The functions of long noncoding RNAs in development and stem cells

Rotem Ben-Tov Perry* and Igor Ulitsky*

ABSTRACT

Eukaryotic genomes are pervasively transcribed, with tens of thousands of RNAs emanating from uni- and bi-directional promoters and from active enhancers. In vertebrates, thousands of loci in each species produce a class of transcripts called long noncoding RNAs (lncRNAs) that are typically expressed at low levels and do not appear to give rise to functional proteins. Substantial numbers of lncRNAs are expressed at specific stages of embryonic development, in many cases from regions flanking key developmental regulators. Here, we review the known biological functions of such lncRNAs and the emerging paradigms of their modes of action. We also provide an overview of the growing arsenal of methods for lncRNA identification, perturbation and functional characterization.

KEY WORDS: Long noncoding RNAs, Regulatory networks, Stem cell differentiation

Introduction

It is thought that virtually all protein-coding genes in vertebrate genomes have already been discovered, and it is established that the key drivers of differences between species, as well as the majority of genetic variants associated with human traits and diseases, map to regions that do not overlap with protein-coding exons. It is also established that much of the genomic sequence between protein-coding genes is transcribed into long noncoding RNAs (lncRNAs) (Clark et al., 2011). However, because the vast majority of the loci transcribed into lncRNAs [up to 50,000 in humans (Iyer et al., 2015)] are expressed at low levels and are poorly conserved in other species, there is uncertainty about how many human lncRNAs are functional. Nevertheless, there are ∼1000 human lncRNAs that are moderately to highly expressed and show signs of evolutionary constraint on their sequences and their transcription, and ∼300 of these are conserved outside mammals in other vertebrates (Hezroni et al., 2015). Other eukaryotes also express hundreds to thousands of lncRNAs, although none has so far been found to be orthologous to a vertebrate lncRNA.

An increasing number of human and mouse lncRNAs have been implicated as key regulators in a variety of cellular processes, including proliferation, apoptosis and response to stress. Many lncRNAs are differentially expressed in human diseases (Huarte, 2015; Lorenzen and Thum, 2016; Zhao and Lin, 2015), an observation that boosts interest in their study as potential biomarkers and therapeutic targets. Several features of lncRNAs make them attractive candidates for having important roles in embryonic development: (1) stem and progenitor cells, which are typically associated with more open and active chromatin, produce numerous lncRNAs (Guttman et al., 2009); (2) lncRNAs are typically expressed in very specific patterns, both spatially and temporally (Cabilio et al., 2011; Ulitsky et al., 2011); and (3) many lncRNAs are transcribed from large regions flanking transcription factor (TF) genes and other regulators that are important during embryonic development (Ulitsky et al., 2011). Indeed, the list of lncRNAs implicated in embryonic development and in the acquisition of cell identity during differentiation is rapidly growing. Lagging behind, but also progressing, is our understanding of how these functions are carried out.

In this Review, we present an overview of the current approaches and methods for identifying and annotating lncRNAs. We then survey the diverse functions of lncRNAs in embryonic development, highlighting the insights into lncRNA functions and modes of action that have recently been obtained and the remaining challenges and open questions in the field.

The identification and annotation of lncRNAs

Approaches for reconstructing transcriptomes and sifting them for the purposes of identifying lncRNAs have been described in detail elsewhere (Housman and Ulitsky, 2015; Ulitsky and Bartel, 2013); here, we provide just an overview. The transcriptome of a tissue or a developmental state is typically reconstructed using RNA sequencing (RNA-seq) libraries prepared from either poly(A)-selected or rRNA-depleted (‘total’) RNA. The use of total RNA has the advantage of capturing non-polyadenylated lncRNAs, but these are generally non-abundant transcripts that are difficult to reconstruct accurately and, in our experience, tools for expression level estimation work better when applied to poly(A)-selected RNA-seq data. The RNA-seq reads are then assembled into transcripts, either with or without the use of a reference genome, and the resulting transcript models are annotated based on known protein-coding genes and small RNAs and the sequences and transcriptomes of related species. We recently developed PLAR – a pipeline for lncRNA annotation from RNA-seq data – (Hezroni et al., 2015), and other tools are also available (Chen et al., 2016). The accuracy of the exon-intron structures reconstructed from short-read RNA-seq data is limited by several factors, including algorithmic challenges, extensive alternative splicing and incomplete genome assembly (in practically all animals, except human and mouse). The use of orthogonal data, such as chromatin marks (Guttman et al., 2009; Ulitsky et al., 2011) or RNA fragments enriched for 5’ or 3’ termini of the transcripts (Brown et al., 2014; Hezroni et al., 2015; Ulitsky et al., 2011), can improve transcriptome quality, but these data are frequently difficult to obtain. Therefore, it is important to treat each reconstructed transcript of interest with caution and validate its structure both by manual inspection of all the available data and experimentally. Particular caution is warranted when dealing with single-exon
transcripts, antisense transcripts and those overlapping pseudogenes, as these are especially prone to errors in read mapping or transcript assembly.

Once the transcriptome of a specific combination of tissues, cell types or developmental stages is available, the next challenge is to identify those transcripts that correspond to IncRNAs. Because IncRNAs are rather loosely defined, different researchers adopt different criteria for inclusion and exclusion of transcripts, in particular those that overlap loci of protein-coding genes on either the sense or antisense strand. A typical pipeline retains intergenic IncRNAs and those that overlap mRNAs on the other strand, and removes those transcripts that: (1) have an ORF that is either long or has a protein-like sequence composition; (2) have sequences similar to known proteins or domains; or (3) overlap regions where sequence evolution suggests a selective pressure to preserve a particular succession of amino acids [for a description of specific tools and algorithms, see Housman and Ulitsky (2015)]. Such a combination of filters is effective in removing candidates that encode large or conserved peptides; in our experience, most of these correspond to pseudogenes. Recent studies have identified a number of conserved peptides of 34-58 amino acids in length encoded by transcripts that were initially annotated by some pipelines as IncRNAs, and identified important physiological roles for those peptides (Anderson et al., 2015; Nelson et al., 2016; Pauli et al., 2014). Also, there are probably species-specific short proteins that remain to be discovered, but we estimate that these are relatively rare, in part because mRNAs and IncRNAs are distinguishable from each other in ribosome-footprinting data (Chew et al., 2013; Guttman et al., 2013). Perhaps counterintuitively, but consistent with the cytoplasmic presence of many IncRNAs, there is evidence for some degree of translation on most well expressed IncRNAs (Ingolia et al., 2014). This translation is similar to that occurring at 5′ UTRs, and the vast majority of peptides produced by such translation events in IncRNAs and 5′ UTRs are likely to be immediately degraded and nonfunctional [as discussed at length by Housman and Ulitsky (2015)]. The presence of some ribosome-protected fragments overlapping an IncRNA, or its localization to the polysome fraction (Carlevaro-Fita et al., 2016; van Heesch et al., 2014), do not, therefore, invalidate the noncoding nature of a transcript.

