REVIEW

CrossMark

Long non-coding RNA in stem cell pluripotency and lineage commitment: functions and evolutionary conservation

Annalisa Fico[1,2] · Alessandro Fiorenzano[1,2,3] · Emilia Pascale[1,2] · Eduardo Jorge Patriarca[1,2] · Gabriella Minchiotti[1,2]

Abstract
LncRNAs have recently emerged as new and fundamental transcriptional and post-transcriptional regulators acting at multiple levels of gene expression. Indeed, lncRNAs participate in a wide variety of stem cell and developmental processes, acting in *cis* and/or in *trans* in the nuclear and/or in the cytoplasmic compartments, and generating an intricate network of interactions with RNAs, enhancers, and chromatin-modifier complexes. Given the versatility of these molecules to operate in different subcellular compartments, via different modes of action and with different target specificity, the interest in this research field is rapidly growing. Here, we review recent progress in defining the functional role of lncRNAs in stem cell biology with a specific focus on the underlying mechanisms. We also discuss recent findings on a new family of evolutionary conserved lncRNAs transcribed from ultraconserved elements, which show perfect conservation between human, mouse, and rat genomes, and that are emerging as new player in this complex scenario.

Introduction

Most of the mammalian genome (> 90%) is transcribed into non-coding RNA (ncRNA), once controversially known as 'junk DNA' because of its inability to encode proteins and the absence of evolutionary conservation [1–3]. In recent decades, however, large-scale genome-wide sequencing analysis has revealed the tissue-specific expression of ncRNAs and their functional importance as essential regulators in fundamental biological processes, dismantling the now obsolete paradigm of RNA as simply an intermediary

between DNA and protein [4–6]. In this expanded view of both genomic and transcriptomic analysis, thousands of long ncRNAs (lncRNAs) have been identified and classified to include any transcript with primary sequence longer than 200 nucleotides [7]. Based on this arbitrary cut-off, lncRNAs are distinct from more extensively studied classes of short ncRNA, such as transfer RNA, microRNA (miRNA), and small nucleolar RNA [8]. LncRNAs share several features with coding mRNAs (both are transcribed by RNA polymerase II and further capped and spliced [9]), but have lower expression levels, are longer in length, and are involved in many different regulatory circuitries, reflecting their multifunctional role in cells [10]. For instance, they can fold into complex three-dimensional structures able to bind DNA, RNA, and protein molecules, thus determining complex regulatory networks in both the cytoplasm and nucleus [11]. Through distinct modes of action, they are also able to (1) regulate chromatin state and methylation, recruiting remodeling factors in *cis* or in *trans*, (2) act as scaffolds for interactions between proteins by tethering them to complexes that enable transcription factors (TFs) and recruit chromatin modifiers, (3) impact on genome targeting by serving as guides, (4) function as decoys (also referred to as sponges) for miRNA target sites able to sequester and inactivate miRNA function, and (5) mediate antisense interference for

Annalisa Fico and Alessandro Fiorenzano have contributed equally to this work.

✉ Annalisa Fico
annalisa.fico@igb.cnr.it

1 Stem Cell Fate Laboratory, Institute of Genetics and Biophysics "A. Buzzati-Traverso", CNR, 80131 Naples, Italy

2 Institute of Genetics and Biophysics "A. Buzzati-Traverso", CNR, 80131 Naples, Italy

3 Developmental and Regenerative Neurobiology, Wallenberg Neuroscience Center, and Lund Stem Cell Centre, Department of Experimental Medical Science, Lund University, 22184 Lund, Sweden

coding mRNA [12]. Very recent studies investigating these particular characteristics showed that lncRNAs are essential to establish developmental patterning and maintain the pluripotency network, further underscoring their important role in stem cell biology/technology, and in particular cellular reprogramming [13, 14]. As the majority of loci transcribed into lncRNA display low expression levels and poor conservation in the other species, the question of how many human lncRNAs are actually functional is still debated. However, about 1000 human lncRNAs show moderate-to-high expression as well as signs of evolutionary constraint, and around 300 of these are conserved in other non-mammalian vertebrates [15]. Increasing interest is emerging on lncRNAs in embryonic stem cell biology [16–18]. Here, we present an overview of current knowledge on the functional mechanisms of lncRNA in stem cell pluripotency and lineage commitment, highlighting the importance of evolutionary conservation in this context. We also focus on a new family of evolutionary conserved lncRNAs transcribed from ultra-conserved elements (UCEs) known as *Transcribed UCEs* (T-UCEs), sequences of DNA which exhibit the unique feature of retaining extended perfect sequence identity between human, mouse, and rat genomes. This high level of conservation suggests a significant role for T-UCEs during embryogenesis and in stem cell lineage commitment, introducing a novel layer of biological regulation and determining new candidate targets for stem cell-based therapy and other clinical applications.

Mode of action and cellular localization of lncRNAs

Although lncRNAs are reported to be involved in several processes related to physiology and/or disease, only a few have been functionally and mechanistically characterized [11]. The intracellular localization of lncRNAs is normally predictive of their mode of action [11].

In general, nuclear lncRNAs guide chromatin-modifying complexes to precise genomic loci and/or act as molecular scaffolds connecting distinct, but functionally related proteins [19]. As they are able to interact with other nucleic acids forming DNA/RNA duplexes, lncRNAs can exert either repressive or promoting activities on target genes by coordinating protein and RNA interactions, both in *cis* (on neighboring genes) and in *trans* (on distant loci) [20–22] (Fig. 1).

Enhancer-derived ncRNAs (eRNAs), a group of *cis*-acting lncRNAs, are functional transcripts associated with active enhancer sequences involved in many gene activation

