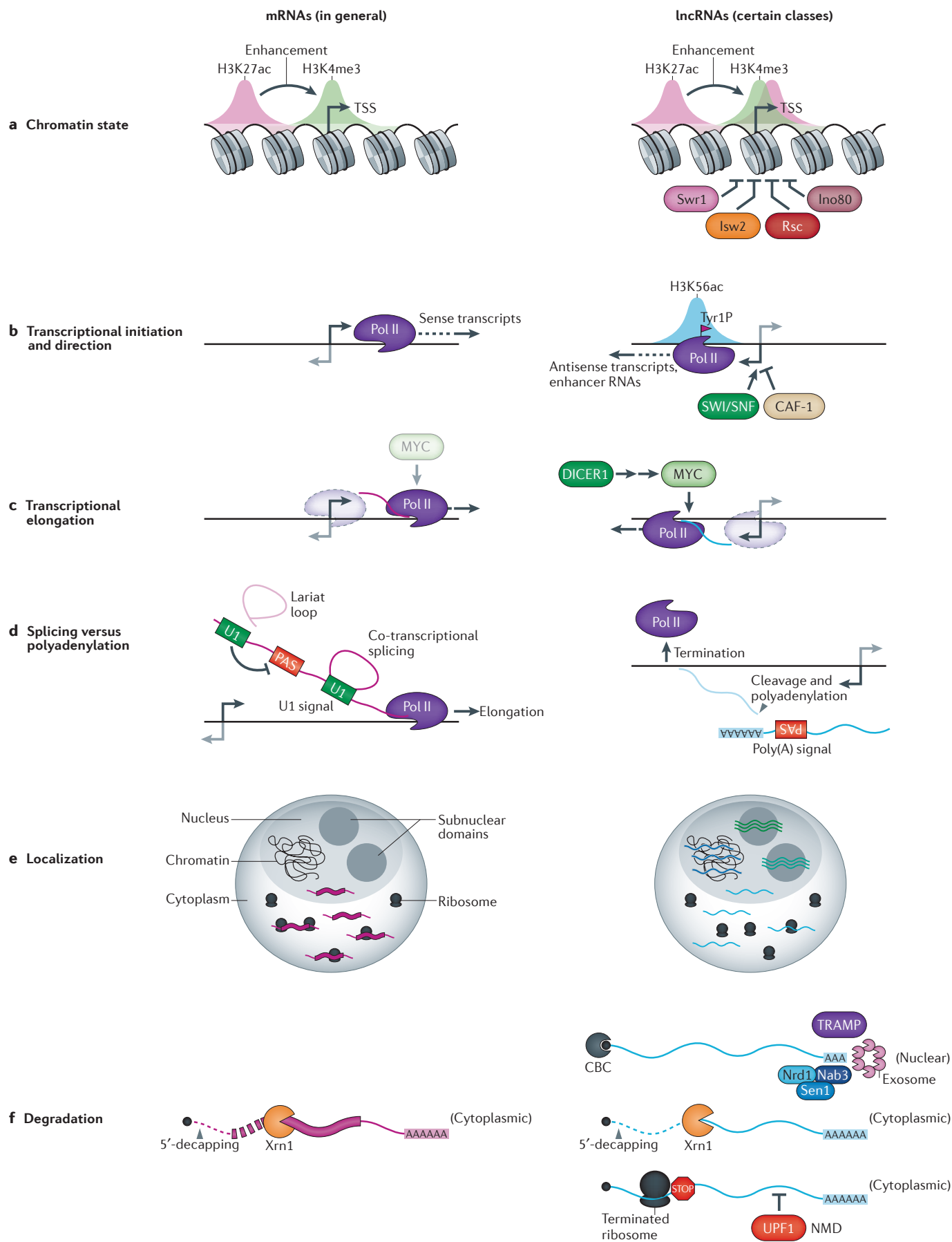
(snoRNAs). A class of small RNAs that guide the chemical modification of ribosomal RNAs, tRNAs and others.

#### Ribosome profiling

A biochemical technique for mapping ribosome-associated and translated RNAs.

#### Chromatin

The compartment of the nucleus that compacts and organizes genomic DNA and regulates gene expression.



identify lncRNAs<sup>3,6</sup> (FIG. 1a). However, recent work has revealed that certain transcription factors and chromatin remodelling enzymes globally regulate lncRNA expression (FIG. 1a–c).

For example, there are quantitative differences in the transcriptional regulation of lncRNAs and mRNAs, suggesting that many lncRNAs are regulated as a class<sup>29</sup>. Knocking out a gene responsible for generating miRNAs, *Dicer1*, in mouse embryonic stem cells resulted in the lower expression of hundreds of lncRNAs, particularly divergent transcripts<sup>29</sup>. Furthermore, the oncogenic transcription factor MYC was shown to be at least partly responsible for this differential expression. Specifically, the transcriptional elongation (and, to a lesser extent, initiation) of lncRNAs seems to be more sensitive to MYC dosage than are mRNAs as a class<sup>29</sup>. This result indicates that the miRNA circuitry and MYC are important for activating and sustaining lncRNA expression in a manner that is decoupled from mRNA regulation (FIG. 1c). Thus, DICER1 is an important factor in both the biogenesis of small ncRNAs and the downstream activation of hundreds of lncRNAs. The exact mechanism that is involved in regulating lncRNA expression is not fully known but probably involves an interplay with chromatin factors that also distinguish the lncRNA gene features outlined here.

In addition, from a genetic RNA interference screen in yeast, four mechanistically distinct chromatin remodelling complexes (Swr1, Isw2, Rsc and Ino80) were identified as global repressors of ncRNA transcription, particularly of lncRNAs that overlap protein-coding genes<sup>30</sup> (FIG. 1a). Disruption of these complexes led to the derepression of antisense lncRNAs and resulted in a marked decrease in levels of their overlapping mRNAs. Therefore, in yeast, these chromatin remodellers may repress such lncRNAs in order to activate the sense-strand mRNAs in *cis*. Other studies using different reporter-based screens in yeast also identified chromatin remodelling and nucleosome assembly factors as key regulators of intragenic cryptic unstable transcripts (CUTs) and divergent lncRNAs<sup>31,32</sup>.

**Divergent transcription.** Most eukaryotic promoters are bidirectional, such that initiating Pol II can generate a transcript in either direction: the sense (mRNA) direction or the upstream, antisense (uaRNA) direction (reviewed in REF. 33). Divergent transcripts account for a large proportion of observed lncRNAs, and are classified by their divergent transcription from shared protein-coding gene promoters. These mRNA–uaRNA pairs have coordinated expression, such that high expression of mRNAs also results in higher levels of the corresponding uaRNA<sup>34</sup>. In most instances, however, transcriptional elongation is only productive in the sense direction<sup>35–37</sup>.

It was recently revealed that the asymmetric distribution of polyadenylation and splicing signal sequences in the sense and antisense directions from a promoter dictates the marked difference in mRNA–uaRNA elongation and stability<sup>38</sup>. Specifically, polyadenylation signals (PASs) are enriched in the nearby antisense direction, whereas the U1 snRNP splicing signal is enriched in the nearby sense direction, thereby defining the U1–PAS

**Figure 2 | Post-transcriptional processing events in special lncRNA classes. a,b** | Many long non-coding RNAs (lncRNAs) undergo special processing events that have not been observed in mRNAs. For example, MALAT1 (metastasis-associated lung adenocarcinoma transcript 1) and NEAT1 (nuclear enriched abundant transcript 1) lncRNAs are processed at their 3' ends by RNase P, which generates tRNA-like small RNA products and the mature lncRNA, which possesses a stabilizing 3'-terminal RNA triplex structure; MALAT1 is localized to nuclear speckles and NEAT1 is localized to nuclear paraspeckles; the tRNA-like structures cleaved from MALAT1 (mascRNAs) are stable and cytoplasmic, whereas those from NEAT1 are unstable. **c** | Canonical splicing of mRNAs produces linear transcripts but back-splicing produces stable circular RNAs (circRNAs) consisting of non-sequential exon–exon junctions. **d** | Intronic lariats are typically unstable after splicing, but some escape debranching and degradation and persist as non-coding circular intronic long non-coding RNAs (ciRNAs). **e** | sno-lncRNAs are derived from the introns of small nucleolar RNA (snoRNA) host genes and are flanked by snoRNAs instead of 5' caps or poly(A)-tails. **f** | Whereas many microRNA (miRNA) genes are found within the introns of protein-coding genes (right), some lncRNAs host miRNA genes, which are processed by Microprocessor instead of the traditional cleavage and polyadenylation pathway (left). Pol II, RNA polymerase II.

axis of bidirectional transcription (FIG. 1d). This bias ensures the early and efficient termination and polyadenylation of antisense transcripts and, conversely, the productive splicing and elongation of sense transcripts. One hypothesis states that divergent transcripts exist as sites for the evolutionary innovation of new genes<sup>39</sup> or for *cis*-regulation of their promoters and corresponding protein-coding genes<sup>40</sup>.

Other features have been shown to differentially affect the expression and stability of divergent transcripts. To screen for possible regulators of sense versus antisense directionality, one group constructed a bidirectional fluorescent reporter and tested for biases in transcriptional direction in yeast<sup>32</sup>. This screen identified the chromatin assembly factor complex CAF-1 as a key genome-wide repressor of divergent transcription; conversely, divergent transcription is enhanced by H3K56ac and the chromatin remodeller SWI/SNF in a nucleosome-recycling-dependent manner (FIG. 1b). Additionally, two groups found that the Pol II carboxy-terminal domain was enriched for a specific phosphorylation mark (Tyr1P) at both uaRNAs and enhancers compared with Pol II at protein-coding genes, further suggesting that these ncRNAs are under the control of distinct transcriptional regulation<sup>41,42</sup>.

### Post-transcriptional processing of lncRNAs

To reach their mature forms, nascent RNA transcripts undergo extensive co-transcriptional and post-transcriptional processing events, such as 5'-capping, splicing, polyadenylation and chemical base modification. During this impressionable period in the life of an RNA, some lncRNAs experience alternative forms of processing that distinguish them from other transcripts (FIG. 2) (reviewed in REFS 43,44).