How many IncRNAs are evolutionarily conserved?

Conservation in other species is currently one of the key indicators for functionality of IncRNAs, and so the fraction of IncRNAs that are conserved is an indicator of the fraction that is functional. Based on the criteria listed above, the human genome encodes tens of thousands of IncRNAs, as do the genomes of non-human primates, and at least thousands of IncRNAs are found in other vertebrate species. Large numbers of IncRNAs have also been found in every multicellular organism subjected to in-depth RNA-seq analysis, including vertebrates, insects, nematodes, sponges and plants (Kapusta and Feschotte, 2014; Ulitsky, 2016). Although there are no clear orthologous IncRNAs between these groups of species, there are some prominent similarities between them. For example, IncRNAs are shorter than mRNAs, span only a few exons (typically 1-3), accumulate to levels at least an order of magnitude lower than those of mRNAs and in a more spatially and temporally specific manner, and evolve much faster than mRNA (Hezroni et al., 2015). This rapid evolution manifests itself on multiple levels. For example, human and mouse share the vast majority of protein-coding genes and highly expressed microRNA (miRNA) genes. By contrast, most human IncRNAs do not have any recognizable homologs in mice, and vice versa. In the homologous pairs that are present, many exons appear in one species but not in the other, and within those exons that do align the sequence similarity is typically much lower than among homologous coding or UTR exons. Overall, the sequences of IncRNA exons are only marginally better conserved than those in the introns of protein-coding genes or in random intergenic regions, and they are only marginally depleted of sequences derived from transposable elements (Kapusta et al., 2013; Kelley and Rinn, 2012). There is some weak correlation between conservation levels and expression (Necsulea et al., 2014), and between conservation and expression breadth, and so IncRNAs discovered in larger and deeper sequencing efforts appear to be less conserved than those discovered by earlier efforts, which used relatively shallow sequencing. Overall, taking into account both genomic features and the conservation of IncRNAs in human and in other species, it appears that many human IncRNAs arose recently through neutral sequence evolution and retrotransposition, and will likely be lost with time through similar mechanisms. In this view, the IncRNAs that we witness in the human genome today are likely to constitute only a small subset of those that appeared in our ancestors. They include a subset that is functional and has been selectively retained, but this subset is likely to be relatively small. Unfortunately, even in relatively well-conserved IncRNAs, the constrained regions are rather short and degenerate, so identifying conserved IncRNAs and the conserved regions within them remains a challenge in need of better computational tools, such as those described recently (Quinn et al., 2016).

Which IncRNAs are more likely to be functionally important?

Researchers interested in a particular developmental system typically need to employ various criteria for selecting IncRNAs, often from hundreds of dynamically expressed candidates, for experimental follow-up. Possible parameters for selection include levels and specificity of expression, the genomic context of the IncRNA, and the extent of its sequence conservation. Because our experience and that of others is still limited, it is difficult to specify the right criteria for ‘successful’ selection, but several factors are important to keep in mind. First, tissue- or stage-specific expression does not necessarily imply function; weak promoters randomly placed in the mouse genome have been shown to have very specific activity patterns, depending on the context into which they are introduced (Ruf et al., 2011) and the presence of active enhancers in the vicinity of the insertion site. The expression levels of an IncRNA are, however, an important consideration and can limit the range of possibilities for its mode of action; for example, an RNA expressed at 1-2 copies per cell is unlikely to directly regulate many targets in trans. Because many IncRNAs are likely to regulate gene expression in cis, genomic proximity to a gene with a known function in the process of interest can also indicate potential functionality and provide a readily testable hypothesis for the mode of action. Close proximity or overlap with promoters of known genes, or with regions bearing chromatin marks characteristic of enhancers, can however complicate the mechanistic dissection of the mode of action of an IncRNA, as genome editing or chromatin modulation can have secondary effects on promoter or enhancer activity. For these reasons, conservation of an IncRNA in other species, which at least in vertebrates is now relatively easy to test with the available data (Chen et al., 2016; Hezroni et al., 2015; Necsulea et al., 2014; Washietl et al., 2014), is arguably the strongest current indicator for functional relevance. The combination of some of these factors typically allows the candidate list to be narrowed down from
hundreds to a number that is feasible for some degree of experimental characterization.

The identification of lncRNA homologs
As reviewed in detail elsewhere (Kapusta and Feschotte, 2014; Ulitsky, 2016), and as touched upon above, studies suggest that lncRNAs evolve rapidly, with gains and losses occurring much more frequently than in other gene classes. This complicates the identification of homologs for study in model organisms. The resources available for addressing this problem include whole-genome alignments (such as those available in the UCSC genome browser or in Ensembl), databases of lncRNAs annotated in various species (Bu et al., 2015; Chen et al., 2016; Hezroni et al., 2015; Necsulea et al., 2014), as well as groups of orthologous protein-coding genes (such as Ensembl Compara) for identifying positionally conserved (syntenic) lncRNAs. Several levels of lncRNA conservation are possible (Ulitsky, 2016). On the highest level, multiple regions within the lncRNA sequence are conserved together with splice sites and the overall exon-intron architecture; the Mid lncRNA is an example that exhibits this level of conservation (Ulitsky, 2016).

However, such cases are scarce between human and mouse, and are very rare when considering more distantly related species (Chen et al., 2016; Hezroni et al., 2015). More common are cases in which just one or two short sequence stretches are conserved, embedded in a rapidly evolving locus that underwent extensive rewiring of exon-intron architecture, typically with substantial contribution from transposable elements. When considering distantly related species (e.g. mammals and birds, or mammals and fish), the number of pairs of lncRNAs that exhibit ‘positional conservation’ – two lncRNAs with the same relative orientation to flanking conserved protein-coding genes – exceeds the number expected by chance and the number of lncRNAs with conserved sequences (Amaral et al., 2016 preprint; Hezroni et al., 2015; Ulitsky et al., 2011).