Fig. 1 Scheme representing the main mode of action of the lncRNAs localized in the nucleus. *TRF2* telomeric repeat-binding factor 2, *DBHS* drosophila behavior human splicing proteins
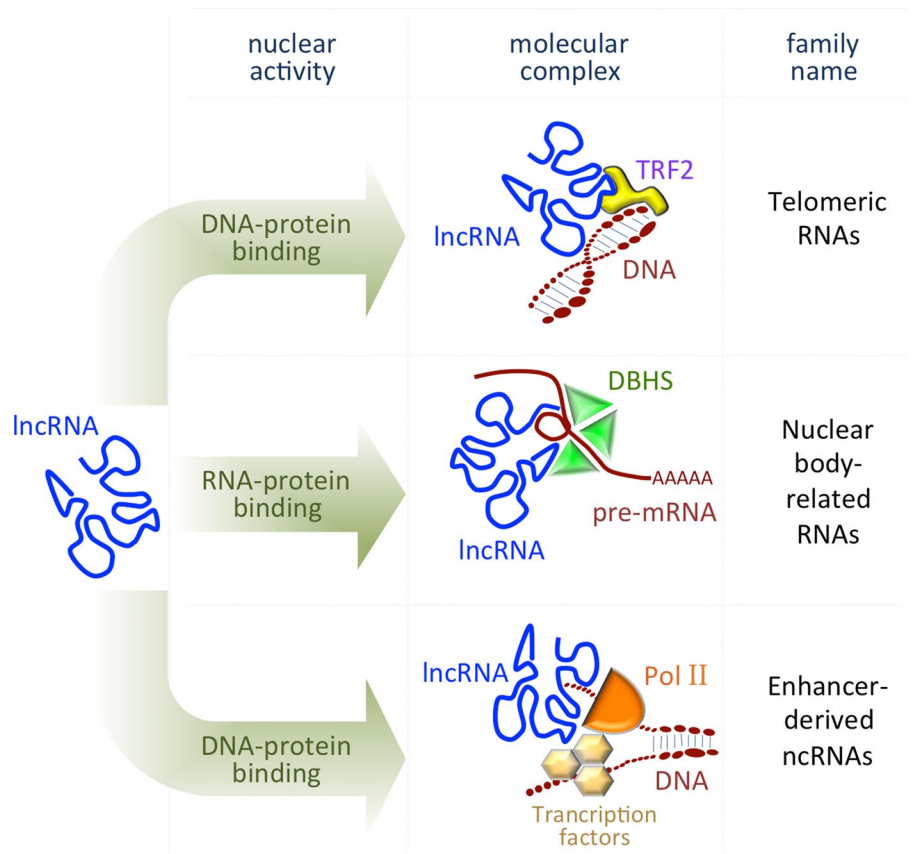
programs. Specifically, they play a fundamental role in guiding chromatin-remodeling complexes to specific promoters and in mediating chromatin loop formation [23]. A recent report identified an eRNA, transcribed at distal super-enhancer 45 kb upstream of the *Nanog* locus, which regulates the two nearest neighbor genes *Nanog* and *Dppa3*, two essential embryonic stem cell (ESC) pluripotency core factors [24]. Emerging evidence indicates that eRNAs are involved in ESC patterning and during differentiation. Using an integrated epigenomic screening approach, a set of eRNAs expressed during cardiac differentiation of ESCs was identified [25]. The expression of these transcripts correlates with the expression of their target genes, located in genomic proximity, which are robustly downregulated after genetic depletion of eRNAs. Intriguingly, eRNAs exhibit distinct expression dynamics, being inhibited in a negative regulatory loop when target mRNAs reach maximal intracellular levels or physiologically increased during stress response in adult heart. Overall, these data highlight a functional role for cardiac eRNAs in heart development and cardiac remodeling after injury. *MesEndoderm Transcriptional Enhancer Organizing Region* (*Meteor*), another eRNA specifically expressed in ESCs, is indispensable during mesendoderm specification and subsequently cardiac differentiation, supporting the involvement of this class of genomic elements in ESC specification during development [26].

Divergent lncRNAs are another group of *cis*-acting lncRNAs that are transcribed in the opposite direction to nearby coding genes [27]. By way of an example, the divergent lncRNA *Evx1as* regulates transcription of its neighbor gene *EVX1*, enhancing ESC mesendoderm differentiation. Mechanistically, *Evx1as* binds to chromatin regulatory sites and, through interaction with Mediator, a multiprotein complex that functions as a transcriptional co-activator, establishes an active chromatin state [28].

Many other lncRNAs act at a distance, regulating gene expression in *trans* via tethering specific protein partners. *Pnky*, an evolutionarily conserved neural-specific transcript, controls mouse and human neurogenesis by preserving neural stem cells in embryonic and post-natal brain. *Pnky* specifically interacts with polypyrimidine tract-binding protein 1, an RNA-splicing factor that is a potent regulator of neural development [29], and this complex regulates a set of transcripts associated with neuronal differentiation. Phenotypically, *Pnky* downregulation promotes neuronal differentiation by increasing the number of cell divisions and depleting the pool of neural progenitors [30].

Although the shuttling mechanisms remain unclear, a large fraction of lncRNAs is exported to the cytoplasm, where they act as important post-transcriptional regulators. As a result of their ability to bind RNA targets through complementary base pairing, lncRNAs regulate gene expression via mRNA degradation or by mediating translational repression (Fig. 2). Cytoplasmic lncRNAs include *Competing Endogenous RNAs* (ceRNAs), which can indirectly enhance protein translation by sequestering miRNAs that would, otherwise, inhibit downstream target mRNAs. This mechanism was shown to be involved in differentiation and several cancer types [31–37]. Circular RNAs are a cryptic class of sponging lncRNAs [38, 39], whose peculiar circular structure provides greater stability than other transcripts. Finally, in humans, several cytoplasmic lncRNAs transactivate Staufen1-mediated mRNA decay by duplexing with 3′-UTRs via Alu elements [40, 41]. However, little is known of the specific molecular functions of these transcripts (Table 1).

Fig. 2 Scheme representing the main mode of action of the lncRNAs localized in the cytoplasm. *STAU1* Staufen 1 protein
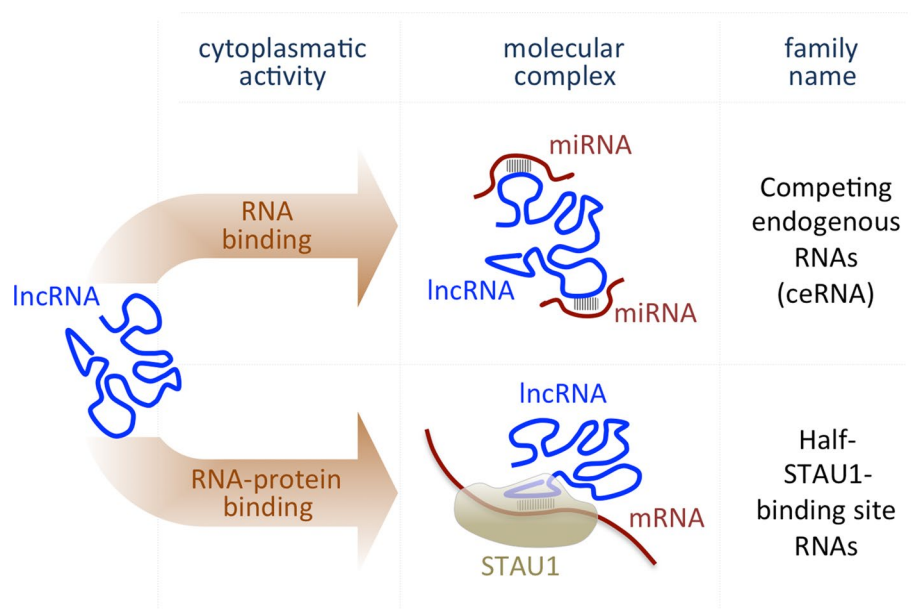
Table 1 Subcellular localization and function of lncRNAs

