NON-CODING RNA

Long non-coding RNAs: new players in cell differentiation and development

Alessandro Fatica[1] *and Irene Bozzoni*[1–4]

Abstract | Genomes of multicellular organisms are characterized by the pervasive expression of different types of non-coding RNAs (ncRNAs). Long ncRNAs (lncRNAs) belong to a novel heterogeneous class of ncRNAs that includes thousands of different species. lncRNAs have crucial roles in gene expression control during both developmental and differentiation processes, and the number of lncRNA species increases in genomes of developmentally complex organisms, which highlights the importance of RNA-based levels of control in the evolution of multicellular organisms. In this Review, we describe the function of lncRNAs in developmental processes, such as in dosage compensation, genomic imprinting, cell differentiation and organogenesis, with a particular emphasis on mammalian development.

microRNAs
(miRNAs). Small non-coding RNAs of ~22 nucleotides that are integral components of RNA-induced silencing complex (RISC) and that recognize partially complementary target mRNAs to induce translational repression, which is often linked to degradation. Among the RISC proteins, AGO binds to miRNA and mediates the repressing activity.

[1]*Department of Biology and Biotechnology "Charles Darwin", Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy.*
[2]*Institute of Molecular Biology and Pathology of the National Research Council, Piazzale Aldo Moro 5, 00185 Rome, Italy.*
[3]*Istituto Pasteur Fondazione Cenci Bolognetti, Piazzale Aldo Moro 5, 00185 Rome, Italy.*
[4]*Center for Life Nano Science @Sapienza, Istituto Italiano di Tecnologia, Sapienza University of Rome, Viale Regina Elena 291, 00161 Rome, Italy.*
Correspondence to I.B.
e-mail:
irene.bozzoni@uniroma1.it
doi:10.1038/nrg3606
Published online
3 December 2013

Although the central role of RNA in cellular functions and organismal evolution has been advocated periodically during the past 50 years, only recently has RNA received a remarkable level of attention from the scientific community. Analyses that compare transcriptomes with genomes of mammalian species (BOX 1) have established that approximately two-thirds of genomic DNA is pervasively transcribed, which is in sharp contrast to the <2% that is ultimately translated into proteins[1,2]. Moreover, the degree of organismal complexity among species better correlates with the proportion of each genome that is transcribed into non-coding RNAs (ncRNAs) than with the number of protein-coding genes, even when protein diversification by both alternative splicing and post-translational regulation are taken into account[3]. This suggests that RNA-based regulatory mechanisms had a relevant role in the evolution of developmental complexity in eukaryotes.

The range of ncRNAs in eukaryotes is vast and exceeds the number of protein-coding genes. Besides the different families of small ncRNAs[4], a large proportion of transcriptomes results in RNA transcripts that are longer than 200 nucleotides, which are often polyadenylated and are devoid of evident open reading frames (ORFs) — these are defined as long ncRNAs (lncRNAs)[5–7]. Many roles are emerging for lncRNAs in ribonucleoprotein complexes that regulate various stages of gene expression[5,7]. Their intrinsic nucleic acid nature confers on lncRNAs the dual ability to function as ligands for proteins (such as those with functional roles in gene regulation processes) and to mediate base-pairing interactions that guide lncRNA-containing complexes to specific RNA or DNA target sites[5,7,8]. This dual activity is shared with small ncRNAs[4], such as microRNAs (miRNAs), small nucleolar RNAs and many other small nuclear ribonucleoprotein particles (BOX 2). However, unlike small ncRNAs, lncRNAs can fold into complex secondary and higher order structures to provide greater potential and versatility for both protein and target recognition[5,7,8]. Moreover, their flexible[8,9] and modular[10,11] scaffold nature enables lncRNAs to tether protein factors that would not interact or functionally cooperate if they only relied on protein–protein interactions[5,8,12–14]. Such combinatorial RNA-mediated tethering activity has enhanced gene regulatory networks to facilitate a wide range of gene expression programmes (FIG. 1) to provide an important evolutionary advantage[5,7,8]. This complexity is likely to be further expanded by differential splicing and the use of alternative transcription initiation sites and polyadenylation sites by lncRNAs, thus increasing the number of tethering-module combinations.

The expression of lncRNAs has been quantitatively analysed in several tissues and cell types by high-throughput RNA sequencing (RNA-seq) experiments, and it was generally found to be more cell type specific than the expression of protein-coding genes[5,6,8,15–17]. Interestingly, in several cases, such tissue specificity has been attributed to the presence of transposable elements that are embedded in the vicinity of lncRNA transcription start sites[18–20]. Moreover, lncRNAs have been shown to be differentially expressed across various stages of differentiation, which indicates that they may be novel 'fine-tuners' of cell fate[5–7]. This specific spatiotemporal expression can be linked to the establishment of both

Box 1 | **Methodologies for lncRNA identification and analyses**

The identification of long non-coding RNAs (lncRNAs) relies on the detection of transcription from genomic regions that are not annotated as protein coding, such as regions that are devoid of open reading frames. This can be achieved by the direct detection of the transcribed RNA. However, conventional gene expression microarrays are only designed to detect the expression of protein-coding mRNAs, and unbiased RNA detection methods are therefore required. These include tiling arrays, serial analysis of gene expression (SAGE), cap analysis of gene expression (CAGE) and high-throughput RNA sequencing (RNA-seq). Alternatively, transcription can be inferred from genomic DNA analyses by detecting specific histone marks (such as H3K4–H3K36 domains) that signify transcriptionally active chromatin.

Tiling arrays
In this technique, cDNA is hybridized to microarray slides carrying overlapping oligonucleotides that cover either specific chromosomal regions or a complete genome. This methodology allows the analysis of global transcription from specific genomic regions and was initially used for both identification and expression analysis of lncRNAs[112].

SAGE
SAGE was the first method to use sequencing for high-throughput analyses of transcriptomes[113]. It is based on the generation of short stretches of unbiased cDNA sequence (that is, SAGE tags) by restriction enzymes. SAGE tags are concatenated before cloning and sequencing. This methodology allows both the quantification of transcripts throughout the transcriptome and the identification of new transcripts, including lncRNAs. Several modifications to the original SAGE strategy have been developed to improve the specificity by generating larger tags, such as LongSAGE and SuperSAGE[114,115].

CAGE
CAGE relies on the isolation and sequencing of short cDNA sequence tags that originate from the 5′ end of RNA transcripts[116]. Similarly to SAGE, tags are concatenated before cloning and sequencing. However, in addition to quantifying expression level, CAGE also identifies the location of each transcription start site.

RNA-seq
Sequencing of transcriptomes by RNA-seq is one of the most powerful methodologies for *de novo* discovery and expression analyses of lncRNAs[117]. In this method, total RNA is converted to a cDNA library that is directly sequenced by high-throughput sequencing instruments. There are several types of sequencing technologies but Illumina platforms are currently the most commonly used for RNA-seq experiments. A single sequencing run produces billions of reads that are subsequently aligned to a reference genome. Following alignment, the data are translated into a quantitative measure of gene expression by specific algorithms.

Chromatin immunoprecipitation (ChIP)
ChIP allows the isolation of DNA sequences that are associated with a chromatin component of interest. When combined with high-throughput readouts such as microarrays (that is, ChIP–chip) and DNA sequencing (that is, ChIP–seq), these methods can infer the genomic distribution of either proteins or histone modifications. Analysis of loci with specific histone modifications that characterize active transcription (such as H3K4–H3K36 marks) allowed an indirect identification of many unknown lncRNAs[77].

Polyadenylation sites
Sequences that are required for the cleavage of primary RNA transcripts that are produced by RNA polymerase II. As a consequence of such cleavage, the 5′ cutoff product becomes polyadenylated, whereas the 3′ product undergoes rapid degradation that induces Pol II release from the DNA and hence transcriptional termination.

Polycomb repressive complex
(PRC). A multiprotein complex that silences target genes by establishing a repressive chromatin state; PRC2 trimethylates histone H3 at lysine 27, which is recognized by PRC1 that mediates chromatin compaction by inducing H2A monoubiquitylation.

MLL1 complex
A multiprotein complex that mediates both histone H3 trimethylation at lysine 4 (H3K4me3) and histone H4 acetylation at lysine 16 (H4K16ac), which are associated with transcriptionally active genes.

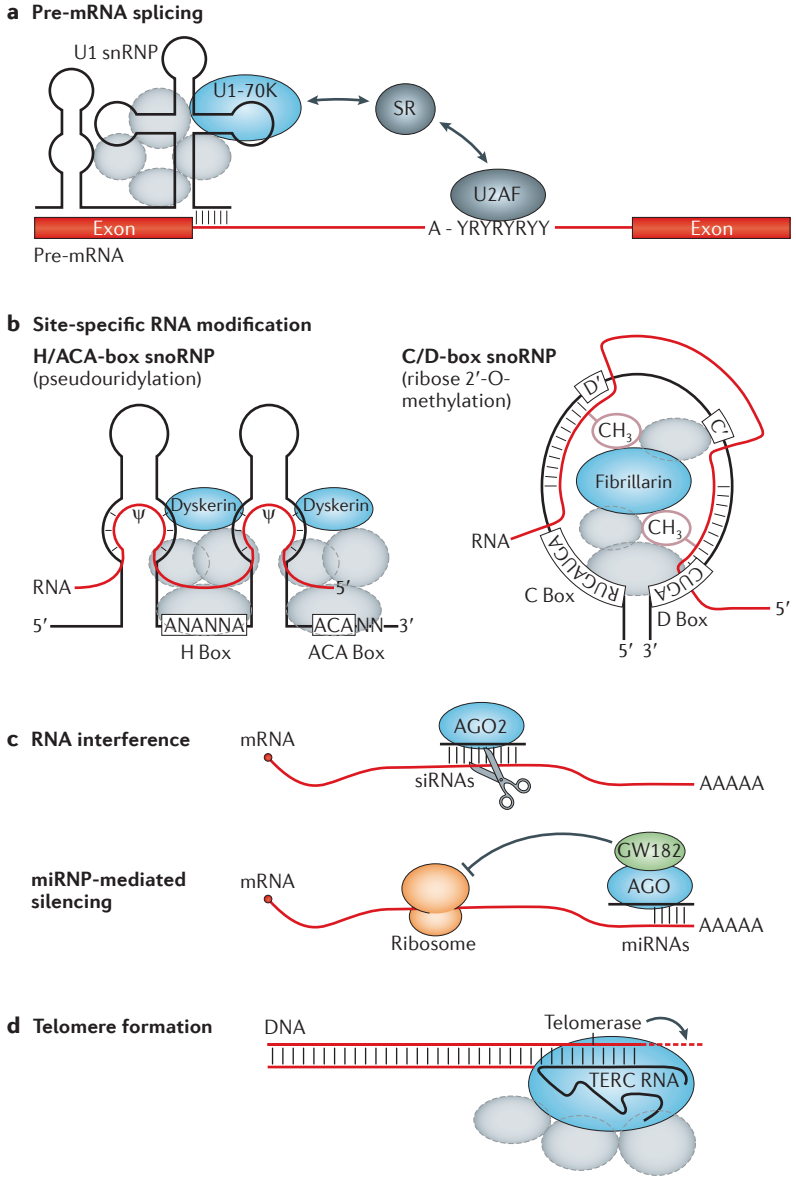
well-defined barriers of gene expression and cell-type-specific gene regulatory programmes. Combined with the involvement of lncRNAs in positive or negative feedback loops, lncRNAs can amplify and consolidate the molecular differences between cell types that are required to control cell identity and lineage commitment[21–23].