The Pou3f3 adjacent noncoding transcript 1 (Pantr1; LINC01158 in human) lncRNA, which is transcribed divergently from the Pou3f3 (Bmrl) gene (Fig. 1) and plays a role in regulating gene expression in the developing mouse brain (Goff et al., 2015), provides an illustrative example of the complexity of lncRNA evolution. Several parts of Pantr1, including its two main promoters, are well conserved among mammals. Similar lncRNAs divergent with Pou3f3, which are expressed specifically in neuronal tissues and in the kidney, are also found throughout vertebrates, but their locus architectures, lengths and sequences are different. Furthermore, the human, mouse and opossum lncRNAs each contain exons derived from independent transposable elements (Fig. 1), and the 3’ end of Pantr1 in mouse is derived from a transposon. The sequence homology between the mammalian lncRNAs and the homologs in chicken and lizard is restricted to the first exon. In more distant species, such as teleost fish and the spotted gar, syntenic, kidney-expressed lncRNAs without detectable sequence homology to amniote Pantr1 lncRNAs are found in the same orientation near Pou3f3 orthologs (Fig. 1). This rapid turnover of sequence and expression status implies that lncRNAs rarely depend on specific long sequences in their loci for function. It is possible that, in many of them, just the act of transcription across a certain region is important, or that functionality is sequence dependent but relies on short sequences that are conserved but difficult to align across large evolutionary distances. Another implication of rapid lncRNA evolution is that some human lncRNAs can only be studied in human cells, and that for those with homologs in model organisms the low sequence and gene structure homology may cast doubts on the conservation of functionality. Few studies have looked into this question [reviewed in detail elsewhere (Ulitsky, 2016)] and, although it is still too early to draw definitive conclusions, it does appear that limited sequence conservation can be sufficient to maintain conserved functionality across large evolutionary distances.

lncRNA mechanisms of action
The potential modes by which lncRNAs function, and the relative prevalence of each ‘mechanistic group’, remain unclear and are still actively debated. However, lncRNAs can be broadly divided into three groups based on their mechanisms of action (Fig. 2): (1) those for which only the act of transcription is important, and the RNA that is produced carries no function (‘transcription only’); (2) those for which the RNA is important, but the activity is linked to the site of transcription (‘cis-acting’); and (3) those for which the RNA acts independently of the site of its transcription (‘trans-acting’). Below, we discuss each of these lncRNA groups, detailing their modes of action and highlighting how they can be distinguished from one another.

Transcription-only lncRNAs
The act of transcription itself can affect regulatory elements overlapping the lncRNA locus. Indeed, one potential functional outcome of transcription is changes in chromatin modifications in the locus. For example, parts of transcribed regions (mostly from the second exon onwards) are typically demarcated with H3K36 trimethylation (Wagner and Carpenter, 2012), which is associated with repression of transcription initiation (Carrozza et al., 2005; Fang et al., 2010). Therefore, if the lncRNA locus overlaps with the promoter of another gene (such as in the case of Airn and Igf2r), transcription of the lncRNA can potentially lead to repression of the promoter. Importantly, in these cases, the RNA product of transcription, and hence its sequence or structure, do not have a functional role.

cis-acting RNAs
The main mechanisms ascribed to the cis-acting group of lncRNAs involve the recruitment of factors, either to the site of lncRNA transcription or to adjacent loci. The recruited factors can be repressive, as in the cases of ANRIL [CDKN2B-AS1; which is reported to recruit CBX7 from the PRC1 complex (Yap et al., 2010) and SUZ12 from the PRC2 complex (Katolke et al., 2011)], Airn [which is reported to recruit G9a (Ehmt2) (Nagano et al., 2008) and H19 [which is reported to recruit MBD1 (Monnier et al., 2013)]. Alternatively, such factors can be activating, as in the case of Hotip, which is reported to recruit WDR5 (Wang et al., 2011; Yang et al., 2014). There can be various consequences to the recruitment of such factors. For example, the lncRNA can increase the local concentration of the factor in the locus, as has been suggested for YY1-binding lncRNAs (Sigova et al., 2015). By contrast, it can act as a local decoy, diverting the factor from binding to DNA elsewhere across the locus (Krawczyk and Emerson, 2014).

For many years, the prototypical example of an lncRNA acting through factor recruitment has been Xist, which has been proposed to recruit the PRC2 complex to the mammalian X chromosome during X-inactivation (Zhao et al., 2008). However, recent studies have challenged this model and shown that the picture is much more complicated, as PRC2 appears to bind RNA non-specifically (Brockdorff, 2013; Davidovich et al., 2013), Xist specifically binds to other factors (Roth and Diederichs, 2015), and the temporal dynamics of silencing are inconsistent with the simple model of PRC2 recruitment leading to gene silencing (Gendrel and Heard, 2014).
trans-acting RNAs

These lncRNAs act independently of their sites of transcription, either by regulating expression from other loci in the nucleus or having transcription-unrelated functions anywhere in the cell. Many well-studied lncRNAs, such as Xist, Neat1 and Malat1, are strongly retained in the nucleus, but most annotated lncRNAs are either both nuclear and cytoplasmic or predominantly cytoplasmic (Cabili et al., 2015; Derrien et al., 2012; Ulitsky and Bartel, 2013). Studies have shown that trans-acting lncRNAs can play important roles in establishing nuclear architecture (Batista and Chang, 2013) by...
dictating the proximity of different loci, as in the case of Firre (Hacisuleyman et al., 2014), or by nucleating subnuclear bodies, as in the case of Neat1 (Clemson et al., 2009).

Another trans-acting function that has been proposed to be widespread is the recruitment of chromatin-altering complexes to specific loci in trans (Koziol and Rinn, 2010). However, the mechanisms by which lncRNAs can specifically recognize elements in trans remain unknown, and it is unclear how many target loci can be efficiently reached by lncRNAs that accumulate at only a few copies per cell. The same stoichiometric concerns exist for the cytoplasmic functions of lncRNAs and, in general, the roles of cytoplasmic lncRNAs have been less explored. Some cytoplasmic lncRNAs, including the circular lncRNA CDR1as (Memczak et al., 2013) and NORAD (Lee et al., 2016; Tichon et al., 2016), are abundant enough to bind to and affect the function of RNA-binding proteins (Argonaute and Pumilio, respectively) and thereby modulate their ability to regulate their other targets. Other cytoplasmic lncRNAs have been shown to bind and modulate the stability and translation of other RNAs by base-pairing with them (Carriero et al., 2012; Gong and Maquat, 2011).