Cellular localization	Group/family	*Cis*- and *trans*-activity	Examples	References
Nuclear	Enhancer-derived ncRNAs (eRNA)	*Cis*	*Meteor*	[42]
	Divergent lncRNAs	*Cis*	*Evx1as*	[28]
	Pseudogenes	*Trans*	*Oct4P4*	[43]
	Telomeric RNAs	*Trans*	TERRA	[44]
	Nuclear body-related lncRNAs	*Cis/Trans*	*Neat1*	[45, 46]
Cytoplasmatic	Competing endogenous RNAs (ceRNAs)	*Trans*	CircularRNA	[38, 39]
	Half-STAU1-binding site RNAs	*Trans*	AF087999	[40, 41]

LncRNAs are key regulators of pluripotency and differentiation

LncRNAs as members of the pluripotency core network

Several studies identified lncRNAs as fundamental regulators of molecular mechanisms orchestrating the intricate gene regulatory network that controls ESC pluripotency and cell lineage determination [20, 47–51]. In a large-scale functional study, individual knockdown of more than 130 lncRNAs was shown to cause a clear perturbation of the transcriptome, often leading to ESC pluripotency loss [52]. Most lncRNAs involved in mESC self-renewal are transcriptionally regulated by pluripotency core TFs Oct4, Nanog, and Sox2. *Tcl1 Upstream Neuron-Associated* lincRNA (*TUNA*, also known as *megamind*) is a very well-characterized lncRNA required for mESC proliferation and maintenance of self-renewal [53]. TUNA forms an RNA–multiprotein complex, which activates transcription of *Nanog* and *Sox2* upon binding to their promoters. Consistent with its expression in neural progenitors and zebrafish, mouse, and human central nervous system (CNS), TUNA depletion also dramatically impairs neural lineage commitment [54]. Further examples of lncRNAs involved in self-renewal include *AK028326* (Oct4-activated) and *AK141205* (Nanog-repressed), direct targets of OCT4 and NANOG, respectively. Knockdown and overexpression of these transcripts lead to substantial changes in *Oct4* and *Nanog* mRNA levels, with consequent alterations in mESC pluripotency and cellular lineage-specific gene expression [55]. The interplay between pluripotency core TFs and lncRNAs has also been described in hESCs. For example, the lncRNAs *ES1*, *ES2*, and *ES3* were found new regulators of pluripotency and neurogenesis through interaction with SOX2 [51].

Linc-RoR (*Long intergenic non-protein-coding RNA, Regulator of Reprogramming*) is well described as one of the few examples of cytoplasmic lncRNAs regulating pluripotency [47]. It was identified by Zhang and colleagues as an lncRNA able to promote cell reprogramming by inhibiting p53-mediated cell cycle arrest and apoptosis [56]. It was subsequently shown that *linc-RoR* also preserves hESC self-renewal by acting as a ceRNA. Specifically, in undifferentiated hESCs, *linc-RoR* sequesters mir-145 (able to inhibit translation of core TFs [57]) de-repress the translation of all pluripotency factors. Upon differentiation, *linc-RoR* expression is downregulated with consequent release of miR-145 and repression of *Oct4*, *Nanog,* and *Sox2* [58]. In addition, OCT4 activates *linc-RoR* and inhibits miR-145 at transcriptional level, underscoring the existence of a network between TFs, lncRNAs and, small RNAs that, in turn, fine-tunes ESC pluripotency/differentiation balance (reviewed by Rosa and Ballarino [16]).

LncRNAs in chromatin modifications

Many nuclear lncRNAs interact with histone modifiers (writers, readers, and erasers) and/or other chromatin-associated proteins [52], thus acting as epigenetic regulators. A pivotal study using RNA immunoprecipitation sequencing identified several Polycomb repressive complex 2 (PRC2)-associated RNAs in ESCs at genome-wide level [59]. Although promiscuous RNA binding to the PRC2 complex is reported [60], the other studies highlight that lncRNA binding to this complex is important in modulating its interaction with cofactors that in turn confer its specificity of action. The lncRNA–PRC2 interaction involving *Maternally Expressed 3* (*Meg3*) and JARID2 in pluripotent stem cells is very well described. *Meg3* is a maternally expressed, imprinted lncRNA belonging to the *Dlk1–Dio3* gene cluster on chromosome 12qF1. Appropriate expression of these lncRNAs is required for embryonic development [61, 62] and to reach full pluripotency during cell reprogramming. Indeed, *induced Pluripotent Stem Cells* (iPSCs) carrying aberrantly silenced *Dlk1–Dio3* cluster genes are unable to contribute to development of chimeric mice and fail to pass tetraploid complementation assay [63]. JARID2 is a catalytically inactive Jumonji family histone demethylase that is essential for PRC2 recruitment in ESCs [64]. JARID2 is able to directly interact with about 100 lncRNAs in mESCs

[65], including *Meg3*, which regulates the activity of PRC2 in ESCs by binding JARID2 [59]. Specifically, *Meg3* stabilizes PRC2 occupancy in *trans* at genomic loci encoding for factors involved in cell differentiation [59]. At a mechanistic level, *Meg3* acts as a scaffold to increase/stabilize the interaction between JARID2 and the PRC2 component EZH2, assembling the Polycomb complex on chromatin at specific JARID2 target sites. The specificity of this mechanism is achieved via RNA–DNA base pairing recognition between the lncRNA and the target gene [65].

Similarly, the lncRNA Braveheart (*Bvht*) directly binds PRC2. *Bvht* was shown to directly interact with SUZ12, a core component of the PRC2 complex, at numerous stages during ESC-cardiac differentiation. Interestingly, SUZ12 and its associated repressive modification H3K27me3 are enriched at the promoters of cardiac-associated genes such as *MesP1* in cells lacking *Bvht* expression. Notably, those genes remain bivalent in *Bvht*-depleted cells, similar to their initial configuration in ESCs, in line with the inability of these cells to trigger cardiac cell commitment. Indeed, *Bvht* has been described as a novel lncRNA that mediates specific cell commitment by epigenetic regulation of gene-expression programs [42].