#### Divergent transcription

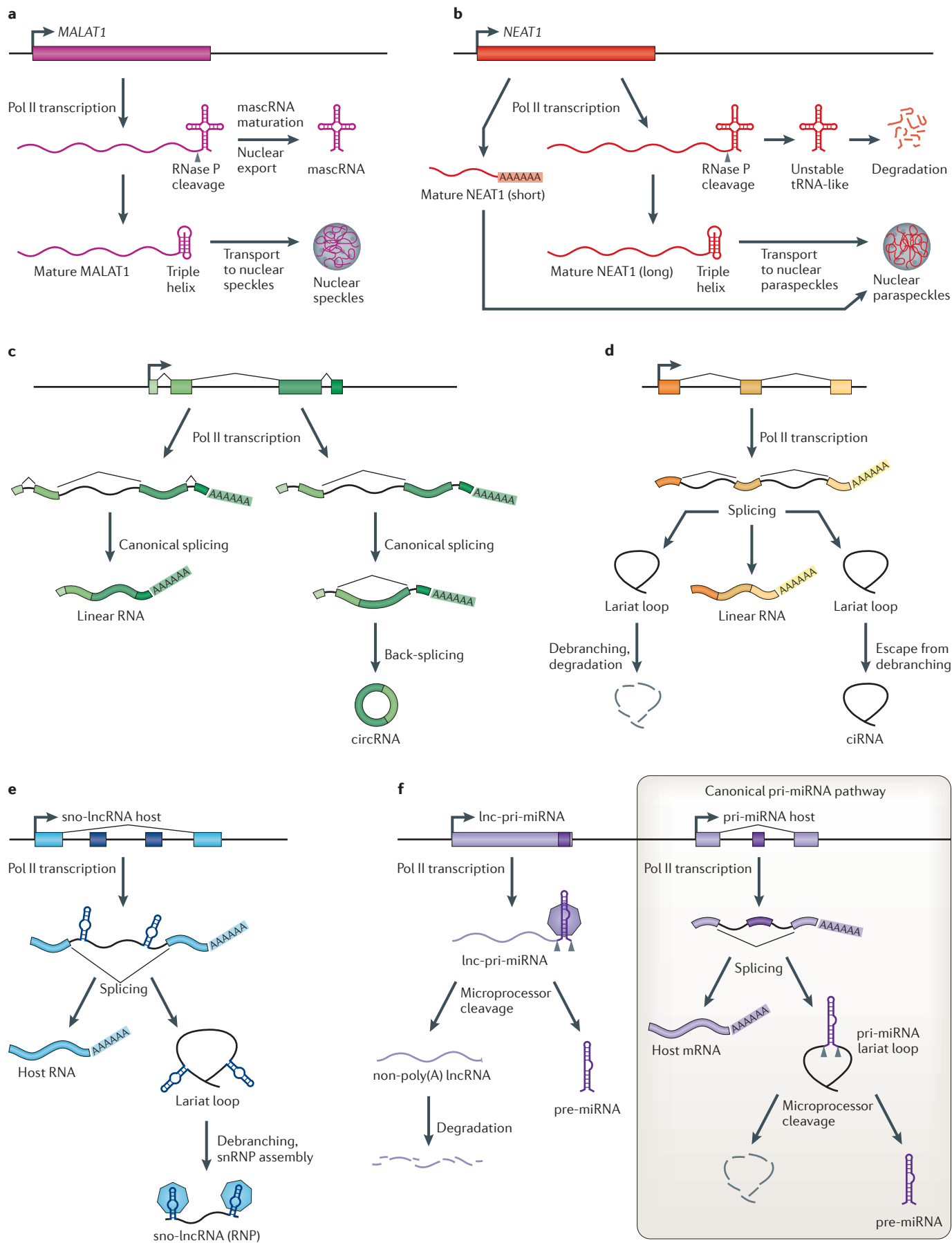
The production of sense and antisense RNAs from bidirectional promoters.

#### U1 snRNP

A ribonucleoprotein complex that coordinates the precise splicing of nascent RNA transcripts.

#### Enhancers

DNA-encoded elements that activate the expression of nearby genes.



**RNase P-processed 3' maturation.** Mature tRNAs are trimmed from pre-tRNAs by the RNase P ribonucleoprotein complex, which itself contains a catalytic ncRNA. In this way, RNase P cleavage is an alternative method to the canonical cleavage and polyadenylation that is used to process the 3' ends for the vast majority of mRNAs and most lncRNAs<sup>44</sup>. A well-known exception to this rule are histone-encoding mRNAs, which lack poly(A)-tails and instead are stabilized by a stem-loop in their 3' untranslated region (UTR)<sup>45</sup>. In addition to tRNAs, some lncRNAs have been identified as substrates for RNase P cleavage<sup>46</sup>, including metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) and nuclear enriched abundant transcript 1 (NEAT1) in mammals (FIG. 2a,b).

MALAT1 and NEAT1 are both highly abundant nuclear lncRNAs that are expressed in many mammalian cell types. MALAT1 and NEAT1 are both restricted to the nucleus where they localize to nuclear speckles and paraspeckles, respectively, and they both share two similar structural elements at their 3' termini, a tRNA-like cloverleaf and A/U-rich tracts that form a highly stable triple helix. These focal elements are conserved across vertebrate evolution at the level of the primary sequence and secondary structure, but the remaining majority of MALAT1 and NEAT1 sequences are not well conserved<sup>47</sup>. RNase P cleaves the tRNA-like structures from MALAT1 and NEAT1, resulting in mature lncRNA transcripts. The MALAT1-derived products, known as MALAT1-associated small cytoplasmic RNAs (mascRNAs), resemble tRNAs and are exported to the cytoplasm, where they may exert their function<sup>48</sup> (FIG. 2a). After the cleavage of the mascRNA subunit, the U-A-U RNA triple helix at the 3' end of mature MALAT1 increases the stability of the transcript — much like a poly(A)-tail — and thus enabling its function in regulating alternative splicing<sup>49–51</sup>. NEAT1, however, is transcribed into two isoforms: a short form that has a canonical poly(A)-tail and a long unspliced form (22.7 kb); only the long form lacks polyadenylation and is instead processed by RNase P<sup>49,52</sup>. Like MALAT1, cleavage of NEAT1 by RNase P results in an RNA-stabilizing triple helix at its 3' terminus and a small tRNA-like by-product<sup>21</sup>. Unlike mascRNAs, the resulting tRNA-like small RNAs are unstable (FIG. 2b). NEAT1 is important for paraspeckle formation, and *Neat1*-knockout mice have impaired lactation and pregnancy<sup>53–55</sup>.

Intriguingly, such 3'-terminal RNA triple helices are not exclusive to mammalian lncRNAs; indeed, similar structures were first identified in viral lncRNAs, such as the expression and nuclear retention element (ENE) of the polyadenylated nuclear (PAN) lncRNA that is expressed by Kaposi's sarcoma-associated herpesvirus (KSHV)<sup>56</sup>. Similar ENE-like elements have been reported in unrelated viral genomes, suggesting that these structures may have widespread application in stabilizing RNA species<sup>57–59</sup>. Despite the structural similarity to those in MALAT1 and NEAT1, viral ENE triple helices are not formed by RNase P processing.

**Other unique examples of lncRNA processing.** Circular RNAs (circRNAs) represent a class of lncRNAs that have

special 5'- and 3'-end processing. One class of circRNAs is formed through the non-sequential splicing of introns (known as back-splicing); this results in a chemically circularized transcript in which 3' sequences are spliced upstream of 5' sequences<sup>19</sup> (FIG. 2c). Such RNA circularization is a regulated process; for example, during the epithelial–mesenchymal transition in humans, circRNA formation is regulated by the alternative splicing factor Quaking (QKI)<sup>182</sup>. These circRNAs are not known to be translated into proteins<sup>60,61</sup>, although engineered circRNAs containing internal ribosome entry sites (IRESs) can be translated<sup>62,63</sup>; circRNAs may seem exotic, however, they have been detected in all domains of life<sup>64</sup>. Some contain miRNA-binding sites and may act as miRNA sponges<sup>19,60</sup>, such as the abundant, cytoplasmic circular non-coding RNA *CDRIAs*, which contains more than 70 miRNA-binding sites<sup>65</sup>. Another class of circRNAs is formed after canonical intron splicing. Splicing typically results in an unstable lariat loop with a 2',5'-phosphodiester linkage; however, circular intronic long non-coding RNAs (ciRNAs) escape classical lariat loop-debranching and degradation and are thus stable by-products of splicing. They may have regulatory roles within the nucleus, particularly at their site of transcription<sup>66</sup> (FIG. 2d). Exon-intron circRNAs (EIciRNAs) are yet another recently described class of circular RNAs, which are back-spliced circRNAs that retain unspliced introns; EIciRNAs may promote the transcription of their parental genes through interaction with Pol II and U1 snRNP<sup>67</sup>.

Some introns can also give rise to another unique form of stable non-coding transcript. snoRNAs are most often encoded within the introns of some protein-coding genes, from which they are trimmed. However, when tandem snoRNAs are encoded within a single intron, trimming can result in a sno-lncRNA, which consists of an intronic lncRNA flanked by two snoRNAs and thereby lacks a 5'-cap or poly(A)-tail<sup>68</sup> (FIG. 2e). A region of chromosome 15 that is specifically deleted in patients with Prader–Willi Syndrome (PWS) encodes a cluster of such sno-lncRNAs<sup>68,69</sup>, implicating them in the molecular pathogenesis of PWS and suggesting that sno-lncRNAs are more than a simple quirk of RNA processing. Similarly, many miRNA genes are harboured within the introns of protein-coding and poorly characterized ncRNA genes<sup>183</sup>, but relatively few are derived from lncRNA exons<sup>70</sup>. For the few lncRNAs that contain miRNAs, the canonical polyadenylation pathway for transcriptional termination and 3' maturation is not used. Instead, Microprocessor — a protein complex that processes miRNAs — cleaves the nascent transcript to terminate transcription in a polyadenylation-independent manner, thereby producing lnc-pri-miRNAs, which are 3'-capped by the Microprocessor complex<sup>70</sup> (FIG. 2f). lnc-pri-miRNAs are further processed into miRNAs and unstable, non-polyadenylated lncRNAs<sup>70</sup>. Post-transcriptional cleavage of other RNAs also generates a diverse set of long and short ncRNAs, such as 3' UTR-associated non-coding RNAs that are processed from the 3' UTRs of mRNAs<sup>71</sup>, or other ncRNAs that are generated by recursive cleavage and 5'-capping of mature mRNAs<sup>72</sup>.

**RNase P**

A ribonucleoprotein complex that cleaves tRNA precursors during their maturation.

**Secondary structure**

The base-pairing interactions that dictate nucleic acid folds.

**Circular RNAs**

(circRNAs). Chemically circular RNAs produced by nonsequential exon–exon back-splicing.

**Circular intronic long non-coding RNAs**

(ciRNAs). Circular, branched intronic RNAs resulting from stabilized introns after canonical splicing.

**Exon-intron circRNAs**

(EIciRNAs). A class of circular RNAs that retain unspliced introns.

**sno-lncRNA**

Long non-coding RNAs that are capped on the 5' and 3' ends by processed small nucleolar RNAs.

lncRNAs transcribed from enhancers are rarely polyadenylated but are instead cleaved and terminated by the Integrator complex<sup>73</sup>.

### Localization of lncRNAs

Many lncRNA species have very well defined subcellular localizations (FIG. 1e), including XIST (on the inactive X), Gomafu (also known as MIAT; subnuclear domains)<sup>74</sup>, BORG (restricted to the nucleus)<sup>75</sup> and GAS5 (exported to the cytoplasm)<sup>76</sup>. More generally, the trend of lncRNA localization relative to mRNA localization is discussed below.

**The nuclear-versus-cytoplasmic debate.** The subcellular localization of lncRNAs has recently been the subject of heated debate, primarily focused on nuclear-versus-cytosolic localization. On interpreting the subtext of this debate, it appears that arguments on either side are predicated on one of two seemingly opposing models: that lncRNAs are functional gene products with predominantly nuclear or chromatin-templating roles; or that lncRNAs are predominantly the product of ‘transcriptional noise’, can in fact engage the ribosome and are therefore not truly non-coding. Neither of these two models is categorically correct on a lncRNA-by-lncRNA basis, nor are they mutually exclusive. There are likely to be many examples of lncRNAs that are misclassified and actually encode functionally relevant proteins<sup>77,78</sup>. Conversely, there is abundant evidence for lncRNAs and even mRNAs acting as important regulators of chromatin state and chromosome conformation<sup>79–81</sup>. Further blurring the boundary between these two arguments are the numerous examples of bifunctional RNAs (reviewed in REF. 82). Indeed, some lncRNAs have established cytoplasmic functions, such as terminal tissue differentiation-inducing ncRNA (TINCR)<sup>83</sup>, competing endogenous RNAs (ceRNAs; although not all are non-coding; reviewed in REF. 84) and BACE1-AS<sup>85</sup>.