In this Review, we discuss our latest understanding of lncRNAs according to their roles in various developmental processes. We focus on lncRNAs for which roles have been confirmed by functional studies in cellular systems, and some of these have been further characterized through *in vivo* loss-of-function approaches in model organisms (TABLE 1). We describe molecular mechanisms of action and roles of lncRNAs in cellular processes such as genomic imprinting, maintenance of pluripotency and development in various organs, and in environment-responsive developmental programmes. Finally, we summarize future challenges in the field.

Modes of action of lncRNAs

Nuclear lncRNAs. Evidence so far indicates that most nuclear lncRNAs function by guiding chromatin modifiers to specific genomic loci[5,7,8,24] (FIG. 1A). In most cases, they recruit DNA methyltransferase 3 (DNMT3) and histone modifiers, such as the Polycomb repressive complex PRC2 (REFS 12,25) and histone H3 lysine 9 (H3K9) methyltransferases[26,27]. The resultant DNA and histone modifications predominantly correlate with the formation of repressive heterochromatin and with transcriptional repression. Furthermore, the act of lncRNA transcription itself can negatively affect gene expression[28,29]. Transcriptional activation has also been shown through the recruitment of chromatin-modifying complexes, such as the histone H3K4 methyltransferase MLL1 complex (REFS 30,31), and by the activation of specific enhancer regions through changes to three-dimensional chromatin conformation[30,32,33]. With respect to the target sites, it is possible to distinguish between *cis*- and *trans*-acting lncRNAs — *cis*-acting lncRNAs control the expression of genes that are positioned in the vicinity of their transcription sites and can sometimes spread their effect to long distances on the same chromosome, whereas *trans*-acting lncRNAs can either repress or activate gene expression at independent loci[5,7,8]. However, for both classes of lncRNAs, the targeting mechanisms is still far from being understood; in particular, it is not known how *cis*-acting lncRNAs

Box 2 | **Examples of ncRNAs with guiding activity for proteins**



Guiding activity of RNAs has been described over many years in different classes of well-characterized non-coding RNAs (ncRNAs). These ncRNAs function in *trans* by bringing specific interactors to specific targets (shown in red) through base pairing with other RNA or DNA molecules. For example, the spliceosomal U1 small nuclear ribonucleoprotein particles (snRNPs) work in the splicing process by recognizing 5′ splice sites in pre-mRNA molecules through the U1 small nuclear RNA component and by contacting splicing factors (such as SR proteins and U2 auxiliary factor (U2AF)), which are bound to the 3′ portion of the intron, through the U1-70K protein component of the U1 snRNP[118] (see the figure, part **a**). Similarly, small nucleolar RNAs (snoRNAs) guide modifying enzymes to produce site-specific pseudouridylation (Ψ; for H/ACA-box snoRNAs, catalysed by dyskerin) and ribose 2′-O-methylation (for C/D-box snoRNAs, catalysed by fibrillarin) on target RNAs[119] (see the figure, part **b**). In the cytoplasm, small interfering RNAs (siRNAs) and microRNAs (miRNAs) direct Argonaute (AGO) proteins to target RNAs. The AGO2–siRNA complex induces RNA cleavage, whereas the AGO–miRNA complex, in combination with GW182, triggers translational repression and mRNA degradation[22] (see the figure, part **c**). Finally, telomerase RNA component (TERC) recognizes single-stranded DNA at telomeres and serves as a template for the associated telomerase[9] (see the figure, part **d**). In this case, the ncRNA molecules provide DNA-binding specificity to a protein complex. In all these examples, target specificity is provided by short complementary sequences. miRNP, microRNA ribonucleoprotein complex.

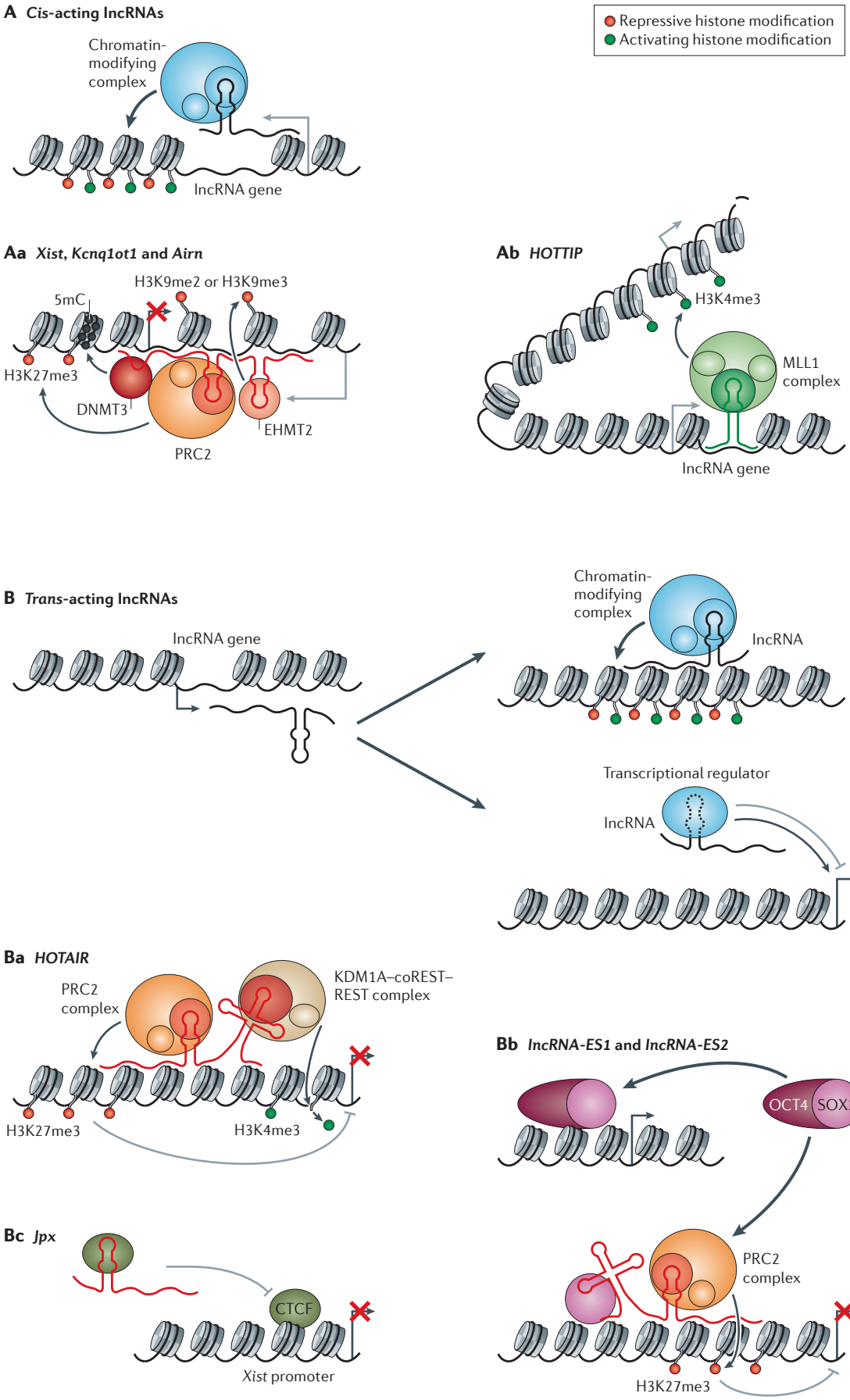
are retained to the sites of their transcription and how *trans*-acting lncRNAs find distantly located targets. Different recognition mechanisms have been proposed, including recruitment by bridging proteins[34], formation of an RNA–DNA triplex[35] and DNA recognition by RNA structures[36]. Nuclear lncRNAs can also have indirect regulatory effects on gene loci (FIG. 1B); for example, by acting as decoys that sequester transcription factors[37,38], by allosterically modulating regulatory proteins[39], and by altering nuclear domains[40] and long-range three-dimensional chromosomal structures[41].

Cytoplasmic lncRNAs. Many lncRNA-mediated mechanisms of gene regulation have been identified in the cytoplasm[7]. These lncRNAs often show sequence complementarity with transcripts that originate from either the same chromosomal locus or independent loci. Upon recognition of the target by base pairing, they can modulate translational control, examples of which include positive regulation by the ubiquitin carboxy-terminal hydrolase L1 antisense RNA 1 (*Uchl1-as1*)[42] and negative regulation by tumour protein p53 pathway corepressor 1 (*Trp53cor1*; also known as *lincRNA-p21*)[43]. Similarly, lncRNAs can modulate mRNA stability; for example, both β-site APP-cleaving enzyme 1-antisense (*BACE1-AS*)[44] and tissue differentiation-inducing non-protein-coding RNA (*TINCR*)[45] increase the stability of their target mRNAs, whereas half-STAU1 (staufen double-stranded RNA-binding protein 1)-binding site RNAs (1/2sbsRNAs)[46,47] decrease target mRNA stability (FIG. 2A).

A peculiar mode of action is that of lncRNAs that function as competing endogenous RNAs (ceRNAs)[48] — by binding to and sequestering specific miRNAs, ceRNAs function as 'miRNA sponges' to protect the target mRNAs from repression. This represents a new type of regulatory circuitry in which different types of RNAs (both coding and non-coding) can crosstalk to each other by competing for shared miRNAs. This activity was initially described in plants[49] and, subsequently, in mammals[50], in which it was shown to be relevant in many processes, including tumorigenesis[51,52], cell differentiation[21] and pluripotency[23]. Recently, an additional example of ceRNA was found in a newly identified class of circular RNAs (circRNAs)[53–55], which function as sponges for miRNAs in neuronal cells (FIG. 2B). It is interesting that, whereas the linear ceRNAs have a short half-life that allows a rapid control of sponge activity, circRNAs have much greater stability and their turnover can be controlled by the presence of a perfectly matched miRNA target site[53–55].

Dosage compensation and genomic imprinting

Among the first and best-characterized examples of lncRNAs that have specific developmental roles and robust loss-of-function phenotypes *in vivo* are those involved in dosage compensation and genomic imprinting (FIG. 1Aa). These two processes are required for normal development and rely on the formation of silenced chromatin to produce monoallelic expression of specific genes in mammals[56].



◀ Figure 1 | **Models of nuclear lncRNA function.** Examples of long non-coding RNAs (lncRNAs) that regulate transcription in *cis* (part **A**) and in *trans* (part **B**), by recruiting specific transcriptional regulators onto specific chromosomal loci, are shown. **Aa** | lncRNAs that are involved in dosage compensation and genomic imprinting include X-inactive specific transcript (*Xist*), *Kcnq1* overlapping transcript 1 (*Kcnq1ot1*) and *Airn* (antisense *Igf2r* (insulin-like growth factor 2 receptor) RNA). These lncRNAs induce the formation of repressive chromatin through the recruitment of DNA methyltransferase 3 (DNMT3), which induces DNA methylation; Polycomb repressive complex 2 (PRC2), which produces histone H3 lysine 27 trimethylation (H3K27me3); and histone lysine *N*-methyltransferase EHMT2, which is responsible for producing H3K9me2 and H3K9me3 (REF. 56). **Ab** | HOXA distal transcript antisense RNA (*HOTTIP*) functions through the recruitment of the MLL1 complex, which drives the formation of the activating H3K4me3 mark[30]. **Ba** | HOXA transcript antisense RNA (*HOTAIR*) is a *trans*-acting regulator of the HOXD genes[12]. It is characterized by a modular scaffold structure that allows the recruitment of two distinct repressive complexes, PRC2 and the H3K4 demethylating complex KDM1A–coREST–REST (lysine-specific histone demethylase 1A–REST corepressor 1–RE1-silencing transcription factor) on the same genomic region[11]. **Bb** | The pluripotency RNAs *lncRNA-ES1* and *lncRNA-ES2* associate with both PRC2 and the transcription factor sex-determining region Y-box 2 (SOX2), which suggests that these lncRNAs control embryonic stem cell pluripotency by silencing SOX2-bound developmental genes[14]; this function is alternative to OCT4- and SOX2-dependent activation of pluripotency genes. **Bc** | The lncRNA *Jpx* (Jpx transcript, Xist activator) that binds to the transcriptional repressor CTCF inhibits its binding to the *Xist* promoter, thus activating *Xist* transcription[38].