Another well-characterized histone modifier that interacts with lncRNAs is WDR5, a component of mixed-lineage leukemia (MLL) complexes [66–68]. WDR5 is particularly important for mammalian ESC self-renewal and maintenance of active chromatin for pluripotency genes, and is required for efficient generation of iPSCs from differentiated somatic cells [69, 70]. In undifferentiated stem cells, it interacts with many lncRNAs [71] including *lincRNA-1592* and *lincRNA-1552*, which are involved in the maintenance of ESC pluripotency. In turn, WDR5 binds the promoters of these two lncRNAs, suggesting a *cis*-regulatory mechanism (detailed in the review by Rosa and Ballarino [16]).

A sense pseudogene–lncRNA-based mechanism of gene regulation at epigenetic level controlling cross-talk between pseudogenes and their ancestral genes is also described, in which the X-linked *Oct4* pseudogene controls mESC self-renewal [43]. Specifically, in differentiating mESCs, the lncRNA *Oct4P4* forms a complex with the histone-lysine *N*-methyltransferase SUV39H1, which translocates to the *Oct4* promoter silencing *Oct4* gene.

Among the lncRNAs functionally characterized, there is also *lncPress1*, identified by analyzing 40 lncRNAs highly expressed in undifferentiated hESCs and repressed during differentiation in a p53-dependent manner. The interaction between *lncPRESS1* and SIRT6, the main deacetylase of histone H3 on lysine 56 (H3K56ac), has been extensively studied in hESCs. Transcriptionally activated pluripotency genes show high H3K56ac levels, which decrease significantly during ESC differentiation [72, 73]. Mechanistically, *lncPRESS1* interacts with SIRT6 and blocks its chromatin localization by maintaining high levels of histone H3K56 acetylation at promoters of pluripotency genes, ensuring the hESC pluripotent state [74].

LncRNAs in maintenance of nuclear architecture

The nuclear organization of hESCs is particularly compartmentalized, and nuclear organization remodeling is linked to epigenomic reprogramming during differentiation [75, 76]. LncRNAs, such as *nuclear-enriched autosomal transcript 1* (*Neat1*), are important components of paraspeckles and are required for the assembly and structural integrity of these nuclear bodies [45, 46, 77]. Undifferentiated hESCs lack *Neat1* expression and paraspeckles within the nucleus; interestingly, they both only appear upon differentiation, suggesting that lncRNAs maintaining nuclear structure integrity have a potential regulatory role in ESCs (detailed in the review by Ng and Stanton [50]).

LncRNA in cell signaling pathways and metabolism

The importance of external stimuli and cell signaling in the tuning of ESC pluripotency/differentiation balance is extensively described [78–80]. In this scenario, lncRNAs are players in the molecular orchestra of the pluripotent regulatory network. Examples include the *Growth arrest-specific transcript 5* (*GAS5*), a known tumor suppressor and growth arrest-related lncRNA, which is highly expressed in hESCs and directly regulated by the pluripotency factors OCT4 and SOX2. Specifically, *GAS5* maintains TGFβ signaling by protecting TGFβ receptor family ligand NODAL expression from miRNA-mediated degradation, thereby promoting hESC and iPSC self-renewal and pluripotency [81, 82].

Another lncRNA involved in cell signaling pathways is the conserved transcript *Divergent to* Goosecoid (GSC) *Induced by TGFβ* family signaling (*DIGIT*). By mapping the genome-wide occupancy of SMAD3 during endoderm differentiation, *DIGIT* was identified as an lncRNA regulated by an enhancer bound by SMAD3 upon Activin stimulation. *DIGIT* controls definitive endoderm specification by positively regulating in *trans* the proximal mesendoderm regulator GSC, with which it is divergently transcribed. Depletion of the DIGIT transcript inhibits the induction of GSC during endoderm differentiation of both hESCs and mESCs [83]. *Telomeric RNA* (*TERRA* or *TelRNA*), another type of lncRNA, is also highly expressed in mESCs, but declines significantly upon differentiation, implying that it may be involved in the maintenance of cell pluripotency. Interestingly, *TERRA* is one of the targets of the Wnt/β-catenin signaling pathway and can mimic its self-renewal-promoting effect when overexpressed. *TERRA* was found to inhibit the transcription of TCF3, and this is likely its key contribution to maintenance of mESC self-renewal, as overexpression of

TCF3 abolishes the self-renewal-promoting effect of *TERRA* [44].

Recent studies have also highlighted the involvement of lncRNAs in the regulation of metabolic pathways contributing to stem cell fate specification. For instance, the lncRNA Lncenc1 has been shown to preserve ESC self-renewal by regulating the transcription of glycolytic genes through interaction with two RNA binding proteins, PTBP1 and HNRNPK. Indeed, the absence of Lncenc1 leads to a significant reduction of glycolysis-associated genes expression, which eventually results in an impaired glycolytic activity [84].

LncRNAs and the importance of evolutionary conservation

Evolutionary conservation is currently considered one of the most powerful and reliable parameters to identify functional sequences in the genome, highlighting their role as potential regulatory elements in key biological processes [85].

Although there are a few experimentally studied lncRNAs that are conserved at sequence level, most exhibit weak or imperceptible primary sequence conservation, reflecting a lack of evolutionary constraints. This opens to debate the functional importance of lncRNAs. However, similarities in canonical sequences cannot be used as the only criterion to measure evolutionary relatedness, as lack of conserved sequences does not imply per se lack of functional conservation; additional dimensions of conservation need to be considered for lncRNAs such as structure, function, and expression from syntenic loci [86]. For instance, the lncRNA *HOTAIR* shows a conserved function and genomic position in the *HOX-C* cluster, although it is barely conserved in primary sequence between human and mouse [87]. Similarly, *GAS5* is another good example of an lncRNA without substantial sequence conservation but unambiguous biological function [81, 82]. In addition, recent in silico studies provided greater insight into interspecies conservation based on secondary structures among lncRNA homologs, underscoring how structure is more important than primary sequence changes in lncRNA functionality (reviewed by Nitsche and Stadler [88]).

However, there are some examples of lncRNA showing some degree of conservation in their primary sequence. For example, *TUNA* is one of the best conserved lncRNAs associated with ESC biology. Discovered in zebrafish and named *megamind*, it is mainly involved in brain development but also expressed in spinal cord and eye tissue [54]. As previously mentioned, it also displays important functions in molecular mechanisms underpinning self-renewal and cell fate determination. The exonic regions of *TUNA* show untypical strong sequence conservation across vertebrates.