When compared with mRNAs, lncRNAs are more enriched in the nucleus relative to the cytoplasm, and within the nucleus they occupy the chromatin fraction (17% of lncRNAs versus 15% of mRNAs are enriched in the nucleus, whereas 4% versus 26%, respectively, are enriched in the cytoplasm)<sup>3</sup>. This observation agrees with a report that lncRNA read density is higher in the nuclear fraction<sup>1</sup>, and further supports the theory that many lncRNAs are engaged in epigenetic regulation on the chromatin<sup>80,81</sup>. It is important to highlight the subtle distinction between relative and absolute enrichment in the nucleus versus the cytoplasm. Even though lncRNAs have a more nuclear-biased localization pattern than mRNAs, more lncRNAs by transcript number are present in the cytoplasm than in the nucleus<sup>13</sup>. Thus, although numerous well-characterized lncRNAs are restricted to and abundant in the nucleus, as a class lncRNAs (and RNAs in general) are more abundant in the cytoplasm.

Conflicting reports are still common, and the debate continues. In one recent study, RNAs were fractionated into ribosome-bound cytosolic, free cytosolic and nuclear pools and then sequenced<sup>86</sup>. This

comprehensive analysis of lncRNA–ribosome interactions found that lncRNAs are present in every subcellular compartment but are particularly enriched in the cytoplasm, with a minority having nuclear enrichment<sup>86</sup>. In the largest scale study of lncRNA localization to date, RNA fluorescence *in situ* hybridization (RNA FISH) of 61 lncRNAs showed at single-cell and single-molecule resolution that many lncRNAs are located in the nucleus<sup>87</sup>. From this report, it was also clear that lncRNAs exhibit many different and interesting nuclear patterns ranging from defined subnuclear points and nuclear retention to diffuse whole-cell spread<sup>87</sup>. It is clear that lncRNAs do not categorically occupy one particular locale; rather, lncRNAs are ubiquitous. So, when discussing the unique localization of RNA species, it is perhaps more appropriate to turn the localization argument on its head: mRNAs are truly special RNAs, in that they possess the unique qualities of strong cytoplasmic enrichment, specific localization to the ribosome and productive translation (FIG. 1e).

**Ribosome association of lncRNAs.** Results from initial ribosome-profiling experiments that analysed the repertoire of transcripts that are engaged with the ribosome have challenged the mounting evidence for nuclear-biased lncRNA localization and have questioned their non-coding classification<sup>88</sup>. In this experiment, ribosome-associated RNAs were profiled in mouse embryonic stem cells and, surprisingly, thousands of annotated lncRNAs were identified as ribosome-bound. This finding immediately roused the suspicion that non-coding RNAs may be translated. Further analysis showed that the majority of annotated lncRNAs contain highly translated ORFs that are bound by elongating ribosomes, including about 50% of the candidate lncRNAs required for pluripotency<sup>89</sup>. This work (perhaps prematurely) reclassified these lncRNAs as short polycistronic ribosome-associated RNAs, though still leaving thousands of ‘true’ lncRNAs that did not seem to be translated (for example, NEAT1). From this experiment alone, it was unclear whether the translated products were functional. Perhaps this nearly ubiquitous RNA–ribosome association is not altogether surprising, considering that numerous ncRNAs are involved in regulating ribosomal processes and that the ribosome itself is composed of the most abundant cellular ncRNAs<sup>90</sup>.

Following this work, proteomic mass spectrometry was used to search for the translated protein products of these supposed lncRNA ORFs<sup>16</sup>. However, sensitive mass spectrometry failed to find any such products, suggesting that lncRNAs are rarely translated into proteins; specifically, approximately 92% of GENCODE lncRNAs are not translated in two human cell lines<sup>16</sup>. Other protein mass spectrometry-based analyses have identified polypeptide products from translation on classically defined lncRNAs (for example, H19) but, more commonly, protein products were not detected from non-coding transcripts<sup>17</sup>. A conclusion from these studies is that ribosomes can exquisitely discriminate between coding and non-coding transcripts. This hypothesis is further supported by a comprehensive reanalysis of the original

Bifunctional RNAs  
RNAs with separable  
non-coding and protein-coding  
functions.

ribosome-profiling experiments, whereby a stringent metric was devised for discriminating between productive and unproductive ribosome-association (that is, translating versus scanning ribosomes), known as the ribosome release score<sup>91</sup>. This metric depends on the release and active disassembly of translating ribosomes when they encounter true stop codons. Importantly, this metric accurately distinguishes between canonical ncRNAs (for example, snoRNAs, TERC and RNase P RNA) and protein-coding RNAs<sup>91</sup>. Furthermore, the ribosome occupancy on lncRNAs (both classical and broadly defined) resembled the scanning ribosome profile on 5' UTRs, and stop codons on lncRNAs were not characterized by ribosome release as with bona fide coding ORFs. Altogether, this result reverses the primary conclusions from the initial ribosome-profiling study<sup>88</sup>, thereby suggesting that most lncRNAs do not encode proteins.

Whether ribosome association and cryptic translation events on lncRNAs serve some functional role remains to be categorically proven — although, ironically, sceptics of lncRNA biology have long made a very similar argument regarding categorical proof of lncRNA function<sup>86,88</sup>. Nevertheless, we now appreciate that ORFs come in more flavours than the canonical start-to-stop codon variety: there are short and polycistronic upstream ORFs that can regulate the translation of the primary downstream ORF<sup>92</sup>, truly tiny ORFs that because of their diminutive size would otherwise be disregarded as not encoding bona fide functional proteins (for example, polished rice<sup>78</sup> and myoregulin micropeptide<sup>77</sup>), ORFs with non-canonical start or stop codons<sup>93</sup>, and bifunctional mRNAs that exhibit separate coding and non-coding functions<sup>82</sup>. All of these examples exist in a complex milieu of promiscuous RNA-ribosome association on a spectrum from translationally productive to translationally inert<sup>91</sup>.

### lncRNA degradation

A wide survey of 800 lncRNAs in mice showed that lncRNAs and mRNAs have comparable stability, although on average the half-lives of lncRNAs are slightly shorter than those of mRNAs<sup>94</sup>; this trend in stability further emphasizes the general similarity between lncRNAs and mRNAs. Nonetheless, several degradation pathways preferentially act on lncRNAs over mRNAs, perhaps to limit the number of transcripts pervasively produced from the genome<sup>33</sup>. Thus, the turnover of lncRNAs has a major role in shaping their repertoire and apparent expression pattern. Several degradation mechanisms, including classical nonsense-mediated decay (NMD), seem to affect lncRNAs distinctly from mRNAs (FIG. 1f).

**Cryptic unstable transcripts.** Eukaryotic genomes are pervasively transcribed, and many of the RNA products are short-lived non-coding RNAs, often referred to as cryptic unstable transcripts (CUTs) in yeast (reviewed in REFS 2,95). Owing to the inherent transience of these transcripts they have been notoriously difficult to detect in wild-type contexts; however, CUTs are abundant after genetic ablation of RNA quality-control pathways,

such as polyadenylation polymerases and nuclear RNases. For example, genetic ablation of exosome subunits in yeast leads to the accumulation of thousands of CUTs<sup>96</sup>. Further characterization revealed that CUTs are degraded through a variety of mechanisms, including digestion by the nuclear exosome, the cytoplasmic decapping complex (Dcp1–Dcp2) and subsequent 5'-to-3' exonuclear degradation by Xrn1, or by nonsense-mediated decay<sup>97,98</sup>. The nuclear exosome is recruited to CUTs by early transcriptional termination through the Nrd1–Nab3–Sen1 complex and the alternative poly(A)-polymerase complex TRAMP, which adds short poly(A)-tails to the ends of CUTs as a degradation tag<sup>99–102</sup>.

These RNAs are hypothesized to be entirely spurious — ‘transcriptional noise’ — and the numerous post-transcriptional quality-control mechanisms that have evolved limit their expression<sup>102</sup>. Perhaps they are the result of trial and error transcription, or are a side effect of changing or sustained chromatin landscapes<sup>2</sup>. Nonetheless, there are some CUTs that have established functions, such as one antisense CUT at the *PHO84* gene that silences genes in *cis* via histone deacetylation<sup>103</sup>, and a class of CUTs that silence Ty1 retrotransposon transcription in yeast<sup>104</sup>.

Whereas CUTs are only appreciably detected in nuclear exosome mutants, stable uncharacterized transcripts (SUTs) can be found in wild-type yeast cells because they escape the nuclear exosome and the Nrd1–Nab3–Sen1 pathway<sup>105</sup>. Instead, SUTs are decapped by the decapping complex and degraded by Xrn1 in the cytoplasm, in a similar way to mRNAs<sup>105</sup>. By some estimates, SUTs account for 12% of identifiable RNA species in wild-type yeast<sup>106</sup>. In addition, Xrn1-sensitive unstable transcripts (XUTs) are similar to CUTs, but are more often antisense to protein-coding genes and are longer<sup>107</sup>. CUTs, SUTs and XUTs have alternative (NRD-dependent) 3' termination, and Pcf1 is required for NRD-dependent termination, thus mutation of Pcf1 delays the degradation of these ncRNAs<sup>108</sup>.

Although CUTs have been predominantly described in yeast, ncRNAs with similar properties have been identified in other organisms, such as upstream antisense RNAs (uaRNAs), transcription start site-associated and promoter-associated ncRNAs, and promoter upstream transcripts (PROMPTs) in humans, which are also rapidly degraded by the exosome<sup>37</sup>. The poly(A)-binding protein nuclear 1 (PABPN1) promotes RNA turnover, and genetic deficiency of PABPN1 leads to the accumulation of many lncRNAs, presumably by escaping poly(A)-dependent degradation by the exosome<sup>109,110</sup>. Many of these lncRNAs are in the same category as divergent transcripts, which are enriched for early polyadenylation on the U1–PAS axis<sup>38</sup>.

**Nonsense-mediated decay.** Computational analysis of mammalian lncRNA sequences indicates that lncRNAs most often contain weak ORFs that would activate NMD pathways, and frequently have similar structural and sequence characteristics to the 3' UTRs of mRNAs<sup>111</sup>. As many cytoplasmic lncRNAs outwardly

**Nonsense-mediated decay (NMD).** A genetic pathway for the elimination of RNAs that are defective in protein coding, primarily owing to premature stop codons.

**Exosome**  
A large protein complex that degrades RNA.

**Upstream antisense RNAs (uaRNAs).** Non-coding RNAs transcribed in the opposite direction from promoters by divergent transcription.



appear very similar to mRNAs, with the key exception of lacking an ORF, it seems likely that the NMD would survey lncRNAs, sample their coding potential (or lack thereof) and destroy them. Indeed, the cytoplasmic lncRNA GAS5 is sensitive to the NMD protein UPF1, suggesting that at least some lncRNAs are subject to degradation through NMD<sup>112</sup>. Mutating key proteins in the NMD pathway in *Arabidopsis thaliana* resulted in the global upregulation of both protein-coding genes and mRNA-like lncRNAs but with significantly greater effects on lncRNAs relative to protein-coding genes<sup>113</sup>. This suggests that NMD has a particularly strong role in dampening the expression of ncRNAs, a hypothesis that is consistent with results suggesting that NMD primarily serves to mute transcriptional noise in mammalian cells<sup>114</sup>. NMD has also been implicated as an RNA quality surveillance mechanism for the regulation of non-coding pseudogenes in *Caenorhabditis elegans*<sup>115</sup>. Estimates of the percentage of non-coding transcripts that undergo NMD in diverse organisms (including yeast, *A. thaliana*, zebrafish and humans) are low but are higher than the percentage of coding transcripts (approximately 4–14% versus 0.3–13%, respectively, depending on the species)<sup>116</sup>.

