Regulation of gene expression by *cis*-acting long non-coding RNAs

Noa Gil and Igor Ulitsky *

Abstract | Long non-coding RNAs (lncRNAs) are diverse transcription products emanating from thousands of loci in mammalian genomes. *Cis*-acting lncRNAs, which constitute a substantial fraction of lncRNAs with an attributed function, regulate gene expression in a manner dependent on the location of their own sites of transcription, at varying distances from their targets in the linear genome. Through various mechanisms, *cis*-acting lncRNAs have been demonstrated to activate, repress or otherwise modulate the expression of target genes. We discuss the activities that have been ascribed to *cis*-acting lncRNAs, the evidence and hypotheses regarding their modes of action, and the methodological advances that enable their identification and characterization. The emerging principles highlight lncRNAs as transcriptional units highly adept at contributing to gene regulatory networks and to the generation of fine-tuned spatial and temporal gene expression programmes.

High-throughput sequencing technologies and chromatin state maps have shown that eukaryotic cells produce a plethora of non-coding transcripts[1–3]. Of these, long non-coding RNAs (lncRNAs) are defined arbitrarily as transcripts of more than 200 nucleotides that do not belong to any other well-defined group of non-coding RNAs, such as ribosomal RNAs. Through various mechanisms, lncRNAs have been implicated in a wide array of cellular processes, including transcriptional regulation, differentiation, cellular reprogramming and many others (reviewed elsewhere[4–6]). With varying levels of evidence, lncRNAs have also been implicated in various human diseases[7–9]. lncRNAs are transcribed by RNA polymerase II (Pol II), and their biogenesis is similar to mRNAs in that they are capped and polyadenylated. lncRNAs are also usually spliced, although their exon number and splicing efficiency are on average lower than those of mRNAs[10–13]. However, as lncRNAs are predominantly defined by exclusion criteria, the set of genes annotated as lncRNAs includes many distinct subgroups, exemplifying diverse structural and, presumably, functional characteristics. Assigning lncRNAs to distinct functional groups is essential to identify common principles, and thus comprises a pivotal step when beginning to elucidate their roles. This step remains very challenging, with limited progress being made in the past decade of lncRNA research.

One type of lncRNA classification is based on the location at which the lncRNA functions relative to its transcription site. *Trans*-acting lncRNAs are transcribed, processed and then vacate their sites of transcription to exert their function elsewhere, akin to mRNAs. Their final destination, be it in the cytoplasm or nucleus, does not depend on their transcription site. Accordingly, as long as their levels are properly maintained, transcribing these lncRNAs from a different genomic location or supplanting them into the system should not interfere with their function (that is, their loss of function can be rescued by their expression from exogenous locations). A few examples of such lncRNAs have now been extensively characterized[14–16], and many additional lncRNAs have been ascribed *trans* activities[17–20].

By contrast, *cis*-acting lncRNAs are those whose activity is based at and dependent on the loci from which they are transcribed. Transcripts with the potential of acting in *cis* likely make up a substantial portion of known lncRNAs: the majority of lncRNAs are enriched in the chromatin fraction, and specifically are tethered to chromatin — presumably at their sites of transcription — through Pol II[21]. If functional, this would indicate that the effects of these lncRNAs are centred at these loci. In addition, the fairly low levels at which lncRNAs are generally expressed, oftentimes just a few molecules per cell[13,22,23], naturally favour a *cis* mechanism of action, as diffusion or transport to other cellular compartments would render these transcripts too diluted to mediate a plausible function. Functions in *cis* regulation are also consistent with the evolutionary trajectories of lncRNAs (BOX 1), which often lack substantial sequence conservation, yet are sometimes found in conserved locations relative to other genes across distant species[24].

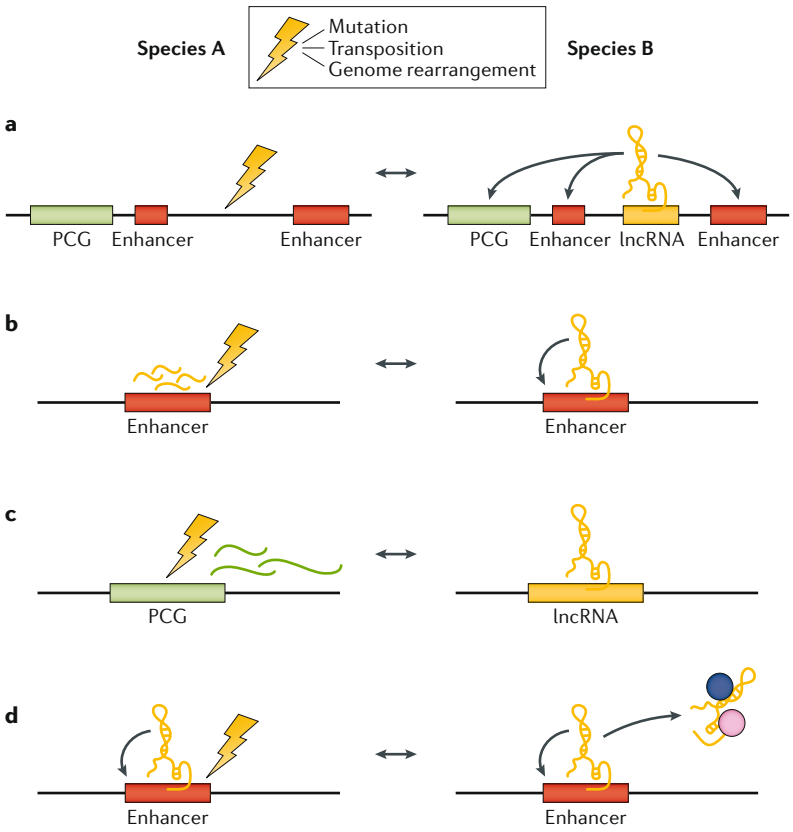
Advancements in our understanding of chromatin organization and transcription regulation, and the development of perturbation techniques relevant to

Department of Biological Regulation, Weizmann Institute of Science, Rehovot, Israel.

**e-mail: igor.ulitsky@weizmann.ac.il*

https://doi.org/10.1038/s41576-019-0184-5

Box 1 | **Evolution of *cis*-acting long non-coding RNAs**



Turnover of long non-coding RNA (lncRNA) loci is a major component of genome evolution. As discussed extensively elsewhere[24], emergence of new lncRNAs genes generally requires a combination of a promoter and polyadenylation sequences as well as potentially some splicing signals. Accumulation of the correct combination of mutations — that is, mutations that enable transcription initiation, recruitment of RNA processing factors and suppression of transcription termination — might lead to lncRNA production in a previously transcriptionally silent locus. This lncRNA, or the act of its transcription, might then acquire *cis* functions (see the figure, part **a**; yellow structure denotes lncRNA transcript). As enhancers already contain elements for the recruitment of RNA polymerase II (Pol II), it might be possible that few sequence changes are sufficient to gain or lose lncRNA production at enhancers (see the figure, part **b**; yellow wavy lines denote enhancer RNAs). Indeed, one study identified hundreds of regulatory regions with repurposed activities among closely-related mammalian species, that is, regions with promoter characteristics in one species and enhancer characteristics in another[160]. Interestingly, most of these regions likely served as enhancers in the ancestral species, which then acquired species-specific promoter abilities, demonstrating that enhancers can serve as fertile ground for the creation of novel transcripts. It is tempting to speculate that enhancers which acquire promoter capabilities do not necessarily or immediately lose their enhancer activity. Rather, other features of the enhancer locus (for example, spatial proximity to a target gene) can favour the creation of a transcript which participates in the same circuit as the enhancer.

Another important, albeit rare, source of new functional lncRNAs is disruption of protein-coding gene (PCG) open reading frames (ORFs)[161,162] (see the figure, part **c**; green wavy lines denote mRNAs). This process may affect preferentially PCGs that have paralogues in the genome, so that the original PCG function is maintained[162]. One could hypothesize that, compared with enhancer-derived lncRNAs, PCG-derived lncRNAs are more likely to act in *trans*, as the sequences they inherit from their PCG ancestors would facilitate higher (and perhaps more cytoplasmic) expression[162]. Nonetheless, such lncRNAs may hold on to *cis* functionality, as some PCG loci seem to induce *cis* effects similar to lncRNA loci. Over time, a *cis*-acting RNA can accumulate sequences that will facilitate *trans*-acting activities and vice versa (see the figure, part **d**; coloured circles denote *trans* factors). As most lncRNAs are found in both the cytoplasm and the nucleus[163,164], it is possible that some lncRNAs carry both functions, although such examples remain scarce[107].

the study of lncRNAs (most notably CRISPR-related technologies), have facilitated the in-depth characterization of several *cis*-acting lncRNAs that operate at various genomic distances and through apparently distinct mechanisms of action. From these studies, a broader picture of the roles of *cis*-acting lncRNAs is starting to emerge. In this Review, we begin with an overview of the prevalence, genomic locations, orientations and conservation of *cis*-acting lncRNAs, as well as the functional clues provided by these attributes. We describe the types of *cis* effects that lncRNAs can have on their targets, supplementing with recent examples of fairly comprehensively studied lncRNAs. Finally, we elaborate on reported and possible mechanisms of action employed by *cis*-acting lncRNAs, as well as the tools currently being used or developed for investigating their functions.

Discerning between *trans*- and *cis*-acting lncRNAs

Although the associations of a lncRNA with its site of transcription, genomic neighbourhood, epigenetic environment and mode of conservation (that is, sequence versus synteny) can all point to a lncRNA acting in *cis*, they do not provide direct proof. For example, some lncRNAs co-expressed with their neighbouring gene seem to have no *cis*-regulatory activity but rather cooperate with the product of this gene[20]. Some experimental indication for *cis* activity can be achieved by expressing the lncRNA from a different genomic location or supplanting it into the system, for example through transfection of a vector. If such *trans* supplementation does not rescue the phenotype, the lncRNA is suspected of working in *cis*. It is noteworthy that some lncRNAs that act near their transcription sites can be rescued by *trans* expression, indicating they are not 'true' *cis*-acting lncRNAs. For example, the *FIRRE* lncRNA, which orchestrates spatial proximity between loci found on different chromosomes, operates from its site of transcription[25], although expressing the lncRNA in *trans* can rescue a *FIRRE*-null phenotype[26]. Similarly, the *NEAT1* lncRNA serves as a scaffold for the assembly of paraspeckles, nuclear membrane-less organelles that form near its transcription site[27]. Paraspeckle-like foci can also be formed when *NEAT1* is expressed exogenously[28], although their functionality has not yet been tested. Nonetheless, the success of *trans* supplementation is highly dependent on exact timing and expression levels, and so failure in such a setting does not prove *cis* activity. Additionally, even if *trans* supplementation does work, it does not unequivocally prove that the lncRNA does not work in *cis*, as it is theoretically possible that when the lncRNA 'floods' the system, a sufficient amount reaches the vicinity of its site of transcription. More direct experimental proof can be gained using cells or tissues generated from a mixed genetic background, such as through crosses between mouse strains[29–33]. This approach allows perturbation of only a single lncRNA allele and then exploitation of single-nucleotide polymorphisms to discern whether the effect is limited to target genes located on the same allele as the perturbed lncRNA, which would confirm a *cis* mechanism of action.