In particular, it contains a sequence element of about 200 bp in length with more than 80% sequence similarity between human and zebrafish [54]. Such a level of conservation exceeds even that of most coding genes. Another possible exception is *TERRA*, which is conserved between human and yeast [8]. Well-studied functionally important lncRNAs with orthologs over a wide phylogenetic range of species include genes such as *DIGIT*, *Braveheart*, *Pnky*, and *Neat1* (Table 2).

The *Transcribed ultraconserved elements* (T-UCEs), known to be the class of lncRNAs with the highest level of evolutionary conservation, are of particular interest in this already complex scenario. In the following sub-sections, we provide an overview of the UCEs and a summary of major findings that support the importance of T-UCEs in controlling ESC and early embryonic lineage commitment, highlighting key molecular pathways, and discussing their potential significance for stem cell biology.

Ultraconserved elements (UCEs): a close look at evolutionary conservation

Computationally identified for the first time by Bejerano and colleagues in 2004, UCEs are 481 genomic segments longer than 200 bp; they are rarely lost across a wide range of mammalian species, showing the signs of evolutionary sequence constraint [89]. UCEs display perfect conservation (100% identity with no insertions or deletions) between human, mouse, and rat, and also retain a high percentage of identity in chicken, dog, and fugu genome [90]. UCEs are found in all chromosomes except chromosome 21 and Y, and are commonly classified according to the most recent human genome assembly (hg20) into five subgroups: intergenic, intronic, exonic, partially exonic, and exon-containing [91]. Although their complete functional characterization is still a long way off, it is known that UCEs are not randomly distributed, and their position within the genome also seems to reflect their function. These constrained sequences are in fact mostly allocated in clusters, flanking, or embedding genes involved in important physiological processes, and act as splicing or enhancer factors [92]. Their extreme conservation could be due to the absence of annotated transposons near many UCEs during evolution [93]. It has also been observed that these transposon-free regions coincide with the so-called chromatin bivalent domains, which mark key regulatory genes in embryo development and ESC pluripotency, implying a potential correlation between UCEs and genome and chromatin architecture [94]. Extreme conservation over such long stretches of DNA indicates strong negative selection pressure, suggesting that UCE depletion may have a dramatic effect on mammalian development [95, 96]. Unexpectedly, the initial findings suggested

Table 2 LncRNAs and their evolutionary conservation

lncRNA	Conservation	Function roles	References
AK028326	Poor conserved	Self-renewal	[8]
AK141205	Conserved	Self-renewal	[52]
Braveheart	Not conserved	Cardiovascular differentiation	[42]
DIGIT	Conserved	Meso-endoderm differentiation	[83]
Evx1as	Conserved	Mesoderm differentiation	[28]
GAS5	Poor conserved	Self-renewal	[81, 82]
Hotair	Poor conserved	Self-renewal Cell proliferation	[87]
LincPRESS1	Poor conserved	Pluripotency Cell cycle regulation	[74]
LincRNA1592-1552	Poor conserved	Pluripotency	[52]
Lin-RoR	Poor conserved	Pluripotency Self-renewal	[47]
Meg3	Conserved	Pluripotency Reprogramming	[62, 63]
Meteor	Conserved	Mesoderm specification	[26]
Neat1	Conserved	Differentiation	[77]
Oct4P4	Poor conserved	Self-renewal Cell proliferation	[43]
Pnky	Conserved	Neuronal differentiation	[30]
TERRA	Conserved	Pluripotency	[8]
TUNA	Conserved	Self-renewal Neural differentiation	[54]

that the deletion of some UCEs was dispensable for mouse viability, and that the phenotype was only detectable over many generations, raising the possibility that they may be phenotypically redundant [97, 98]. However, this apparent discrepancy was successfully resolved in a recent study by Dickel and colleagues. The authors clarified the role of ultraconserved sequences in the early development, re-establishing the intriguing hypothesis that extreme levels of UCE conservation mirror their functional importance in genomic regulation for the acquisition of cell identity [99]. Emerging evidence shows that a significant fraction of UCEs act both as gene-expression enhancer promoters and lncRNAs, postulating their dual function during mouse development. This new layer of regulation makes the biology of UCEs even more challenging and adds further complexity to their functional annotation, while reinforcing the concept that the extraordinary constraint on their sequence implies simultaneous multiple roles [100, 101].

Ultraconserved enhancer

Ultraconserved loci are normally located near key developmental coding genes and can work as *cis*-acting regulatory genomic elements directing the expression of neighbor genes. By comparing next-generation sequencing data sets with readouts on the function of UCEs as enhancers in mouse embryonic development, it emerges that a large number of UCEs can be transcribed and act as enhancers concomitantly at specific developmental stages [102]. This indicates that transcription and enhancer functions overlap within the same DNA sequence. A genome-wide high-throughput in vivo screening study using transgenic mouse reporter assay systematically tested hundreds of ultraconserved DNA elements annotated at E11.5 [102]. Remarkably, on examining a single time point of mouse development, almost 50% were found to act as tissue-specific enhancers. Although it is still unclear whether UCEs are also able to enhance in *trans* expression of multiple genes located further away along a chromosome, these findings warrant in-depth exploration of UCEs as enhancers in the other stages of embryo development, including the earliest stages of embryogenesis when cell lineage segregation is established [99, 102].

Deletion of UC.248, UC.329, UC.467, and UC.482 in knockout (KO) mice was found to be compatible with life and development with no obvious deleterious defects, challenging the paradigm that in vivo depletions would lead to a lethal phenotype, thus pointing to a functional role for these non-coding ultraconserved sequences [98]. The hypothesis that extreme sequence constraint does not necessarily reflect key functions required for viability was only recently validated [99]. Using CRISPR/Cas9 genome editing, a series of KO mice was generated lacking individual

or different combinations of seven UCEs (UC.463–UC.465 and UC.467–UC.470). These UCEs display transcriptional enhancer activity in telencephalon or diencephalon, and are located along the X chromosome within regions of DNA spanning the *Arx* gene encoding an important neuronal TF, whose mutation causes neurological and sexual development disorders. Although KO mice were still viable and fertile, in almost all cases every single deletion induced an aberrant phenotype with dramatic neurological and growth abnormalities, undermining the hypothesis of functional redundancy [99]. Moreover, pair-wise losses in various combinations of these seven UCEs increased the severity of the mutant phenotype. After post-natal mouse brain dissection, the rodents showed abnormal alterations of neuron populations and substantial structural brain defects accompanied by decreasing fitness over long time periods [99]. These findings go some way to revealing the important role of non-coding UC sequences in neural development, shedding light, for the first time, on the significant impact of ultraconserved 'dark matter' on embryo development.