Distinguishing between IncRNA modes of action

In many cases, it is relatively straightforward to experimentally distinguish between the cis-acting and the trans-acting mechanisms. For example, if the phenotype caused by loss of a particular IncRNA can be rescued by its expression from an exogenous locus [as in the case of roX IncRNAs (Park et al., 2008; Quinn et al., 2014; Quinn et al., 2016)], the IncRNA is most likely trans-acting. By contrast, it is difficult to distinguish between transcription-only and cis-acting lncRNAs, as this requires eliminating the possibility that the RNA product is important. For example, in the case of the IncRNA Airn, it was shown that piecewise replacement of all parts of the Airn transcript does not compromise its allele-specific repression of Igf2r, strongly implying that only the act of transcription and not the product of Airn transcription is important for Airn function. However, in most cases, such studies require extensive genome engineering and precise phenotypic recoding, which often remain very difficult despite the growing toolbox for lncRNA manipulation (Box 1). It should also be noted that the same locus can have multiple independent activities. For example, the Airn IncRNA regulates the expression of imprinted genes in the vicinity of its locus via at least two mechanisms: the Airn RNA product acts on the Slc22a3 promoter by recruiting the G9a histone methyltransferase (Nagano et al., 2008) in a mechanism shared by other lncRNAs implicated in imprinting (Pandey et al., 2008), but only transcription through the Airn locus is required for silencing Igf2r (Latos et al., 2012). Furthermore, even the transcription of mRNAs (which all ‘act in trans’ to produce proteins) can have a cis-regulatory function and affect the transcription of nearby genes (Ebisuya et al., 2008).

Roles for IncRNAs during mammalian development

In recent years, loss-of-function (LOF) and gain-of-function (GOF) studies have revealed that many IncRNAs are involved in a wide variety of biological processes during development (Ponting et al., 2009). The different techniques used in these studies are presented in Box 1, although it should be noted that each of these has pros and cons [see Box 2 and the recent review by Bassett et al. (2014)]. In this section, we focus on the functional roles of IncRNAs in mammalian embryonic development and differentiation, as revealed mostly by in vivo LOF studies in mice.

A general survey of mammalian IncRNA functions in vivo

The largest systematic surveys of IncRNA knockout (KO) mouse models have revealed important roles for IncRNAs in regulating organism viability as well as many specific developmental processes. Three complementary publications characterized 20 IncRNA KO mouse strains using expression pattern characterization and LOF experiments (Goff et al., 2015; Lai et al., 2015; Sauvageau et al., 2013). The 20 candidates were selected based on a combination of stringent filters (e.g. the absence of protein-coding capacity, a lack of overlap with protein-coding genes), expression levels and conservation; all 20 IncRNAs overlap sequences conserved in the human genome, and at least 14 are homologous to annotated human IncRNAs (Sauvageau et al., 2013). Notably, these candidate IncRNAs are not necessarily a representative sample of mammalian IncRNAs. KO mouse strains in which a lacZ reporter cassette replaced each IncRNA locus were generated using VelociGene technology (Valenzuela et al., 2003). These lacZ reporters revealed a wide spectrum of spatiotemporal and tissue-specific IncRNA transcription patterns, both in mouse embryos and in adult mice (Goff et al., 2015; Lai et al., 2015; Sauvageau et al., 2013). Importantly, since most of the locus is excised in this approach, the phenotypes might result from the removal of DNA elements in the locus rather than from the loss of RNA expression (Bassett et al., 2014).

The initial characterization of 18 of these KO strains revealed the functional relevance of specific IncRNAs in mouse embryonic development, viability and growth. Indeed, three out of the 18 IncRNA KO strains – Peril (Perl), Mdgt (Haglr) and Fendr...
Box 1. Tools for studying lncRNA function

Transcript characterization. Initial reconstruction of IncRNA transcripts is typically performed using RNA-seq data. RT-PCR, 3' and 5' RACE and northern blots can be used to validate IncRNA transcript structure. Single-molecule RNA fluorescence in situ hybridization (smFISH) can then be used to examine expression patterns, detect even low-abundance IncRNAs, and enable absolute quantification of the number and location of IncRNA molecules within cells (Dunagin et al., 2015).

Binding partners. The dissection of IncRNA mechanisms typically requires the identification of protein, DNA and RNA binding partners. Biotinylated RNA transcribed in vitro from a cDNA template can be incubated with cell lysates to identify protein binding partners (Hämmerle et al., 2013). Alternatively, endogenous RNAs can be enriched using biotinylated antisense oligos and the resulting material can be subjected to mass spectrometry, DNA or RNA sequencing (Simon, 2016). RNA immunoprecipitation can also be used to identify IncRNA species that bind to a protein of interest (Cozzorto et al., 2015).

Transient LOF. Methods to transiently perturb IncRNA expression without changing the underlying DNA sequence include RNA interference (RNAi), antisense oligonucleotides (ASOs; which includes morpholinos) and CRISPR interference (CRISPRi). RNAi has been used to successfully knock down IncRNAs in many studies (Guttman et al., 2011). Recently, it was found that nuclear IncRNAs are more effectively suppressed using ASOs, cytoplasmic IncRNAs are more effectively suppressed using RNAi, and dual-localized IncRNAs are suppressed using either method (Lennox and Behlke, 2016). CRISPRi is a much newer technology that has already been used successfully for targeting IncRNAs (Gosh et al., 2016).

Constitutive LOF via genome engineering. Genome engineering can be performed using both traditional recombination-based techniques and TALEN/CRISPR technologies to target IncRNA genes. Small insertions or deletions caused by Cas9-mediated double-strand breaks and inducing frameshifts in ORFs are the current method of choice for inactivating protein-coding genes. These are not well suited to knock out IncRNAs, and several other targeting strategies are used: deletion of the full-length IncRNA locus (or its replacement by reporter genes or selection cassettes (Lai et al., 2015; Sauvageau et al., 2013); deletion of the promoter sequence; mutation of putative functional domains; engineered inversions (Li and Chang, 2014); or insertion of transcriptional terminator sequences (i.e. STOP signals) (Bond et al., 2009; Grote et al., 2013). In all cases, it is important to minimize the removal or reorganization of regulator factor binding sites or other regulatory elements within the DNA locus, and to control for the addition of novel DNA regulatory elements (Bassett et al., 2014). To prove that an IncRNA molecule has a direct functional role, rescue or GOF assays involving transgene expression can be used (Grote et al., 2013).

GOF assays. Exogenous overexpression of the IncRNA from a plasmid or a viral vector is the most common and simplest GOF approach, but it has limited efficacy if the IncRNA acts in cis. Instead, CRISPR-on, which is a variant of CRISPRi that combines a catalytically inactive Cas9 (dCas9) with transcriptional activators such as the VP64 activator domain, can be used to increase IncRNA production from endogenous loci (Gilbert et al., 2014; Luo et al., 2016). Another possibility is to localize the exogenously expressed IncRNA to a specific genomic locus by fusing it with the CRISPR gRNA targeted to the specific locus and to co-introduce this chimeric RNA into cells together with dCas9 (Luo et al., 2016; Shechner et al., 2015). Lastly, genome editing can be used to knock-in a strong promotor upstream of the IncRNA to increase gene expression (Luo et al., 2016) or to knock-in the cDNA of an IncRNA for rescue experiments (Yin et al., 2015).