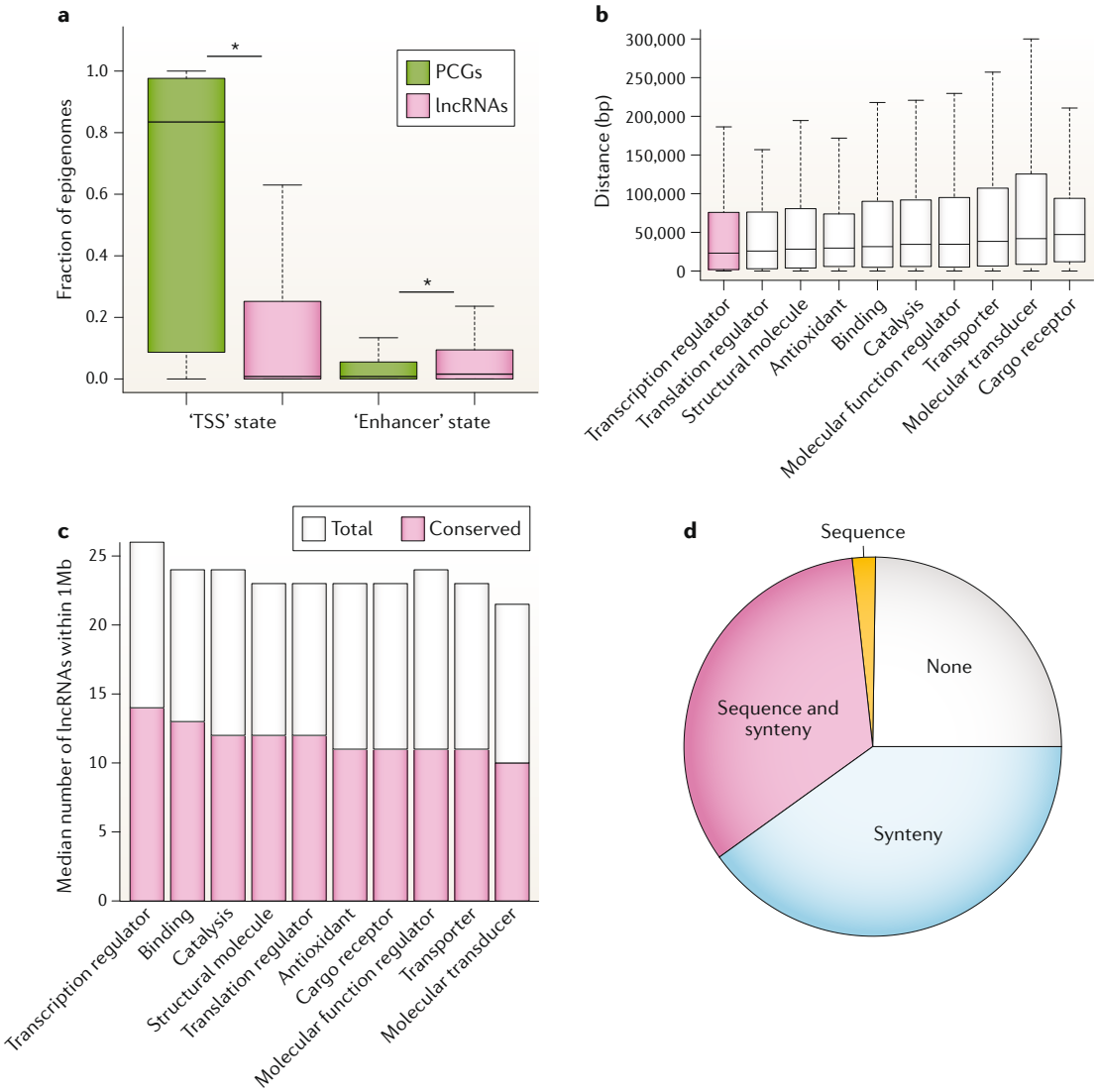


Fig. 1 | **Genomic locations and prevalence of *cis*-acting long non-coding RNAs. a** | The fraction of transcription start sites (TSSs) of RefSeq protein-coding gene (PCG) transcripts or GENCODE v30 long non-coding RNA (lncRNA) transcripts that overlap regions classified as 'TSSs' or, conversely, as 'enhancers' in a 15-state chromHMM model of 127 distinct epigenomes[165,166]. Asterisks indicate $P < 2.2 \times 10^{-16}$, calculated using the two-sided Wilcoxon test. **b** | The distribution of the distances between the TSSs of human PCGs belonging to the indicated categories of gene ontology (GO) terms and the closest lncRNA transcript. Shown are all GO term categories under 'molecular function' that contain at least 50 genes. Highlighted in pink is the GO term category 'transcription regulator activity'. **c** | The median number of lncRNA transcripts found within 1 Mb of the TSS of PCGs that belong to the indicated categories of GO terms (pink: the median number of sequence-conserved lncRNAs within that same distance). Shown are all GO term categories under 'molecular function' that contain at least 50 genes. **d** | The conservation type for the lncRNA found closest to TSSs of PCGs belonging to the 'transcription regulator activity' GO category.

Genomic locations of *cis*-acting lncRNAs

If functional, lncRNAs that are enriched around their own sites of transcription are expected to participate in chromatin-related processes, such as the modulation of chromatin structure, chromatin modifications or transcription control. The contribution of lncRNAs to these processes in *cis* is evidenced by the enrichment of lncRNA genes in the vicinity of regulatory sequences of the genome. Most notably, various studies have indicated that 30–60% of lncRNAs are transcribed from regions showing characteristics of enhancers[21,34–36], with the variability arising from the different methodologies and parameters used for annotating both lncRNAs and enhancers. When considering the commonly used annotations of enhancers by chromHMM[37], lncRNA transcription start sites (TSSs) tend to overlap regions classified as 'enhancers' more than protein-coding gene (PCG) TSSs do (FIG. 1a). This observation is in agreement with findings showing that many lncRNA promoters exhibit DNA motifs and protein-binding characteristics associated with both promoters and enhancers[36,38].

An additional indication for the widespread *cis*-regulatory function of lncRNAs is their enrichment in the vicinity of genes belonging to specific categories. Various

Enhancers
DNA regulatory elements that activate gene transcription. Enhancers can operate from nearby or within their targets, or across large genomic distances.

analyses have shown that the protein products of genes that flank lncRNA genes are involved in transcription-related processes[1,23,39–41]. A similar analysis of more recent lncRNA annotations shows that genes involved in processes related to transcription regulation, such as transcription factors (TFs) or chromatin remodellers, tend to have a closer lncRNA neighbour than do other genes (FIG. 1b). Genes involved in transcription regulation also tend to be surrounded by more lncRNA genes, suggesting that multiple lncRNAs might cooperate in their regulation; this trend is more pronounced when considering the number of conserved lncRNAs[41] (FIG. 1c). Importantly, although somewhat weaker, this trend is maintained when normalizing to the number of PCGs within the same genomic distance, or to the number of annotated enhancer elements (not shown), indicating that this enrichment does not reflect merely an increase in regulatory intergenic space. The nearest lncRNA neighbours of transcription-related genes often have conserved positions rather than sequences (FIG. 1d), pointing to the potential importance of lncRNA presence — sometimes independent of the specific sequence — at their vicinity. The enrichment of lncRNA transcriptional units in close proximity to genes involved in transcriptional regulation could be attributed to a requirement for precise and robust expression levels at such loci, as evidenced by the relative instability of mRNA products of such genes[42], which enables quick response to stimuli and establishment of distinct cell states. Alternatively, these lncRNA genes could provide transcriptional units that are adept at responding to the levels of the protein products of adjacent PCGs, thereby allowing the establishment of feedback or autoregulatory loops (discussed below).

Distances of cis-acting lncRNAs. *Cis*-acting lncRNAs can be positioned at various distances and orientations relative to their target genes, in a manner reminiscent of enhancers. A genomic position next to an expressed lncRNA gene is associated with increased expression of a PCG relative to other PCGs[43,44], suggesting that the *cis* effects of lncRNA production are oftentimes contained within fairly short genomic distances. However, *cis*-acting lncRNAs are certainly capable of acting over longer genomic distances. The repressive effects of *Xist*, for example, span the entire X chromosome (with the exception of specific regions that escape repression) and are not limited to the vicinity of its transcription site; this is achieved through a proximity-guided mechanism, whereby the 3D architecture of the chromosome allows *Xist* to spread to increasingly distal regions and extend its effects[45]. Even *cis*-acting lncRNAs that have a more contained effect on select genes can function over large genomic distances[46,47]: for example, *Peril*, which is transcribed from a *Sox2*-related super-enhancer in mouse embryonic stem cells, positively regulates the expression of two genes found in a separate topologically associating domain (TAD) ~1.5 Mb away from its transcription site[48]. Additionally, multiple lncRNA genes have been shown to overlap TAD boundaries and loop anchors[49,50], endowing these lncRNAs with the potential of regulating long-range chromatin interactions, thus affecting gene regulation over long distances.

Orientations of cis-acting lncRNAs. The orientation of *cis*-acting lncRNAs relative to their genetic targets is also variable, as has been reviewed elswhere[51,52]. Possible architectures of lncRNA-target units include lncRNAs that act from a distance[53,54]; antisense lncRNAs, which overlap their target genes in the antisense orientation[55–57]; sense lncRNAs, which are typically contained within introns of the target genes[58]; divergent lncRNAs, which are transcribed divergently from the target in the antisense orientation, oftentimes using a shared promoter[56,59,60]; and lncRNAs arranged in tandem units (that is, transcribed in the same orientation, either upstream or downstream) with their targets[29,61]. Generally, it seems that when a lncRNA regulates a *cis* target from a distance, there is no preference for a particular orientation, whereas when lncRNA genes are very close to or even overlap their targets, the relative orientation of the two genes can be consequential.

Functions of *cis*-acting lncRNAs

Activating cis-acting lncRNAs. The arguably largest group of *cis*-acting lncRNAs are those that function to augment the expression of target genes, akin to the function of enhancers. Indeed, most studied examples of activating *cis*-acting lncRNAs describe lncRNAs transcribed from regions demarcated as enhancers. Several characteristics of lncRNAs render them appealing candidates for activating gene expression through participating in, or mediating, the activity of the enhancers from which they are transcribed. First, as described above, lncRNA genes are enriched at enhancer elements. Second, lncRNA expression is highly tissue-specific and cell type-specific[13,22,23,62], as has been shown for many enhancer elements[63,64].

In this context, it is important to distinguish between lncRNAs transcribed from enhancers (sometimes referred to as enhancer lncRNAs or e-lncRNAs) and another species of non-coding RNAs produced at enhancers, termed enhancer RNAs (eRNAs) (see recent reviews[65–67]). Although the two terms are often conflated, and although some enhancers produce both lncRNAs and eRNAs, the main distinctions between eRNAs and e-lncRNAs are size and stability: eRNAs are shorter RNA species (~1 kb on average), which are generally unspliced and non-polyadenylated, and therefore constitute very unstable transcripts, typically invisible in RNA-sequencing (RNA-seq) data sets that are not enriched for nascent transcripts. It is noteworthy that eRNAs are also generally transcribed in a bidirectional manner, although some unidirectional eRNAs have been described[68], which may actually be e-lncRNAs.

Activating cis-acting lncRNAs that modulate enhancer availability. Initial studies of enhancer-transcribed lncRNAs suggested that these transcripts modulate enhancer activity by contributing to the formation or maintenance of chromatin loops between the underlying enhancers and the genes they regulate[44,47,54,69]. This process occurs mostly through the recruitment of proteins involved in establishing such loops (FIG. 2a). For example, depletion of a set of activating lncRNAs, termed ncRNA-a, caused a reduction in expression of nearby PCGs;

Transcription factors
(TFs). Proteins that bind to specific DNA sequence motifs found within regulatory DNA elements — either promoters or enhancers — to modulate gene expression.

Chromatin remodellers
Proteins that regulate gene expression by altering chromatin structure. Two main groups exist: those that mediate post-translational modifications of histones; and ATP-dependent chromatin remodellers, which regulate the association with and location of nucleosomes along the DNA.

Super-enhancer
A particularly active and cell type-specific enhancer. Demarcated by high levels of chromatin modifications such as histone 3 lysine 27 acetylation (H3K27ac) and long sequence stretches bound by transcription factors and coactivators such as Mediator.

Topologically associating domain
(TAD). A genomic region with an average size of ~1 Mb characterized by high-density chromatin interactions. Sequences within TADs tend to form interactions with one another but less so with sequences in other TADs.

Enhancer RNAs
(eRNAs). A species of bidirectional, unstable non-coding RNAs produced at enhancers. Considered a hallmark of active enhancers and sometimes used for enhancer annotation.

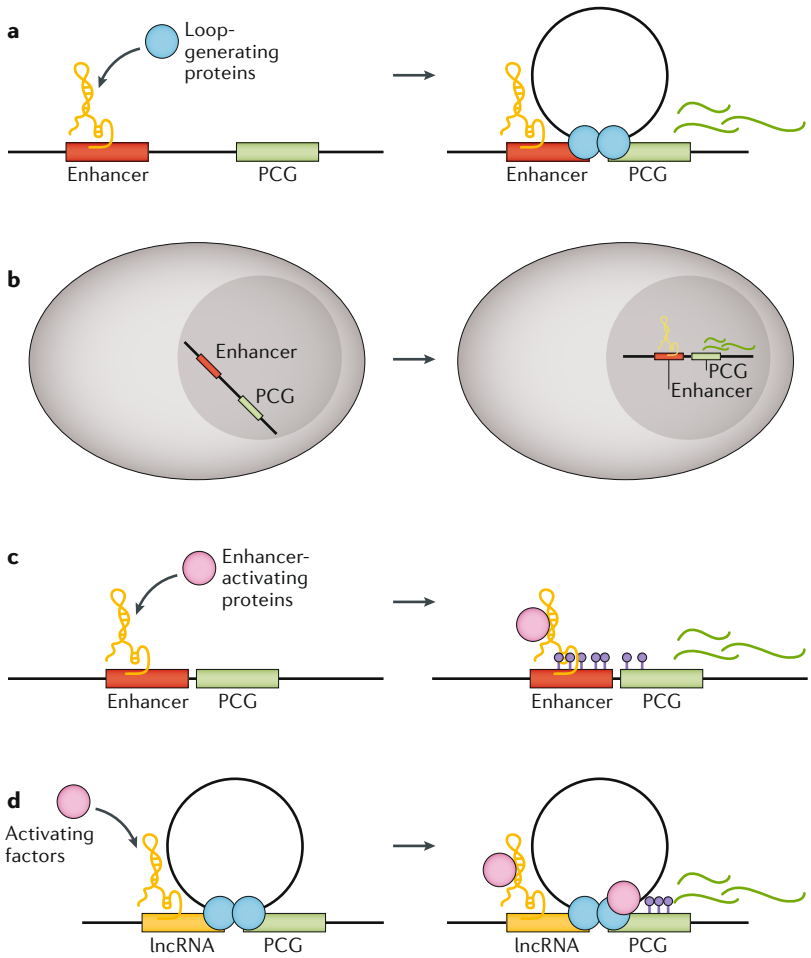


Fig. 2 | **Mechanisms of action of activating *cis*-acting long non-coding RNAs.** *Cis*-acting long non-coding RNAs (lncRNAs) can activate target protein-coding genes (PCGs) through various mechanisms. **a** | lncRNA transcripts (yellow) can act by recruiting proteins that modulate chromatin loops, thus bringing lncRNA-proximal enhancers into spatial proximity of the target genes (for example, *CCAT1-L*). The proteins may be recruited by direct interaction with the lncRNA or, alternatively, the act of its transcription would affect DNA accessibility. Green wavy lines denote mRNAs. **b** | lncRNA transcripts can also affect the nuclear positioning of their underlying enhancer, thus potentiating the enhancer to activate target genes (for example, *ThymoD*). **c** | Alternatively, lncRNAs can potentiate target gene-proximal enhancers (for example, *Hand2os1*), possibly through direct recruitment of proteins that enhance gene expression. Small purple circles denote histone modifications. **d** | Similar to part **c**, except pre-formed chromatin loops bring the lncRNA into the proximity of target genes, where it can recruit activating proteins (for example, *UMLILO*).

further study showed that these transcripts act by mediating contacts between their own genomic loci and the promoters of proximal PCGs through recruitment of the Mediator complex[69]. Similarly, the lncRNA *CCAT1-L*, which is highly expressed in colorectal cancer cells from a super-enhancer region located ~500 kb upstream of *MYC*, is localized to its site of transcription and directly interacts with CTCF; possibly through this interaction, chromatin loops are formed between the *CCAT1-L* and *MYC* loci, leading to increased *MYC* transcription and enhanced tumorigenicity[54].

lncRNAs can also contribute to the formation of contacts between their underlying enhancers and target genes through additional, less direct mechanisms, as in the case of *ThymoD*, a lncRNA expressed from an enhancer region ~700 kb from the *Bcl11b* gene in mouse developing T cell progenitors[30]. Interference with *ThymoD* transcription by insertion of a polyadenylation (polyA) sequence increased the methylation of CTCF motifs and reduced occupancy of Pol II and cohesin throughout the region. Combined, these effects interfered with the chromatin contacts between the *Bcl11b* enhancer and promoter, maintaining the *Bcl11b* enhancer at the nuclear periphery and hindering *Bcl11b* expression (FIG. 2b).