Transcribed UCEs

Several studies demonstrated that a large subset of UCEs is actively transcribed, though without any protein-coding potential [103–106]. These transcripts, known as T-UCEs, act as lncRNAs regulating other RNAs, and include (and often extend beyond) the conserved sequences initially described by Bejerano et al. [89]. Interestingly, members of this new family of transcripts are expressed in a tissue-specific manner and exhibit aberrant expression levels in several human cancers, as reported for the first time by Calin and colleagues [104]. T-UCEs are preferentially located in the cytoplasm, where they are able to influence gene-expression levels and regulate several biological processes, such as cell proliferation and differentiation. The main molecular mechanism of T-UCE activity described to date is its 'decoy' function. By acting as natural sponges, these lncRNAs can sequester miRNAs via sequence complementarity [107–110]. Decoy binding sites within T-UCEs also act through chromatin remodeling by interplaying with distinct regulatory molecules including epigenetic modifiers, transcription factors, and catalytic proteins [111, 112]. Although the recent technological advances have revealed the molecular circuitries involving T-UCEs in human diseases and in development, their functional activities still remain largely unexplored. For a functional genome-wide characterization of T-UCEs, precise and rigorous gene annotation would require enhanced bioinformatics and high-throughput RNA sequencing approaches [13]. T-UCEs can in fact be transcribed from either the sense or antisense strand of the encoding gene, corresponding to the sense and complementary sequence, respectively, and are generally referred to as '+' and '+A'. To date, a total of 962 T-UCEs have been annotated, further supporting independent transcriptional regulation by neighboring protein-coding genes [89, 113]. Although growing evidence has underscored the importance of T-UCEs during the early stages of development, and in stem cell biology, the physiological role of this specific class of lncRNAs and their mechanism(s) of action is only recently emerging [114, 115].

Role of T-UCEs during embryonic development

Due to their highly conserved nature, T-UCEs are emerging as new and critical players of key developmental gene regulation orchestrating the patterning of cells into tissues and organs during development [116–118].

The majority of UCEs are transcribed into single-stranded transcripts during development, and exhibit regional and cell-specific localization. They are involved in the formation of a broad range of cell types, and in ensuring correct embryonic patterns. By combining large-scale genome-wide expression analysis and in situ hybridization detection, several groups showed that T-UCEs are differentially expressed in both time and space with highly restricted expression in selected regions of the mammalian embryo [117–119]. High-resolution screening of RNA sequence signals derived from annotated ultraconserved sequences revealed that 76 UCEs are actively transcribed and specifically enriched in the nervous system of developing brain at embryonic day E14.5 [119]. In particular, T-UC.77, T-UC.338, T-UC.377, and T-UC.359 show dynamic expression profiles increasing during E12.5–E18.5 of mouse brain development. In contrast, T-UC.138, T-UC.189, and T-UC.376 display decreasing expression patterns in the same time frame, indicating physiological significance in earlier stages of mouse brain development [117] (Fig. 3, top panel). Of note, some T-UCEs remain expressed in adult brain, functioning in homeostasis in the cerebral cortex [119].

Intriguingly, most T-UCEs also retain high conservation of gene expression among different species. Indeed, a large spectrum of T-UCEs exhibit a similar expression pattern and cell-specific localization during development across mice, macaques, and humans, indicating that DNA sequence conservation may also correspond to conserved expression and function during evolution [117]. T-UCEs can also cooperate with TFs, promoting the differential and fine-tuned regulation of nearby genes controlled by the same DNA regulatory elements. One example is the ultraconserved lncRNA *Evf-2*, localized downstream of the *Dlx5* locus and transcribed antisense to *Dlx6*, which regulates neural cell fates in the brain. *Evf-2* controls GABAergic interneuron activity in *cis* by regulating cellular levels of the TFs Dlx5 and Dlx6, while it inhibits DNA methylation by interacting simultaneously in
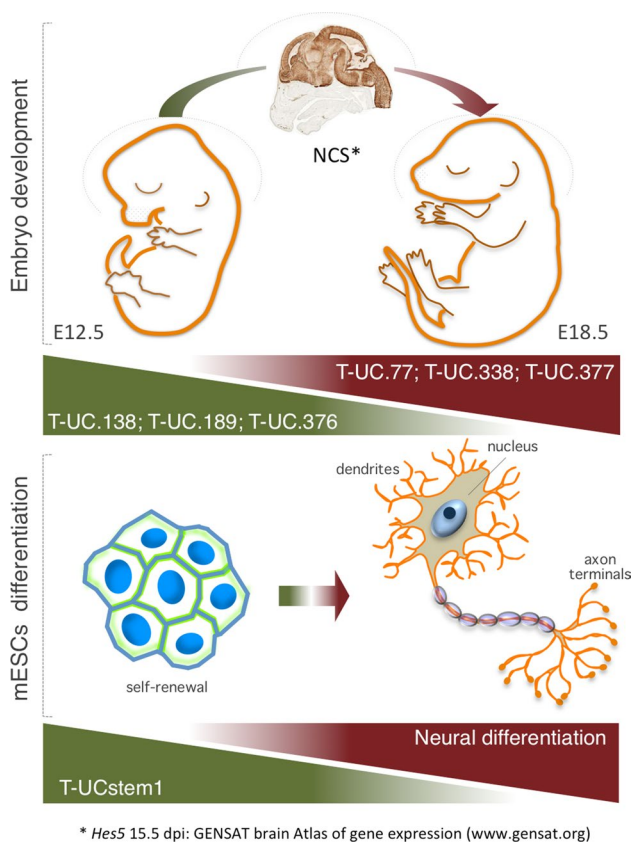
Fig. 3 T-UCEs showing dynamic expression profiles during E12.5–E18.5 of mouse brain development (top panel). T-Ucstem1 expression decreases during ESC neural differentiation (bottom panel)

trans with the transcriptional activator Dlx2 and the repressor Mecp2. In particular, *Evf-2* prevents site-specific CpG DNA methylation of the Dlx5/6 enhancer by modulating competition between Dlx2 and Mecp2 in the medial ganglionic eminence at E13.5 [120, 121].

T-UCE-mediated regulation of ESC self-renewal and differentiation

Although the previous findings mostly linked T-UCEs to human diseases such as cancer, more recent evidence provides a new dimension to the understanding of T-UCE functions in regulating basal cellular physiology, demonstrating a direct involvement of T-UCEs in stem cell biology. Due to their polyhedral nature, T-UCEs can interplay with different types of molecular factors, DNA, RNA, and proteins, and this confers the capacity to transduce higher order regulatory networks maintaining the balance between self-renewal and multi-lineage differentiation. T-UCE family members are dynamic and temporally regulated in gene expression during stem cell differentiation [122].