Box 2. Potential caveats to IncRNA perturbation approaches

When considering LOF approaches, the main distinction is between methods that target the transcript and leave the DNA intact (e.g. RNAi, antisense oligonucleotides) and those that alter the DNA in the locus using genome engineering (e.g. TALEN, CRISPR). Both can be efficient at dramatically reducing transcript levels, but the former methods suffer from off-targeting while the latter can interfere with other activities encoded in the same locus, such as those exerted by enhancer elements and/or elements regulating chromatin architecture. Methods based on a catalytically inactive CRISPR/Cas9 system (dCas9) have the potential of harnessing the targeting specificity of CRISPR without altering the genome (Dominguez et al., 2016), but they still may have IncRNA-unrelated effects on chromatin, in particular when the dCas9 is coupled with a chromatin-modifying domain [as in dCas9-KRAB fusions (Gilbert et al., 2014)]. In the case of rescue experiments following GOF or LOF approaches, the available techniques differ in their relevance for IncRNAs acting in cis, which requires the expression of the IncRNA in the proximity of its endogenous site of transcription. Here too, the use of dCas9 can assist in recruiting the IncRNA to the target site of interest, which can be sufficient for eliciting regulatory effects (Luo et al., 2016; Shechner et al., 2015).

LncRNAs implicated in dosage compensation: Xist and Tsix

One of the first-discovered and best-characterized lncRNAs shown to have a specific developmental role and a robust LOF phenotype in vivo is X-inactive specific transcript (Xist). Xist is directly involved in the process of X chromosome inactivation, which is initiated by the induction of Xist expression. Xist is absolutely required for X-inactivation to occur in cis, and an extended Xist locus is sufficient for silencing when placed on an autosome (Lee and Bartolomei, 2013). Accordingly, the deletion of Xist in mice causes a loss of X-inactivation and female-specific lethality (Marahrens et al., 1997). The expression of Xist itself is controlled by other IncRNAs. For example, Tsix lncRNA, which is the antisense partner of Xist RNA, represses Xist expression (Lee and Bartolomei, 2013). Tsix LOF in vivo results in ectopic Xist expression, aberrant X-inactivation and early embryonic lethality (Sado et al., 2001). Xist also plays a role beyond early embryonic development. Because Xist activity is followed by epigenetic changes, transient repression of Xist does not result in immediate X reactivation (Wutz and Jaenisch, 2000). However, in Xist-deficient mouse hematopoietic stem cells, which undergo a large number of cell divisions, Xist loss does cause X reactivation and subsequent genome-wide changes that lead to cancer, thus potentially linking the X chromosome to cancer in mice (Yildirim et al., 2013).

Imprinting-associated lncRNAs: Kcnq1, Airn and H19

A number of lncRNAs are associated with the process of genomic imprinting, which is crucial for normal development. During this event, genes are epigenetically silenced on the basis of their parental origin, resulting in monoallelic expression. Many imprinted clusters contain protein-coding genes and lncRNAs that are expressed from reciprocal alleles. The best-characterized examples of lncRNAs that regulate imprinting are Kcnq1ot1 and Airn; both are paternally expressed and repress flanking protein-coding genes in cis. Thus, although the loss of these lncRNAs in the embryo is not lethal,
paternal inheritance of a LOF allele causes a loss of imprinting and hence gives rise to growth defects, whereas maternal inheritance of this allele has no effects on imprinting or growth (Fitzpatrick et al., 2002; Sleutels et al., 2002).

The lncRNA H19 is encoded by a conserved imprinted gene that is expressed exclusively from the allele of maternal origin. H19 is strongly expressed in both mesoderm- and endoderm-derived tissues during embryogenesis in mice, then becomes fully repressed after birth except in skeletal muscle and heart (Poirier et al., 1991). This expression pattern is similar to that of the major fetal growth factor gene Igf2, which is paternally expressed (DeChiara et al., 1991). Two KO models have been established to examine H19 function – one in which only the 3 kb transcription unit is deleted (H19−/− mice) and another in which 10 kb upstream of that region is also deleted (H19+/− mice). In both cases, maternal heterozygotes are viable and fertile but exhibit an overgrowth phenotype. In H19+/− mice, the maternal Igf2 allele is totally reactivated in all expressing tissues (Leighton et al., 1995). In H19+/− mice, the maternal Igf2 allele is also reactivated, although its expression is only 25% of that of the paternal allele in wild-type mice and is only observed in mesoderm-derived tissues (skeletal muscle, tongue, diaphragm and heart) (Ripoche et al., 1997). The overgrowth phenotype can be rescued in H19+/− mice expressing an H19 transgene, with expression of Igf2 and other imprinted genes returning to wild-type levels. This activity was recently associated with the ability of H19 to recruit MBD1 (Monnier et al., 2013), suggesting that H19 itself can act in trans to control imprinting (Gabory et al., 2009). Together, these findings highlight that lncRNAs in imprinted gene clusters are not just imprinted in their expression patterns, but can also regulate the imprinting process.

Hox cluster IncRNAs: Hotair, Hottip and Mdgdt

Hox genes encode TFs that orchestrate the embryo body plan and contribute to several adult cell fate specification processes (Barber and Rastegar, 2010). In addition to containing protein-coding and miRNA genes, these clusters produce numerous lncRNAs that exhibit spatiotemporal expression patterns resembling those of their neighboring protein-coding genes. For example, Hotair is a 2.2 kb lncRNA expressed from the HoxC cluster that can repress the HoxD locus in trans in mammalian cells (Rinn et al., 2007). In line with the key function of HoxD genes, homeotic transformation of the fourth caudal vertebra is observed in Hotair−/− mice (Lai et al., 2015). The Hotair homeotic phenotype is also observed in mice carrying an alternative Hotair KO allele (Li et al., 2013). In this model, disruption of Hotair also leads to the derepression of hundreds of genes, including those within the HoxD cluster. By contrast, no major skeletal transformations are observed in a mouse strain in which the entire HoxC gene cluster, including the Hotair gene, is deleted (Schorderet and Duboule, 2011). However, it should be noted that the HoxC cluster produces a multitude of noncoding RNAs, and removal of the entire gene cluster may remove protein-coding genes and lncRNAs that could oppose Hotair activity. Thus, although the phenotypes of Hotair mutants are not severe, they provide compelling evidence that lncRNAs in Hox clusters can regulate the expression patterns of Hox genes, and other genes, during development. Indeed, another lncRNA – Hottip – has been shown to regulate the expression of posterior HoxA genes. It does so by interacting with the activating histone-modifying MLL1 complex and via the formation of chromatin loops that connect distally expressed Hottip transcripts with posterior HoxA gene promoters (Wang et al., 2011). Hottip−/− mice exhibit hind limb abnormalities, including muscle weakness and skeletal malformations (Lai et al., 2015). Furthermore, the transfection of short hairpin RNAs (shRNAs) targeting a region with sequence similarity to Hottip in chick embryos alters limb morphology (Wang et al., 2011). The lncRNA Mdgdt [also called Haglr (Lai et al., 2015; Yarmishyn et al., 2014)], which is transcribed from a bi-directional promoter that is shared with Hoxd1, has also been implicated in the control of Hox gene expression. Homozygous Mdgdt mutants die within 2 weeks after birth, with Mdgdt−/− pups displaying a severe growth retardation phenotype that may contribute to their lethality (Sauvageau et al., 2013).