Perhaps relatedly, it was recently demonstrated that binding of the nuclear matrix factor hnRNPU to chromatin-associated RNAs — specifically to C_0T-1 RNA, which is enriched for repetitive sequences — is required for hnRNPU-mediated chromatin decompaction[70]. Although the authors did not note any substantial transcriptional changes following chromatin compaction mediated by hnRNPU depletion, the fact that hnRNPU has also been shown to bind various types of RNA species, including multiple lncRNAs[71], coupled with the relative enrichment of lncRNAs in the chromatin fraction, suggests that the transcription of some *cis*-acting lncRNA genes might function to recruit hnRNPU or similar proteins that promote local chromatin decompaction.

Activating cis-acting lncRNAs that function at target gene promoters. *Cis*-acting lncRNAs have also been demonstrated to contribute to enhancer activity through various mechanisms that do not seem to involve modulating spatial interactions. Such activity would nonetheless likely require proximity to the target promoter, entailing either that the lncRNA is transcribed proximally to its target gene or that lncRNA-independent chromatin interactions would bring it to the spatial vicinity of its target. One such *cis*-functioning, enhancer-overlapping lncRNA is *Hand2os1* (also known as *Upperhand* or *Hand2as*), which is transcribed in mouse cardiac tissues divergently to the gene encoding the *Hand2* TF[59]. Two crucial enhancers that control *Hand2* expression in various cardiac tissues are located within the second intron of *Hand2os1*. Perturbation of *Hand2os1* transcription through insertion of a polyA sequence upstream of these enhancers reduced the levels of histone 3 lysine 27 acetylation (H3K27ac) and histone 3 lysine 4 monomethylation (H3K4me1) at enhancer-proximal regions. This led to reduced *Hand2* levels and a phenotype resembling that of the embryonic-lethal *Hand2* knockout[59]. The authors proposed that transcription through the *Hand2os1* locus is required for proper activation of the enhancers contained within that locus, which in turn control transcription (and particularly elongation) rates of *Hand2* (FIG. 2c). Transcription through other enhancer-overlapping lncRNA genes has been suggested to similarly affect the expression of their targets[72]. Importantly, however, it was recently reported that deletion of the *Hand2os1* promoter, although leading to an almost complete ablation of *Hand2os1* transcription, did not reduce *Hand2* levels in cardiac tissues[73]. Conversely, deletion of the entire *Hand2os1* locus led to a subtle increase in *Hand2* levels, as well as an increase in the levels of several additional genes located in its

CTCF
A transcription factor that acts primarily in chromatin 3D architecture regulation, through anchoring long-range chromatin loops and demarcating topologically associating domain boundaries.

vicinity, and to various cardiac-related morphological and functional abnormalities. The interplay between *Hand2os1*, the DNA elements that regulate *Hand2* and the expression of *Hand2* is thus more complex than previously appreciated.

Although the above examples describe lncRNAs that activate gene expression by making their underlying enhancers more available to augment distal gene expression, other lncRNAs activate gene expression in *cis* irrespective of an identifiable active enhancer. For example, the human lncRNA *UMLILO* is required for the induction of several chemokine genes located within its TAD[53]. At steady state, the *UMLILO* locus is found in close, *UMLILO* RNA-independent spatial proximity to these neighbouring chemokine genes, which are not being transcribed. Upon tumour necrosis factor stimulation, the *UMLILO* RNA binds to and recruits the WDR5–MLL complex, leading to the deposition of histone 3 lysine 4 trimethylation (H3K4me3) at the chemokine promoters, thus inducing their expression. *UMLILO* knockdown by either small interfering RNAs or antisense oligonucleotides (ASOs) substantially hinders this activation, indicating that the mature RNA mediates this activity. In this way, the pre-formed, Pol II-anchored chromatin loop between *UMLILO* and its target genes enables speedy induction of these genes upon stimulation, and similar lncRNAs likely act to induce the expression of chemokines located in other TADs. Thus, although *UMLILO* is transcribed from a region denoted as a super-enhancer, the enhancer is seemingly required for *UMLILO* functionality by delivering the lncRNA to close spatial proximity of its target locus, rather than the other way around.

Through similar suggested mechanisms, other activating lncRNAs have been proposed to function in *cis* independently of the underlying DNA enhancer or without overlapping a known active DNA enhancer[74–76]. The emerging general principle is that an apparently pre-formed chromatin loop places the lncRNA product in the proximity of the target genes where, likely through protein recruitment, it exerts its activating effect (FIG. 2d). However, as the resolution of available methods for measuring spatial proximity as well as their ability to report on the frequency of contacts are both limited, it is presently difficult to rule out a lncRNA contribution to quantitative changes in spatial chromatin organization around the target gene. Naturally, in those cases where the lncRNA is transcribed from a region with enhancer characteristics, the close spatial proximity that enables lncRNA functionality does not preclude additional activating roles of the enhancer. Instead, the lncRNA can be thought of as another factor — alongside TFs, the Pol II machinery and RNA processing proteins — found at enhancers and relayed to target genes through chromatin loops.

Global consequences of lncRNA production on enhancer activity. Alongside individual examples, recent studies are beginning to converge on global principles of the consequences of lncRNA transcription from enhancers. We recently reported that in enhancers which produce lncRNAs, the characteristics of enhancer activity — such as histone acetylation and activity in enhancer assays — are stronger than in enhancers that do not[38]. The differential DNA sequence composition of enhancers that produce lncRNAs, and particularly the presence of conserved motifs that dictate lncRNA splicing, points to a causal role of lncRNA processing in enhancer activity and suggests that the ability to transcribe and process lncRNAs is maintained in evolution so as to allow increased enhancer activity (BOX 1). Relatedly, another study showed that splicing of enhancer-transcribed lncRNAs correlates with enhancer activity, and single-nucleotide polymorphisms that decrease the splicing efficiency of lncRNAs are associated with decreased expression of not only the lncRNAs but also the enhancers' putative target genes[77]. Interestingly, although regions with sequence conservation are generally scarce in lncRNAs, those that are identified are oftentimes regions containing splicing motifs[78,79]. This observation suggests that lncRNA maturation contributes to the activity of many *cis*-acting lncRNAs. Intriguingly, a recent study found that reporter loci found in physical proximity to nuclear speckles — nuclear suborganelles enriched in splicing factors[80] — are associated with a strong transcriptional boost upon activation[81], plausibly implicating lncRNA splicing with repositioning to nuclear speckles and increased transcription of neighbouring loci. Similarly, transcription was found to be correlated with proximity to the nuclear speckles and with co-transcriptional splicing efficiency[82]. Combined, these findings indicate that lncRNA splicing at enhancers drives enhancer functionality. It remains to be determined whether this effect is direct (that is, mediated by recruitment of splicing factors that can then promote target gene expression through various mechanisms[83]) or indirect (splicing promotes the generation of mature, stable transcripts or dissociation of the transcripts from chromatin).

Repressive cis-acting lncRNAs. In addition to activating functions, *cis*-acting lncRNAs can act to repress the expression of target genes. The epitome of a lncRNA that represses gene activity is *Xist*, which coats and silences the inactive X chromosome in placental mammals to achieve dosage compensation. *Xist* has been the subject of extensive study, thoroughly reviewed elsewhere[84–86]. Briefly, the A repeat sequence of *Xist* was proposed to be important for histone 3 lysine 27 trimethylation (H3K27me3) deposition and gene silencing through recruitment of transcriptional repressors such as SPEN and Polycomb repressive complex 2 (PRC2), although the exact chronology of protein recruitment and whether these interactions are direct are both a matter of ongoing debate. Whereas the A repeat is required for the initiation of *Xist*-mediated gene silencing, additional *Xist* repeats have been proposed to be responsible for its spreading and coating of the to-be-inactivated X chromosome, and for maintaining the inactive transcriptional state once established[86]. Although it is not clear to what extent *Xist* biology is applicable to other repressive lncRNAs, many of the principles discovered through the study of *Xist*, as well as the methods developed and calibrated for this purpose (particularly for detecting interactions between lncRNAs and proteins and/or DNA), have been instrumental for lncRNA research.

Polycomb repressive complex 2
(PRC2). A histone methyltransferase protein complex that induces trimethylation of histone 3 lysine 27 (H3K27), a histone modification associated with long-term epigenetic silencing.

Vernalization
The process of induction of plant flowering, brought on by exposure to prolonged cold temperatures.

Enhancer competition
Two (or more) transcriptional units that can be activated by the same enhancer, and which compete over direct binding to and activation by that enhancer.

Transcriptional interference
A process whereby transcription through one genomic region interferes with transcription of a nearby (often overlapping) locus, for example, by curbing the recruitment of *trans* factors such as transcription factors or chromatin remodellers, or through deposition of chromatin modifications incompatible with transcription initiation.

Other lncRNAs repress their *cis* targets using similar mechanisms but on smaller scales. For example, the Arabidopsis *COOLAIR* lncRNA, which is transcribed antisense to the *FLOWERING LOCUS C* (*FLC*) region, represses *FLC* transcription during the process of vernalization[87]. Prolonged exposure to cold leads to an increase in *COOLAIR* transcription. The resulting transcript coats the *FLC* locus and causes a switch to a repressed chromatin state, possibly mediated through a reduction in histone 3 lysine 36 trimethylation (H3K36me3) levels, which in turn enables PRC2 recruitment and H3K27me3 deposition[88] (the detailed studies of the interplay between *COOLAIR* and *FLC* were recently reviewed elsewhere[89]). Another example is the *Morrbid* lncRNA, which represses nearby *Bcl2l11* expression in mouse immune cells by recruiting PRC2 to its locus[32]. Using a mechanism reminiscent of that described for the activating lncRNA *UMLILO*, a pre-formed chromatin loop brings the *Morrbid* gene to the vicinity of the *Bcl2l11* locus, to which *Morrbid* lncRNA then recruits PRC2, leading to deposition of H3K27me3 and suppression of *Bcl2l11* expression. Similar target repression through PRC2 recruitment is a suggested modus operandi for additional lncRNAs[57,90] (FIG. 3a; TABLE 1), although the specificity of PRC2–RNA interactions has been a subject of debate (see below).

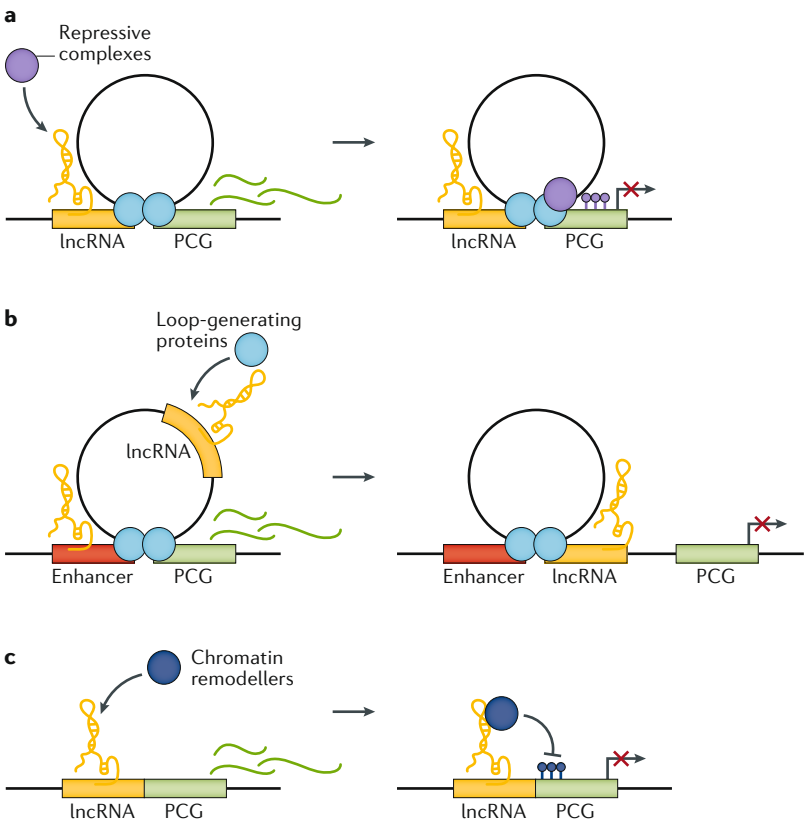


Fig. 3 | **Functions of repressive *cis*-acting long non-coding RNAs.** *Cis*-acting long non-coding RNAs (lncRNAs) can repress target genes through various mechanisms. **a** | Recruitment of proteins that repress gene expression. The lncRNAs can be transcribed from near the target gene or be brought to its proximity through pre-formed chromatin loops (for example, *Morrbid*). Yellow structure denotes lncRNA transcripts, green wavy lines denote mRNAs and small purple circles denote histone modifications. **b** | Competition over available enhancers in the vicinity (for example, *PVT1*). **c** | Transcriptional interference, whereby transcription of the lncRNA near to or (usually) overlapping the target gene represses target expression, for example, through nucleosome remodelling or deposition of epigenetic modifications (for example, *Airn*). PCG, protein-coding gene.