X chromosome inactivation. The identification of X-inactive specific transcript (*XIST*) as a regulator of X chromosome inactivation in mammals provided one of the first examples of a lncRNA that is directly involved in the formation of repressive chromatin[56]. *Xist* deletion in mice causes a loss of X chromosome inactivation and female-specific lethality[57]. Various studies both in mice and in mouse embryonic stem cells (ESCs) — a major model system for X chromosome inactivation — have demonstrated that, in female cells, *Xist* acts in *cis* by inducing the formation of transcriptionally inactive heterochromatin on the X chromosome from which it is transcribed[56]. *Xist* is required only for the initiation and not for the maintenance of X inactivation, and its spatiotemporal expression must be properly controlled[56]. *Xist* induces the formation of repressive heterochromatin, at least in part, by tethering PRC2 to the inactive X chromosome[25]. However, parallel PRC2-independent pathways have been recently demonstrated in both mouse and human ESCs[58,59].

The interaction between *Xist* and chromatin may involve, among others, transcriptional repressor protein YY1 that is thought to function as a recruitment platform for *Xist* by binding to its first exon[34]. Moreover, it has been recently shown that *Xist* itself is able to recognize the three-dimensional conformation of the X chromosome[41]. Notably, *Xist* expression is itself controlled by other lncRNAs in both a positive and a negative manner[56]. One of the best-characterized *Xist* regulators is its natural antisense non-coding transcript *Tsix*. *Tsix* counteracts *Xist* expression by inducing repressive epigenetic modifications at the *Xist* promoter[56]. The loss of *Tsix* function *in vivo* resulted in ectopic *Xist* expression, aberrant X inactivation and early embryonic lethality[60,61]. These mouse models showed, for the first time, an important role for a naturally occurring antisense transcript in gene expression regulation. Furthermore, *Xist* activation also requires the lncRNA *Jpx*[62], which induces *Xist* transcription through the sequestration of transcriptional repressor CTCF[38] (FIG. 1Bc).

Xist, which is transcriptionally regulated by a network of pluripotency factors, may also have an important role in differentiation. Indeed, both the homozygous and heterozygous conditional deletion of *Xist* in mouse haematopoietic stem cells produced an aberrant maturation of haematopoietic progenitors in females[63], which resulted in the development of blood cell cancers and in accelerated death. Aberrant *XIST* expression has been observed in human cancers, which further suggests that alteration in the X inactivation process contributes to tumorigenesis.

Genomic imprinting. Imprinted genes generally associate in clusters and are epigenetically marked in sex-dependent ways during male and female gametogenesis; they are subsequently silenced on only one parental chromosome in the embryo. Imprinted regions encode different species of ncRNAs, including lncRNAs that, in many cases, bind to imprinted regions and are directly involved in silencing[56]. These lncRNAs are generally long (more than 100 kb) and function in *cis*. The best-characterized example at both the genetic and the molecular levels are the lncRNAs *Kcnq1* overlapping transcript 1 (*Kcnq1ot1*) and *Airn* (antisense *Igf2r* (insulin-like growth factor 2 receptor) RNA). These lncRNAs are paternally expressed; they function by repressing flanking protein-coding genes in *cis* and are involved in early development in mice[56]. The loss of function of these lncRNAs in the embryo is not lethal — paternal inheritance of a loss-of-function allele results in a loss of imprinting and in growth defects, whereas maternal inheritance of this allele does not affect imprinting or growth[64–66]. These studies showed that multiple repressive pathways regulate imprinted gene silencing by lncRNAs during development, and that the extent of silencing along the chromosome varies in different tissues[26,27,66]. For example, during embryonic development, *Kcnq1ot1* functions by establishing and maintaining repressive DNA methylation on surrounding genes, whereas, in the placenta, it functions by recruiting the repressive histone modifiers PRC2 and the H3K9 methyltransferase EHMT2 (also known as G9a) on genes that are located further away from the imprinted region[66]. It is worth noting that, in the establishment of transcriptional gene silencing by *cis*-acting lncRNAs, continuous transcription might be more important than the production of mature RNA. This has been elegantly shown for *Airn*, which is expressed from the paternal chromosome and is antisense to the *Igf2r* gene. *Airn* functions in *cis* to silence the paternal *Igf2r* allele, whereas the maternal *Igf2r* allele remains expressed. In embryonic tissues, *Airn* silences paternal *Igf2r* through a mechanism that does not require a stable RNA product but that is based on continuous *Airn* transcription, which interferes with the recruitment of RNA polymerase II[29]. By contrast,

Dosage compensation
The process that ensures equal levels of X-linked gene expression in males (XY) and females (XX).

Genomic imprinting
Epigenetic silencing of genes on the basis of their parental origin, which results in monoallelic expression.

EHMT2
A histone lysine methyltransferase that is responsible for dimethylation and trimethylation at histone H3 lysine 9, which creates epigenetic marks that predominantly correlate with transcriptional repression.

Table 1 | **lncRNA manipulation and resulting phenotypes in model animal systems**

lncRNA	Process	Site of action	Loss-of-function methods	Phenotype	Refs
Mouse					
Xist	Dosage compensation	Nucleus	Gene disruption in embryo	Embryonic lethality	57
			Conditional disruption in haematopoietic stem cells	Aberrant haematopoiesis and blood cell cancer	63
Tsix	Dosage compensation	Nucleus	Embryonic gene inactivation by either promoter deletions or premature termination	Embryonic lethality	60,61
Kcnq1ot1	Genomic imprinting	Nucleus	Embryonic gene inactivation by either promoter deletions or premature termination	Growth defects	64,66
Airn	Genomic imprinting	Nucleus	Embryonic gene inactivation by either promoter deletions or premature termination	Growth defects	29,65
Fendrr	Gene expression regulation in mesoderm	Nucleus	Gene disruption in embryo	Embryonic lethality	95
			60% reduction in embryo by RNA interference	Normal development	95
Hotair	Hox gene regulation	Nucleus	Gene disruption in embryo	Defects in skeletal system development	72
Dlx1os	Homeodomain transcription factor regulation in developing forebrain	Nucleus	Embryonic gene inactivation by premature termination	Morphologically normal with mild skull and neurological defects	92
Dlx6os1	Homeodomain transcription factor regulation in developing forebrain	Nucleus	Embryonic gene inactivation by premature termination	Morphologically normal with altered GABAergic interneuron development	90
Malat1	Tumorigenesis	Nucleus	Gene disruption in embryo	Normal development	82
Miat	Retina development	Nucleus	Knockdown and overexpression in neonatal retina	Defects in specification of retina cell types	126
Six3os1	Retina development	Nucleus	Knockdown and overexpression in neonatal retina	Defects in specification of retina cell types	127
Tug1	Retina development	Nucleus	Knockdown in neonatal retina	Defects in differentiation of photoreceptor progenitor cells	128
Vax2os	Retina development	Nucleus	Overexpression in neonatal retina	Defects in differentiation of photoreceptor progenitor cells	129
Zebrafish					
Cyrano	Embryogenesis	Not analysed	Knockdown and functional inactivation in embryo by morpholino oligonucleotides	Developmental defects	110
Megamind	Embryogenesis	Not analysed	Knockdown and functional inactivation in embryo by morpholino oligonucleotides	Defects in brain morphogenesis and in eye development	110
Chicken					
HOTTIP	HOXA regulation	Nucleus	Knockdown in chick embryos by RNA interference	Altered limb morphology	30

Airn, antisense *Igf2r* (insulin-like growth factor 2 receptor) RNA; *Dlx1os*, distal-less homeobox 1, opposite strand; *Dlx6os1*, *Dlx6* opposite strand transcript 1; *Fendrr*, *Foxf1* adjacent non-coding developmental regulatory RNA; GABA, γ-aminobutyric acid; *Hotair*, HoxA transcript antisense RNA; *HOTTIP*, HOXA distal transcript antisense RNA; *Kcnq1ot1*, *Kcnq1* overlapping transcript 1; lncRNA, long non-coding RNA; *Malat1*, metastasis-associated lung adenocarcinoma transcript 1; *Miat*, myocardial infarction-associated transcript (also known as *Rncr2*); *Six3os1*, *Six3* opposite strand transcript 1; *Tsix*, X (inactive)-specific transcript, opposite strand; *Tug1*, taurine upregulated gene 1; *Vax2os*, *Vax2* opposite strand transcript; *Xist*, X-inactive specific transcript.

in the placenta, mature *Airn* recruits EHMT2 to induce the formation of repressive chromatin[65]. Altogether, these studies showed that a single lncRNA could work by different mechanisms depending on the cell type, which might reflect the presence of either different interactors or chromatin modifications that influence lncRNA functions in diverse cellular contexts. These examples also show the advantages of using *cis*-acting lncRNAs to regulate a gene cluster. The *in situ* production of regulators at their site of function is intrinsically more robust than dedicated *trans*-acting proteins. Thus, it is not surprising that the use of *cis*-acting lncRNAs to silence gene transcription is an evolutionarily conserved mechanism and is not restricted to complex and multicellular organisms, as in the case of yeast cryptic unstable transcripts[67].

Regulation of HOX genes

HOX genes encode an evolutionary conserved family of transcription factors that regulate the embryo body plan and that contribute to cell specification in several adult differentiation processes[68]. In mammals, there are 39 HOX genes that are grouped in four clusters (HOXA, HOXB, HOXC and HOXD), which allow precise spatiotemporal coordination of expression. In addition to protein-coding genes, these clusters produce hundreds of lncRNAs that show similar spatiotemporal windows of expression to their neighbouring protein-coding genes[12]. Some of these lncRNAs have been shown to be directly involved in the regulation of HOX genes.

Cis-*acting lncRNAs.* The *cis*-acting lncRNA HOXA distal transcript antisense RNA (*HOTTIP*), which is produced from the 5′ end of the human HOXA locus upstream of *HOXA13*, was identified in human primary fibroblasts. The downregulation of *HOTTIP* levels in primary fibroblasts induced the transcription of several downstream 5′-HOXA genes. *HOTTIP* is conserved in vertebrates, and its knockdown by short hairpin RNAs in chick embryos altered limb morphology[30]. The mechanism by which *HOTTIP* regulates HOXA expression relies on its interaction with the activating histone-modifying MLL1 complex and on the formation of chromatin loops that connect distally expressed *HOTTIP* transcripts with various HOXA gene promoters[30] (FIG. 1Ab). Notably, many DNA enhancer elements produce enhancer RNAs, which might work by a similar mechanism[32,33].

An identical mode of action has been described for mistral lncRNA (*Mira*), which is a mouse-specific lncRNA that is transcribed from the HOXA locus[31]. *Mira* was identified in retinoic acid-induced differentiation of mouse ESCs, in which it positively controls the transcription of two adjacent genes, *HOXA6* and *HOXA7*. The knockdown of *Mira* in mouse ESCs inhibited the activation of germ layer specification genes, which suggests a role for *Mira* in early mouse ESC differentiation[31]. However, it is not clear how this is linked to the regulation of *HOXA6* and *HOXA7*, as the deletion of these genes in mouse embryos indicated that they are involved in later developmental stages[69].