LncRNAs required for neuronal development

The detailed analysis of lncRNA KO mouse models has revealed key roles for three lncRNAs – Peril, Evf2 (Dlx6os1) and linc-Brn1b – during neural development. The Peril transcript is derived from an 18.2 kb genomic locus that is located 110 kb downstream of Sox2, which encodes a key pluripotency factor. Peril is highly enriched in mouse embryonic stem cells (ESCs) but is also expressed at lower levels in the mouse adult brain and testes (Sauvageau et al., 2013). The deletion of Peril in mice leads to reduced viability; 50% of Peril−/− pups die within 2-20 days of birth (Sauvageau et al., 2013). The expression of Sox2 and its overlapping lncRNA Sox2ot is not significantly affected in Peril KO brains, suggesting that the KO phenotype is not due to a defect in Sox2 function (Sauvageau et al., 2013). In a follow-up study of Peril KO mice (Goff et al., 2015), using RNA-seq analyses and β-gal staining, it was shown that Peril is expressed in neural stem cells and may affect their biology. Furthermore, the neural stem cell-specific expression of Peril is maintained in Peril−/− adult mice, with β-gal staining observed in the ependymal lining of the ventricles and in the dentate gyrus of the hippocampus, both of which are regions associated with adult neurogenesis. Consistently, RNA-seq analyses of Peril−/− embryonic brains revealed a misregulation of cell cycle genes that are known to be important for the correct maintenance and differentiation of neural progenitors. These results clearly demonstrate biological activity of the Peril transcript, although additional work needs to be done to understand the molecular mechanism by which Peril functions and the extent to which it operates independently of Sox2.

The lncRNA Evf2 also appears to play a role in neural development. Evf2 is transcribed antisense to Dlx6 and is located immediately downstream of the Dlx5 genomic locus. Dlx5 and Dlx6, which are related to the Drosophila melanogaster Distal-less (Dll) gene, encode TFs that are expressed in the developing ventral forebrain and have been implicated in both forebrain and craniofacial development. Evf2 KO mice have been generated by inserting a transcriptional terminator consisting of three polyadenylation signals into the first exon. In these mice, the numbers of GABAergic interneurons in the early postnatal hippocampus are reduced (Bond et al., 2009). Although GABAergic interneuron numbers and levels of Gad1 mRNA (which encodes an enzyme involved in GABA synthesis) return to normal in the adult hippocampus of Evf2 mutants, defects in synaptic inhibition are observed, indicating a crucial role for Evf2 in neuronal activity in vivo (Bond et al., 2009). To determine whether Evf2 controls Dlx5/6 CpG methylation and hence expression through trans or cis mechanisms, an Evf2 rescue transgenic model was developed. Using this model, it was shown that transcription through the Evf2 locus controls the levels of Dlx6 in cis; after disengaging the polymerase, Evf2 then acts in trans to modulate methylation of the Dlx5/6 enhancer and transcription of Dlx5. Therefore, by regulating the cellular levels of the Dlx5 and Dlx6 TFs, Evf2 controls GABAergic interneuron activity (Berghoff et al., 2013; Bond et al., 2009).
Finally, recent studies suggest a role for the lncRNA *linc-Brn1b* (also known as *Pantr2*) in neural development. *linc-Brn1b* is located near the *Pou3f3* (*Brn1*) gene, which is also adjacent to the lncRNA *Pantr1* (described above) (Sauvageau et al., 2013). *linc-Brn1b*−/− mice exhibit distinct growth defects as well as defects in the cerebral cortex, especially in the development of upper layer II/III-IV neurons (Sauvageau et al., 2013).

### LncRNAs required for the development of other organs

Several lncRNAs have been shown to play roles in the development of specific organs and tissues during embryogenesis. *Fendrr* is a 2397 nt transcript consisting of seven exons transcribed divergently from the TF gene *Foxf1*. Two approaches have been used to perturb *Fendrr* in mice: KO mice (Lai et al., 2015; Sauvageau et al., 2013) have been obtained by genomic deletion (starting from the second exon to the last annotated exon), and knock-in (KI) mice have been generated by replacing the first exon of *Fendrr* with a polyadenylation cassette without a reporter gene (Grote et al., 2013). KO pups exhibit multiple defects in the lung, heart and gastrointestinal tract (Sauvageau et al., 2013). Further investigations highlighted that, at E13.5, the developing lungs of the KO are small with globular and disorganized lobes. Accordingly, these KO mice survive to birth but die shortly thereafter due to breathing problems (Lai et al., 2015). By contrast, the KI mice display lethality at E13.75 due to heart and body wall (omphalocele) defects. Notably, resorbed embryos or omphalocele were not observed when analyzing E14.5 KO embryos. The differences between the studies extended to the expression domain of *Fendrr*. Using whole-mount *in situ* hybridization and qPCR analysis, Grote et al. (2013) observed that endogenous *Fendrr* expression is restricted to nascent lateral plate mesoderm and cannot be detected in any other tissues or organs. However, expression profiling of the knocked-in *lacZ* reporter in E14.5 and E18.5 embryos showed expression of *Fendrr* in lung, colon, liver, spleen and brain, as well as in the trachea and the gastrointestinal tract (Lai et al., 2015). Given that both studies used similar genetic background strains, the different targeting strategies used to remove the *Fendrr* gene provide the most probable explanation for the phenotypic discrepancies. Discrepancies in the reported expression patterns might be a result of the different methods used to study the endogenous *Fendrr* expression pattern. Regardless, both studies confirm that *Fendrr* LOF is lethal in mice. Mechanistically, *Fendrr* was shown to act by binding to the PRC2 and TrxG/MLL complexes, in turn modifying the chromatin signatures of genes involved in lateral mesoderm lineage formation and differentiation (Grote et al., 2013).