Interestingly, lncRNA loci that repress their neighbouring genes can also overlap and/or function through DNA enhancers, via enhancer competition (FIG. 3b). For example, silencing of *PVT1*, a highly conserved lncRNA found downstream of the gene encoding the MYC TF in all organisms from fish to mammals, was recently described to result in an increase in cell proliferation through *MYC* induction[91]. The study found that the *PVT1* and *MYC* promoters compete for binding to DNA enhancers found within the *PVT1* gene body. Silencing of the *PVT1* promoter using CRISPR interference (CRISPRi) allowed the *MYC* promoter to form more pronounced contacts with these enhancers, thereby increasing *MYC* expression and promoting cell proliferation. Thus, *PVT1* transcription represses *MYC* by appropriating the available enhancers. This functionality seems to depend exclusively on the DNA in the *PVT1* locus, as early arrest of *PVT1* transcription using dCas9 does not affect it[86]. Of note, additional functions have been attributed to both the *PVT1* locus and its lncRNA product, with various and sometimes seemingly contradictory effects on cell proliferation[92–94].

An additional mechanism by which *cis*-acting lncRNA genes can repress their targets is transcriptional interference (FIG. 3c). This is the suggested mode of action for the *Airn* lncRNA gene, which is transcribed from the paternal allele of the *Igf2r* imprinted cluster and is required for silencing the paternal *Igf2r* allele as well as additional genes in the cluster[55]. *Airn* is transcribed from a promoter within the imprinted *Igf2r* gene in the antisense orientation, so that the *Airn* transcript overlaps the *Igf2r* promoter. Curtailing *Airn* transcription by inserting several polyA sequences throughout the *Airn* sequence led to *Igf2r* de-repression only when the polyA sequence was inserted upstream of the *Igf2r* promoter[95]. This result indicates that transcription through the *Igf2r* promoter, rather than the mature *Airn* sequence or transcription through the rest of the locus, is important for *Igf2r* silencing. Several yeast lncRNA genes have also been suggested to function via transcriptional interference[96–98], including *SRG1*, which is expressed upstream of the gene encoding *SER3* (REF.[61]). *SRG1* deletion causes upregulation of *SER3* in *cis*, with *SRG1* termination just upstream of the *SER3* TSS being important for this repression. When the *SRG1* transcription termination signal is deleted and *SRG1* transcription continues into the *SER3* region, *SER3* repression is abolished[61]. A further study showed that the repressive effect of *SRG1* on *SER3* is mediated through nucleosome positioning in the *SRG1* locus; in the wake of Pol II transcription through *SRG1*, nucleosomes are repositioned so as to block transcription from the nearby *SER3* TSS[99]. Lack of *SRG1* transcription results in reduced nucleosome occupancy, thereby enabling *SER3* expression.

Complex transcriptional units involving cis-acting lncRNAs. The capability of *cis*-acting lncRNAs to activate or repress their target genes through various mechanisms is also being utilized in the regulation

Table 1 | **Representative proteins or protein complexes suggested to interact with *cis*-acting long non-coding RNAs**

Protein or complex	Bound *cis*-acting lncRNA	Organism	Method of identification and fixation	Proposed outcome of interaction
PRC2	*Xist*[167,168]	Mouse	RIP (native)/iDRiP (UV C-L)	PRC2 recruitment and maintenance on the Xi
	Morrbid[32]	Mouse	RIP (native) + PAR-CLIP (UV C-L)	PRC2 recruitment to the target gene
	Kcnq1ot1 (REF.[57])	Mouse	RIP (native)	PRC2 recruitment to target genes
	ANRASSF1 (REF.[90])	Human	RIP (native)	PRC2 recruitment to the target gene
ATRX	*Xist*[169]	Mouse	RIP (UV C-L)	Remodelling of *Xist* RNA, thereby enhancing PRC2 binding
SPEN	*Xist*[168,170,171]	Mouse	iDRiP (UV C-L)/ChIRP-MS (formaldehyde C-L) + CLIP-qRT-PCR (UV C-L)/RAP-MS (UV C-L) + RIP (UV C-L)	Recruitment of SPEN to the Xi
hnRNPU	*Xist*[168,170–172]	Mouse	iDRiP (UV C-L)/ChIRP-MS (formaldehyde C-L)/RAP-MS (UV C-L)/RIP (UV C-L)	Accumulation of *Xist* on the Xi
WDR5–MLL	*HOTTIP*[76]	Human	RIP (native)	Recruitment of WDR5–MLL to target genes
	UMLILO[53]	Human	RIP (formaldehyde C-L) + biotin-RNA pull-down (native)	Recruitment of WDR5–MLL to the target gene
	Evx1as[173]	Mouse	RNA-ChIP (formaldehyde C-L)	Recruitment of WDR5–MLL to the target gene
	Hoxb5/6as[173]	Mouse	RNA-ChIP (formaldehyde C-L)	Recruitment of WDR5–MLL to the target gene
G9a	*Airn*[108]	Mouse	RIP (C-L)	G9a recruitment to the target gene
	Kcnq1ot1 (REF.[57])	Mouse	RIP (native)	PRC2 recruitment to target genes
CTCF	*CCAT1-L*[54]	Human	RIP (UV C-L) + biotin-RNA pull-down (native)	Recruitment of CTCF to maintain chromatin looping with target locus
	Jpx[121]	Mouse	RIP (UV C-L) + RNA pull-down (native)	Extrication of CTCF from the *Xist* promoter
Mediator	*ncRNA-a7* (REF.[69])	Human	RIP (UV-CL) + gel filtration (native)	Recruitment of Mediator to promote chromatin looping with target locus
	Evx1as[174]	Mouse	RIP (native) + biotin-RNA pull-down (native)	Recruitment of Mediator to the target gene
Cohesin	*Xist*[168]	Mouse	iDRiP (UV C-L)	Repulsion of cohesin from the Xi
YY1	*Sox2ot*[175]	Mouse	RIP (formaldehyde C-L)	Recruitment of YY1 to the target gene
	Xist[176]	Mouse	RIP (UV C-L) + in vitro RNA pull-down (native)	*Xist* loading onto the Xi
HNRNPK	*Xist*[46,170]	Mouse	RIP (formaldehyde C-L)/ChIRP-MS (formaldehyde C-L) + CLIP-qRT-PCR (UV C-L)	Recruitment of HNRNPK
	Airn[46]	Mouse	RIP (formaldehyde C-L)	Recruitment of HNRNPK
	Kcnq1ot1 (REF.[46])	Mouse	RIP (formaldehyde C-L)	Recruitment of HNRNPK
DNMT3B	*yylncT*[60]	Human	CHOP-MS (UV C-L) + RIP (C-L)	Local inhibition of DNMT3B

ChIP, chromatin immunoprecipitation; ChIRP, chromatin isolation by RNA purification; CHOP, chromatin oligo affinity precipitation; C-L, crosslinking; CLIP, crosslinking and immunoprecipitation; DNMT3B, DNA (cytosine-5)-methyltransferase 3B; iDRiP, identification of direct RNA-interacting proteins; lncRNA, long non-coding RNA; MS, mass spectrometry; PAR-CLIP, photoactivatable ribonucleoside-enhanced CLIP; PRC2, Polycomb repressive complex 2; qRT-PCR, quantitative PCR with reverse transcription; RAP, RNA antisense purification; RIP, RNA immunoprecipitation; UV, ultraviolet; Xi, inactive X chromosome.

of highly complex transcriptional units. For example, lncRNA transcription was shown to contribute to the creation of diverse transcripts from the complex *Protocadherin a* locus, where an alternative first exon — out of 13 possible exons — is stochastically chosen to be expressed in individual neurons[100]. Stochastic transcription of an antisense lncRNA from one of the alternative first exons resulted in demethylation of a CTCF binding site contained within that exon, which allowed CTCF binding and formation of a chromatin loop between the exon and the downstream enhancer. Such looping promotes sense transcription from the selected promoter, resulting in the production of transcripts with different first exons in individual neurons, a process which contributes to neuronal self-identity[100].

Multiple lncRNA genes can cooperate together to fine-tune target gene expression. For example, the yeast *ICR1* and *PWR1* lncRNAs cooperate to regulate the expression of the gene encoding Flo11 cell surface protein[98]. *ICR1* is transcribed upstream to and on the same strand as *FLO11*, and transcription of *ICR1* represses *FLO11*, likely through transcriptional interference. *ICR1* is itself regulated by transcription of the overlapping *PWR1* lncRNA, likely also through transcriptional interference. The proposed model is that competitive binding of the *trans*-acting Flo8 or Sfl1 to the vicinity of the *FLO11* promoter determines which lncRNA is expressed: either *ICR1*, in which case *FLO11* transcription is repressed, or *PWR1*, in which case *ICR1* transcription is repressed and *FLO11* is activated. This system establishes heterogeneous expression of

Auto-regulatory feedback loops
A type of transcriptional regulation network in which a gene product regulates its own levels, for example, a transcription factor which binds its own locus and activates (or represses) transcription.

FLO11 in clonal cell populations[98,101]. Additional regulatory networks consisting of multiple *cis*-acting lncRNA genes have been described[59,102], most notably multiple lncRNA genes involved in mammalian X chromosome inactivation (reviewed elsewhere[103]).

lncRNAs are also particularly appealing candidates for forming auto-regulatory feedback loops or sensing of transcriptional outputs. We recently characterized *Chaserr*, a highly conserved lncRNA located immediately upstream of the gene encoding the chromatin remodeller CHD2 (REF.[29]). Loss of *Chaserr*, which is lethal in mice, results in upregulation of the adjacent CHD2. This CHD2 overexpression causes downregulation of a subset of genes with particular genomic characteristics: they are the downstream gene in a two-member tandem transcriptional unit, where the upstream neighbour is expressed at fairly high levels. The suggested model is thus that CHD2 overexpression induced by loss of *Chaserr* leads to CHD2 accumulation at transcription termination sites of highly expressed genes. When the promoter of another gene happens to be found immediately downstream, this CHD2 accumulation interferes with its transcription, resulting in downregulation of the downstream neighbour. Intriguingly, the *Chaserr*–CHD2 transcriptional unit is itself arranged in a highly conserved tandem organization, presenting a way for CHD2 to regulate its own levels: when CHD2 levels are high, its increased binding near the *Chaserr* transcription termination site could interfere with the transcription of its own gene downstream, thereby keeping CHD2 levels at bay. Although it is not yet clear how exactly *Chaserr* depletion causes CHD2 overexpression, it is tempting to speculate that this involves aberrant nucleosome positioning at the CHD2 promoter, as described for the *SRG1–SER3* transcriptional unit[61,99]. This possibility is particularly appealing as the CHD2 protein is itself implicated in nucleosome repositioning[104], providing an elegant way for self-regulation of its levels in a *Chaserr*-dependent manner.

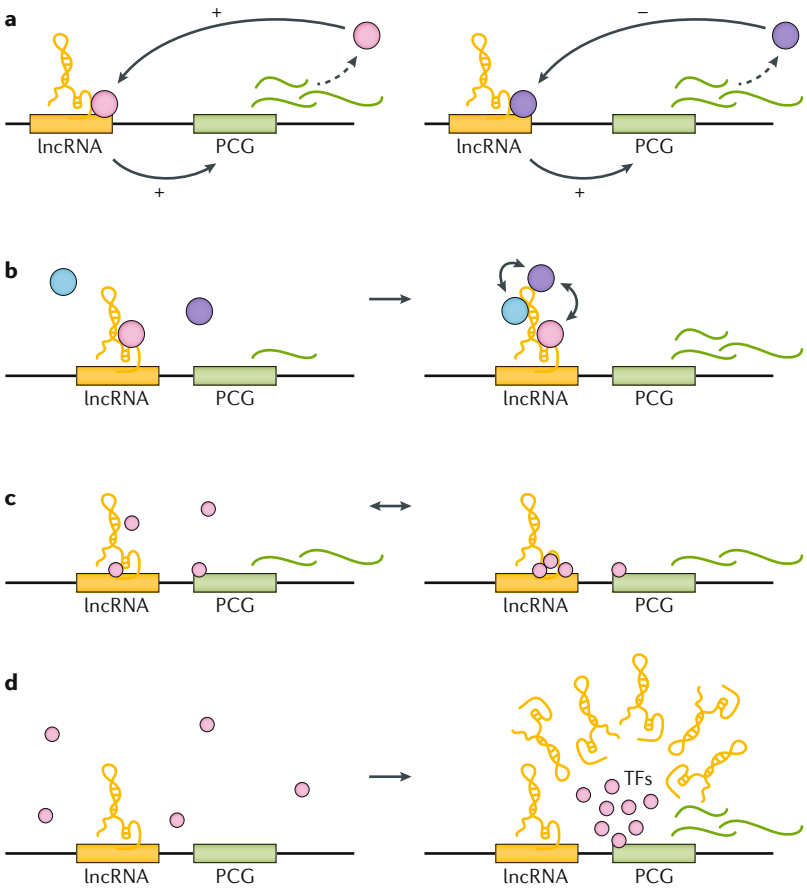


Fig. 4 | **Plausible functions of *cis*-acting long non-coding RNAs.** Various additional mechanisms by which *cis*-acting long non-coding RNAs (lncRNAs) activate or repress their targets are expected to be uncovered as lncRNA research advances. **a** | lncRNAs could form auto-regulatory loops together with transcription factors (TFs) and chromatin modifiers. These may include both positive and negative feedback loops. Yellow structures denote lncRNA transcripts, green wavy lines denote mRNAs, pink circles denote activating factors and purple circles denote repressive factors. **b** | *Cis*-acting lncRNAs could synchronize between the activity of different TFs, for example, by binding a factor that can then recruit another or by serving as a scaffold on which factors can interact with each other. Coloured circles represent various protein factors. **c** | *Cis*-acting lncRNAs transcribed in the vicinity of protein-coding genes (PCGs) could buffer the effects of TFs or transcriptional repressors, for example, by sequestering these factors so that a steady local concentration is maintained that allows steady target gene expression. **d** | lncRNAs may nucleate phase-separated structures around their transcription loci, which could maintain a high local concentration of transcription-related factors — similar to the effects proposed for enhancer hubs.