HOX genes are also involved in cell differentiation, and their deregulation is associated with different types of human disease, including cancer[67]. HOXA transcript antisense RNA myeloid-specific 1 (*HOTAIRM1*) was identified as a lncRNA that is produced from the 3′ end of the HOXA locus specifically in myeloid lineages[70]. The knockdown of *HOTAIRM1* in myeloid leukaemia cell lines inhibited the expression of 3′-HOXA genes by a currently uncharacterized mechanism, which indicates a positive role in gene expression, but it did not produce substantial effects on granulocytic differentiation. By contrast, *linc-Hoxa1* RNA was recently identified in mouse ESCs, in which it functions in *cis* to repress *Hoxa1* transcription by recruiting transcriptional activator protein Pur-β (PURB)[71].

Trans-*acting lncRNAs.* HOX transcript antisense RNA (*HOTAIR*) was one of the first *trans*-acting lncRNAs to be identified[12]. *HOTAIR* is transcribed from the HOXC gene cluster but acts as a repressor of the HOXD cluster, which is located on a different chromosome. *HOTAIR* interacts with the PRC2 and KDM1A–coREST–REST (lysine-specific histone demethylase 1A–REST corepressor 1–RE1-silencing transcription factor) histone-modifying complexes[11,12], and it is proposed to function in *trans* through the recruitment of these two repressive complexes to specific target genes (FIG. 1Ba). Indeed, the knockdown of *HOTAIR* in human fibroblasts led to decreased activity of these repressor complexes and to an increase in the expression of HOXD genes. Mouse models that carry targeted knockout of *Hotair* have been recently produced[71]. *Hotair* deletion did not affect viability but led to developmental defects and to homeotic transformations in the skeletal system[71]. Consistent with its association with repressive histone-modifying complexes, derepression of several genes, including HoxD components, was reported upon *Hotair* knockout[72]. Interestingly, the deletion of the entire mouse HoxC locus is perinatally lethal and does not show skeletal transformations, whereas individual knockouts of its components, including *Hotair*, are viable but show developmental defects[73,74]. This may reflect the presence either of compensatory mechanisms among members of the HoxD cluster or of genes with functions that are antagonistic to *Hotair*[72], thus emphasizing the importance of using appropriate *in vivo* models to define lncRNA function. Finally, *HOTAIR* was found to be upregulated in different cancers; in breast cancer metastasis, such upregulation was shown to result in the re-targeting of PRC2 to silence tumour suppressor genes[75].

Pluripotency versus differentiation commitment

Several lncRNAs that are associated with pluripotency have been identified either as species that are induced upon the reprogramming of fibroblasts to induced pluripotent stem cells (iPSCs)[76] or as species that are expressed in mouse[13,77,78] and human ESCs[14]. Notably, these lncRNA species show expression profiles that correlate well with those of OCT4 (also known as POU5F1), homeobox protein NANOG and sex-determining region Y-box 2 (SOX2), which are core components of the transcriptional network that controls pluripotency[78], and the promoters of these lncRNA species are bound by at least one of these core pluripotency transcription factors[13]. Loss-of-function experiments resulted in either exit from the pluripotent state or the upregulation of lineage commitment gene expression programmes, which was comparable to the knockdown of well-known ESC regulators[13,79]. Consistent with the 'modular scaffold' hypothesis[13], both the *lncRNA-ES1* (also known as LINC01108) and *lncRNA-ES2* pluripotency-associated lncRNAs were found to interact with Polycomb protein SUZ12 and SOX2, which suggests a model whereby pluripotency-associated lncRNAs function as scaffolds to recruit SUZ12 — part of the repressive PRC2 — to silence neural targets of SOX2 in pluripotent human ESCs[14] (FIG. 1Bb).

Germ layer
Primary germ layers (that is, ectoderm, endoderm and mesoderm) are specified during vertebrate embryogenesis and, through further differentiation, give rise to the organs and tissues of the body.

Pluripotency
The ability of a cell to differentiate into one of many cell types.

Induced pluripotent stem cells
(iPSCs). *In vitro*-derived pluripotent cells that originate from non-pluripotent cells in a process called reprogramming.

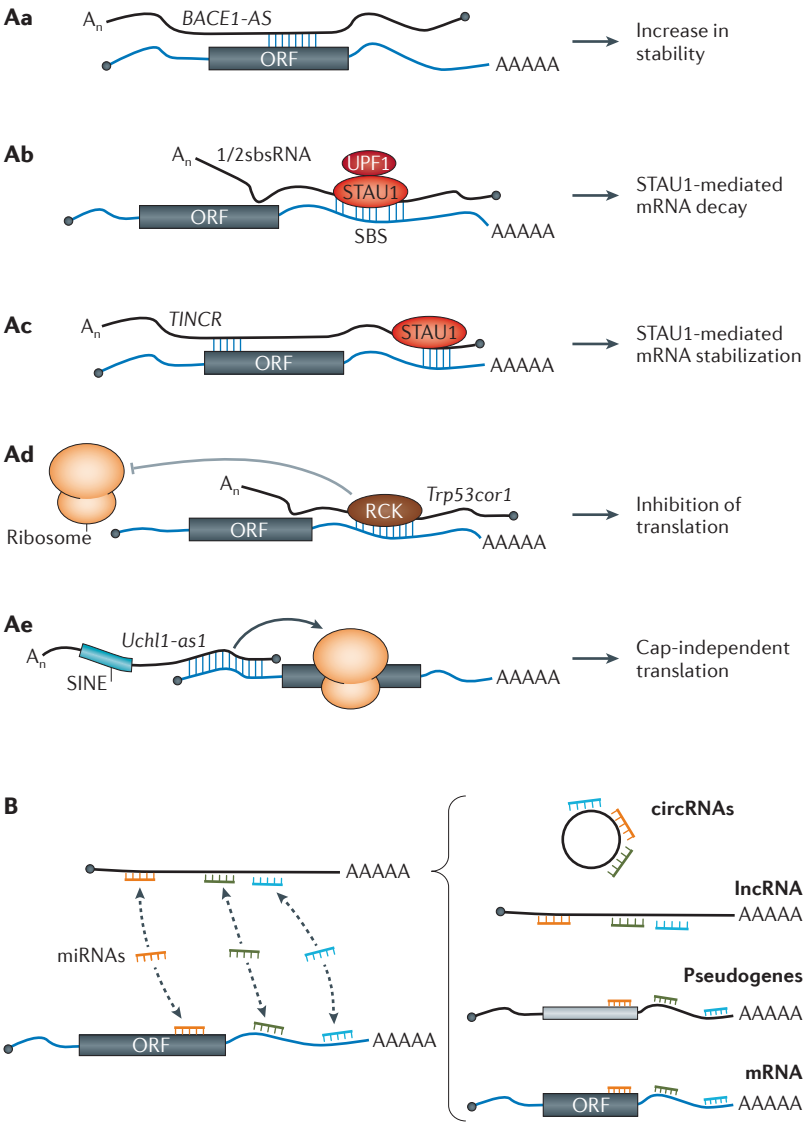


Figure 2 | **Models of cytoplasmic lncRNA function.** **A** | A common recognition mechanism is through base pairing of complementary regions between the long non-coding RNA (lncRNA) and their target RNA sequence. **Aa** | Base pairing between specific regions of the human β-site APP-cleaving enzyme 1 (*BACE1*) mRNA and its antisense transcript *BACE1-AS* induces stabilization of the target mRNA and increases BACE1 protein expression[44]. **Ab** | Staufen double-stranded RNA-binding protein 1 (STAU1)-mediated mRNA decay is induced when intermolecular base pairing is formed between an Alu element (or short interspersed element (SINE) in mice) in the 3′ untranslated region of the mRNA and an Alu element within a long half-STAU1-binding site RNA (1/2sbsRNA)[46,47]. This mRNA decay mechanism also involves the RNA helicase up-frameshift 1 (UPF1). **Ac** | By contrast, STAU1-mediated mRNA stabilization has been described in the case of tissue differentiation-inducing non-protein coding RNA (*TINCR*), which recognizes its target mRNAs through a 25 nucleotide-long motif[45]. Antisense recognition has been shown to also control translation. **Ad** | A repressive effect on translation was shown for the targets of tumour protein p53 pathway corepressor 1 (*Trp53cor1*) lncRNA[43], which functions with the RNA helicase RCK. **Ae** | Translation is induced upon stress induction of ubiquitin carboxy-terminal hydrolase L1 antisense RNA 1 (*Uchl1-as1*)[42]. **B** | Base pairing is also the mode of action of competing endogenous RNAs. In this case, however, the complementarity is between microRNAs (miRNAs) and different targets (including circular RNAs (circRNAs)[54,55], lncRNAs[21], pseudogene transcripts[50] and mRNAs[48]). The effect of these interactions is that protein-coding RNAs and non-coding RNAs can crosstalk to each other by competing for miRNA binding through their miRNA recognition motifs. ORF, open reading frame; SBS, STAU1-binding site.

Among pluripotency-associated lncRNAs, *LINC-ROR* (long intergenic non-protein coding RNA, regulator of reprogramming) is consistently enriched in human iPSCs, regardless of the cell of origin of these iPSCs[76]. *LINC-ROR* functions as a ceRNA to regulate the expression of the core pluripotency transcription factors by competing for miR-145 binding[23] (FIG. 3). Thus, it is a powerful example of how regulatory networks among transcriptional factors, miRNAs and lncRNAs are relevant for controlling the alternative fate between proliferation and differentiation commitments.

Brain and CNS development

Roles have been identified for lncRNAs in establishing and maintaining cell-type-specific gene expression patterns during organ development. In particular, the central nervous system (CNS), which is characterized by a vast range of neuronal and glial subtypes, is by far the most complex and diversified organ in terms of ncRNAs[80]. Cells in the CNS show intense transcription of lncRNAs, and the increase in number of lncRNAs has been linked to evolutionary complexity. Below, we describe several examples of lncRNAs that have roles in neurogenesis.

Abundant lncRNAs. Among the most abundant lncRNAs of the nervous system, metastasis-associated lung adenocarcinoma transcript 1 (*Malat1*) is by far the most extensively studied. This lncRNA was described as a highly expressed species in different types of mouse neurons[81]. *Malat1* is localized in nuclear speckles and has been shown to regulate synapse formation by modulating a subset of genes that have roles in nuclear and synapse function[81]. Indeed, the knockdown of *Malat1* in cultured mouse hippocampal neurons produced decreased synapse density and decreased dendrite growth[80]. However, a loss-of-function genetic model of *Malat1* indicated that it is not essential for mouse prenatal and postnatal development[82]. These mice showed only a minor effect of deregulation of several genes, such as *Malat1*-neighbouring genes, thus indicating a potential *cis*-regulatory role for *Malat1* in gene transcription.