### Cis-regulatory circuits of lncRNAs

Some protein-coding genes exhibit autoregulation — that is, a protein may return to and regulate its own genomic locus (for example, transcription factors such as OCT4). This protein-enacted autoregulatory loop must necessarily follow a circuitous route from transcription, nuclear mRNA export, translation and protein maturation to nuclear import, diffusion to its original genomic locus and finally action at *cis*-regulatory elements. Thus, the fact that proteins must be translated in a different cytological compartment expressly prohibits their ability to truly self-regulate their own transcription in *cis*. In contrast to this paradigm, lncRNAs are bestowed with the unique capacity for *cis*-regulatory action (reviewed in REF. 40). lncRNAs are physically linked to the locus from which they are encoded, thus they may exert their function immediately following — or even during — their transcription without the need for processing or shuttling. Examples of *cis*-acting lncRNAs include enhancer RNAs (eRNAs), lncRNAs from imprinted loci, dosage compensation lncRNAs, antisense RNAs and autoregulatory RNAs (FIG. 3).

**Enhancer RNAs and chromosomal looping.** Enhancers are *cis*-encoded DNA elements that tightly regulate genes within their own chromosomal neighbourhood (reviewed in REF. 117). Enhancers are abundant and particularly essential to developmental patterning and cellular identity. For some years it has been appreciated that enhancers are transcriptionally active, producing a rich range of bidirectional, Pol II-transcribed ncRNAs that are known as enhancer RNAs (eRNAs)<sup>118,119</sup> (reviewed in REF. 120). eRNA expression is often correlated with the mRNAs that are targeted by the enhancer, and enhancers and target promoters are physically associated<sup>118,119</sup>, and in some cases the eRNA

transcript is required for target gene regulation<sup>121</sup>, suggesting that enhancer activity may depend on these *cis*-encoded ncRNAs.

Anterior–posterior positional identity during animal development is determined by sets of transcription factor clusters at the *HOX* loci; in mammals, these *HOX* clusters are also hotbeds of enhancer-associated lncRNAs that regulate the *HOX* genes, such as HOTAIR and *HOXA* distal transcript antisense RNA (HOTTIP)<sup>122,123</sup>. HOTTIP is encoded at the 5' tip of the *HOXA* locus, where it facilitates chromosome looping and is brought into proximity to distant *HOXA* genes. Mechanistically, HOTTIP recruits WDR5 and the MLL complex, which in turn deposits the H3K4me3 mark and activates these *HOXA* genes<sup>123</sup>. The expression of HOTTIP has been implicated in the progression of several carcinomas<sup>124</sup>. This phenomenon of a lncRNA inducing chromosome looping to regulate nearby protein-coding genes is widespread. For example, CCAT1-L regulates chromosome looping at the *MYC* locus and is upregulated in human colorectal cancers<sup>125</sup>. LUNAR1 is a T cell acute lymphoblastic leukaemia (T-ALL)-specific lncRNA that is transcribed from the insulin-like growth factor 1 receptor (*IGF1R*) locus and that exhibits pro-oncogenic characteristics, such as stimulating T-ALL cell growth. LUNAR1 itself activates the *IGF1R* locus in *cis* via chromosome looping, thus leading to sustained IGF1 signalling in T-ALL cells<sup>126</sup> (FIG. 3a).

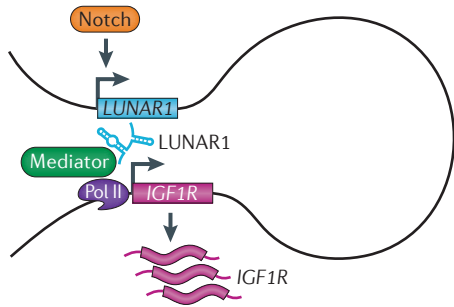
HOTTIP, CCAT1-L and LUNAR1 may loosely fit into a larger class of lncRNAs known as activating ncRNAs (ncRNA-a) that mediate DNA looping and chromatin enhancement via Mediator, a large transcriptional co-activating complex<sup>126,127</sup>. A close investigation of two such eRNAs, ncRNA-a3 and ncRNA-a7, showed that they recruit and activate Mediator as a bridge between enhancers and their target genes<sup>127</sup>. In this study, knocking down Mediator components attenuated the activity of the eRNAs; conversely, depletion of the ncRNA-a led to the decreased recruitment of Mediator and Pol II to the target genes, indicating that enhancement is dependent on ncRNA. Using multiple biochemical techniques, this study showed that Mediator subunits contact the eRNAs, and that these contacts are essential for chromosomal looping between the enhancer and the target gene, as demonstrated by chromosome conformation capture<sup>127</sup>. In this way, such enhancer-associated lncRNAs may facilitate enhancement or may act as the functional output of enhancers for establishing chromosomal looping and subsequent *cis* regulation. In a recent study in murine cortical neurons, it was shown that some eRNAs function as decoys for the negative elongation factor (NELF), thus competitively derepressing paused Pol II and enabling productive elongation of the target RNA<sup>128</sup>; similarly, eRNAs may also 'trap' certain RNA-binding transcription factors at enhancers, thereby sustaining transcription factor-mediated regulation<sup>184</sup>.

In both mammals and flies, the Polycomb group (PcG) and Trithorax group (TrxG) proteins modify the chromatin landscape of numerous developmental genes by acting through nearby enhancers known as PcG response elements (PREs) and TrxG response elements

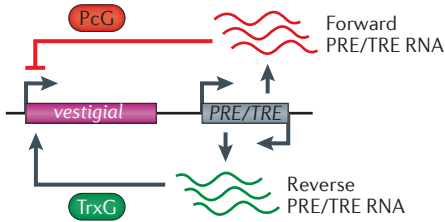
**Enhancer RNAs**  
(eRNAs). Non-coding RNAs transcribed from enhancers.

**Dosage compensation**  
The epigenetic process of balancing gene expression from sex chromosomes between males and females.

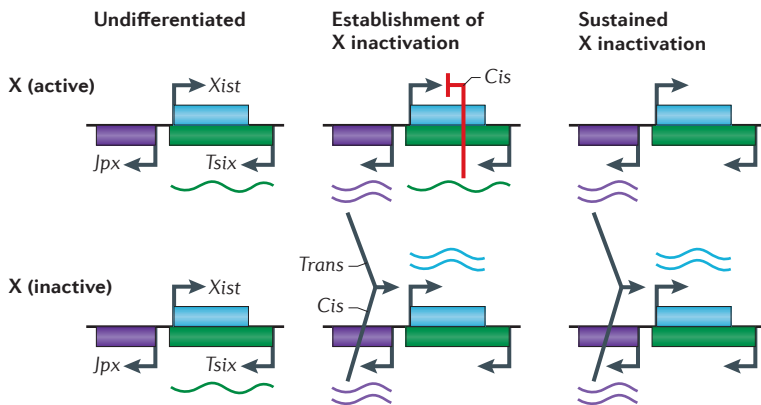
**a Enhancer RNAs and chromosome looping**



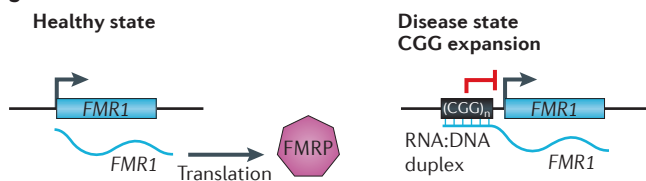
**b PRE/TRE enhancer RNA switching**



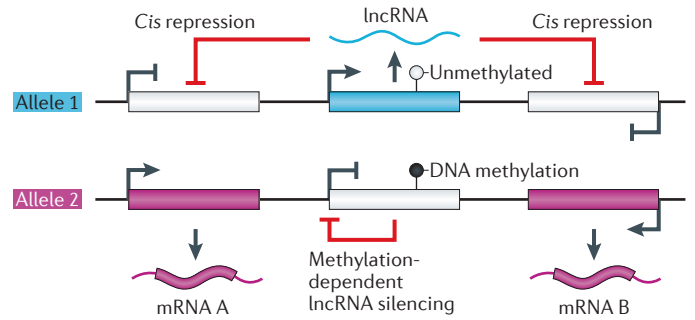
**d Mammalian dosage compensation**



**g FMR1 auto-inhibition**

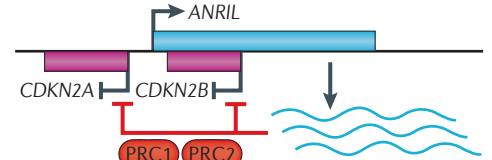


**c Imprinted gene clusters**

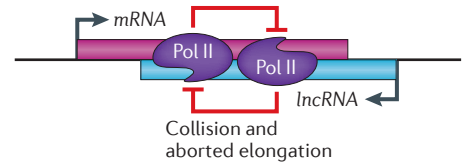


lncRNA	Expression	mRNAs	Expression	Refs
Air	Paternal	<i>Igf2r</i> , <i>Slc22a2</i> and <i>Slc22a3</i>	Maternal	136
Kcnq1ot1	Paternal	<i>Kcnq1</i> , <i>Cdkn1c</i> , <i>Ascl2</i> and others	Maternal	139
Nespas	Paternal	<i>Gnas</i> and <i>Nesp</i>	Maternal	140
H19	Maternal	<i>Igf2</i>	Paternal	132
Gtl2	Maternal	<i>Dlk1</i> and <i>Rtl1</i>	Paternal	141

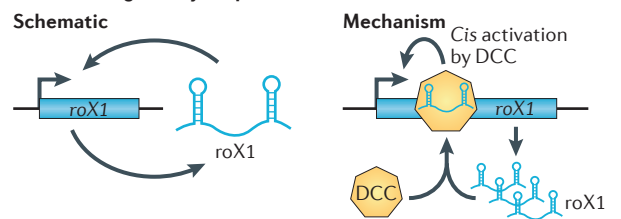
**e Antisense inhibition by ANRIL**



**f Antisense inhibition by Pol II collision**



**h roX1 autoregulatory loop**



**Figure 3 | Cis-regulatory mechanisms of lncRNA function. a** | Long non-coding RNAs (lncRNAs) are uniquely poised to regulate their genomic neighbourhoods in cis. Some enhancer RNAs, such as LUNAR1 near the insulin-like growth factor 1 receptor (*IGF1R*) locus, mediate chromosome looping between enhancers and nearby target genes via Mediator or MLL protein complexes. **b** | PcG response element/TrxG response element (PRE/TRE) enhancer RNAs can switch between silencing and activating states by switching bidirectional transcription; forward transcription of one such PRE/TRE represses *vestigial* expression via Polycomb group (PcG), whereas transcription in the reverse direction activates *vestigial* expression via Trithorax group (TrxG). **c** | Allele-specific DNA methylation at imprinted genomic loci silences the expression of lncRNAs within the imprinted gene cluster, thereby allowing neighbouring protein-coding genes to be expressed; conversely, on the other allele the lncRNA is expressed in the

absence of DNA methylation, thereby repressing protein-coding genes in cis. **d** | The mammalian dosage compensation lncRNA, *Xist*, is silenced on the active X chromosome in cis by the antisense lncRNA *Tsix*; meanwhile, *Xist* is activated on the inactive X chromosome in cis and in trans by the lncRNA *Jpx*. **e** | ANRIL antisense lncRNA represses the cyclin-dependent kinase inhibitor 2A (*CDKN2A*)–*CDKN2B* locus in cis by recruiting PRC1 and PRC2. **f** | When protein-coding genes and antisense lncRNA genes overlap, processing RNA polymerase II (Pol II) particles may collide and thus abort transcription, effectively inhibiting the expression of both genes. **g** | *FMR1* (fragile X mental retardation 1) binds and silences its own promoter via RNA–DNA hybrids at CGG repeat expansions that are characteristic of disease. **h** | *roX1* in male *Drosophila* spp. autoregulates its own locus and sustains its own transcription by recruiting the activating dosage compensation complex (DCC).