To date, an unbiased genome-wide expression analysis approach has been the starting point for identifying tissue-specific lncRNAs in many studies. Dinger et al. described, for the first time, the developmentally regulated expression of T-UCEs during differentiation of mouse ESCs. By integrating genomic context analysis with expression profiling of transcripts, they identified novel candidates with a potential role in self-renewal and cell fate choice. Among the T-UCEs examined, *Evf-1/2* and the antisense *Dlx1as* showed progressively increasing expression levels during cardiovascular differentiation, and exhibited coordinated expression profiles with their respective genomically associated coding genes, *Dlx5/Dlx6* and *Dlx1/Dlx2* [49]. Notably, *Dlx1as* is expressed in adult mice in brain regions associated with neurogenesis including anterior sub-ventricular zone and olfactory bulb acting as a crucial neural development regulator in the glial–neuronal lineage specification of multipotent adult stem cells [123].

More recent large-scale efforts, employing genome-wide sequencing of multiple tissues, identified T-UC.283+ as the T-UCE most expressed in ESCs, iPSCs and in several embryonic and extra-embryonic tissues. Although no specific function of T-UC283+ has been elucidated, it has been suggested that it may be involved in pluripotency maintenance but also during the first-cell lineage specification [107].

At genomic level, by examining histone methylation in mouse ESCs across 56 large UCE-rich loci, most highly conserved non-coding elements in mammalian genomes were found grouped within regions enriched for key genes of differentiation and developmental patterning. The association of conserved non-coding sequences with 'bivalent domains', which represent a chromatin-based mechanism for maintaining pluripotency and silencing developmental genes in mouse ESCs, revealed their potential role in epigenetic regulation. Indeed, most of the ultraconserved DNA sequences contained specific binding sites of SUZ12 enabling trimethylation at histone H3K27 in pluripotent stem cells [124, 125]. These findings highlighted the crucial functions of ultraconserved loci in chromatin state regulation for the correct induction of cell differentiation programming.

One of the most functionally characterized T-UCEs in stem cell biology is *T-UCstem1*, which fine-tunes the balance between ESC self-renewal and differentiation (Fig. 3, bottom panel) by exerting a dual but distinct role in the nucleus and in the cytosol (Fig. 4). Nuclear *T-UCstem1* directly interacts with PRC2, facilitating its recruitment to chromatin on bivalent domain-associated genes. Indeed, in the absence of *T-UCstem1*, PRC2 is displaced resulting in an increased H3K4me3/H3K27me3 ratio at bivalent domains, driving ESCs to rapidly exit from pluripotency and undergo differentiation. In the cytosol, *T-UCstem1* controls cell cycle progression in ESCs by acting as a

Fig. 4 Scheme representing the T-UCE mode of action in ESCs

WRE: writers, readers and erasers of DNA epigenetics marks
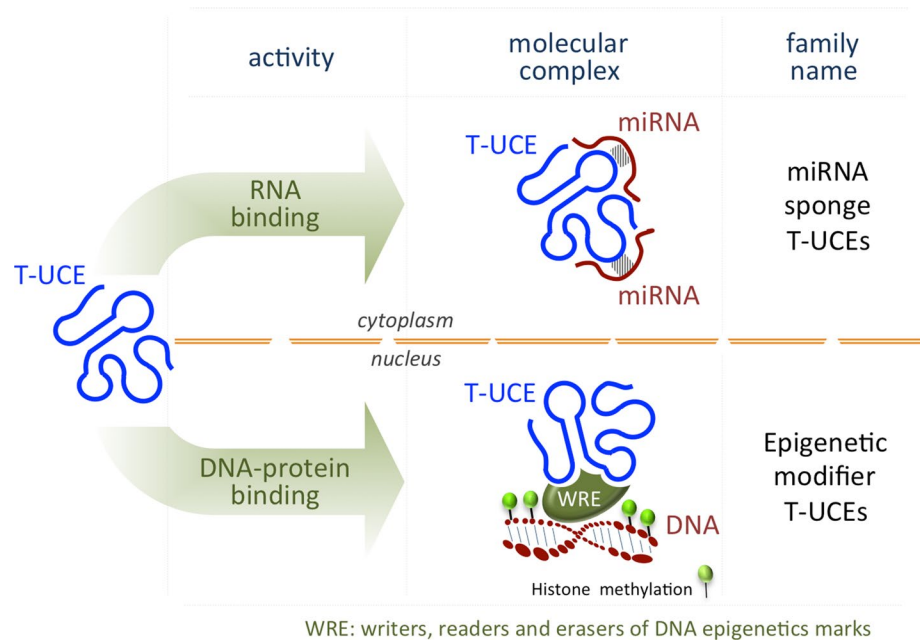
sponge, sequestering miR-9, and de-repressing its mRNA targets *TLx1* and *Lin28* [126] (Fig. 4).

Given the well-known function of miR-9 in the embryonic and post-natal neurogenesis [127], it will be interesting to investigate in the future whether T-UCstem1 may exert a regulatory role also in adult neurogenesis, particularly in the neural stem cells of the sub-ventricular zone.

Together, all these studies explore the role of conserved lncRNAs in stem cell biology and provide a better understanding of their modes of action within an already complex regulatory landscape, which may be useful in developing more effective stem cell-based clinical strategies (Table 3).

Conclusions and future perspectives

Despite the increasing interest, our understanding of the biological functions and molecular mechanism of lncRNAs is still in its infancy. Major challenges towards this goal are the lack of functional annotation for the majority of lncRNAs and low intracellular expression levels. In addition, their mostly weak sequence conservation precludes the use and comparison of vertebrate model organisms (e.g., zebrafish, mouse, and human) for studying fundamental biological processes including stem cell differentiation and embryonic development. This has prompted scientists to focus on a specific subgroup of conserved/ultraconserved lncRNAs, such

Table 3 UCE and T-UCE

	Biological context	Function	References
UCE			
UC.248, 329, 463, 465, 467, 470, 482	Brain development	Enhancer activity	[98, 99]
T-UCE			
Evf-1/2	ESC-cardiac differentiation (EBs)	Chromatin modification	[49]
Dlx1as	Differentiation	Chromatin modification Competitive endogenous RNA	[123].
T-UC.283+	Embryonic and extra-embryonic tissues/iPSCs	Not defined	[107]
T-UCstem1	Undifferentiated ESCs	Competitive endogenous RNA and chromatin modification	[126]
T-UC.77, 338, 377, 359, 138, 189, 376	Brain development	Not defined	[117]

as T-UCEs, which show a high level of interspecies conservation to decipher their molecular basis.