Consistent with its original identification in cancer, the downregulation of *MALAT1* by 1,000-fold in human lung tumour cells, which was achieved by integrating RNA-destabilizing elements into *MALAT1* using zinc-finger nucleases, revealed a *MALAT1*-controlled metastatic gene expression programme[83]. In this setting, *MALAT1* was found in nuclear speckles and interacted with E3 SUMO-protein ligase CBX4 (also known as PC2), which is a component of PRC1. Overall, we still await clarification of the exact mechanisms of *MALAT1* activity in the distinct processes of tumour progression and neuronal differentiation.

lncRNAs with connections to key neural developmental protein-coding genes. lncRNAs have been profiled by complementary genome-wide techniques and *in situ* hybridizations on different adult mouse brain regions, which indicated that lncRNAs are associated with distinct brain cell types and are expressed in a more

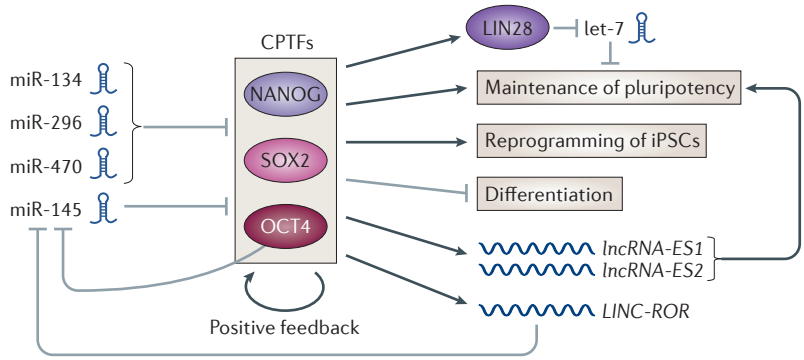


Figure 3 | **Pluripotency control.** A schematic representation of the regulatory circuitries that involve the homeobox protein NANOG, sex-determining region Y-box 2 (SOX2) and OCT4 core pluripotency transcription factors (CPTFs), microRNAs (miRNAs) and long non-coding RNAs (lncRNAs) in pluripotency control is shown. Several miRNAs were described as being necessary and sufficient to control self-renewal and pluripotency in human embryonic stem cells or to trigger differentiation[120,121] by a direct link with CPTFs. *LINC-ROR* (long intergenic non-protein coding RNA, regulator of reprogramming) contributes to this circuitry by maintaining high levels of CPTFs — it competes for miR-145 binding through its miR-145-recognition motif. When present, *LINC-ROR* prevents miR-145 from repressing the translation of CPTFs and therefore ensures the stem cell fate; when *LINC-ROR* is downregulated, the synthesis of CPTFs is repressed. Notably, within this circuitry, CPTFs activate their own synthesis through a positive feedback loop, thus reinforcing the regulatory circuit[122]. Additionally, *lncRNA-ES1* and *lncRNA-ES2* contribute to maintaining pluripotency of embryonic stem cells by repressing SOX2 neural targets[14]. iPSCs, induced pluripotent stem cells.

tissue-specific manner than mRNAs[15,84]. Additionally, they show specific temporal expression patterns during brain development[15,84].

Interestingly, transcriptome analysis indicated that many brain-expressed lncRNAs are primate- or human-specific lncRNAs[85]; along with the identification of signatures of positive selection and the accelerated evolution in lncRNA regions, these findings indicate that lncRNAs may be crucial effectors in human brain evolution and, possibly, in cognitive and behavioural repertoires[86,87]. A lncRNA that has rapidly evolved since the divergence of humans from the other great apes is highly accelerated region 1A (*HAR1A*). Its expression level correlates with that of reelin, a protein that is crucial for brain development, which suggests that it could coordinate the establishment of regional forebrain organization in a similar manner.

By contrast, other classes of brain-expressed lncRNAs seem to be highly conserved from birds to mammals and have similar spatiotemporal expression profiles, which indicate ancient roles for these lncRNAs in brain development[88]. Moreover, brain-expressed lncRNAs that originate from ultraconserved regions (UCRs) of DNA have been shown to be transcribed from complex genetic loci, where they often overlap or are antisense to genes that encode key developmental regulator proteins[89,90]. Such lncRNAs modulate the activity of their nearby genes by acting as molecular scaffolds to recruit specific factors[90,91]. For example, the co-activator lncRNA *Dlx6* opposite strand transcript 1 (*Dlx6os1*; also known as *Evf2*)[90] is located in a UCR; it is an antisense RNA to distal-less homeobox 6 (*Dlx6*) and is located downstream of *Dlx5* in mice. These Dlx genes are related to the *Drosophila melanogaster Distal-less* (*Dll*) gene; they encode homeodomain transcription factors that are expressed in the developing ventral forebrain and have been postulated to have a role in both forebrain and craniofacial development. *Dlx6os1* controls the expression of *Dlx5*, *Dlx6* and the glutamate decarboxylase 1 gene (*Gad1*; also known as *Gad67*) through both *cis*- and *trans*-acting mechanisms[90]. In *cis*, the transcription of *Dlx6os1* negatively regulates *Dlx6* expression. By contrast, in *trans*, *Dlx6os1* recruits the transcription factors homeobox protein DLX2 (which is an activator) and methyl-CpG-binding protein 2 (MECP2, which is a repressor) to regulate the expression of *Dlx5* and *Gad1* (which encodes an enzyme that is responsible for γ-aminobutyric acid (GABA) synthesis)[90,91]. The loss of *Dlx6os1* function in mice produced a specific neural phenotype that had reduced numbers of GABAergic interneurons in the early postnatal hippocampus. Although the number of GABAergic interneurons and *Gad1* RNA levels returned to normal in the adult hippocampus of *Dlx6os1* mutants, defects in synaptic inhibition were observed, which indicates a crucial role for *Dlx6os1* in neuronal activity *in vivo*[90,91]. The characterization of *Dlx6os1* has opened the way to the identification of a large number of brain lncRNAs that are transcribed from UCRs, which can constitute a new class of transcriptional regulators[90].

Dlx1os shares some functional properties with *Dlx6os1* in that it regulates, in *cis*, the antisense *Dlx1* gene by modulating the level and stability of its transcript[92]. The loss of *Dlx1os* function produced viable and fertile mice that had mild skeletal and neurological phenotypes, which essentially replicated a *Dlx1* gain-of-function phenotype[92].

lncRNAs as inducers of neurogenesis. A large screen for lncRNAs that are involved in neurogenesis identified various lncRNAs for which knockdown blocked the differentiation of human ESCs into mature neurons[14]. Interestingly, the nuclear localized *lncRNA-N1* (also known as *LINC01109*) and *lncRNA-N3* were shown to bind to SUZ12 and REST, which suggests a model in which neuronal lncRNAs reinforce REST-repressing activity by recruiting PRC2 to specific glial lineage genes, thereby promoting neurogenesis. By contrast, the cytoplasmic *lncRNA-N2* (also known as *MIR100HG*) seemed to function as precursor molecules for let-7 and for the neurogenic miR-125b, which are miRNAs that promote proliferation arrest and neuronal differentiation, respectively[22,93].

lncRNAs in the retina. Interestingly, several lncRNAs were found to be specifically expressed in the retina, which is a specialized part of the CNS. The retina is a tractable tissue type for *in vivo* studies because loss of gene functions can be achieved by locally administered RNA interference (RNAi) reagents, in contrast to germline genetic modifications that are required for many *in vivo* studies of the CNS. TABLE 1 shows several examples for which a clear function of lncRNAs in retinal patterning and specification has been established.

Nuclear speckles
A class of nuclear body that is located in interchromatin regions of the nucleoplasm of mammalian cells, which are enriched in pre-mRNA splicing factors.

Zinc-finger nucleases
Artificial proteins that contain a zinc-finger DNA-binding element fused to an endonuclease domain. Double-stranded breaks are produced at specific DNA sequences to induce natural DNA repair. This strategy allows targeted gene deletions, integrations or modifications.

GABAergic interneurons
Neurons of the central nervous system that form a connection between other types of neurons and use the neurotransmitter γ-aminobutyric acid (GABA), which inhibits excitatory responses.

Hippocampus
A part of the brain that is specifically responsible for storing and retrieving memories.

lncRNAs with non-canonical structures. A circRNA that is derived from non-canonical splicing of an antisense transcript (*CDR1AS*; also known as *ciRS-7*) to the cerebellar degeneration-related protein 1 (*CDR1*) mRNA was recently identified in the human brain[53], as well as in mouse cortical pyramidal neurons and interneurons[53]. Interestingly, this circRNA functions as a sponge for miR-7 through 70 selectively conserved miR-7 target sites, thus regulating endogenous miR-7 targets[54,55]. Zebrafish was used to study the *in vivo* function of this circRNA because it has lost the *cdr1* locus while maintaining miR-7 expression in the embryonic brain during evolution. Embryos that expressed ectopic *CDR1AS* developed brain defects and had a smaller midbrain region, which is similar to the phenotype of the loss of miR-7 function obtained by treatment with morpholino oligonucleotides[55]. Therefore, circRNAs may also have roles in neuronal function and in neurological disorders[54,55].

In conclusion, the multifaceted functions of lncRNAs seem appropriate for the complex regulatory demands of the CNS, and further studies of lncRNAs may uncover details of even more complex brain function and of the pathogenetic events that underlie neurodegenerative disorders. However, deeper analyses of the differences that are often found between *in vivo* and *in vitro* systems, as well as those between different knockdown strategies, are required for a more reliable understanding of lncRNA functions in the development of the brain and the CNS.

Development of other organs

In addition to extensive roles in brain development, lncRNAs are known to function in the development of diverse organs and tissue types, which are described below.

Heart. One of the best examples of the importance of lncRNAs in organ development is provided by two lncRNAs that are involved in mouse cardiac development — braveheart (*Bvht*; also known as *Gm20748*)[94] and *Foxf1* adjacent non-coding developmental regulatory RNA (*Fendrr*)[95]. These lncRNAs were identified from the mesoderm, from which the heart originates[94,95].

The knockdown of *Bvht* by RNAi in mouse ESCs and neonatal cardiomyocyte cultures affected cardiac-specific gene expression and altered development into mature cardiomyocytes[94], thus suggesting a possible role for *Bvht* in cardiac tissue regeneration after injuries. *Bvht* was shown to interact with PRC2, which suggests that it functions by mediating epigenetic regulation of cardiac commitment[94]. Notably, *Bvht* is specific to mice and is not expressed in rats or humans; whether alternative molecular components carry out roles that are equivalent to *Bvht* in other mammals is currently unclear.

In the case of *Fendrr*, a 60% reduction of expression by RNAi *in vivo* did not show any apparent phenotypes[95]. By contrast, the knockout of *Fendrr* resulted in embryonic lethality owing to impaired heart function and to deficits in the body wall, thus indicating the importance of null-mutant models for uncovering roles for lncRNAs[95]. Although *Fendrr* was suggested to interact with components of both repressive chromatin-associated complexes (such as PRC2) and activating chromatin-associated complexes (such as MLL1) in mouse embryos, chromatin immunoprecipitation (ChIP) analysis following *Fendrr* deletion showed a change in occupancy at *Fendrr*-target genes only for the repressive PRC2 (REF. 95). Unlike *Bvht*, *Fendrr* has a human orthologous transcript *FENDRR* that, similarly to murine *Fendrr*, is also associated with PRC2 (REF. 24).

Skeletal muscle. One of the first lncRNAs that was identified with a role in myogenesis was *Linc-MD1* (long non-coding RNA, muscle differentiation 1). This lncRNA is expressed in a specific temporal window during *in vitro* muscle differentiation of mouse myoblasts and was shown to control the progression from early to late phases of muscle differentiation by functioning as a ceRNA. Through competition for the binding of miR-133 and miR-135, it regulates the expression of mastermind-like protein 1 (MAML1) and myocyte-specific enhancer factor 2C (MEF2C), which are transcription factors that activate late-differentiation muscle genes[21]. *LINCMD1* is conserved between mice and humans[96], and its expression is strongly reduced in myoblasts of patients with Duchenne muscular dystrophy[21]. Interestingly, in these cells, the recovery of *LINCMD1* levels rescued the correct timing of *in vitro* differentiation, which suggests a relevant conserved role in the control of muscle differentiation[21] (FIG. 4).