Mechanisms of action of *cis*-acting lncRNAs

Other criteria by which to categorize *cis*-acting lncRNAs, alongside their effect on their target genes, is by their mechanism of action. A growing number of *cis*-acting lncRNAs, some of which are discussed above, have been studied in depth in recent years, which allows initial conclusions about the general principles by which they function to be drawn. Intriguingly, *cis*-acting lncRNAs, whether activating or repressing, seem to operate through highly similar mechanisms.

The functional feature. An initial question when assigning a mechanism of action to a *cis*-acting lncRNA is which feature of the lncRNA confers the function: the DNA element that encodes the lncRNA[105] (rendering the lncRNA transcript itself irrelevant for the *cis*-acting regulatory function); the RNA product; or the act of RNA transcription or RNA processing[95]. Identifying the functional feature requires careful probing using

Disordered regions
Proteins or regions within proteins that do not adopt an ordered or well-defined 3D structure. These regions can serve as linkers between structured regions, or be functional themselves.

various perturbation techniques. One type of functional feature does not exclude another; an active DNA element, such as an enhancer, can also encode a functional lncRNA. Moreover, a region whose transcription is important can also produce a mature lncRNA transcript with a distinct function. Few such cases of multiple functional features have been described, with the evidence that supports them typically consisting of discordant effects of locus removal and RNA depletion on the expression of genes in *cis* and on additional phenotypes[106,107].

The distinction between the functional features of a lncRNA is not always clear-cut. For example, an elegant study showed that the process of dissociation of the *A-ROD* lncRNA from the chromatin is important for activating transcription of the nearby *DKK1* gene[74]. *A-ROD* is located ~130 kb downstream of *DKK1*, with the two loci being in close, *A-ROD*-independent spatial proximity. Small interfering RNA-mediated *A-ROD* knockdown, which targets only the mature, chromatin-released fraction of the lncRNA, repressed *DKK1* transcription. However, similar *DKK1* repression was achieved when using ASOs targeted to *A-ROD* introns, which specifically target the chromatin-bound fraction without immediately affecting the nucleoplasmic levels. Repression of *DKK1* was also achieved when blocking *A-ROD* splicing or polyadenylation, which presumably results in inhibition of *A-ROD* chromatin dissociation. Together, these results imply that the process of dissociation of mature *A-ROD* from chromatin, rather than strictly its act of transcription or the mature RNA, renders the lncRNA accessible to bind and recruit regulatory proteins, thereby enhancing the expression of *DKK1* (REF.[74]). The conserved splicing signals identified in lncRNAs in general, and in enhancer-associated lncRNAs in particular, might indicate that the process of dissociation from chromatin, which relies on RNA maturation steps, could be important for the activity of many *cis*-acting lncRNAs.

Recruitment or repelling of specific proteins. The majority of functional *cis*-acting lncRNAs seem to function through binding to RNA-binding proteins (RBPs). lncRNAs have been demonstrated to bind to silencing complexes, such as PRC2; activating complexes, such as WDR5–MLL; proteins involved in genome topology, such as cohesin or CTCF; and a multitude of other proteins and protein complexes (TABLE 1) — with some of these interactions validated to be functionally important. Under this scenario, the highly cell type-specific expression of lncRNAs[13,22,23,62] renders them attractive candidates for guiding ubiquitously expressed TFs, chromatin-modifying proteins or genome topology-related proteins to their own transcription locus or its vicinity[108], thereby establishing cell type-specific chromatin states. RNA binding has also been suggested to reinforce, rather than initiate, protein recruitment. For example, a positive feedback loop has been proposed to guide the DNA enrichment of the YY1 TF, whereby nascent RNA captures the protein as it dissociates from the chromatin, promoting its re-binding to nearby DNA elements and augmenting local transcription[109].

Importantly, however, the specificity of protein–RNA interactions of most RBPs is unclear. Up to 25% of all RNAs were discovered to be bound by PRC2 (REFS[110,111]), with the selectivity and specificity of these interactions being a subject of ongoing debate[112–115]. In addition, recent studies that characterized the RNA interactome of cells identified hundreds of proteins not previously known to interact with RNA[116,117]. RBPs include many proteins that lack canonical RNA-binding domains; rather, binding to RNA seems to be mediated by other domains, such as DNA-binding domains that moonlight as RNA-binding domains or disordered regions[118–120]. Whether these non-canonical RBPs can identify specific RNA sequences or structures, or whether they bind RNA indiscriminately, remains unclear. Interestingly, many of the non-canonical RBPs have known functions in chromatin organization and transcriptional regulation, emphasizing the potential that *cis*-acting lncRNAs have in shaping various aspects of chromatin structure and regulation by refining the chromatin association of proteins.

Of note, some *cis*-acting lncRNAs have been proposed to act not by recruiting but rather by repelling protein binding to their locus[121]. For example, upon lipopolysaccharide stimulation, the *PACER* lncRNA, which is transcribed divergently from *COX-2*, binds to the inhibitory p50 and sequesters its binding to the *COX-2* promoter, thereby enabling activation of *COX-2* (REF.[75]). Similarly, it has been proposed that PRC2 binding to RNA prevents it from binding to chromatin, thereby hindering the acquisition of repressive marks[122–125]. Whether the type of protein domain involved could dictate the outcome of the interaction remains to be investigated; for example, if the DNA-binding domain is responsible for both DNA and RNA binding, it is plausible that these events compete with one another, possibly resulting in protein sequestration rather than recruitment.

Local chromatin changes through the act of lncRNA production. In addition to direct binding of the lncRNA transcript by proteins, multiple studies point to a connection between the act of transcription and locus characteristics such as chromatin modifications, topology, motility and accessibility, suggesting that transcription through a lncRNA gene, rather than the lncRNA product, could influence target gene expression. Both Pol II complex members as well as a multitude of Pol II interacting factors have been demonstrated to recruit chromatin-modifying complexes during transcriptional elongation (reviewed elsewhere[126]), which can depose epigenetic modifications at the transcribed locus or in its vicinity. Transcription has also been shown to affect chromatin folding[127,128] and to reposition DNA into the active compartment by disrupting chromatin interactions[30,129]. Relatedly, Pol II transcription has been shown to increase the relative mobility of transcribed loci[130]. Interestingly, another study found that Pol II-mediated transcription hinders locus mobility[131], with the discrepancy possibly due to the length scales at which mobility was assayed. Nonetheless, the joint conclusion is that transcription or transcription-coupled processes bear implications on the rigidity and mobility of the transcribed loci. Thus, the act of transcription through a lncRNA gene

CRISPR–Cas9
A bacterial immune mechanism whereby a Cas9 protein uses short guide RNA (gRNA) sequences to target and cleave foreign DNA. CRISPR–Cas9 can be used for gene editing, by ectopic expression of both Cas9 and a gRNA that targets the gene of interest.

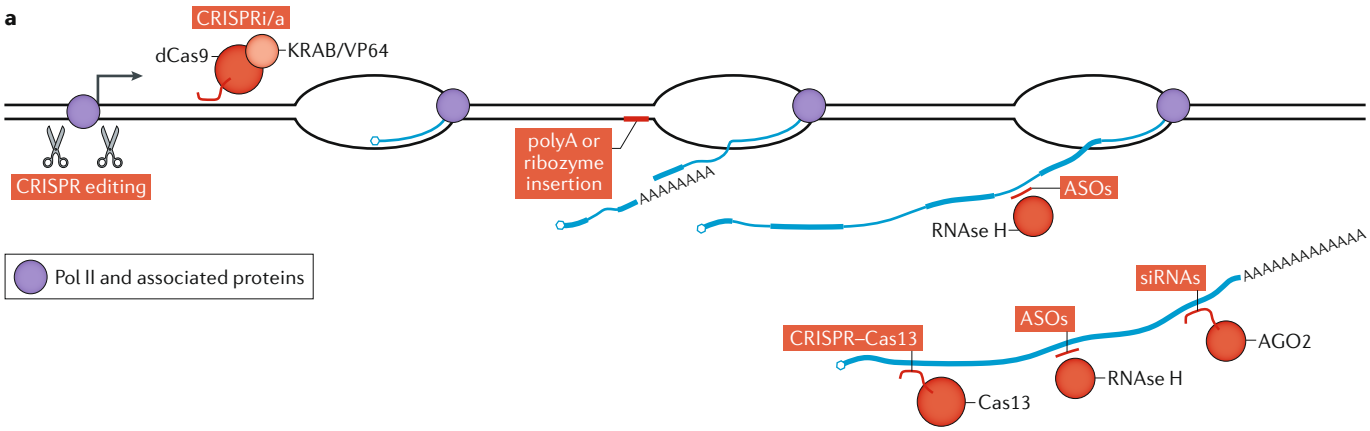
(or any gene) could lead to pronounced effects on the surrounding chromatin, ultimately influencing the transcriptional output of neighbouring genes.

Studying *cis*-acting lncRNA function

An initial indication about the function of a particular *cis*-acting lncRNA can usually be gathered from its expression patterns and genomic location. For example, a lncRNA located adjacent to, and co-expressed with, a key regulatory gene would be a likely candidate for regulating that gene (although sometimes lncRNA perturbation coupled with expression analysis of genes over larger distances identifies other targets)[48,106]. Similarly, a lncRNA transcribed from an enhancer would be a candidate for modulating the activity of that enhancer. Therefore, the first step often consists of perturbing the expression of the lncRNA, followed by assessment of the levels of the putative target genes, as well as assaying the chromatin environment and 3D structure of its genomic locus in perturbed cells, so as to identify changes that might explain the effects on target genes.

Mapping functional features in cis-acting lncRNAs. Various perturbation techniques can be used to study both the function and the functional feature of the lncRNA. As each method has both advantages and disadvantages, and as different and even contradictory results can be reached through different perturbations (as described for the perturbation of *Hand2os1* expression through deletion versus polyA insertion)[59,73], an informative combination of perturbation techniques should be employed, as well as careful interpretation of their respective results[132].

The use of CRISPR–Cas9-related techniques[133,134] has substantially improved our ability to study lncRNAs. CRISPR knockouts aimed at abolishing the expression of a lncRNA usually target a region encompassing the first exon of the lncRNA (FIG. 5). Care should be taken to ensure that no alternative promoters come into play[135]. Although such knockouts should target as small a genomic region as possible — generally not beyond a few hundred base pairs — the resulting phenotype could nonetheless be due to disruption of the DNA element rather than prevention of lncRNA transcription or mature lncRNA production. Instead, CRISPRi can be used to prevent transcription of the lncRNA — by either steric hindrance of the Pol II machinery or local modulation of the chromatin environment — without changing the underlying DNA sequence. CRISPR activation (CRISPRa) is also a particularly useful tool for



b

	DNA	Transcription initiation	Elongation and RNA processing	Mature RNA
CRISPR knockout (full body or promoter)	✓	✓	✓	✓
CRISPRi or CRISPRa	✗[a]	✓	✓	✓
polyA or ribozyme insertion	✗	✗	?[b]	✓
RNAi or CRISPR–Cas13	✗	✗	✗[b]	✓

[a]Unless it affects the chromatin environment. [b]Depending on the location of the insertion or targeting.

Fig. 5 | **Methods for mapping functional features in *cis*-acting long non-coding RNAs. a** | Various types of perturbations can be used to probe the function and identify the functional feature of *cis*-acting long non-coding RNAs (lncRNAs), which target the lncRNA at different stages. CRISPR editing entails deletion of the first exon or a particular region within the target sequence by a Cas9 protein, which is recruited by guide RNAs (gRNAs). For CRISPR interference (CRISPRi), a catalytically dead Cas9 (dCas9) is used to reduce expression of a gene rather than induce DNA breaks. This is oftentimes achieved by fusing dCas9 to a repressive effector protein, such as a KRAB domain, although recruitment of dCas9 alone can sometimes be sufficient for inducing silencing, likely through steric inhibition of RNA polymerase II (Pol II). Similar to CRISPRi, CRISPR activation (CRISPRa) entails recruitment of a dCas9 fusion protein to a target genomic locus, except the Cas9 is fused to an activator domain, commonly the VP64 activation domain. Insertion of a polyadenylation (polyA) sequence causes premature, co-transcriptional cleavage of the nascent transcript and halts transcription. Self-cleaving ribozyme sequences degrade the RNA and may also affect continuance of transcription. RNA knockdown can be achieved by RNA interference (RNAi) using either small interfering RNAs (siRNAs), antisense oligonucleotides (ASOs) or the RNA-targeting enzyme Cas13. **b** | Correspondence between the functional feature of a lncRNA locus and the expected outcome of various perturbation techniques. AGO2, protein argonaute 2.

studying the effects of *cis* activation of lncRNAs (FIG. 5). However, a major caveat of CRISPRi and CRISPRa is that the observed phenotypes might be due to chromatin changes rather than prevention of lncRNA transcription. This is especially true if the lncRNA overlaps a regulatory element, which can itself be affected by the recruitment of activating or repressing proteins, although spreading of epigenetic modifications can also lead to changes in the chromatin environment of the target genes. Indeed, the catalytically inactive ('dead') Cas9 (dCas9) fused with KRAB or VP64 that are most commonly used for CRISPRi and CRISPRa, respectively, are also used for mapping enhancer elements[136,137].