The lncRNA *Neat1* has been implicated in mammary gland development and pregnancy in mice. *Neat1* is an essential constituent of paraspeckles—nuclear substructures that are found in all primary cells and cell lines examined to date, with the exception of ESCs (Clemson et al., 2009). It has been proposed that paraspeckles control key cellular processes, including differentiation, via their ability to integrate transcriptional and post-transcriptional events (Hirose and Nakagawa, 2012). *Neat1* KO mice, which were generated by insertion of *lacZ* and polyadenylation signals immediately downstream of the transcription start site, develop normally and are indistinguishable from their wild-type littermates with respect to growth, viability and apparent behavior (Nakagawa et al., 2011). However, two follow-up studies on these mice revealed important functions for *Neat1* and paraspeckles in *vivo*. First, it was demonstrated that, during mammary gland development in mice, paraspeckles containing *Neat1* are present in mammary gland luminal cells and that *Neat1* is required for branching morphogenesis, lobular–alveolar development and lactation (Standaert et al., 2014). Second, it was shown that, despite exhibiting normal ovulation, *Neat1* KO mice stochastically fail to become pregnant, potentially owing to corpus luteum dysfunction and concomitant low progesterone levels (Nakagawa et al., 2014).

Finally, studies suggest that the lncRNA *Pint* (*Lncpint*) might also play a role in organ growth and development. *Pint*, which is ubiquitously expressed in mice, is a direct transcriptional target of p53 (Tip53) (Marín-Béjar et al., 2013) and positively regulates cell proliferation and survival by affecting the expression of hundreds of genes. Accordingly, *Pint* KO mice are smaller and exhibit lower body weights than their wild-type littermates (Sauvageau et al., 2013).

### Roles for lncRNAs during zebrafish embryogenesis

While KO studies in mice have provided key insights into the function of lncRNAs, it should be noted that zebrafish studies have also been fruitful with regards to understanding the biology of lncRNAs in both developing embryos and adult tissues (Pauli et al., 2011). Indeed, several groups have shown that lncRNAs are expressed dynamically in a spatiotemporal manner in zebrafish embryos and adult fish (Kaushik et al., 2013; Pauli et al., 2012; Uilitsky et al., 2011). Hundreds of zebrafish lncRNAs are found in syntenic positions to human lncRNAs, and a few dozen of them display short sequences of high homology across vertebrate evolution (Ulitsky et al., 2011). Morpholino (MO)-mediated knockdown of two lncRNAs that display such short stretches of sequence conservation across vertebrates — *cyrua* (*oip5-as1*) and *megamind* (*birch-6-as2*) — results in defects in the central nervous system (Ulitsky et al., 2011). Importantly, the zebrafish, human and mouse homologs of these lncRNAs are able to rescue these phenotypes, suggesting that the functionality of the lncRNA sequence has been conserved despite drastic changes in most of the sequence. However, a recent study reported that genetic deletion of *megamind* does not give rise to the same phenotype, although the mutant is still susceptible to *megamind* MO (Kok et al., 2015). Although this suggests that MO off-target effects might be involved, this appears to be unlikely given that the same phenotype is observed with different MOs. One possible explanation is that the KO of *megamind* is compensated by the presence of two homologs with related sequences in the zebrafish genome (Ulitsky et al., 2011).

In a different study, three highly conserved lncRNAs – *TERMINATOR, ALIEN* [also known as *lnc-FOXA2*, LL35 (Herriges et al., 2014) and *DEANRI* (or *LINC00261*) (Jiang et al., 2015)] and *PUNISHER* – were identified and functionally characterized in zebrafish embryos as well as in mouse ESCs (Kurian et al., 2015). These lncRNAs are specifically expressed in pluripotent stem cells, cardiovascular progenitors and differentiated endothelial cells. LOF analyses, using shRNAs in mouse ESCs and MOs against zebrafish orthologs in *vivo*, demonstrate that all three lncRNAs are involved in cardiovascular development (Kurian et al., 2015).

### LncRNAs with cellular functions but no apparent in vivo phenotype

The lncRNA KO models discussed above indicate that some have essential roles in development in *vivo*. However, this is not always the case. Some lncRNAs, despite exhibiting reasonably high expression levels and conservation, fail to show a clear phenotype when deleted in mice (Oliver et al., 2015; Sauvageau et al., 2013). Interestingly, several lncRNA KO animals show subtle or no
phenotypes during development, whereas in vitro studies of the same IncRNA suggest an essential function in the cell. An example of such an IncRNA is *Malat1*, which is among the most abundant, widely expressed and conserved IncRNAs in vertebrate cells. *Malat1* localizes to nuclear speckles and has been shown to regulate synapse formation by modulating a subset of genes that have roles in nuclear and synapse function (Bernard et al., 2010), to regulate splicing in a human cell line (Tripathi et al., 2010), and to have several important roles in regulating metastasis potential in cancer cells (Gutschner et al., 2013). To address the physiological function of *Malat1* in a living organism, three KO mouse models have been generated independently (Eissmann et al., 2012; Nakagawa et al., 2012; Zhang et al., 2012). Unexpectedly, these mice are viable and generated independently (Eissmann et al., 2012; Nakagawa et al., 2012), although it is likely that these changes are a result of removing the strong *Malat1* promoter rather than the *Malat1* RNA itself. Therefore, despite its high expression and conservation between human and fish (Ulitsky et al., 2011), which, based on the criteria described above, makes it an excellent candidate for having an important function, *Malat1* does not appear to be required for the proper development of whole-organism physiology. However, it is possible that *Malat1* is important under suboptimal conditions or in specific cell types, and there is strong evidence for its roles in cancer and metastasis (Schmitt and Chang, 2016).

**Roles for IncRNAs during stem cell differentiation**

A large number of studies carried out in the past five years have implicated IncRNAs in regulating stem cell maintenance and differentiation. For example, specific IncRNAs have now been implicated in neuronal differentiation [e.g. *RMDST* (Ng et al., 2013), *Mat* (Aprea et al., 2013), *Tunarr* (also known as meganoud or TUNA) (Lin et al., 2014)], epidermal differentiation [e.g. *DANCER* (Kretz et al., 2012)], cardiac differentiation [e.g. *Braveheart* (Klltenhoff et al., 2013)], endoderm differentiation [e.g. *ALIEN* (Jiang et al., 2015; Kurian et al., 2015)], endothelial differentiation [e.g. *SENCR* (Boulierdaa et al., 2016)], adipocyte differentiation [e.g. *Inc-RAP1-10* (Sun et al., 2013)] and hematopoietic differentiation [e.g. *hoxBlnc* (Deng et al., 2016)]. These studies have shown that IncRNAs play important roles in establishing and maintaining cell identity throughout the mammalian body. In these various studies, transcriptional profiling is typically first used to find the IncRNAs that are differentially expressed during the differentiation process. Perturbation approaches, most commonly using RNai, are then typically used to probe IncRNA functions during differentiation. More recent studies have employed a combination of post-transcriptional repression using RNAi and antisense oligonucleotides with CRISPR/Cas9 genome editing. For instance, this combination has recently been used to study the roles of *Haunt* (*Haatr1* or *linc-HoxA1*) in mouse ESCs and during differentiation into embryoid bodies (Yin et al., 2015). Using an extensive allelic series of edited mouse ESC lines, it was shown that *Haunt* IncRNA negatively modulates the activity of enhancers found in the same locus, thus acting as a repressor during the early activation of proximal genes in the HoxA cluster. This use of a mixture of techniques and a combination of strategies in the context of CRISPR-based editing is now expected to become standard in the field.