More recently, the imprinted *H19* lncRNA, which is highly expressed in the developing embryo and in adult muscle, was shown to work as a ceRNA for let-7 and to control muscle differentiation. Indeed, the depletion of *H19* caused precocious muscle differentiation — a phenotype that is recapitulated by let-7 overexpression[97]. As high let-7 levels are generally associated with increased cellular differentiation, it was hypothesized that *H19* inhibits let-7 activity, thereby preventing precocious differentiation[97].

Another lncRNA that is linked to neuromuscular disease is D4Z4-binding element transcript (*DBE-T*), which is selectively expressed in patients with facioscapulohumeral muscular dystrophy (FSHD). *DBE-T* recruits histone-lysine *N*-methyltransferase ASH1L — a component of the MLL1 complex — which results in H3K36 dimethylation and in aberrant transcriptional activation of the *FSHMD1A* (also known as *FSHD*) locus in patients with FSHD[98].

Moreover, lncRNAs that regulate gene expression by driving STAU1-mediated mRNA decay have also been recently linked to myogenesis — sbsRNAs induce mRNA degradation by recruiting STAU1 to target mRNAs through base pairing with short interspersed elements (SINEs) in the 3′ untranslated region of target mRNAs (FIG. 2Ab). Remarkably, downregulating the abundance of three of the four sbsRNAs that were tested altered the rate of mouse myoblast differentiation *in vitro*[47].

Morpholino oligonucleotides
Oligonucleotides that are modified to be highly stable in the cell; they are used as antisense RNA to block cell components from accessing the target site for which they are designed.

Chromatin immunoprecipitation
(ChIP). A method used to determine whether a given protein binds to, or is localized to, specific chromatin loci *in vivo*.

Duchenne muscular dystrophy
A severe genetic disorder that is characterized by the rapid progression of muscle degeneration, which leads to a loss of ambulation and death. It is due to mutations in the dystrophin gene that prevent its production.

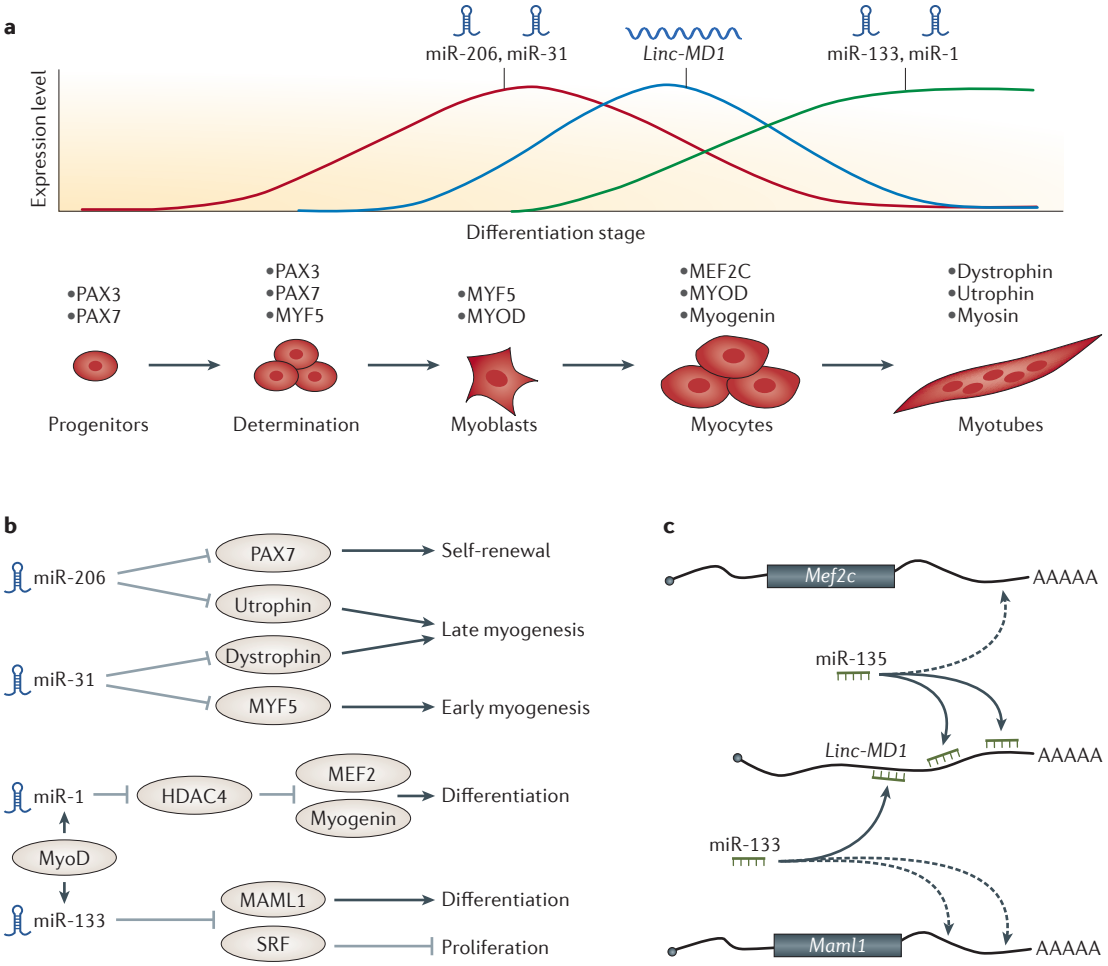


Figure 4 | **ncRNAs and muscle differentiation.** **a** | A schematic representation of the differentiation stages from progenitor muscle cells to terminally differentiated fibres is shown. The cells are labelled with the characteristic proteins that are expressed at each stage. These include master transcription factors that regulate the switch from one stage to the following one — such as paired box protein Pax-3 (PAX3), PAX7, myogenic factor 5 (MYF5), myoblast determination protein (MYOD), myocyte enhancer factor 2C (MEF2C) and myogenin — as well as the late myogenic proteins dystrophin, utrophin and myosin[123]. The graph shows the corresponding temporal expression patterns of selected non-coding RNAs (ncRNAs). **b** | MicroRNAs (miRNAs) cooperate with transcription factors to sharpen their temporal expression pattern[124]; for example, miR-206 and miR-31 repress expression of the self-renewal factor PAX7 and the early myogenic factor MYF5, respectively. The same miRNAs prevent the early activation of late myogenic proteins, such as utrophin and dystrophin[125]. By contrast, late myogenic miRNAs reinforce late differentiation stages; for example, miR-1 controls the expression of later myogenic transcription factors MEF2C and myogenin through the repression of histone deacetylase 4 (HDAC4). **c** | In these circuitries, the role of *Linc-MD1* (long non-coding RNA, muscle differentiation 1) is crucial. It further reinforces the switch from early to late differentiation gene expression by acting as a 'sponge' to limit the repressive effect of miR-133 on mastermind-like 1 (*Maml1*) and of miR-135 on *Mef2c*. SRF, serum response factor.

Skin, haematopoietic and adipose development. Roles for lncRNAs have been identified in the epidermis. Transcriptome sequencing of progenitor and differentiating human keratinocytes identified *TINCR* as the most highly induced lncRNA during keratinocyte differentiation[45]. *TINCR*-deficient epidermis lacked terminal differentiation ultrastructure, including keratohyalin granules and intact lamellar bodies. Interestingly, *TINCR* also binds to STAU1; however, unlike the sbsRNAs described above, the *TINCR*–STAU1 complex targets mRNAs that have a 25-nucleotide 'TINCR box' motif, which results in the stabilization of differentiation-associated mRNAs, such as keratin 80 (*KRT80*), to ensure their expression and cellular differentiation[45] (FIG. 2Ac).

Relevant lncRNAs have also been identified in haematopoiesis and adipogenesis[99,100]. The analysis of lncRNAs during erythroid differentiation of mouse fetal liver progenitors allowed the identification of *lincRNA-EPS* (erythroid prosurvival). The knockdown of *lincRNA-EPS* in mouse erythroid progenitors blocked differentiation and promoted apoptosis by inhibiting the expression of the pro-apoptotic PYD and CARD domain-containing gene (*Pycard*) through a mechanism that is still undefined[99]. More recently, lncRNAs

were profiled in mice during differentiation to white and brown adipose tissue. Loss-of-function studies identified ten lncRNAs that have specific roles in adipogenesis[100].

lncRNAs in environmental and stress responses

An emerging function for lncRNAs is their contribution to various genetic programmes that enable response to different environmental conditions. One of the first and best-studied examples is the regulation of flowering in plants. In *Arabidopsis thaliana*, the transcriptional repressor gene *FLOWERING LOCUS C* (*FLC*) has an important role in this process by blocking the expression of genes that are required for the switch to flowering. lncRNAs have been shown to function in *FLC* regulation in various ways[101]. The long exposure to cold during winter — a process known as vernalization — seems to induce the expression of a sense transcript from *FLC* called *COLD-ASSISTED INTRONIC NON-CODING RNA* (*COLDAIR*). *COLDAIR* is thought to function similarly to animal lncRNAs in the formation of repressive heterochromatin through a physical association with PRC2 (REF. 102). *FLC* is also regulated by a set of antisense lncRNAs called *COLD-INDUCED LONG ANTISENSE INTRAGENIC RNA* (*COOLAIR*) that encompass the whole *FLC* sense transcription unit[101]. These antisense RNAs are upregulated in response to cold temperatures, whereas they are alternatively polyadenylated in warm temperatures[103]. The use of the proximal polyadenylation site in warm temperatures is linked to histone demethylation in the gene body and leads to reduced *FLC* transcription[104]. *COOLAIR* transcription is repressed in warm temperatures by a mechanism that involves the stabilization of an R-loop (that is, an RNA–DNA hybrid structure) in its promoter region by the NDX1 homeobox protein homologue[105].

More recently, a novel lncRNA has been identified in mice as being activated by a stress signalling pathway that controls the activity of the mammalian target of rapamycin (mTOR) kinase, which is an important regulator of translation[42]. The lncRNA *Uchl1-as1* is an antisense transcript to the neuron-specific *Uchl1* gene, which functions in protein ubiquitylation and has roles in brain function and various neurodegenerative diseases. *Uchl1-as1* contains an embedded SINEB2 element that stimulates *Uchl1* translation and thus UCHL1 protein expression under stress conditions[42]. In particular, upon stress-induced inhibition of mTOR activity and the resulting repression of cap-dependent translation, *Uchl1-as1* is exported from the nucleus to the cytoplasm, where it can base pair with the *Uchl1* mRNA and stimulate its cap-independent translation. As this activation of UCHL1 expression does not require *de novo* RNA synthesis, it provides a rapid response to environmental changes.

Conclusions and perspectives

The discoveries linked to lncRNA function go far beyond the identification of new mechanisms that regulate gene expression. The organization of lncRNA-coding loci, which are often finely intertwined with protein-coding ones, has added a high degree of complexity in the comprehension of the structure, function and evolution of our genome. Moreover, despite the burst of interest in identifying new lncRNAs and in setting up new methodologies to characterize their function, a future topic of interest will be the origin and evolution of lncRNAs.

One interesting feature relates to the contribution of transposable elements to the genesis and regulation of lncRNAs[18,20]. Their relevance is supported by the discovery that, in vertebrates, transposable elements occur in more than two-thirds of mature lncRNAs, whereas they seldom occur in protein-coding transcripts. Moreover, transposable elements were found in biased positions and orientations within lncRNAs, particularly at their transcription start sites, which suggests a role in the regulation of lncRNA transcription[18,20]. Therefore, it has been proposed that transposable elements may contribute to lncRNA evolution and that they function by conferring on lncRNAs tissue-specific expression from existing transcriptional regulatory signals[18,20].