Novel CRISPR techniques are continually being developed, including some that are expected to be useful for lncRNA research. For example, the recently discovered Cas13, a CRISPR family member that serves as an RNA targeting enzyme, has been demonstrated to be useful for RNA knockdown through guide RNA-mediated target identification and cleavage[138] (FIG. 5). Importantly, Cas13-mediated knockdown was suggested to be as efficient as, but much more specific than, antisense-based knockdown methods, reducing the potential for off-target effects, although this report awaits support from independent studies. Similarly to dCas9 fusion proteins, future studies can be imagined that utilize catalytically inactive Cas13 proteins fused to various proteins that modulate target RNA activity, such as RNA editing proteins[139], which will facilitate identification of active elements within the lncRNA sequence. In addition, CRISPR display can be used for probing the functionality of mature lncRNA transcripts by recruiting them to ectopic sites and measuring their effect on gene expression. This method has been successfully used to show that, among others, the A repeat of *Xist* and the *HOTTIP* lncRNA have a repressive and an activating effect, respectively, when recruited to the promoter of a reporter gene[140]. Nonetheless, these effects were very modest, as were similar attempts at ectopically tethering eRNAs[140,141], suggesting either that (some) lncRNAs lose functionality when detached from their transcription sites or that this technique requires further calibration and adaptation so as to allow full reconstitution of lncRNA functionality.

To study the consequences of lncRNA transcription more directly and separate them from both the underlying DNA and the function of the mature lncRNA, several studies have made use of insertion of polyA sequences or, more recently, self-cleaving ribozyme sequences[30,47,142] (FIG. 5). Both methods entail insertion of a sequence — generally in the order of 50–800 bp — that causes premature, co-transcriptional cleavage of the nascent transcript and discontinuance of transcription, the main benefit being no potential for off-target effects. Self-cleaving ribozymes may prove to be a particularly useful tool for the study of lncRNA biology, as their effects can potentially be induced and/or reversed by addition of a signal or by using ASOs that block the ribozyme sequence, respectively[142,143]. However, the efficacy of ribozymes is currently variable[142], and the rules for the design of effective ribozyme insertions require further study.

It is noteworthy that the effects of 'the act of transcription' may be specific to transcription through a particular region within the lncRNA gene. Therefore, an additional consideration when analysing results of polyA or ribozyme sequence insertions is the exact insertion place, as well as the distance to which Pol II continues transcribing after a polyA or ribozyme sequence has been encountered, and the dynamics at which the nascent transcripts are then cleaved. Pol II continues to transcribe for ~2.5–3 kb after encounter of a polyA site[144]. The effect of ribozyme-mediated cleavage on the continuance of transcription is less clear and might depend on local sequence attributes that dictate folding dynamics. Integration sites should thus be chosen with awareness of these constraints. The importance of proper placement of polyA sequences is elegantly demonstrated in a study of the *Airn* lncRNA[95].

RNA interference (RNAi)-related techniques can also be used to study lncRNA functionality (FIG. 5). Small interfering RNAs or short hairpin RNAs are thought to operate mainly in the cytoplasm, although there are some reports of the presence of Argonaute proteins in the nucleus and their utility in targeting of nuclear RNAs, including lncRNAs[74,145]. Nonetheless, it is unclear how sensitive the chromatin-associated fraction of an RNA is to such perturbations, which in the case of *cis*-acting lncRNAs is likely the active fraction. Various types of ASOs, by contrast, have been shown to be effective in reducing levels of nuclear RNAs, including chromatin-associated RNAs[29,47,74,146], especially when targeted to intronic regions of the RNA. However, it is important to note that in cases where RNAi-based interference targets the RNA co-transcriptionally, cleavage of nascent transcripts might lead to Pol II drop-off some distance after the cleavage site, and so similar considerations and interpretations to those described above for polyA or ribozyme sequence insertions are relevant. In addition, a major caveat of antisense-based techniques is the as yet unclear propensity for and rules of off-target hybridizations, as well as the possible activation of a cellular immune response[147]. Although these shortcomings can be somewhat abrogated by using various types of chemically modified oligonucleotides, or using multiple ASOs targeting the same gene[147,148], the results should nonetheless be interpreted with caution. Generally, when employing any type of RNAi-based techniques, it is important to also ensure appropriate controls, such as multiple non-targeting sequences, are used.

Mapping interactions of cis-acting lncRNAs. Methods being used to study the mechanisms of action of *cis*-acting lncRNAs focus on interactions of the lncRNA with DNA (so as to identify sites at which it could be active) or with proteins (so as to identify potential interacting partners), such as RNA antisense purification[149]. Such methods have been thoroughly discussed in a recent review[150]; briefly, they rely on probes that target an RNA of interest, followed by identification of bound proteins through mass spectrometry or of bound DNA regions through PCR or high-throughput sequencing. Such methods should be complemented by reciprocal pull-downs of identified proteins so as to validate interaction with the lncRNA.

CRISPR display
Utilization of CRISPR–Cas9 for the recruitment of non-protein components. For example, long non-coding RNA sequences can be fused to the guide RNA and be brought to the target locus via 'dead' Cas9.

Self-cleaving ribozyme sequences
RNA sequences that can catalyse a reaction that would cut their own RNA.

RNA interference
(RNAi). Short non-coding RNA molecules — either microRNAs or short interfering RNAs — bind to complementary sequences in the target genes, leading to translation inhibition or target RNA degradation.

Similar methods have recently been developed for the investigation of multiple *cis*-acting lncRNA genes in parallel. For example, global RNA interaction with DNA sequencing (GRID-seq)[151] entails global RNA–DNA ligation followed by fragmentation and sequencing, so as to identify the subset of chromatin-bound RNAs as well as the loci to which they bind. Mapping RNA–genome interactions (MARGI), chromatin-associated RNA sequencing (ChAR-seq) and RNA and DNA interacting complexes ligated and sequenced (RADICL-seq) are based on similar principles[152–154] (reviewed elsewhere[150]). All four techniques uncovered substantial chromatin enrichment of many RNA species, including both coding and non-coding RNAs. Improvements in the sensitivity and resolution of such techniques, alongside complementary approaches, should help identify the fraction of *cis*-acting RNAs whose association with focal target loci is functionally significant, and does not merely reflect spreading to loci found in close spatial proximity to their sites of transcription. Nonetheless, a few general observations can already be made from existing data. For example, GRID-seq identified a particular enrichment of chromatin-associated RNAs around super-enhancers, with the level of RNA association correlating with increased expression of neighbouring genes[151], reinforcing the positive association between enhancer activity and non-coding RNA production described above. Combined with individual mechanistic studies, such experiments could begin to shed light on general *cis*-acting lncRNA characteristics such as their expression levels, their distance from their target genes, shared features of the regions they bind and more, thus enabling general conclusions about the functionality of this class of lncRNAs to be drawn.

Conclusions and perspectives

Deep sequencing of transcriptomes from diverse tissues and cell types has facilitated the annotation of tens of thousands of lncRNAs in mammalian genomes. Although the fraction of lncRNAs that are functional – that is, confer any type of fitness advantage – is not yet known, even the most modest estimates place this number at hundreds of transcripts. *Cis*-acting lncRNAs likely comprise a substantial subgroup of these, and as such plausibly affect gene regulation through various mechanisms (FIGS 2–4). Regulating the release of lncRNAs from chromatin, as well as their transport within or export out of the nucleus, can all help fine-tune such *cis* activities, and likely affect the abilities of different *cis*-acting lncRNAs to act on genes within their immediate neighbourhood, their TAD, across a few megabases or across an entire chromosome. Some progress has been made towards identifying the sequence elements that dictate the enrichment of long RNAs on chromatin or in the nucleus[155–157], but the overall rules governing the post-transcriptional fate of such RNAs remain largely unknown.

Within gene regulatory networks, lncRNAs can act alongside *cis* factors, such as DNA regulatory elements and epigenetic modifications, as well as *trans* factors, such as TFs and small non-coding RNAs. Compared with these *trans* factors, which are translated or loaded into active complexes in the cytoplasm, lncRNAs have a unique ability to exert their function in *cis*, as they are targeted by the act of their biogenesis to a specific locus. As other functional *cis*-acting elements such as enhancers clearly outnumber functional *cis*-acting lncRNAs, and as most lncRNAs are poorly conserved in evolution (BOX 1), it seems that lncRNAs are used for gene control only in specific scenarios. It is interesting to speculate what relative advantages lncRNA-mediated regulation might entail: for example, the ability to coordinate regulation of larger, cross-TAD regions; the ability to confer transient effects via regulation of lncRNA stability; and the ability to nucleate phase-separated organelles via multivalent binding of other factors (FIG. 4). By contrast, the limited abundance of most lncRNAs, coupled with the stochasticity of gene expression, makes them less suitable for performing 'housekeeping' gene regulatory activities. For example, it is unlikely that cells would often rely on the presence of a specific low-abundance lncRNA molecule for the formation of a critical chromatin interaction. As lncRNAs are produced by the same machinery that transcribes and processes mRNAs, they also possess unique abilities to sense the state of different components of this system, and so are particularly well suited to participate in regulatory feedback loops with products of neighbouring chromatin-related genes. Importantly, the unique features of *cis*-acting lncRNAs, combined with the ability to effectively degrade RNA using antisense or CRISPR-based tools, render them attractive candidates for targeted therapeutics through highly specific activation or repression of target genes[158].

As we attempted to illustrate above, the complex web of interdependencies between gene regulatory processes makes it particularly difficult to pinpoint the direct consequences of lncRNA transcription. Although the notion of a chromatin-tethered lncRNA recruiting general TFs or chromatin modifiers to the vicinity of its locus is an appealing candidate for a common and straightforward mechanism of action, the apparent lack of specificity and selectivity exhibited by most proteins implicated in such functions points to a more complicated scenario. lncRNA production could assist recruitment of activating or repressive complexes to target genes indirectly, such as by affecting nuclear position or motility of the locus, which in turn affects the availability of the loci to form chromatin or protein interactions. Thus, lncRNA binding might follow or reinforce protein recruitment, rather than serving a main or a direct role in this process. Distinguishing between the direct and indirect effects of lncRNA production is particularly challenging as these are often highly correlated with each other: inhibiting the transcription of a locus could affect its nuclear positioning, chromatin looping, epigenetic modifications and interacting proteins, with each of these processes affecting each other, making a direct order of events hard to discern.

Furthermore, the study of lncRNA functionality is complicated by the high dependency of the observed effects on the method by which lncRNA production is perturbed. Most lncRNA studies to date have employed only one or few techniques, making it hard to identify with certainty the functional feature of a lncRNA as well

as the consequences of its production. The increasing number of studies which utilize multiple perturbation techniques to study individual lncRNAs, as well as more in-depth characterizations of previously described lncRNAs, are expected to help unravel general principles of lncRNA functionality.

Cis effects on transcription of nearby genes imposed by the production of a lncRNA – regardless of the functional feature or order of events that mediate them – are not limited to lncRNAs. For example, some mRNA loci have been demonstrated to serve as enhancers that modulate the expression of genes found in *cis*[33,159], presumably through similar mechanisms to those employed by *cis*-acting lncRNAs. Relatedly, the GRID-seq technique[151] identified that the majority of chromatin-associated RNAs are in fact pre-mRNAs, which are enriched around their sites of transcription, suggesting that these pre-mRNAs may participate in gene regulation prior to export and translation. All in all, these observations highlight that the interplay between transcription of neighbouring genes is a highly prevalent mechanism for the control of gene expression. Although the study of *cis*-acting lncRNAs provides a 'cleaner' setting by which to study such processes, not complicated by perturbation of additional mRNA functionalities, the conclusions are likely applicable to additional types of transcripts.