(TREs). PREs and TREs are known to reversibly switch between repressive and activating epigenetic states and, similar to other enhancers, they are sites of extensive non-coding transcription<sup>120,129</sup>. It was recently shown that this switch is controlled by lncRNAs that are encoded by opposite strands of the PRE/TRE<sup>130</sup>. Close investigation of a PRE/TRE at the *vestigial* locus in *Drosophila melanogaster* revealed that the PRE/TRE expresses tissue-specific and mutually exclusive bidirectional lncRNAs that were positively or negatively correlated with *vestigial* expression (FIG. 3b). The ‘forward’ PRE/TRE lncRNA recruits PcG through pairing-sensitive silencing with related PRE/TREs, thereby repressing *vestigial*. On the other side, the ‘reverse’ PRE/TRE lncRNA directly inhibits E(Z) enzymatic activity, evicts PcG and diminishes the repressive H3K27me3 mark, thereby activating *vestigial*. Thus, transcription in the forward or reverse PRE/TRE orientations produces distinct lncRNAs with opposing functions. Furthermore, there is evidence that such PRE/TRE lncRNA switches are widespread throughout both the fly and mouse genomes<sup>130</sup>.

***lncRNAs at imprinted loci.*** Some genes are epigenetically imprinted, meaning that they are expressed in a mutually exclusive, parent-of-origin, allele-specific pattern (for example, the maternally inherited allele is expressed while the paternally inherited allele is repressed (reviewed in REF. 131)). It has long been appreciated that imprinted gene clusters are sites of rich non-coding transcription; in fact, the imprinted gene *H19* was among the first described lncRNA loci<sup>132</sup>. We now understand that this is no coincidence, as lncRNAs are inherently essential to the allele-specific expression that is observed at imprinted loci (reviewed in REFS 133,134). The reason for this lncRNA dependence is, by all accounts, the innate competence of lncRNAs for *cis* action, and imprinting must categorically be carried out in *cis* because the two alleles exist in the same nuclear environment but are subject to different epigenetic conditions. A well-established principle governing imprinting is DNA methylation: the methylated allele is established during gender-specific gametogenesis and silences the transcription of nearby imprinted lncRNA genes and thereby activate target mRNA genes within the imprinted locus (FIG. 3c). Numerous studies have shown that imprinted lncRNAs themselves (rather than just their transcription) are required for imprinting<sup>135,136</sup>.

A prime example of imprinted lncRNA–mRNA gene clusters is the paternally expressed lncRNA *Air*, which is encoded antisense to the insulin-like growth factor 2 receptor (*Igf2r*) gene in a cluster of imprinted, maternally expressed protein-coding genes (*Igf2r*, *Slc22a2* and *Slc22a3*)<sup>136</sup> in mice and in other mammals. The promoter of *Air* is located within a CpG island that is hypermethylated on the maternal allele. Hypermethylation suppresses the transcription of *Air* from the maternal locus, thereby allowing the expression of flanking protein-coding genes. The paternal allele, conversely, is not methylated, such that *Air* is paternally transcribed and thus silences the imprinted locus<sup>136</sup>. This lncRNA is unusually long, resulting in a mature transcript of

108 kb in mice<sup>20</sup>. Loss of the differentially methylated CpG island results in the unregulated expression of *Air*, indicating the dependence of *Air* on the imprinting status of the locus<sup>137</sup>. Truncation of the *Air* transcript to 3 kb by the insertion of a polyadenylation site results in the paternal reactivation (and therefore biallelic expression) of protein-coding genes in the *Igf2r* cluster, indicating that the full-length *Air* lncRNA is necessary for imprinting<sup>136</sup>. *Air* may also silence *Igf2r* and flanking genes through antisense transcriptional interference, a *cis*-regulatory mechanism that is unique to antisense lncRNAs<sup>138</sup>. Other examples of imprinted lncRNA loci include *Kcnq1ot1* at the *Kcnq1* locus, *Nespas* at the *Gnas–Nesp* gene cluster and *H19* at the *Igf2* locus<sup>132,139–141</sup>.

***lncRNAs in dosage compensation.*** Dosage compensation — that is, sex chromosome-wide regulation — also shares some phenomenological commonalities with imprinting, in that two alleles in the same nuclear context must be differentially regulated. In dosage compensation, gene expression from sex chromosomes is modulated by a lncRNA-dependent epigenetic process; for example, in female eutherian mammals (XX), one of the two X chromosomes is epigenetically silenced by the XIST lncRNA, thereby equalizing their output with that of the single X in males (XY)<sup>142</sup> (FIG. 3d). Curiously, many evolutionarily diverse dosage-compensation mechanisms use *cis*-encoded lncRNAs (for example, beyond XIST in eutherian mammals, *Rsx* is used in metatherian mammals and *roX* is used in *Drosophila* spp.)<sup>142–144</sup>. Perhaps dosage compensation and epigenetic imprinting both require *cis*-encoded lncRNAs for the same mechanistic reasons. Further accentuating the comparison between these two phenomena is *Rsx*, which differs from XIST in that it does not randomly inactivate one of the two X chromosomes, but rather it always inactivates the paternally inherited X chromosome in female marsupials<sup>144,145</sup>, in a mechanism that is essentially chromosome-wide epigenetic imprinting. Interestingly, XIST only acts in *cis*, as placing the XIST gene on an autosome causes ectopic autosomal inactivation<sup>146</sup>, illustrating how chromosomal context can be key to lncRNA functions.

***Repression by antisense lncRNA transcription.*** When antisense lncRNAs overlap with a protein-coding gene, antisense transcription can interfere with the sense transcription of the protein-coding gene through direct Pol II transcriptional interference in *cis* (reviewed in REF. 147). *Cis*-regulatory antisense lncRNAs are common and act in myriad ways<sup>40</sup>. For example, the above-mentioned *Air* antisense lncRNA silences the overlapping sense gene *Igf2r* by interfering with the recruitment of the transcriptional machinery to the *Igf2r* promoter<sup>138</sup>. This form of repression is thought to occur when Pol II complexes that are transcribing in the sense and antisense directions collide with one another (FIG. 3f). Other examples include the lncRNA ANRIL recruiting PcG repressors to its gene cluster, which contains the tumour suppressor genes cyclin-dependent kinase inhibitor 2A (*CDKN2A*) and *CDNK2B*<sup>148</sup> (FIG. 3e); the antisense lncRNA SCAANT1 repressing the *ATXN7* gene in *cis*

#### DNA methylation

The chemical modification of cytosine residues on DNA that can be epigenetically inherited.

#### CpG island

DNA sites of the sequence CG where the cytosine may be methylated in a regulated manner.

via CTCF interaction and chromatin remodelling<sup>149</sup>; the yeast lncRNA *IME4-antisense* repressing *IME4* through transcriptional interference<sup>150</sup>; and the antisense lncRNAs *COOLAIR* and *COLDAIR* controlling the flowering locus (*FLC*) in plants by recruiting PcG-repressive complexes<sup>151,152</sup>. *Xist* lncRNA, which coordinates mammalian X inactivation (as discussed above), is itself regulated by other lncRNAs also encoded within the X inactivation centre (XIC), including the divergent lncRNA *Jpx* and the antisense lncRNA *Tsix*<sup>153</sup> (reviewed in REF. 142). *Xist* and *Tsix* exhibit mutually exclusive expression: *Xist* is expressed from the inactive X, whereas *Tsix* is expressed from the active X during the onset of X inactivation; meanwhile, *Jpx* transactivates *Xist* in *cis* and in *trans*<sup>153</sup> (FIG. 3d). *Tsix* inhibits *Xist* activity in *cis* via altering its chromatin state, modulating DNA methylation and interfering with its transcription<sup>154</sup>.

In addition to these *cis* activities, antisense lncRNAs may also transactivate sense mRNAs in a sense-antisense hybridization-dependent manner, as in the cases of ZEB2-AS (which mediates the retention of an intron of ZEB2 that is necessary for its translation and involvement in epithelial-to-mesenchymal transition)<sup>155</sup> and the Alzheimer's disease-associated antisense transcript BACE1-AS (which blocks miRNA recognition and stabilizes the *BACE1* mRNA)<sup>85</sup>.

**Autoregulatory lncRNAs.** Biological systems commonly use autoregulation as a strategy to produce stably biphasic states: 'off' and 'on'. In the same way that many protein-coding genes often produce proteins that regulate their own transcription, autoregulation has also been shown to be a regulatory mechanism used by lncRNAs. In several cases, disrupted autoregulatory loops can give rise to pathological conditions, as in the cases of *FMRI* (fragile X mental retardation 1), *DHFR* and *DMPK*, which are RNAs with separable coding and non-coding roles (bifunctional RNAs) that repress their own loci (reviewed in REF. 82).

Although *FMRI* encodes a protein that is essential to normal cognitive development, the *FMRI* transcript engages in a remarkably lncRNA-like process that is entirely separable from its coding role. Using chromatin isolation by RNA purification (ChIRP) of the *FMRI* RNA, it was discovered that nascent *FMRI* mRNA forms an RNA-DNA duplex at a key CGG trinucleotide repeat at its promoter, thereby driving epigenetic silencing of its own locus<sup>79</sup> (FIG. 3g). Expansion of these CGG repeats is observed in nearly all cases of fragile X mental retardation, implying a molecular role for this aberrant interaction between *FMRI* RNA and its promoter<sup>156</sup>. This is an example of RNA-mediated *cis* autoregulation, in which an RNA inhibits its own transcription. Importantly, this nuclear role for the *FMRI* transcript precludes its nuclear export and translation in the cytoplasm, thus *FMRI* sometimes does not encode a protein and may therefore be considered a conditional lncRNA (or bifunctional RNA, more strictly speaking<sup>82</sup>). Also in this category of *cis*-autoregulatory bifunctional RNAs are *DHFR* non-coding, encoded within the 5' UTR of *DHFR*<sup>157</sup>; *DHFR* non-coding represses the *DHFR*

locus by inhibiting the Pol II pre-initiation complex via RNA-DNA triplex formation at the *DHFR* promoter<sup>158</sup>. Similarly, *DMPK* mRNA can bind to CTG-repeat expansion at its own locus in a pathologically relevant manner similar to *FMRI* and *DHFR*<sup>159</sup>.

In *D. melanogaster* dosage compensation, the single X chromosome of males is epigenetically upregulated — exactly the opposite strategy to that used in the mammalian system (reviewed in REF. 143). However, similar to mammalian dosage compensation, flies use X-encoded lncRNAs, roX1 and roX2 (RNAs on X). The roX lncRNAs coordinate the assembly of the dosage compensation ribonucleoprotein complex and target it to hundreds of specific sites on the male X. In addition to these sites, roX1 lncRNA directly binds to the *roX1* locus, thereby recruiting the dosage compensation machinery to its own locus<sup>160</sup> (FIG. 3h). Activation of *roX1* late in development requires pre-existing expression of either roX RNA; thereafter, *roX1* self-regulates in a positive feedback loop that results in sustained roX1 expression<sup>161</sup>.