Phylogenetic analysis is generally one of the first approaches to be considered when searching for lncRNA function. However, bioinformatic analysis tools should be implemented to account for the differential evolutionary pressure that operates on the various lncRNA subdomains; such pressure acts either on the primary sequence of lncRNAs (for antisense effectors against RNA or DNA targets) or through their secondary structure (for protein-binding domains). In this respect, the modular scaffold hypothesis suggests that lncRNAs have undergone extensive molecular bricolage by the gain or loss of different modules, which provides alternative and more complex functions that might be subjected to evolutionary selection[8,9,13,14]. Moreover, the degree of lncRNA conservation often does not indicate functional relevance; for example, non-coding genes such as *XIST* and nuclear paraspeckle assembly transcript 1 (*NEAT1*) have undergone rapid sequence evolution while preserving their functional roles[106,107], and highly accelerated evolution in ncRNA regions has been suggested to contribute to the development of complex structures, such as the brain[86,87].

Another relevant question concerns the non-coding definition of a transcript. In fact, it is possible that specific lncRNAs have previously uncharacterized coding potential for small peptides (<50 amino acids) with biological function. Even if lncRNAs are bound by ribosomes[108], it has been recently observed that they show patterns of ribosome occupancy that are similar to those typical of non-coding sequences, which indicates that this assay is not sufficient to classify transcripts as coding or non-coding[109]. Therefore, additional efforts are required to define the functional implications of the association between lncRNAs and ribosomes, and to establish whether specific subclasses of lncRNAs with coding potential do indeed exist.

Although mechanistic models are starting to emerge, at the core of lncRNA functional studies is the need for appropriate model systems for *in vivo* studies, which should allow a better understanding of the evolution and functions of lncRNAs, and their roles in both development and differentiation. However, owing to the great

Phylogenetic analysis
Comparison of DNA, RNA or protein sequences in different organisms that enables one to establish their evolutionary relationships.

Bricolage
Construction or creation from a diverse range of available things.

variability in the evolutionary conservation or diversification of such RNAs, appropriate animal model systems are not always available. Notably, in a large screen carried out in zebrafish, although many lncRNAs shared characteristics with their mammalian orthologues, only a few of them had detectable sequence similarity[110]. Even among mammals, conservation might be weak; hence, mouse models might not always reflect functions in humans. Moreover, given the highly cell-type-specific expression pattern of many lncRNAs[15–17], they are likely to elicit differential developmental or differentiation programmes in different organs, as is the case for *MALAT1* (REFS 81–83). Therefore, a more exhaustive knowledge of their activity in different cells and tissues of the body is required to elucidate possible tissue-specific functions.

One of the most powerful techniques to study the function of a gene *in vivo* is to disrupt its expression through targeted recombination. However, this methodology requires special consideration when it is applied to lncRNA loci — their complex structure and frequent overlap with other transcripts mean that the disruption of lncRNA loci might interfere with the function of nearby genes, thus confounding the interpretation of the molecular causes of any resultant phenotype. Therefore, gene targeting should be carefully conceived to ensure a truncation of the lncRNA of interest while leaving the surrounding locus unaffected. Recent designs for lncRNA inactivation have successfully used the targeted insertion of multiple polyadenylation sites, which prevents the transcription of full-length lncRNAs[29,91,92,111]. Additional novel strategies need to be developed for generating suitable conditional and loss-of-function model systems for lncRNA studies. An important issue to consider when analysing loss-of-function phenotypes of lncRNAs *in vivo* is the possibility of functional redundancy or of compensatory circuitries that would hide their direct activity, similar to what has been observed for *in vivo* miRNA depletions[22].

The regulation of lncRNA expression is also a relevant topic that has so far been poorly addressed. Besides transcriptional control, post-transcriptional regulation will also be a relevant aspect to investigate. Major issues are related to understanding how polyadenylated lncRNAs are retained in the nucleus and to dissecting which protein interactions control the maturation and subcellular localization of lncRNAs. For example, it remains to be determined how some lncRNAs — such as circRNAs or polyadenylated lncRNAs that overlap with primary miRNA sequences — accumulate in the cytoplasm[21,53,54]. Such lncRNAs are much more abundant than previously thought, and the nature of the *cis*- and *trans*-acting factors that regulate their biogenesis and cellular localization are interesting new issues to be studied.

1. Maeda, N. *et al.* Transcript annotation in FANTOM3: mouse gene catalog based on physical cDNAs. *PLoS Genet.* **2**, e62 (2006).
2. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
3. Taft, R. J., Pheasant, M. & Mattick, J. S. The relationship between nonprotein-coding DNA and eukaryotic complexity. *Bioessays* **29**, 288–299 (2007).
4. Mattick, J. S. & Makunin, I. V. Non-coding RNA. *Hum. Mol. Genet.* **1**, 17–29 (2006).
5. Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **8**, 145–166 (2012).
6. 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).
7. Batista, P. J. & Chang, H. Y. Long noncoding RNAs: cellular address codes in development and disease. *Cell* **152**, 1298–1307 (2013).
8. Guttman, M. & Rinn, J. L. Modular regulatory principles of large non-coding RNAs. *Nature* **482**, 339–346 (2012).
9. Zappulla, D. C. & Cech, T. R. Yeast telomerase RNA: a flexible scaffold for protein subunits. *Proc. Natl Acad. Sci. USA* **101**, 10024–10029 (2004).
10. Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nature Genet.* **30**, 167–174 (2002).
11. Tsai, M. C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
12. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
13. Guttman, M. *et al.* lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**, 295–300 (2011).
This study reports an exhaustive identification and characterization of lncRNAs that are involved in the control of pluripotency and differentiation.
14. Ng, S. Y., Johnson, R. & Stanton, L. W. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J.* **31**, 522–533 (2012).
This paper characterizes several lncRNAs that are involved in the control of pluripotency and neural differentiation; examples of lncRNAs that interact with multiple transcriptional modulators were found, which supports the modular scaffold hypothesis.
15. Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F. & Mattick, J. S. Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl Acad. Sci. USA* **105**, 716–721 (2008).
16. Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).
17. Pauli, A. *et al.* Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* **22**, 577–591 (2012).
18. Kelley, D. & Rinn, J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol.* **13**, R107 (2012).
19. Liang, Q., Xu, Z., Xu, R., Wu, L. & Zheng, S. Expression patterns of non-coding spliced transcripts from human endogenous retrovirus HERV-H elements in colon cancer. *PLoS ONE* **7**, e29950 (2012).
20. Kapusta, A. *et al.* Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* **9**, e1003470 (2013).
21. Cesana, M. *et al.* A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147**, 358–369 (2011).
This paper describes a lncRNA that controls the translation of late myogenic factors and the progression to late differentiation stages through competition for common miRNAs.
22. Ebert, M. S. & Sharp, P. A. Roles for microRNAs in conferring robustness to biological processes. *Cell* **149**, 515–524 (2012).
23. Wang, Y. *et al.* Endogenous miRNA sponge *lincRNA-RoR* regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. *Dev. Cell* **25**, 69–80 (2013).
This paper describes a lncRNA that controls the expression of core pluripotency transcriptional factors by competing for common miRNAs.
24. 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).
25. 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).
26. Nagano, T. *et al.* The *Air* noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717–1720 (2008).
27. Pandey, R. *et al. Kcnq1ot1* antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* **32**, 232–246 (2008).
28. Martianov, I., Ramadass, A., Serra Barros, A., Chow, N. & Akoulitchev, A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* **445**, 666–670 (2007).
29. Latos, P. A. *et al. Airn* transcriptional overlap, but not its lncRNA products, induces imprinted *Igf2r* silencing. *Science* **338**, 1469–1472 (2012).
30. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
31. Bertani, S., Sauer, S., Bolotin, E. & Sauer, F. The noncoding RNA *Mistral* activates *Hoxa6* and *Hoxa7* expression and stem cell differentiation by recruiting MLL1 to chromatin. *Mol. Cell* **43**, 1040–1046 (2011).
32. Ørom, U. A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46–58 (2010).
33. Li, W. *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516–520 (2013).
34. Jeon, Y. & Lee, J. T. YY1 tethers *Xist* RNA to the inactive X nucleation center. *Cell* **146**, 119–133 (2011).
35. Schmitz, K. M., Mayer, C., Postepska, A. & Grummt, I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev.* **24**, 2264–2269 (2010).
36. Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA–chromatin interactions. *Mol. Cell* **44**, 667–678 (2011).
37. Hung, T. *et al.* Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nature Genet.* **43**, 621–629 (2011).
38. Sun, S. *et al. Jpx* RNA activates *Xist* by evicting CTCF. *Cell* **153**, 1537–1551(2013).