Published online: 15 November 2019

1. Guttman, M. et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
2. Mercer, T. R. et al. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.* **30**, 99–104 (2011).
3. Carninci, P. et al. The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).
4. Samata, M. & Akhtar, A. Dosage compensation of the X chromosome: a complex epigenetic assignment involving chromatin regulators and long noncoding RNAs. *Annu. Rev. Biochem.* **87**, 323–350 (2018).
5. Sherstyuk, V. V., Medvedev, S. P. & Zakian, S. M. Noncoding RNAs in the regulation of pluripotency and reprogramming. *Stem Cell Rev.* **14**, 58–70 (2018).
6. Quinodoz, S. & Guttman, M. Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. *Trends Cell Biol.* **24**, 651–663 (2014).
7. Lin, C. & Yang, L. Long noncoding RNA in cancer: wiring signaling circuitry. *Trends Cell Biol.* **28**, 287–301 (2018).
8. Scheuermann, J. C. & Boyer, L. A. Getting to the heart of the matter: long non-coding RNAs in cardiac development and disease. *EMBO J.* **32**, 1805–1816 (2013).
9. Guo, J., Liu, Z. & Gong, R. Long noncoding RNA: an emerging player in diabetes and diabetic kidney disease. *Clin. Sci.* **133**, 1321–1339 (2019).
10. Zuckerman, B. & Ulitsky, I. Predictive models of subcellular localization of long RNAs. *RNA* **25**, 557–572 (2019).
11. Tilgner, H. et al. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* **22**, 1616–1625 (2012).
12. Melé, M. et al. Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Res.* **27**, 27–37 (2017).
13. 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).
14. Tichon, A. et al. A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nat. Commun.* **7**, 12209 (2016).
15. Lee, S. et al. Noncoding RNA *NORAD* regulates genomic stability by sequestering PUMILIO proteins. *Cell* **164**, 69–80 (2016).
16. Musgrove, C., Jansson, L. I. & Stone, M. D. New perspectives on telomerase RNA structure and function. *WIREs RNA* **9**, e1456 (2018).
17. Vendramin, R. et al. SAMMSON fosters cancer cell fitness by concertedly enhancing mitochondrial and cytosolic translation. *Nat. Struct. Mol. Biol.* **25**, 1035–1046 (2018).
18. Grote, P. et al. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev. Cell* **24**, 206–214 (2013).
19. Huarte, M. et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409–419 (2010).
20. Andersen, R. E. et al. The long noncoding RNA Pnky is a *trans*-acting regulator of cortical development in vivo. *Dev. Cell* **49**, 632–642.e7 (2019).
21. Werner, M. S. & Ruthenburg, A. J. Nuclear fractionation reveals thousands of chromatin-tethered noncoding RNAs adjacent to active genes. *Cell Rep.* **12**, 1089–1098 (2015).
This study uses nuclear fractionations to show that the majority of lncRNAs, including many previously unannotated intergenic transcripts, are enriched in the chromatin fraction. These chromatin-enriched lncRNAs are dependent on Pol II elongation for their association with chromatin.
22. 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).
23. Pauli, A. et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* **22**, 577–591 (2012).
24. Ulitsky, I. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat. Rev. Genet.* **17**, 601–614 (2016).
25. Hacisuleyman, E. et al. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat. Struct. Mol. Biol.* **21**, 198–206 (2014).
26. Lewandowski, J. P. et al. The *Firre* locus produces a *trans*-acting RNA molecule that functions in hematopoiesis. Preprint at *bioRxiv* https://doi.org/10.1101/648279 (2019).
27. Clemson, C. M. et al. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol. Cell* **33**, 717–726 (2009).
28. Naganuma, T. et al. Alternative 3′-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. *EMBO J.* **31**, 4020–4034 (2012).
29. Rom, A. et al. Regulation of CHD2 expression by the *Chaserr* long noncoding RNA is essential for viability. *Nat. Commun.* **10**, 5092 (2019).
30. Isoda, T. et al. Non-coding transcription instructs chromatin folding and compartmentalization to dictate enhancer–promoter communication and T cell fate. *Cell* **171**, 103–119.e18 (2017).
This study characterizes the enhancer-transcribed *ThymoD* lncRNA, whose transcription promotes the activity of the underlying enhancer through repositioning it away from the nuclear periphery, demonstrating the interdependence between (lncRNA) transcription, nuclear positioning and enhancer activity.
31. Perry, R. B., Hezroni, H., Goldrich, M. J. & Ulitsky, I. Regulation of neuroregeneration by long noncoding RNAs. *Mol. Cell* **72**, 553–567 (2018).
32. Kotzin, J. J. et al. The long non-coding RNA *Morrbid* regulates Bim and short-lived myeloid cell lifespan. *Nature* **537**, 239–243 (2016).
33. Engreitz, J. M. et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452–455 (2016).
This study demonstrates that promoters of some lncRNAs, but also of some mRNAs, affect the expression of neighbouring genes, emphasizing that studies of the functions of *cis*-acting lncRNAs are also applicable to other transcriptional units.
34. Vučićević, D., Corradin, O., Ntini, E., Scacheri, P. C. & Ørom, U. A. Long ncRNA expression associates with tissue-specific enhancers. *Cell Cycle* **14**, 253–260 (2015).
35. De Santa, F. et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLOS Biol.* **8**, e1000384 (2010).
36. Marques, A. C. et al. Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs. *Genome Biol.* **14**, R131 (2013).
Through analysis of chromatin modifications at regions of transcription initiation, this study notes two distinct classes of intergenic lncRNAs, one with characteristics of enhancers and one with characteristics of promoters.
37. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **12**, 2478–2492 (2017).
38. Gil, N. & Ulitsky, I. Production of spliced long noncoding RNAs specifies regions with increased enhancer activity. *Cell Syst.* **7**, 537–547.e3 (2018).
39. 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).
40. Wamstad, J. A. et al. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* **151**, 206–220 (2012).
41. Ponjavic, J., Oliver, P. L., Lunter, G. & Ponting, C. P. Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. *PLOS Genet.* **5**, e1000617 (2009).
42. Schwanhausser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
43. Kutter, C. et al. Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLOS Genet.* **8**, e1002841 (2012).
44. Werner, M. S. et al. Chromatin-enriched lncRNAs can act as cell-type specific activators of proximal gene transcription. *Nat. Struct. Mol. Biol.* **24**, 596–603 (2017).
45. Engreitz, J. M. et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
46. Schertzer, M. D. et al. lncRNA-induced spread of polycomb controlled by genome architecture, RNA abundance, and CpG island DNA. *Mol. Cell* **75**, 523–537.e10 (2019).
47. Ballarino, M. et al. Deficiency in the nuclear long noncoding RNA causes myogenic defects and heart remodeling in mice. *EMBO J.* **37**, e99697 (2018).
48. Groff, A. F., Barutcu, A. R., Lewandowski, J. P. & Rinn, J. L. Enhancers in the Peril lincRNA locus regulate distant but not local genes. *Genome Biol.* **19**, 219 (2018).
49. Amaral, P. P. et al. Genomic positional conservation identifies topological anchor point RNAs linked to developmental loci. *Genome Biol.* **19**, 32 (2018).
50. Tan, J. Y. et al. *Cis*-acting complex-trait-associated lincRNA expression correlates with modulation of chromosomal architecture. *Cell Rep.* **18**, 2280–2288 (2017).
51. Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166 (2012).
52. Ma, L., Bajic, V. B. & Zhang, Z. On the classification of long non-coding RNAs. *RNA Biol.* **10**, 925–933 (2013).
53. Fanucchi, S. et al. Immune genes are primed for robust transcription by proximal long noncoding RNAs located in nuclear compartments. *Nat. Genet.* **51**, 138–150 (2019).
This article presents a well-studied example of a lncRNA that activates expression of its target gene

through recruitment of activating factors — in this case, the WDR5–MLL complex — to the target promoter, utilizing pre-existing spatial proximity.
54. 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).
This work is one of the first examples of an enhancer-transcribed lncRNA that affects the expression of its target gene through promoting the formation of enhancer–promoter chromatin loops, in this case through recruitment of CTCF.
55. Sleutels, F., Zwart, R. & Barlow, D. P. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **415**, 810–813 (2002).
56. Kherdjemil, Y. et al. Evolution of Hoxa11 regulation in vertebrates is linked to the pentadactyl state. *Nature* **539**, 89–92 (2016).
57. Pandey, R. R. et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* **32**, 232–246 (2008).
58. Tahira, A. C. et al. Long noncoding intronic RNAs are differentially expressed in primary and metastatic pancreatic cancer. *Mol. Cancer* **10**, 141 (2011).
59. Anderson, K. M. et al. Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. *Nature* **539**, 433–436 (2016).
60. Frank, S. et al. yylncT defines a class of divergently transcribed lncRNAs and safeguards the T-mediated mesodermal commitment of human PSCs. *Cell Stem Cell* **24**, 318–327.e8 (2019).
61. Martens, J. A., Laprade, L. & Winston, F. Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature* **429**, 571–574 (2004).
62. Bergmann, J. H. et al. Regulation of the ESC transcriptome by nuclear long noncoding RNAs. *Genome Res.* **25**, 1336–1346 (2015).
63. Heintzman, N. D. et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
64. Xi, H. et al. Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLOS Genet.* **3**, e136 (2007).
65. Chen, H., Du, G., Song, X. & Li, L. Non-coding transcripts from enhancers: new insights into enhancer activity and gene expression regulation. *Genomics Proteomics Bioinformatics* **15**, 201–207 (2017).
66. Ding, M. et al. Enhancer RNAs (eRNAs): new insights into gene transcription and disease treatment. *J. Cancer* **9**, 2334–2340 (2018).
67. Li, W., Notani, D. & Rosenfeld, M. G. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat. Rev. Genet.* **17**, 207–223 (2016).
68. Koch, F. et al. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.* **18**, 956–963 (2011).
69. Lai, F. et al. Activating RNAs associate with mediator to enhance chromatin architecture and transcription. *Nature* **494**, 497–501 (2013).
70. Nozawa, R.-S. et al. SAF-A regulates interphase chromosome structure through oligomerization with chromatin-associated RNAs. *Cell* **169**, 1214–1227. e18 (2017).
71. Xiao, R. et al. Nuclear matrix factor hnRNP U/SAF-A exerts a global control of alternative splicing by regulating U2 snRNP maturation. *Mol. Cell* **45**, 656–668 (2012).
72. Elling, R. et al. Genetic models reveal *cis* and *trans* immune-regulatory activities for lincRNA-Cox2. *Cell Rep.* **25**, 1511–1524.e6 (2018).
73. Han, X. et al. The lncRNA *Hand2os1/Uph* locus orchestrates heart development through regulation of precise expression of *Hand2*. *Development* **146**, dev176198 (2019).
74. Ntini, E. et al. Long ncRNA A-ROD activates its target gene DKK1 at its release from chromatin. *Nat. Commun.* **9**, 1636 (2018).
75. Krawczyk, M. & Emerson, B. M. p50-associated COX-2 extragenic RNA (PACER) activates COX-2 gene expression by occluding repressive NF-κB complexes. *eLife* **3**, e01776 (2014).
76. Wang, K. C. et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
77. Tan, J. Y., Biasini, A., Young, R. S. & Marques, A. An unexpected contribution of lincRNA splicing to enhancer function. Preprint at *bioRxiv* https://doi.org/10.1101/287706 (2018).
78. Haerty, W. & Ponting, C. P. Unexpected selection to retain high GC content and splicing enhancers within exons of multiexonic lncRNA loci. *RNA* **21**, 333–346 (2015).
79. Schüler, A., Ghanbarian, A. T. & Hurst, L. D. Purifying selection on splice-related motifs, not expression level nor RNA folding, explains nearly all constraint on human lincRNAs. *Mol. Biol. Evol.* **31**, 3164–3183 (2014).
80. Lamond, A. I. & Spector, D. L. Nuclear speckles: a model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.* **4**, 605–612 (2003).
81. Kim, J., Khanna, N. & Belmont, A. S. Transcription amplification by nuclear speckle association. Preprint at *bioRxiv* https://doi.org/10.1101/604298 (2019).
82. Ding, F. & Elowitz, M. B. Constitutive splicing and economies of scale in gene expression. *Nat. Struct. Mol. Biol.* **26**, 424–432 (2019).
83. Fong, Y. W. & Zhou, Q. Stimulatory effect of splicing factors on transcriptional elongation. *Nature* **414**, 929–933 (2001).
84. Sahakyan, A., Yang, Y. & Plath, K. The role of xist in X-chromosome dosage compensation. *Trends Cell Biol.* **28**, 999–1013 (2018).
85. Robert Finestra, T. & Gribnau, J. X chromosome inactivation: silencing, topology and reactivation. *Curr. Opin. Cell Biol.* **46**, 54–61 (2017).
86. da Rocha, S. T. & Heard, E. Novel players in X inactivation: insights into Xist-mediated gene silencing and chromosome conformation. *Nat. Struct. Mol. Biol.* **24**, 197–204 (2017).
87. Swiezewski, S., Liu, F., Magusin, A. & Dean, C. Cold-induced silencing by long antisense transcripts of an arabidopsis polycomb target. *Nature* **462**, 799–802 (2009).
88. Rosa, S., Duncan, S. & Dean, C. Mutually exclusive sense–antisense transcription at FLC facilitates environmentally induced gene repression. *Nat. Commun.* **7**, 13031 (2016).
89. Whittaker, C. & Dean, C. The FLC locus: a platform for discoveries in epigenetics and adaptation. *Annu. Rev. Cell Dev. Biol.* **33**, 555–575 (2017).
90. Beckedorff, F. C. et al. The intronic long noncoding RNA ANRASSF1 recruits PRC2 to the RASSF1A promoter, reducing the expression of RASSF1A and increasing cell proliferation. *PLOS Genet.* **9**, e1003705 (2013).
91. Cho, S. W. et al. Promoter of lncRNA gene PVT1 is a tumor-suppressor DNA boundary element. *Cell* **173**, 1398–1412.e22 (2018).
This article presents an example of a lncRNA gene that represses its target through enhancer competition: the *PVT1* promoter competes with the *MYC* promoter on association with and activation by proximally situated enhancers.
92. Wang, F. et al. Oncofetal long noncoding RNA PVT1 promotes proliferation and stem cell-like property of hepatocellular carcinoma cells by stabilizing NOP2. *Hepatology* **60**, 1278–1290 (2014).
93. Tseng, Y. Y. et al. PVT1 dependence in cancer with MYC copy-number increase. *Nature* **512**, 82–86 (2014).
94. Xiao, M., Feng, Y., Liu, C. & Zhang, Z. Prognostic values of long noncoding RNA PVT1 in various carcinomas: an updated systematic review and meta-analysis. *Cell Prolif.* **51**, e12519 (2018).
95. Latos, P. A. et al. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* **338**, 1469–1472 (2012).
This study is a compelling demonstration that transcription through specific regions of a lncRNA gene can be functionally significant, and thus demonstrates the importance of careful selection of integration sites for sequences that interfere with transcription.
96. Hongay, C. F., Grisafi, P. L., Galitski, T. & Fink, G. R. Antisense transcription controls cell fate in *Saccharomyces cerevisiae*. *Cell* **127**, 735–745 (2006).
97. van Werven, F. J. et al. Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. *Cell* **150**, 1170–1181 (2012).
98. Bumgarner, S. L., Dowell, R. D., Grisafi, P., Gifford, D. K. & Fink, G. R. Toggle involving *cis*-interfering noncoding RNAs controls variegated gene expression in yeast. *Proc. Natl Acad. Sci. USA* **106**, 18321–18326 (2009).
99. Thebault, P. et al. Transcription regulation by the noncoding RNA SRG1 requires Spt2-dependent chromatin deposition in the wake of RNA polymerase II. *Mol. Cell. Biol.* **31**, 1288–1300 (2011).
100. Canzio, D. et al. Antisense lncRNA transcription mediates DNA demethylation to drive stochastic protocadherin α promoter choice. *Cell* **177**, 639–653. e15 (2019).
101. Bumgarner, S. L. et al. Single-cell analysis reveals that noncoding RNAs contribute to clonal heterogeneity by modulating transcription factor recruitment. *Mol. Cell* **45**, 470–482 (2012).
102. Ritter, N. et al. The lncRNA locus *Handsdown* regulates cardiac gene programs and is essential for early mouse development. *Dev. Cell* **50**, 644–657.e8 (2019).
103. Richard, J. L. C. & Ogawa, Y. Understanding the complex circuitry of lncRNAs at the X-inactivation center and its implications in disease conditions. *Curr. Top. Microbiol. Immunol.* **394**, 1–27 (2016).
104. Tajul-Arifin, K. et al. Identification and analysis of chromodomain-containing proteins encoded in the mouse transcriptome. *Genome Res.* **13**, 1416–1429 (2003).
105. Paralkar, V. R. et al. Unlinking an lncRNA from its associated *cis* element. *Mol. Cell* **62**, 104–110 (2016).
106. Alexanian, M. et al. A transcribed enhancer dictates mesendoderm specification in pluripotency. *Nat. Commun.* **8**, 1806 (2017).
107. Carmona, S., Lin, B., Chou, T., Arroyo, K. & Sun, S. lncRNA Jpx induces xist expression in mice using both *trans* and *cis* mechanisms. *PLOS Genet.* **14**, e1007378 (2018).
108. Nagano, T. et al. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717–1720 (2008).
109. Sigova, A. A. et al. Transcription factor trapping by RNA in gene regulatory elements. *Science* **350**, 978–981 (2015).
110. 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).
111. Zhao, J. et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell* **40**, 939–953 (2010).
112. Davidovich, C., Zheng, L., Goodrich, K. J. & Cech, T. R. Promiscuous RNA binding by Polycomb repressive complex 2. *Nat. Struct. Mol. Biol.* **20**, 1250–1257 (2013).
This study finds that PRC2 binds RNA with high affinity but low specificity both in vitro and in vivo, with its promiscuous binding patterns correlating with RNA lengths or expression levels, and with no apparent specificity for biologically significant interactions. This casts doubt on the proposed role of many lncRNAs in recruiting PRC2 to their transcription sites.
113. Cifuentes-Rojas, C., Hernandez, A. J., Sarma, K. & Lee, J. T. Regulatory interactions between RNA and polycomb repressive complex 2. *Mol. Cell* **55**, 171–185 (2014).
This study notes that whereas the EZH2 subunit of PRC2 displays high-affinity RNA binding but with low specificity, other subunits – mostly EED – increase the selectivity of the complex's RNA interactions.
114. Wang, X. et al. Targeting of polycomb repressive complex 2 to RNA by short repeats of consecutive guanines. *Mol. Cell* **65**, 1056–1067.e5 (2017).
115. Colognori, D., Sunwoo, H., Kriz, A. J., Wang, C.-Y. & Lee, J. T. Xist deletional analysis reveals an interdependency between Xist RNA and polycomb complexes for spreading along the inactive X. *Mol. Cell* **74**, 101–117.e10 (2019).
116. Baltz, A. G. et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **46**, 674–690 (2012).
117. Perez-Perri, J. I. et al. Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. *Nat. Commun.* **9**, 4408 (2018).
118. He, C. et al. High-resolution mapping of RNA-binding regions in the nuclear proteome of embryonic stem cells. *Mol. Cell* **64**, 416–430 (2016).
119. Castello, A. et al. Comprehensive identification of RNA-binding domains in human cells. *Mol. Cell* **63**, 696–710 (2016).
In this study, the authors develop RBDmap, a strategy for identifying RNA-binding regions in proteins through ultraviolet crosslinking, oligo(dT) capture and identification of crosslinked peptides using mass spectrometry. This and similar methods detailed in Baltz et al., Perez-Perri et al and He et al. unravelled a myriad of known and novel RNA-binding proteins and domains.