### Structure-encoded functions of lncRNAs

**Conservation of structure.** Further distinguishing lncRNAs from protein-coding genes are their patterns of evolutionary conservation. Indeed, lncRNAs exhibit exceptionally poor conservation at the level of the primary nucleotide sequence, especially relative to protein-coding sequences, although lncRNAs are more conserved than neutrally evolving genetic elements<sup>3,18,162</sup>. This low evolutionary conservation probably arises from the fact that non-coding sequences are constrained around different parameters from those of their protein-coding counterparts. Rather than maintaining ORF register or optimal codon usage, lncRNAs are more often conserved along dimensions of genomic position (synteny), short sequence motifs or secondary structure<sup>13,47,163,164</sup>. Although RNA structure is an important feature of both protein-coding and non-coding RNAs<sup>165</sup>, the secondary structure of a lncRNA is occasionally better conserved than its primary sequence, a feature that has been exploited to identify orthologous lncRNAs by searching for covariant sequences in putative secondary structures<sup>166</sup>. Recent estimates of evolutionary conservation found that up to 14% of the human genome shows evidence of purifying selection on RNA structure<sup>167</sup>, suggesting that much of the non-coding human genome encodes function at the level of the RNA structure, although some argue that these estimates are overrepresented by false positives<sup>168</sup>.

**Secondary structures and functional domains.** Various functional lncRNAs are organized into modular functional domains that are capable of combinatorially coordinating RNA-RNA, RNA-protein and RNA-DNA interactions, similar to how proteins are organized into functional subunits, but instead encoded at the RNA level (reviewed in REF. 169). Evidence of this domain-level strategy is widespread. For example, the *HOX*-encoded HOTAIR lncRNA can tether together multiple histone-modifying complexes through modular RNA domains<sup>170,171</sup>. Functional dissection of the roX1

**Chromatin isolation by RNA purification (ChIRP).** A biochemical technique for mapping the genomic binding sites of an RNA of interest.

**Synteny**  
The physical proximity of genes within a single genomic locus.

**CRISPRi and CRISPRa**

Sequence-specific interference (i) or activation (a) of gene expression using CRISPR technology. Typically, this involves a nuclease-deficient Cas9-mutant protein fused to transcriptional activator or repressor proteins; this fusion protein is then directed to a target genomic locus by an engineered guide RNA.

**Box 1 | The challenges of studying lncRNA functions**

Currently, we know of fewer essential long non-coding RNAs (lncRNAs) than essential protein-coding genes. There are several possible explanations for this disparity: the lncRNA field is relatively young; there are fewer RNA-centric tools for querying molecular functions and mechanisms; and perhaps there are simply fewer fundamentally essential lncRNAs. Several lncRNAs are known to have organismal functions, including XIST, TERC, NEAT1, roX, COOLAIR, HOTAIR and many others. New lncRNA functions are reported frequently, and the list of functional lncRNAs — essential or otherwise — continues to grow. A screen of 18 lncRNA knockouts in mice identified three lethal knockouts (*Fendrr*, *Peril* and *Mdgt*) and two more with growth defects (*linc-Brn1b* and *linc-Pint*), as further evidence of lncRNAs with fundamental organismal function<sup>178</sup>.

Conclusive evidence of the functional effects of lncRNAs is hampered by difficulties in study design. Deletions that remove a lncRNA gene may also alter other transcripts or DNA-encoded features within the locus (this is especially true for divergent transcripts, antisense RNAs and enhancer RNAs, but less so for intergenic lncRNAs). Additionally, unlike protein-coding genes, lncRNAs are not inactivated by introducing premature stop codons or frameshift mutations. The current gold standards for disrupting lncRNA functions are antisense oligonucleotides that can deplete lncRNAs and methods for genetically inactivating lncRNA transcription, such as premature polyadenylation. New strategies based on CRISPRi and CRISPRa to modulate lncRNA transcription are promising for dissecting lncRNA functions<sup>179,180</sup>.

The time is ripe for systematic studies of unique lncRNA biogenesis towards understanding their function. The biogenesis of small RNA species has been known for a long time (for example, microRNAs (miRNAs), small nucleolar RNAs (snoRNAs) and tRNAs), and evidence for lncRNAs is being discovered piecemeal. The current assumption is that most lncRNAs (or at least those that are stable at steady-state) are similar to mRNAs, but closer examination may reveal otherwise. Our knowledge of the overall importance of miRNAs was not achieved by only knocking out one miRNA at a time (which is complicated by the fact that individual miRNAs are commonly redundant with one another), but rather by knocking out the biogenesis machinery. Similarly, experimental disruption of other biogenesis factors has elucidated new lncRNAs and lncRNA classes. This approach may prove more powerful than individual lncRNA knockout experiments, in part because functional redundancy between lncRNAs may also be common. Functional redundancy is appreciated in some lncRNA systems, such as the roX lncRNAs (roX1 and roX2) in *Drosophila melanogaster* dosage compensation; they were discovered by screening for genes with male-specific biogenesis pathways<sup>181</sup>. As such, studying unique forms of lncRNA biogenesis will probably aid in their continued functional characterization.

lncRNA in *D. melanogaster* revealed that its functions are encoded in repeated stem-loop structures that are organized into three distinct and functionally redundant RNA domains<sup>160,172</sup>. Additionally, 7SK is a highly abundant nuclear ncRNA that serves as a structural assembly scaffold and regulatory switch for the positive transcriptional elongation factor p-TEFb, which is sequestered and inactivated by 7SK; a conformational switch in 7SK secondary structures releases p-TEFb, thereby enhancing Pol II elongation of mRNAs genome-wide<sup>173</sup>. Such functional structures of RNAs can be profiled using high-throughput analyses of RNA-protein interaction and RNA structure<sup>174,175</sup>.

**lncRNAs as nucleic acid templates**

The biochemical roles of proteins are unrivalled in their diversity by any other biomolecule. However, there are some functions that simply cannot be carried out by a protein (or, at least, no such proteins have yet been discovered). TERC is one such example, in that it serves as a template for RNA-dependent DNA polymerase activity (reverse transcription) carried out by the enzyme TERT (telomerase reverse transcriptase) to elongate telomeres, thus combatting the shortening of chromosome ends that inevitably results from DNA replication<sup>176</sup>.

In another example, certain ciliated protists such as *Oxytricha trifallax* undergo complex genome rearrangements during somatic macronucleus development, during which the germline micronuclear DNA is fragmented, ~95% of the germline genome is eliminated, and finally a new genome is reconstructed. Interestingly, this reconstruction depends on lncRNA transcripts that unscramble DNA fragments into the correct order and

orientation; thus, lncRNAs can even template the programmed wholesale reassembly of genomic DNA in a heritable, non-Mendelian manner<sup>177</sup>.

**Conclusions**

RNA has myriad roles in the cell — from controlling the architecture of whole chromosomes and acting as a regulator of gene expression to templating the translation of genetic codons into protein sequences. lncRNAs are a diverse class of molecules defined by a lack of protein-coding potential (often determined computationally), and we now appreciate several molecular features that distinguish certain lncRNAs or lncRNA classes from other genes. These lncRNA features include unique regulatory mechanisms, alternative forms of biogenesis, *cis*-regulatory activities and functional structured RNA domains. With current deep RNA-sequencing and advanced epigenomic technologies, the rate of discovering new lncRNA genes is rapidly outpacing the rate of characterizing them. This gap between lncRNA discovery and lncRNA characterization is widened by numerous experimental challenges in studying lncRNA genes relative to protein-coding genes (BOX 1). Despite these hurdles, the continuous development of a toolkit for studying lncRNAs (for example, ChIRP, crosslinking immunoprecipitation (CLIP), ribosome profiling, RNA structure mapping, phylogenetic lineage tracing, targeted genome engineering by CRISPR and advanced genetic screens) has yielded a greater appreciation for their varied and essential roles in biological phenomena. Moving forward, with these and other future techniques we will surely uncover even more fascinating and unique features and functions of lncRNAs.

1. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
2. Berretta, J. & Morillon, A. Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Rep.* **10**, 973–982 (2009).
3. 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).
- Comprehensive comparative analysis between mRNAs and lncRNAs in humans.**
4. Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).
5. Mattick, J. S., Taft, R. J. & Faulkner, G. J. A global view of genomic information — moving beyond the gene and the master regulator. *Trends Genet.* **26**, 21–28 (2010).
6. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
7. Clark, M. B. *et al.* Quantitative gene profiling of long noncoding RNAs with targeted RNA sequencing. *Nat. Methods* **12**, 339–342 (2015).
8. Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166 (2012).
9. Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **136**, 629–641 (2009).
10. Flynn, R. A. & Chang, H. Y. Long noncoding RNAs in cell-fate programming and reprogramming. *Cell Stem Cell* **14**, 752–761 (2014).
11. Batista, P. J. & Chang, H. Y. Long noncoding RNAs: cellular address codes in development and disease. *Cell* **152**, 1298–1307 (2013).
12. Esteller, M. Non-coding RNAs in human disease. *Nat. Rev. Genet.* **12**, 861–874 (2011).
13. Ulitsky, I. & Bartel, D. P. lincRNAs: genomics, evolution, and mechanisms. *Cell* **154**, 26–46 (2013).
14. Geisler, S. & Coller, J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat. Rev. Mol. Cell Biol.* **14**, 699–712 (2013).
15. Kornienko, A. E., Guenzl, P. M., Barlow, D. P. & Pauler, F. M. Gene regulation by the act of long non-coding RNA transcription. *BMC Biol.* **11**, 59 (2013).
16. Banfai, B. *et al.* Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res.* **22**, 1646–1657 (2012).
17. Gascoigne, D. K. *et al.* Pinstripe: a suite of programs for integrating transcriptomic and proteomic datasets identifies novel proteins and improves differentiation of protein-coding and non-coding genes. *Bioinformatics* **28**, 3042–3050 (2012).
18. 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).
19. Salzman, J., Gawad, C., Wang, P. L., Lacayo, N. & Brown, P. O. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS ONE* **7**, e30733 (2012).
20. Lyle, R. *et al.* The imprinted antisense RNA at the *Igf2r* locus overlaps but does not imprint *Mas1*. *Nat. Genet.* **25**, 19–21 (2000).
21. Hutchinson, J. N. *et al.* A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* **8**, 39 (2007).
22. Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F. & Mattick, J. S. Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl Acad. Sci. USA* **105**, 716–721 (2008).
23. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
24. Brunner, A. L. *et al.* Transcriptional profiling of long non-coding RNAs and novel transcribed regions across a diverse panel of archived human cancers. *Genome Biol.* **13**, R75 (2012).
25. Gupta, R. A. *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010).
26. Lee, C. S. *et al.* Transcriptome sequencing in Sezary syndrome identifies Sezary cell and mycosis fungoides-associated lincRNAs and novel transcripts. *Blood* **120**, 3288–3297 (2012).
27. Hu, G. *et al.* Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. *Nat. Immunol.* **14**, 1190–1198 (2013).
28. Amaral, P. P. & Mattick, J. S. Noncoding RNA in development. *Mamm. Genome* **19**, 454–492 (2008).
29. Zheng, G. X., Do, B. T., Webster, D. E., Khavari, P. A. & Chang, H. Y. Dicer-microRNA-Myc circuit promotes transcription of hundreds of long noncoding RNAs. *Nat. Struct. Mol. Biol.* **21**, 585–590 (2014).
- Demonstrates that lncRNAs can be regulated as a class differently from protein-coding genes.**
30. Alcidi, E. A. & Tsukiyama, T. ATP-dependent chromatin remodeling shapes the long noncoding RNA landscape. *Genes Dev.* **28**, 2348–2360 (2014).
31. Cheung, V. *et al.* Chromatin- and transcription-related factors repress transcription from within coding regions throughout the *Saccharomyces cerevisiae* genome. *PLoS Biol.* **6**, e2777 (2008).
32. Marquardt, S. *et al.* A chromatin-based mechanism for limiting divergent noncoding transcription. *Cell* **157**, 1712–1723 (2014).
33. Grzechnik, P., Tan-Wong, S. M. & Proudfoot, N. J. Terminate and make a loop: regulation of transcriptional directionality. *Trends Biochem. Sci.* **39**, 319–327 (2014).
34. Sigova, A. A. *et al.* Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc. Natl Acad. Sci. USA* **110**, 2876–2881 (2013).
35. Seila, A. C. *et al.* Divergent transcription from active promoters. *Science* **322**, 1849–1851 (2008).
36. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
37. Preker, R. *et al.* RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**, 1851–1854 (2008).
38. Almada, A. E., Wu, X., Kriz, A. J., Burge, C. B. & Sharp, P. A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**, 360–363 (2013).
39. Wu, X. & Sharp, P. A. Divergent transcription: a driving force for new gene origination? *Cell* **155**, 990–996 (2013).
40. Guil, S. & Esteller, M. *Cis*-acting noncoding RNAs: friends and foes. *Nat. Struct. Mol. Biol.* **19**, 1068–1075 (2012).
- A review of the cis-encoded functions of lncRNAs.**
41. Descostes, N. *et al.* Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *eLife* **3**, e02105 (2014).
42. Hsin, J. P., Li, W., Hoque, M., Tian, B. & Manley, J. L. RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells. *eLife* **3**, e02112 (2014).
43. Zhang, Y., Yang, L. & Chen, L. L. Life without a tail: new formats of long noncoding RNAs. *Int. J. Biochem. Cell Biol.* **54**, 338–349 (2014).
44. Wilusz, J. E. Long noncoding RNAs: re-writing dogmas of RNA processing and stability. *Biochim. Biophys. Acta* <http://dx.doi.org/10.1016/j.bbaggm.2015.06.003> (2015).
45. Williams, A. S. & Marzluff, W. F. The sequence of the stem and flanking sequences at the 3' end of histone mRNA are critical determinants for the binding of the stem-loop binding protein. *Nucleic Acids Res.* **23**, 654–662 (1995).
46. Marvin, M. C. *et al.* Accumulation of noncoding RNA due to an RNase P defect in *Saccharomyces cerevisiae*. *RNA* **17**, 1441–1450 (2011).
47. Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H. & Bartel, D. P. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147**, 1537–1550 (2011).
48. 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).
49. Wilusz, J. E. *et al.* A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev.* **26**, 2392–2407 (2012).
50. 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).
51. Brown, J. A., Valenstein, M. L., Yario, T. A., Tycowski, K. T. & Steitz, J. A. Formation of triple-helical structures by the 3'-end sequences of MALAT1 and MENβ noncoding RNAs. *Proc. Natl Acad. Sci. USA* **109**, 19202–19207 (2012).
52. Sunwoo, H. *et al.* MEN εβ nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res.* **19**, 347–359 (2009).
53. Sasaki, Y. T., Ideue, T., Sano, M., Mituyama, T. & Hirose, T. MENεβ noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc. Natl Acad. Sci. USA* **106**, 2525–2530 (2009).
54. Nakagawa, S. *et al.* The lncRNA *Neat1* is required for corpus luteum formation and the establishment of pregnancy in a subpopulation of mice. *Development* **141**, 4618–4627 (2014).
55. Standaert, L. *et al.* The long noncoding RNA *Neat1* is required for mammary gland development and lactation. *RNA* **20**, 1844–1849 (2014).
56. Tycowski, K. T., Shu, M. D., Borah, S., Shi, M. & Steitz, J. A. Conservation of a triple-helix-forming RNA stability element in noncoding and genomic RNAs of diverse viruses. *Cell Rep.* **2**, 26–32 (2012).
57. Brown, J. A. *et al.* Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nat. Struct. Mol. Biol.* **21**, 633–640 (2014).
58. Conrad, N. K. & Steitz, J. A. A Kaposi's sarcoma virus RNA element that increases the nuclear abundance of intronless transcripts. *EMBO J.* **24**, 1831–1841 (2005).
59. Mitton-Fry, R. M., DeGregorio, S. J., Wang, J., Steitz, T. A. & Steitz, J. A. Poly(A) tail recognition by a viral RNA element through assembly of a triple helix. *Science* **330**, 1244–1247 (2010).
60. Jeck, W. R. *et al.* Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**, 141–157 (2013).
61. Guo, J. U., Agarwal, V., Guo, H. & Bartel, D. P. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol.* **15**, 409 (2014).
62. Chen, C. Y. & Sarnow, P. Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science* **268**, 415–417 (1995).
63. Wang, Y. & Wang, Z. Efficient backsplicing produces translatable circular mRNAs. *RNA* **21**, 172–179 (2015).
64. Danan, M., Schwartz, S., Edelleit, S. & Sorek, R. Transcriptome-wide discovery of circular RNAs in *Archaea*. *Nucleic Acids Res.* **40**, 3131–3142 (2012).
65. Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338 (2013).
66. Zhang, Y. *et al.* Circular intronic long noncoding RNAs. *Mol. Cell* **51**, 792–806 (2013).
67. Li, Z. *et al.* Exon-intron circular RNAs regulate transcription in the nucleus. *Nat. Struct. Mol. Biol.* **22**, 256–264 (2015).
68. Yin, Q. F. *et al.* Long noncoding RNAs with snoRNA ends. *Mol. Cell* **48**, 219–230 (2012).
69. Cassidy, S. B., Schwartz, S., Miller, J. L. & Driscoll, D. J. Prader-Willi syndrome. *Genet. Med.* **14**, 10–26 (2012).
70. Dhir, A., Dhir, S., Proudfoot, N. J. & Jopling, C. L. Microprocessor mediates transcriptional termination of long noncoding RNA transcripts hosting microRNAs. *Nat. Struct. Mol. Biol.* **22**, 319–327 (2015).
71. Mercer, T. R. *et al.* Expression of distinct RNAs from 3' untranslated regions. *Nucleic Acids Res.* **39**, 2393–2403 (2011).
72. Affymetrix & 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).
73. Lai, F., Gardini, A., Zhang, A. & Shiekhattar, R. Integrator mediates the biogenesis of enhancer RNAs. *Nature* **525**, 399–403 (2015).
74. Tsuiji, H. *et al.* Competition between a noncoding exon and introns: Gomafu contains tandem UACUAC repeats and associates with splicing factor-1. *Genes Cells* **16**, 479–490 (2011).
75. Zhang, B. *et al.* A novel RNA motif mediates the strict nuclear localization of a long noncoding RNA. *Mol. Cell Biol.* **34**, 2318–2329 (2014).
76. Coccia, E. M. *et al.* Regulation and expression of a growth arrest-specific gene (*gas5*) during growth, differentiation, and development. *Mol. Cell Biol.* **12**, 3514–3521 (1992).
77. Anderson, D. M. *et al.* A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* **160**, 595–606 (2015).
- Functional characterization of a micropeptide encoded by a transcript previously annotated as a lncRNA, thus illustrating the rich diversity of ORFs and cautioning against computational definitions of lncRNAs.**