Large-scale functional screens have also revealed roles for IncRNAs in the process of stem cell differentiation. For example, shRNA libraries have been used to target hundreds of shRNAs in mouse ESCs (Guttman et al., 2011; Lin et al., 2014), and have implicated numerous IncRNAs in the maintenance of pluripotency and differentiation towards various lineages. In the future, CRISPR-based screens are expected to be of use in identifying functionally important IncRNAs in various differentiation systems, as well as to delineate regions within IncRNAs that are functionally important.

**Characteristic roles of IncRNAs in gene regulatory networks**

As the number of IncRNAs that have been studied in the context of development, stem cells and cell differentiation is still small, realistically it is too early to speak of their ‘characteristic’ roles. Furthermore, in contrast to TFs and miRNAs (Box 3), the few experimentally characterized IncRNAs appear to be very diverse with regards to their modes of action and regulation. Nonetheless, an emerging theme is that IncRNAs play predominantly local roles, acting on just a few targets that are mostly located in cis to the site of IncRNA transcription. In line with this, IncRNAs are enriched in the vicinity of genes encoding transcriptional regulators and, even more so, TFs involved in embryonic development (Ulitsky et al., 2011). In addition, some IncRNAs, such as *CDR1as*, can affect the activity of specific miRNAs. Thus, various IncRNAs are likely to influence development and differentiation by affecting only a few specific master regulators. Several specific IncRNA activities can be currently foreseen in this context (Fig. 3). In the progenitor state, prior to cell fate transition, IncRNAs can facilitate local repression and establish a threshold for the activation of specific targets required for differentiation. During the early stages of cell fate transition, transcription through the IncRNA locus and/or the IncRNA product itself can change the local chromatin environment to facilitate the binding of TFs and their co-factors to yield increased gene activation or repression. At later stages, IncRNAs can also play roles in positive-feedback loops, which are known to be important for high and stable TF expression. In this case, the IncRNA would be

**Box 3. Transcription factors and miRNAs: general modes of action**

Transcription factors (TFs) and microRNAs (miRNAs) are established key players in developmental pathways, and they have both shared and differential characteristics (Hobert, 2004). TFs and miRNAs regulate expression in *trans* via short cis-regulatory elements present in their targets. Each target integrates diverse signals coming from multiple regulators to yield precise protein expression levels. This combinatorial regulation facilitates a myriad of transcriptional programs essential for establishing cell fate in multicellular organisms. In many cases, TFs act as master regulators that dictate the expression of genes supporting a particular cell fate while repressing the expression of other genes. By contrast, miRNAs typically play a more refined role, modulating the expression levels of genes that should be expressed in the same cells as the miRNA, and repressing the expression of genes that should be expressed in alternative lineages (Bartel, 2009).

At least in model organisms, a substantially larger fraction of TFs than of miRNAs is essential for proper development in laboratory conditions (Hobert, 2008). TFs and miRNAs also typically act on hundreds to thousands of target genes and, accordingly, are expressed at hundreds to tens of thousands of copies per cell.
LncRNAs are emerging as an important class of regulators that function during embryonic development and are endowed with certain features that set them apart from other classes of regulators. The currently annotated lncRNAs are likely to constitute an agglomeration of multiple functional classes, each with their own characteristics and relative importance. To resolve these classes, researchers have to overcome several hurdles, including the need to distinguish between lncRNA functions and those of DNA elements in their loci, and the need to use multiple perturbation strategies to overcome inherent limitations and off-target effects of the various available perturbation methods. Still, the field is progressing rapidly, and we expect that future studies will uncover a multitude of important roles for lncRNAs in virtually any central differentiation process. As lncRNAs are emerging as crucial players in both maintaining progenitor identity and driving differentiation, it is also expected that the manipulation of lncRNA genes will allow the development of more efficient cell differentiation protocols for regenerative medicine. Lastly, understanding the roles of lncRNAs in development and cell differentiation will undoubtedly be instrumental for untangling their functions in those human diseases for which lncRNA dysregulation is reported.

Fig. 3. Potential roles for IncRNAs in cell fate decisions during development. LncRNAs can play various roles during the establishment of cell identity. In the progenitor state, IncRNAs can repress differentiation programs to prevent precocious differentiation (e.g. by establishing a repressive chromatin environment). At the onset of differentiation, the transcription of IncRNAs (i.e. the act of transcription alone and/or the IncRNA product itself) can alter the chromatin environment to facilitate the binding of TFs. If the lncRNA target gene is also a TF, lncRNA activity can be reinforced by expression of the target gene resulting in a positive-feedback loop. In addition, lncRNAs expressed during differentiation can repress regulatory programs required for the establishment of alternative cell fates.

Conclusions
LncRNAs are emerging as an important class of regulators that function during embryonic development and are endowed with certain features that set them apart from other classes of regulators. The currently annotated lncRNAs are likely to constitute an agglomeration of multiple functional classes, each with their own characteristics and relative importance. To resolve these classes, researchers have to overcome several hurdles, including the need to distinguish between lncRNA functions and those of DNA elements in their loci, and the need to use multiple perturbation strategies to overcome inherent limitations and off-target effects of the various available perturbation methods. Still, the field is progressing rapidly, and we expect that future studies will uncover a multitude of important roles for lncRNAs in virtually any central differentiation process. As lncRNAs are emerging as crucial players in both maintaining progenitor identity and driving differentiation, it is also expected that the manipulation of lncRNA genes will allow the development of more efficient cell differentiation protocols for regenerative medicine. Lastly, understanding the roles of lncRNAs in development and cell differentiation will undoubtedly be instrumental for untangling their functions in those human diseases for which lncRNA dysregulation is reported.

Acknowledgements
We thank Neta Degani and Binyamin Zuckerman for comments on the manuscript and members of the I.U. lab for fruitful discussions.

Competing interests
The authors declare no competing or financial interests.

Funding
Work in the authors’ laboratories is funded by Israel Centers for Research Excellence [1796/12]; Israel Science Foundation [1242/14 and 1984/14]; European Research Council IncSAFARI; Minerva Foundation; Fritz Thyssen Stiftung; Lapon Raymond; and the Abramson Family Center for Young Scientists. I.U. is incumbent of the Sygnet Career Development Chair for Bioinformatic and recipient of an Aon Fellowship.

References