120. Hansen, A. S. et al. Distinct classes of chromatin loops revealed by deletion of an RNA-binding region in CTCF. *Mol. Cell* **76**, 395–411.e13 (2019).
121. Sun, S. et al. Jpx RNA activates Xist by evicting CTCF. *Cell* **153**, 1537–1551 (2013).
122. Kaneko, S., Son, J., Bonasio, R., Shen, S. S. & Reinberg, D. Nascent RNA interaction keeps PRC2 activity poised and in check. *Genes Dev.* **28**, 1983–1988 (2014).
123. Beltran, M. et al. The interaction of PRC2 with RNA or chromatin is mutually antagonistic. *Genome Res.* **26**, 896–907 (2016).
124. Hosogane, M., Funayama, R., Shirota, M. & Nakayama, K. Lack of transcription triggers H3K27me3 accumulation in the gene body. *Cell Rep.* **16**, 696–706 (2016).
125. Beltran, M. et al. G-tract RNA removes polycomb repressive complex 2 from genes. *Nat. Struct. Mol. Biol.* **26**, 899–900 (2019).
126. Sims, R. J. Elongation by RNA polymerase II: the short and long of it. *Genes Dev.* **18**, 2437–2468 (2004).
127. Brandão, H. B. et al. RNA polymerases as moving barriers to condensin loop extrusion. *Proc. Natl Acad. Sci. USA* **116**, 20489–20499 (2019).
128. Hsieh, T.-H. S. et al. Resolving the 3D landscape of transcription-linked mammalian chromatin folding. Preprint at *bioRxiv* https://doi.org/10.1101/638775 (2019).
129. Heinz, S. et al. Transcription elongation can affect genome 3D structure. *Cell* **174**, 1522–1536.e22 (2018).
130. Gu, B. et al. Transcription-coupled changes in nuclear mobility of mammalian *cis*-regulatory elements. *Science* **359**, 1050–1055 (2018).
131. Nagashima, R. et al. Single nucleosome imaging reveals loose genome chromatin networks via active RNA polymerase II. *J. Cell Biol.* **218**, 1511–1530 (2019).
132. Kopp, F. & Mendell, J. T. Functional classification and experimental dissection of long noncoding RNAs. *Cell* **172**, 393–407 (2018).
133. Sander, J. D. & Joung, J. K. CRISPR–Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32**, 347–355 (2014).
134. Dominguez, A. A., Lim, W. A. & Qi, L. S. Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. *Nat. Rev. Mol. Cell Biol.* **17**, 5–15 (2016).
135. Lavalou, P. et al. Strategies for genetic inactivation of long noncoding RNAs in zebrafish. *RNA* **25**, 897–904 (2019).
136. Fulco, C. P. et al. Systematic mapping of functional enhancer–promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016).
137. Simeonov, D. R. et al. Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature* **549**, 111–115 (2017).
138. Abudayyeh, O. O. et al. RNA targeting with CRISPR–Cas13. *Nature* **550**, 280–284 (2017).
139. Cox, D. B. T. et al. RNA editing with CRISPR–Cas13. *Science* **358**, 1019–1027 (2017).
140. Shechner, D. M., Hacisuleyman, E., Younger, S. T. & Rinn, J. L. Multiplexable, locus-specific targeting of long RNAs with CRISPR–Display. *Nat. Methods* **12**, 664–670 (2015).
141. Carullo, N. V. N. et al. Enhancer RNAs are necessary and sufficient for activity-dependent neuronal gene transcription. Preprint at *bioRxiv* https://doi.org/10.1101/270967 (2018).
142. Tuck, A. C. et al. Distinctive features of lincRNA gene expression suggest widespread RNA-independent functions. *Life Sci. Alliance* **1**, e201800124 (2018).
143. Nomura, Y., Zhou, L., Miu, A. & Yokobayashi, Y. Controlling mammalian gene expression by allosteric hepatitis delta virus ribozymes. *ACS Synth. Biol.* **2**, 684–689 (2013).
144. Nojima, T. et al. Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* **161**, 526–540 (2015).
145. Sarshad, A. A. et al. Argonaute–miRNA complexes silence target mRNAs in the nucleus of mammalian stem cells. *Mol. Cell* **71**, 1040–1050.e8 (2018).
146. Wee, K. B. et al. Dynamics of co-transcriptional pre-mRNA folding influences the induction of dystrophin exon skipping by antisense oligonucleotides. *PLOS ONE* **3**, e1844 (2008).
147. Sioud, M. Promises and challenges in developing RNAi as a research tool and therapy. *Methods Mol. Biol.* **703**, 173–187 (2011).
148. Stojic, L. et al. Specificity of RNAi, LNA and CRISPRi as loss-of-function methods in transcriptional analysis. *Nucleic Acids Res.* **46**, 5950–5966 (2018).
149. Engreitz, J., Lander, E. S. & Guttman, M. RNA antisense purification (RAP) for mapping RNA interactions with chromatin. *Methods Mol. Biol.* **1262**, 183–197 (2015).
150. Li, X. & Fu, X.-D. Chromatin-associated RNAs as facilitators of functional genomic interactions. *Nat. Rev. Genet.* **20**, 503–519 (2019).
This work is an excellent recent review, particularly useful for those looking for detailed information about techniques used to identify chromatin binding sites of lncRNAs.
151. Li, X. et al. GRID-seq reveals the global RNA–chromatin interactome. *Nat. Biotechnol.* **35**, 940–950 (2017).
152. Bell, J. C. et al. Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts. *eLife* **7**, e27024 (2018).
153. Sridhar, B. et al. Systematic mapping of RNA–chromatin interactions in vivo. *Curr. Biol.* **27**, 610–612 (2017).
154. Bonetti, A. et al. RADICL-seq identifies general and cell type-specific principles of genome-wide RNA–chromatin interactions. Preprint at *bioRxiv* https://doi.org/10.1101/681924 (2019).
155. Shao, W. et al. U1 snRNP regulates chromatin retention of noncoding RNAs. Preprint at *bioRxiv* https://doi.org/10.1101/310433 (2018).
156. Lubelsky, Y. & Ulitsky, I. Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. *Nature* **555**, 107–111 (2018).
157. Shukla, C. J. et al. High-throughput identification of RNA nuclear enrichment sequences. *EMBO J.* **37**, e9845 (2018).
158. Matsui, M. & Corey, D. R. Non-coding RNAs as drug targets. *Nat. Rev. Drug Discov.* **16**, 167–179 (2017).
159. Dao, L. T. M. et al. Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nat. Genet.* **49**, 1073–1081 (2017).
160. Carelli, F. N., Liechti, A., Halbert, J., Warnefors, M. & Kaessmann, H. Repurposing of promoters and enhancers during mammalian evolution. *Nat. Commun.* **9**, 4066 (2018).
This study identifies prevalent repurposing of enhancers to promoters (or vice versa) during evolution, demonstrating the large potential that enhancers serve in creating novel transcriptional units, including lncRNAs.
161. Marques, A. C. et al. Evidence for conserved post-transcriptional roles of unitary pseudogenes and for frequent bifunctionality of mRNAs. *Genome Biol.* **13**, R102 (2012).
162. Hezroni, H. et al. A subset of conserved mammalian long non-coding RNAs are fossils of ancestral protein-coding genes. *Genome Biol.* **18**, 162 (2017).
163. Cabili, M. N. et al. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* **16**, 20 (2015).
164. Carlevaro-Fita, J. & Johnson, R. Global positioning system: understanding long noncoding RNAs through subcellular localization. *Mol. Cell* **73**, 869–883 (2019).
165. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
166. Roadmap Epigenomics Consortium. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
167. Zhao, J., Sun, B. K., Erwin, J. A., Song, J.-J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756 (2008).
168. Minajigi, A. *et al*. Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* https://doi.org/10.1126/science.aab2276 (2015).
169. Sarma, K. et al. ATRX directs binding of PRC2 to Xist RNA and polycomb targets. *Cell* **159**, 869–883 (2014).
170. Chu, C. et al. Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404–416 (2015).
171. McHugh, C. A. et al. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).
172. Hasegawa, Y. et al. The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev. Cell* **19**, 469–476 (2010).
173. Dinger, M. E. et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* **18**, 1433–1445 (2008).
174. Luo, S. et al. Divergent lncRNAs regulate gene expression and lineage differentiation in pluripotent cells. *Cell Stem Cell* **18**, 637–652 (2016).
175. Knauss, J. L. et al. Long noncoding RNA *Sox2ot* and transcription factor YY1 co-regulate the differentiation of cortical neural progenitors by repressing *Sox2*. *Cell Death Dis.* **9**, 799 (2018).
176. Jeon, Y. & Lee, J. T. YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* **146**, 119–133 (2011).

Acknowledgements
The authors would like to thank A. Shkumatava, S. Nakagawa, L. Chen, C. Ross, H. Hezroni, M. Goldrich and members of the Ulitsky laboratory for helpful discussions and comments on the manuscript.

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

Peer review information
Nature Reviews Genetics thanks C. Ponting and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.