39. Wang, X. *et al.* Induced ncRNAs allosterically modify RNA-binding proteins in *cis* to inhibit transcription. *Nature* **454**, 126–130 (2008).
40. Yang, L. *et al.* ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell* **147**, 773–788 (2011).
41. Engreitz, J. M. *et al.* The *Xist* lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
42. Carrieri, C. *et al.* Long non-coding antisense RNA controls *Uchl1* translation through an embedded SINEB2 repeat. *Nature* **491**, 454–457 (2012).
This study shows that a brain-specific lncRNA activates translation of an antisense mRNA through an embedded repetitive element.
43. Yoon, J. H. *et al. LincRNA-p21* suppresses target mRNA translation. *Mol. Cell* **47**, 648–655 (2012).
44. Faghihi, M. A. *et al.* Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β-secretase. *Nature Med.* **14**, 723–730 (2008).
45. Kretz, M. *et al.* Control of somatic tissue differentiation by the long non-coding RNA *TINCR*. *Nature* **493**, 231–235 (2013).
This paper shows that STAU1 can be tethered by a lncRNA on specific target mRNAs that share a 25-nucleotide conserved box. However, different from STAU1-mediated decay, such tethering confers stability on the target mRNAs.
46. Gong, C. & Maquat, L. E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature* **470**, 284–288 (2011).
47. Wang, J., Gong, C. & Maquat, L. E. Control of myogenesis by rodent SINE-containing lncRNAs. *Genes Dev.* **27**, 793–804 (2013).
48. Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P. P. The *ceRNA* hypothesis: the Rosetta stone of a hidden RNA language. *Cell* **146**, 353–358 (2011).
49. Franco-Zorrilla, J. M. *et al.* Target mimicry provides a new mechanism for regulation of microRNA activity. *Nature Genet.* **39**, 1033–1037 (2007).
50. Poliseno, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033–1038 (2010).
This paper proposes the ceRNA hypothesis, in which coding and non-coding RNAs can crosstalk through competition for shared miRNA-binding motifs.
51. Karreth, F. A. *et al. In vivo* identification of tumor-suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. *Cell* **147**, 382–395 (2011).
52. Sumazin, P. *et al.* An extensive microRNA-mediated network of RNA–RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell* **147**, 370–381 (2011).
53. Hansen, T. B. *et al.* miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J.* **30**, 4414–4422 (2011).
54. Hansen, T. B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384–388 (2013).
55. Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338 (2013).
This paper identifies a large class of cellular circRNAs; for one of these, a ceRNA function is shown in the control of neuronal functions.
56. Lee, J. T. & Bartolomei, M. S. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* **152**, 1308–1323 (2013).
57. Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for *Xist* in X chromosome inactivation. *Nature* **379**, 131–137 (1996).
58. Schoeftner, S. *et al.* Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J.* **25**, 3110–3122 (2006).
59. Nozawa, R. S. *et al.* Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. *Nature Struct. Mol. Biol.* **20**, 566–573 (2013).
60. Lee, J. T. Disruption of imprinted X inactivation by parent-of-origin effects at *Tsix*. *Cell* **103**, 17–27 (2000).
61. Sado, T., Wang, Z., Sasaki, H. & Li, E. Regulation of imprinted X-chromosome inactivation in mice by *Tsix*. *Development* **128**, 1275–1286 (2001).
62. Tian, D., Sun, S. & Lee, J. T. The long noncoding RNA, *Jpx*, is a molecular switch for X chromosome inactivation. *Cell* **143**, 390–403 (2010).
63. Yildirim, E. *et al. Xist* RNA is a potent suppressor of hematologic cancer in mice. *Cell* **152**, 727–742 (2013).
64. Fitzpatrick, G. V., Soloway, P. D. & Higgins, M. J. Regional loss of imprinting and growth deficiency in mice with a targeted deletion of *KvDMR1*. *Nature Genet.* **32**, 426–431 (2002).
65. Sleutels, F., Zwart, R. & Barlow, D. P. The non-coding *Air* RNA is required for silencing autosomal imprinted genes. *Nature* **415**, 810–813 (2002).
66. Mancini-Dinardo, D., Steele, S. J., Levorse, J. M., Ingram, R. S. & Tilghman, S. M. Elongation of the *Kcnq1ot1* transcript is required for genomic imprinting of neighboring genes. *Genes Dev.* **20**, 1268–1282 (2006).
67. Colin, J., Libri, D. & Porrua, O. Cryptic transcription and early termination in the control of gene expression. *Genet. Res. Int.* **2011**, 653494 (2011).
68. Barber, B. A. & Rastegar, M. Epigenetic control of Hox genes during neurogenesis, development, and disease. *Ann. Anat.* **192**, 261–274 (2010).
69. Kostic, D. & Capecchi, M. R. Targeted disruptions of the murine *Hoxa-4* and *Hoxa-6* genes result in homeotic transformations of components of the vertebral column. *Mech. Dev.* **46**, 231–247 (1994).
70. Zhang, X. *et al.* A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood* **113**, 2526–2534 (2009).
71. Maamar, H., Cabili, M.N., Rinn, J. & Raj, A. *linc-HOXA1* is a noncoding RNA that represses *Hoxa1* transcription in *cis*. *Genes Dev.* **27**, 1260–1271 (2013).
72. Li, L. *et al.* Targeted disruption of *Hotair* leads to homeotic transformation and gene derepression. *Cell Rep.* **5**, 3–12 (2013).
73. Suemori, H. & Noguchi, S. Hox C cluster genes are dispensable for overall body plan of mouse embryonic development. *Dev. Biol.* **220**, 333–342 (2000).
74. Schorderet, P. & Duboule, D. Structural and functional differences in the long non-coding RNA *Hotair* in mouse and human. *PLoS Genet.* **7**, e1002071 (2011).
75. Gupta, R. A. *et al.* Long non-coding RNA *HOTAIR* reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010).
76. Loewer, S. *et al. Large intergenic non-coding RNA-RoR* modulates reprogramming of human induced pluripotent stem cells. *Nature Genet.* **42**, 1113–1117 (2010).
77. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
78. Dinger, M. E. *et al.* Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* **9**, 1433–1445 (2008).
79. Sheik Mohamed, J., Gaughwin, P. M., Lim, B., Robson, P. & Lipovich, L. Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells. *RNA* **16**, 324–337 (2010).
80. Qureshi, I. A. & Mehler, M. F. Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. *Nature Rev. Neurosci.* **13**, 528–541 (2012).
81. Bernard, D. *et al.* A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J.* **29**, 3082–3093 (2010).
82. Zhang, B. *et al.* The lncRNA *Malat1* is dispensable for mouse development but its transcription plays a *cis*-regulatory role in the adult. *Cell Rep.* **2**, 111–123 (2012).
83. Gutschner, T. *et al.* The noncoding RNA *MALAT1* is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res.* **73**, 1180–1189 (2013).
84. Ramos, A. D. *et al.* Integration of genome-wide approaches identifies lncRNAs of adult neural stem cells and their progeny *in vivo*. *Cell Stem Cell.* **12**, 616–628 (2013).
85. Xu, A. G. *et al.* Intergenic and repeat transcription in human, chimpanzee and macaque brains measured by RNA-seq. *PLoS Comput. Biol.* **6**, e1000843 (2010).
86. Pollard, K. S. *et al.* An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**, 167–172 (2006).
87. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).
88. Dorus, S. *et al.* Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. *Cell* **119**, 1027–1040 (2004).
89. Amaral, P. P. *et al.* Complex architecture and regulated expression of the *Sox2ot* locus during vertebrate development. *RNA* **11**, 2013–2027 (2009).
90. Feng, J. *et al.* The *Evf-2* noncoding RNA is transcribed from the *Dlx-5/6* ultraconserved region and functions as a *Dlx-2* transcriptional coactivator. *Genes Dev.* **20**, 1470–1484 (2006).
91. Bond, A. M. *et al.* Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nature Neurosci.* **12**, 1020–1027 (2009).
92. Kraus, P. *et al.* Making sense of *Dlx1* antisense RNA. *Dev. Biol.* **376**, 224–235 (2013).
93. Le, M. T. *et al. MicroRNA-125b* promotes neuronal differentiation in human cells by repressing multiple targets. *Mol. Cell. Biol.* **29**, 5290–5305 (2009).
94. Klattenhoff, C. A. *et al. Braveheart*, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* **152**, 1–14 (2013).
95. 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).
References 94 and 95 highlight the importance of lncRNAs during lineage commitment and in providing a new layer of regulation that is involved in determining cardiac cell fate.
96. Twayana, S. *et al.* Biogenesis and function of non-coding RNAs in muscle differentiation and in Duchenne muscular dystrophy. *Bioch. Soc. Trans.* **41**, 844–849 (2013).
97. Kallen, A. N. *et al.* The imprinted *H19* lncRNA antagonizes *let-7* microRNAs. *Mol. Cell* **52**, 101–112 (2013).
98. Cabianca, D. S. *et al.* A long ncRNA links copy number variation to a Polycomb/Trithorax epigenetic switch in FSHD muscular dystrophy. *Cell* **149**, 819–831 (2012).
99. Hu, W., Yuan, B., Flygare, J. & Lodish, H. F. Long noncoding RNA-mediated anti-apoptotic activity in murine erythroid terminal differentiation. *Genes Dev.* **25**, 2573–2578 (2011).
100. Sun, L. *et al.* Long noncoding RNAs regulate adipogenesis. *Proc. Natl Acad. Sci. USA* **110**, 3387–3392 (2013).
101. Ietswaart, R., Wu, Z. & Dean, C. Flowering time control: another window to the connection between antisense RNA and chromatin. *Trends Genet.* **28**, 445–453 (2012).
102. Heo, J. B. & Sung, S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* **331**, 76–79 (2011).
103. Liu, F. *et al.* The *Arabidopsis* RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate *FLC*. *Mol. Cell* **28**, 398–407 (2007).
104. Liu, F., Marquardt, S., Lister, C., Swiezewski, S. & Dean, C. Targeted 3′ processing of antisense transcripts triggers *Arabidopsis FLC* chromatin silencing. *Science* **327**, 94–97 (2010).
105. Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N. J. & Dean, C. R-loop stabilization represses antisense transcription at the *Arabidopsis FLC* locus. *Science* **340**, 619–621 (2013).
106. 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).
107. Nesterova, T. B. *et al.* Characterization of the genomic *Xist* locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Res.* **11**, 833–849 (2001).
108. Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M. & Weissman, J. S. The ribosome profiling strategy for monitoring translation *in vivo* by deep sequencing of ribosome-protected mRNA fragments. *Nature Protoc.* **7**, 1534–1550 (2012).
109. Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S. & Lander, E. S. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* **154**, 240–251 (2013).
110. 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).
111. Ohhata, T., Hoki, Y., Sasaki, H. & Sado, T. Crucial role of antisense transcription across the *Xist* promoter in *Tsix*-mediated *Xist* chromatin modification. *Development* **135**, 227–235 (2008).
112. Rinn, J. L. *et al.* The transcriptional activity of human chromosome 22. *Genes Dev.* **17**, 529–540 (2003).

113. Velculescu, V. E., Zhang, L., Vogelstein, B. & Kinzler, K. W. Serial analysis of gene expression. *Science* **270**, 484–487 (1995).
114. Wei, C. L. *et al.* 5′ long serial analysis of gene expression (LongSAGE) and 3′ LongSAGE for transcriptome characterization and genome annotation. *Proc. Natl Acad. Sci. USA* **101**, 11701–11706 (2004).
115. Matsumura, H., Krüger, D. H., Kahl, G. & Terauchi, R. SuperSAGE: a modern platform for genome-wide quantitative transcript profiling. *Curr. Pharm. Biotechnol.* **9**, 368–374 (2008).
116. Shiraki, T. *et al.* Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc. Natl Acad. Sci. USA* **100**, 15776–15781 (2003).
117. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nature Methods* **5**, 621–628 (2008).
118. Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).
119. Kiss, T. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell* **109**, 145–148 (2002).
120. Tay, Y., Zhang, J., Thomson, A. M., Lim, B. & Rigoutsos, I. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* **455**, 1124–1128 (2008).
121. Melton, C., Judson, R. L. & Blelloch, R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature* **463**, 621–626 (2010).
122. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
123. Buckingham, M. & Vincent, S. D. Distinct and dynamic myogenic populations in the vertebrate embryo. *Curr. Opin. Genet. Dev.* **19**, 444–453 (2009).
124. Marrone, A. K. & Shcherbata, H. R. Dystrophin orchestrates the epigenetic profile of muscle cells via miRNAs. *Front. Genet.* **2**, 64–72 (2011).
125. Cacchiarelli, D. *et al. miR-31* modulates dystrophin expression: new implications for Duchenne muscular dystrophy therapy. *EMBO Rep.* **12**, 136–141 (2011).
126. Rapicavoli, N. A., Poth, E. M. & Blackshaw, S. The long noncoding RNA *RNCR2* directs mouse retinal cell specification. *BMC Dev. Biol.* **10**, 49 (2010).
127. Rapicavoli, N. A., Poth, E. M., Zhu, H. & Blackshaw, S. The long noncoding RNA *Six3OS* acts in *trans* to regulate retinal development by modulating *Six3* activity. *Neural Dev.* **6**, 32 (2011).
128. Young, T. L., Matsuda, T. & Cepko, C. L. The noncoding RNA *taurine upregulated gene 1* is required for differentiation of the murine retina. *Curr. Biol.* **15**, 501–512 (2005).
129. Meola, N., Pizzo, M., Alfano, G., Surace, E. M. & Banfi, S. The long noncoding RNA *Vax2os1* controls the cell cycle progression of photoreceptor progenitors in the mouse retina. *RNA* **18**, 111–123 (2012).

Acknowledgements
The authors thank members of the Bozzoni and Fatica laboratories for discussion, and J. Rinn, C. Dean and P. Avner for their suggestions. They apologize for papers that are not discussed owing to space limitations. They acknowledge support from FP7-PEOPLE-2011-ITN Project HemID (289611), Telethon (GPP11149), Parent Project Italia, Associazione Italiana per la Ricerca sul Cancro, Italian Institute of Technology "SEED", Fondo per gli Investimenti per la Ricerca di Base, Programmi di Ricerca di Interesse Nazionale and the EPIGEN Project.

Competing interests statement
The authors declare no competing interests.