78. Kondo, T. *et al.* Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis. *Science* **329**, 336–339 (2010).
79. Colak, D. *et al.* Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. *Science* **343**, 1002–1005 (2014).
80. 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).
81. Mondal, T., Rasmussen, M., Pandey, G. K., Isaksson, A. & Kanduri, C. Characterization of the RNA content of chromatin. *Genome Res.* **20**, 899–907 (2010).
82. Ulveling, D., Francastel, C. & Hube, F. When one is better than two: RNA with dual functions. *Biochimie* **93**, 633–644 (2011).
83. Kretz, M. *et al.* Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* **493**, 231–235 (2013).
84. Tay, Y., Rinn, J. & Pandolfi, P. P. The multilayered complexity of ceRNA crosstalk and competition. *Nature* **505**, 344–352 (2014).
85. Faghihi, M. A. *et al.* Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biol.* **11**, R56 (2010).
86. van Heesch, S. *et al.* Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol.* **15**, R6 (2014).
87. Cabili, M. N. *et al.* Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* **16**, 20 (2015). **A wide survey of the diverse localization patterns of lncRNAs.**
88. 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).
89. Guttman, M. *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295–300 (2011).
90. Pircher, A., Bakowska-Zywicka, K., Schneider, L., Zywicki, M. & Polacek, N. An mRNA-derived noncoding RNA targets and regulates the ribosome. *Mol. Cell* **54**, 147–155 (2014).
91. Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S. & Lander, E. S. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* **154**, 240–251 (2013). **Reanalyses the original ribosome-profiling experiments to find that lncRNAs are not productively translated and lack the signature of ribosome release.**
92. Lee, S. *et al.* Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl Acad. Sci. USA* **109**, E2424–E2432 (2012).
93. Ivanov, I. P., Loughran, G. & Atkins, J. F. uORFs with unusual translational start codons autoregulate expression of eukaryotic ornithine decarboxylase homologs. *Proc. Natl Acad. Sci. USA* **105**, 10079–10084 (2008).
94. Clark, M. B. *et al.* Genome-wide analysis of long noncoding RNA stability. *Genome Res.* **22**, 885–898 (2012).
95. Houseley, J. Form and function of eukaryotic unstable non-coding RNAs. *Biochem. Soc. Trans.* **40**, 836–841 (2012).
96. Davis, C. A. & Ares, M. Jr. Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **103**, 3262–3267 (2006).
97. Houseley, J. & Tollervy, D. The many pathways of RNA degradation. *Cell* **136**, 763–776 (2009).
98. Thompson, D. M. & Parker, R. Cytoplasmic decay of intergenic transcripts in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **27**, 92–101 (2007).
99. Arigo, J. T., Eyster, D. E., Carroll, K. L. & Corden, J. L. Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol. Cell* **23**, 841–851 (2006).
100. LaCava, J. *et al.* RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**, 713–724 (2005).
101. Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J. & Libri, D. Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Mol. Cell* **23**, 853–864 (2006).
102. Wyers, F. *et al.* Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**, 725–737 (2005).
103. Cambong, J. *et al.* Trans-acting antisense RNAs mediate transcriptional gene co-suppression in *S. cerevisiae*. *Genes Dev.* **23**, 1534–1545 (2009).
104. Berretta, J., Pinskaya, M. & Morillon, A. A cryptic unstable transcript mediates transcriptional silencing of the Ty1 retrotransposon in *S. cerevisiae*. *Genes Dev.* **22**, 615–626 (2008).
105. Marquardt, S., Hazelbaker, D. Z. & Buratowski, S. Distinct RNA degradation pathways and 3' extensions of yeast non-coding RNA species. *Transcription* **2**, 145–154 (2011).
106. Xu, Z. *et al.* Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**, 1033–1037 (2009).
107. van Dijk, E. L. *et al.* XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* **475**, 114–117 (2011).
108. Grzechnik, P., Gdula, M. R. & Proudfoot, N. J. Pcf11 orchestrates transcription termination pathways in yeast. *Genes Dev.* **29**, 849–861 (2015).
109. Beaulieu, Y. B., Kleinman, C. L., Landry-Voyer, A. M., Majewski, J. & Bachand, F. Polyadenylation-dependent control of long noncoding RNA expression by the poly(A)-binding protein nuclear 1. *PLoS Genet.* **8**, e1003078 (2012).
110. Flynn, R. A., Almada, A. E., Zamudio, J. R. & Sharp, P. A. Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc. Natl Acad. Sci. USA* **108**, 10460–10465 (2011).
111. Niazi, F. & Valadkhan, S. Computational analysis of functional long noncoding RNAs reveals lack of peptide-coding capacity and parallels with 3' UTRs. *RNA* **18**, 825–843 (2012).
112. Tani, H., Torimura, M. & Akimitsu, N. The RNA degradation pathway regulates the function of GAS5 a non-coding RNA in mammalian cells. *PLoS ONE* **8**, e55684 (2013).
113. Kurihara, Y. *et al.* Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **106**, 2453–2458 (2009).
114. Mendell, J. T., Sharifi, N. A., Meyers, J. L., Martinez-Murillo, F. & Dietz, H. C. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutates genomic noise. *Nat. Genet.* **36**, 1075–1078 (2004).
115. Mitrovich, Q. M. & Anderson, P. mRNA surveillance of expressed pseudogenes in *C. elegans*. *Curr. Biol.* **15**, 963–967 (2005).
116. Ruiz-Orera, J., Messeguer, X., Subirana, J. A. & Alba, M. M. Long non-coding RNAs as a source of new peptides. *eLife* **5**, e03523 (2014).
117. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).
118. De Santa, F. *et al.* A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* **8**, e1000384 (2010).
119. Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
120. Natoli, G. & Andrau, J. C. Noncoding transcription at enhancers: general principles and functional models. *Annu. Rev. Genet.* **46**, 1–19 (2012).
121. Orom, U. A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46–58 (2010).
122. 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).
123. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
124. Quagliata, L. *et al.* Long noncoding RNA HOTTIP/HOXA13 expression is associated with disease progression and predicts outcome in hepatocellular carcinoma patients. *Hepatology* **59**, 911–923 (2014).
125. Xiang, J. F. *et al.* Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res.* **24**, 513–531 (2014).
126. Trimarchi, T. *et al.* Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. *Cell* **158**, 593–606 (2014).
127. Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**, 497–501 (2013).
128. Schaukowitz, K. *et al.* Enhancer RNA facilitates NELF release from immediate early genes. *Mol. Cell* **56**, 29–42 (2014).
129. Di Croce, L. & Helin, K. Transcriptional regulation by Polycomb group proteins. *Nat. Struct. Mol. Biol.* **20**, 1147–1155 (2013).
130. Herzog, V. A. *et al.* A strand-specific switch in noncoding transcription switches the function of a Polycomb/Trithorax response element. *Nat. Genet.* **46**, 973–981 (2014).
131. Barlow, D. P. & Bartolomei, M. S. Genomic imprinting in mammals. *Cold Spring Harb. Perspect. Biol.* **6**, a018382 (2014).
132. Zhang, Y. & Tycko, B. Monoallelic expression of the human *H19* gene. *Nat. Genet.* **1**, 40–44 (1992).
133. Adalsteinsson, B. T. & Ferguson-Smith, A. C. Epigenetic control of the genome—lessons from genomic imprinting. *Genes (Basel)* **5**, 635–655 (2014).
134. Royo, H. & Cavaille, J. Non-coding RNAs in imprinted gene clusters. *Biol. Cell* **100**, 149–166 (2008).
135. Mancini-Dinardo, D., Steele, S. J., Levorse, J. M., Ingram, R. S. & Tilghman, S. M. Elongation of the *Kcnq1ot1* transcript is required for genomic imprinting of neighboring genes. *Genes Dev.* **20**, 1268–1282 (2006).
136. Sleutels, F., Zwart, R. & Barlow, D. P. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **415**, 810–813 (2002).
137. Wutz, A. *et al.* Imprinted expression of the *Igf2r* gene depends on an intronic CpG island. *Nature* **389**, 745–749 (1997).
138. Latos, P. A. *et al.* *Airn* transcriptional overlap, but not its lncRNA products, induces imprinted *Igf2r* silencing. *Science* **338**, 1469–1472 (2012).
139. Pandey, R. R. *et al.* *Kcnq1ot1* antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* **32**, 232–246 (2008).
140. Williamson, C. M. *et al.* Uncoupling antisense-mediated silencing and DNA methylation in the imprinted *Gnas* cluster. *PLoS Genet.* **7**, e1001347 (2011).
141. Lin, S. P. *et al.* Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the *Dlk1-Gtl2* imprinted cluster on mouse chromosome 12. *Nat. Genet.* **35**, 97–102 (2003).
142. Lee, J. T. Epigenetic regulation by long noncoding RNAs. *Science* **338**, 1435–1439 (2012).
143. Conrad, T. & Akhtar, A. Dosage compensation in *Drosophila melanogaster*: epigenetic fine-tuning of chromosome-wide transcription. *Nat. Rev. Genet.* **13**, 123–134 (2011).
144. Grant, J. *et al.* Rxsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation. *Nature* **487**, 254–258 (2012).
145. Wang, X., Douglas, K. C., Vandenberg, J. L., Clark, A. G. & Samollow, P. B. Chromosome-wide profiling of X-chromosome inactivation and epigenetic states in fetal brain and placenta of the opossum, *Monodelphis domestica*. *Genome Res.* **24**, 70–83 (2014).
146. Lee, J. T. & Jaenisch, R. Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. *Nature* **386**, 275–279 (1997).
147. Pelechano, V. & Steinmetz, L. M. Gene regulation by antisense transcription. *Nat. Rev. Genet.* **14**, 880–893 (2013).
148. 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).
149. Sopher, B. L. *et al.* CTCF regulates ataxin-7 expression through promotion of a convergently transcribed, antisense noncoding RNA. *Neuron* **70**, 1071–1084 (2011).
150. Hongay, C. F., Grisafi, P. L., Galitski, T. & Fink, G. R. Antisense transcription controls cell fate in *Saccharomyces cerevisiae*. *Cell* **127**, 735–745 (2006).
151. Heo, J. B. & Sung, S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* **331**, 76–79 (2011).
152. Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N. J. & Dean, C. R-loop stabilization represses antisense transcription at the *Arabidopsis* FLC locus. *Science* **340**, 619–621 (2013).
153. Tian, D., Sun, S. & Lee, J. T. The long noncoding RNA, *Jpx*, is a molecular switch for X chromosome inactivation. *Cell* **143**, 390–403 (2010).
154. 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).

155. Beltran, M. *et al.* A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial–mesenchymal transition. *Genes Dev.* **22**, 756–769 (2008).
156. Oberle, I. *et al.* Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* **252**, 1097–1102 (1991).
157. Blume, S. W., Meng, Z., Shrestha, K., Snyder, R. C. & Emanuel, P. D. The 5'-untranslated RNA of the human dhfr minor transcript alters transcription pre-initiation complex assembly at the major (core) promoter. *J. Cell Biochem.* **88**, 165–180 (2003).
158. Martianov, I., Ramadass, A., Serra Barros, A., Chow, N. & Akoulitchev, A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* **445**, 666–670 (2007).
159. Jiang, H., Mankodi, A., Swanson, M. S., Moxley, R. T. & Thornton, C. A. Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Hum. Mol. Genet.* **13**, 3079–3088 (2004).
160. Quinn, J. J. *et al.* Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. *Nat. Biotechnol.* **32**, 933–940 (2014).
161. Lim, C. K. & Kelley, R. L. Autoregulation of the *Drosophila* noncoding *roX1* RNA gene. *PLoS Genet.* **8**, e1002564 (2012).
162. Ponjavic, J., Ponting, C. P. & Lunter, G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.* **17**, 556–565 (2007).
163. Diederichs, S. The four dimensions of noncoding RNA conservation. *Trends Genet.* **30**, 121–123 (2014).
164. Hezroni, H. *et al.* Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep.* **11**, 1110–1122 (2015).
165. Mortimer, S. A., Kidwell, M. A. & Doudna, J. A. Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.* **15**, 469–479 (2014).
166. Byron, K., Cervantes, M. C., Wang, J. T. L., Lin, W. C. & Park, Y. Mining *roX1* RNA in *Drosophila* genomes using covariance models. *Int. J. Comp. Biosci.* **1**, 22–32 (2010).
167. Smith, M. A., Gesell, T., Stadler, P. F. & Mattick, J. S. Widespread purifying selection on RNA structure in mammals. *Nucleic Acids Res.* **41**, 8220–8236 (2013).
168. Eddy, S. R. Computational analysis of conserved RNA secondary structure in transcriptomes and genomes. *Annu. Rev. Biophys.* **43**, 433–456 (2014).
169. Guttman, M. & Rinn, J. L. Modular regulatory principles of large non-coding RNAs. *Nature* **482**, 339–346 (2012).
170. Somarowthu, S. *et al.* HOTAIR forms an intricate and modular secondary structure. *Mol. Cell* **58**, 353–361 (2015).
171. Tsai, M. C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
172. Ilik, I. A. *et al.* Tandem stem-loops in *roX* RNAs act together to mediate X chromosome dosage compensation in *Drosophila*. *Mol. Cell* **51**, 156–173 (2013).
173. Krueger, B. J., Varzavand, K., Cooper, J. J. & Price, D. H. The mechanism of release of P-TEFb and HEXIM1 from the 7SK snRNP by viral and cellular activators includes a conformational change in 7SK. *PLoS ONE* **5**, e12335 (2010).
174. Buenrostro, J. D. *et al.* Quantitative analysis of RNA–protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat. Biotechnol.* **32**, 562–568 (2014).
175. Spitale, R. C. *et al.* Structural imprints *in vivo* decode RNA regulatory mechanisms. *Nature* **519**, 486–490 (2015).
176. Feng, J. *et al.* The RNA component of human telomerase. *Science* **269**, 1236–1241 (1995).
177. Nowacki, M. *et al.* RNA-mediated epigenetic programming of a genome-rearrangement pathway. *Nature* **451**, 153–158 (2008).
178. Sauvageau, M. *et al.* Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife* **2**, e01749 (2013).  
**Genetic screen of 18 lincRNA knockouts in mice that demonstrates lincRNA roles in development.**
179. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR–Cas9 complex. *Nature* **517**, 583–588 (2015).
180. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
181. Amrein, H. & Axel, R. Genes expressed in neurons of adult male *Drosophila*. *Cell* **88**, 459–469 (1997).
182. Conn, S. J. *et al.* The RNA binding protein Quaking regulates formation of circRNAs. *Cell* **160**, 1125–1134 (2015).
183. Chang, T.-C. *et al.* Genome-wide annotation of microRNA primary transcript structures reveals novel regulatory mechanisms. *Genome Res.* **25**, 1401–1409 (2015).
184. Sigova, A. A. *et al.* Transcription factor trapping by RNA in gene regulatory elements. *Science* **350**, 978–981 (2015).

### Acknowledgements

The authors thank P. J. Batista, R. A. Flynn and G. X. Zheng for constructive discussions and comments on this Review. Supported by US National Institutes of Health grants, HHMI (H.Y.C.), and the Stanford Bio-X Fellowship (J.J.Q.). The authors apologize to colleagues whose work could not be cited or discussed owing to space constraints.

### Competing interests statement

The authors declare no competing interests.

### FURTHER INFORMATION

GENCODE: <http://www.genencodegenes.org/>

ALL LINKS ARE ACTIVE IN THE ONLINE PDF