Despite extensive knowledge of lncRNAs biology, their applications in diagnosis and therapy still require further investigations, including better elucidation of their mode of action. For instance, given their cell-type- and organ-specific expression patterns, lncRNAs might be employed as potential selection markers for screening suitable stem cells/iPSCs or progenitor cells [128]. In this context, gaining a greater insight into ultraconserved lncRNAs may help frame and guide further studies into less conserved classes of lncRNAs.

References

1. International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. Nature 431(7011):931–945
2. Wang J et al (2004) Mouse transcriptome: neutral evolution of 'non-coding' complementary DNAs. Nature 431(7010):1 **(following 757; discussion following 757)**
3. Struhl K (2007) Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat Struct Mol Biol 14(2):103–105
4. Huarte M (2015) The emerging role of lncRNAs in cancer. Nat Med 21(11):1253–1261
5. Ling H et al (2015) Junk DNA and the long non-coding RNA twist in cancer genetics. Oncogene 34(39):5003–5011
6. Lorenzen JM, Thum T (2016) Long noncoding RNAs in kidney and cardiovascular diseases. Nat Rev Nephrol 12(6):360–373
7. Carninci P et al (2005) The transcriptional landscape of the mammalian genome. Science 309(5740):1559–1563
8. Ulitsky I, Bartel DP (2013) lincRNAs: genomics, evolution, and mechanisms. Cell 154(1):26–46
9. Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. Annu Rev Biochem 81:145–166
10. Derrien T et al (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 22(9):1775–1789
11. Batista PJ, Chang HY (2013) Long noncoding RNAs: cellular address codes in development and disease. Cell 152(6):1298–1307
12. Wang KC, Chang HY (2011) Molecular mechanisms of long noncoding RNAs. Mol Cell 43(6):904–914
13. Cabili MN et al (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 25(18):1915–1927
14. Perry RB, Ulitsky I (2016) The functions of long noncoding RNAs in development and stem cells. Development 143(21):3882–3894
15. Hezroni H et al (2015) Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. Cell Rep 11(7):1110–1122
16. Rosa A, Ballarino M (2016) Long noncoding RNA regulation of pluripotency. Stem Cells Int 2016:1797692
17. Hu S, Shan G (2016) LncRNAs in stem cells. Stem Cells Int 2016:2681925
18. Pal D, Rao MRS (2017) Long noncoding RNAs in pluripotency of stem cells and cell fate specification. Adv Exp Med Biol 1008:223–252
19. Tsai MC et al (2010) Long noncoding RNA as modular scaffold of histone modification complexes. Science 329(5992):689–693
20. Rinn JL et al (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129(7):1311–1323
21. Wang KC et al (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature 472(7341):120–124
22. Morlando M et al (2014) The role of long noncoding RNAs in the epigenetic control of gene expression. ChemMedChem 9(3):505–510
23. Lai F et al (2013) Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature 494(7438):497–501
24. Blinka S et al (2016) Super-enhancers at the nanog locus differentially regulate neighboring pluripotency-associated genes. Cell Rep 17(1):19–28
25. Ounzain S et al (2014) Functional importance of cardiac enhancer-associated noncoding RNAs in heart development and disease. J Mol Cell Cardiol 76:55–70
26. Alexanian M et al (2017) A transcribed enhancer dictates mesendoderm specification in pluripotency. Nat Commun 8(1):1806
27. Sigova AA et al (2013) Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. Proc Natl Acad Sci USA 110(8):2876–2881
28. Luo S et al (2016) Divergent lncRNAs regulate gene expression and lineage differentiation in pluripotent cells. Cell Stem Cell 18(5):637–652
29. Keppetipola NM et al (2016) Multiple determinants of splicing repression activity in the polypyrimidine tract binding proteins, PTBP1 and PTBP2. RNA 22(8):1172–1180
30. Ramos AD et al (2015) The long noncoding RNA Pnky regulates neuronal differentiation of embryonic and postnatal neural stem cells. Cell Stem Cell 16(4):439–447
31. Cesana M et al (2011) A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell 147(2):358–369
32. Lu W et al (2015) A 3′UTR-associated RNA, FLJ11812 maintains stemness of human embryonic stem cells by targeting miR-4459. Stem Cells Dev 24(9):1133–1140
33. Smith KN et al (2017) Long noncoding RNA moderates microRNA activity to maintain self-renewal in embryonic stem cells. Stem Cell Reports 9(1):108–121
34. Tay Y, Rinn J, Pandolfi PP (2014) The multilayered complexity of ceRNA crosstalk and competition. Nature 505(7483):344–352
35. Weng R et al (2018) Long noncoding RNA-1604 orchestrates neural differentiation through the miR-200c/ZEB axis. Stem Cells 36(3):325–336
36. Zhou Y et al (2016) AK048794 maintains the mouse embryonic stem cell pluripotency by functioning as an miRNA sponge for miR-592. Biochem J 473(20):3639–3654
37. Salmena L et al (2011) A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell 146(3):353–358

38. Ivanov A et al (2015) Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. Cell Rep 10(2):170–177

39. Memczak S et al (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 495(7441):333–338

40. Gong C, Maquat LE (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. Nature 470(7333):284–288

41. Kretz M (2013) TINCR, staufen1, and cellular differentiation. RNA Biol 10(10):1597–1601

42. Klattenhoff CA et al (2013) Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. Cell 152(3):570–583

43. Scarola M et al (2015) Epigenetic silencing of Oct4 by a complex containing SUV39H1 and Oct4 pseudogene lncRNA. Nat Commun 6:7631

44. Xu X et al (2018) Telomeric noncoding RNA promotes mouse embryonic stem cell self-renewal through inhibition of TCF3 activity. Am J Physiol Cell Physiol 314(6):C712–C720

45. Chen LL, Carmichael GG (2009) Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. Mol Cell 35(4):467–478

46. Sunwoo H et al (2009) MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. Genome Res 19(3):347–359

47. Loewer S et al (2010) Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat Genet 42(12):1113–1117

48. Chalei V et al (2014) The long non-coding RNA Dali is an epigenetic regulator of neural differentiation. Elife 3:e04530

49. Dinger ME et al (2008) Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res 18(9):1433–1445

50. Ng SY, Stanton LW (2013) Long non-coding RNAs in stem cell pluripotency. Wiley Interdiscip Rev RNA 4(1):121–128

51. Ng SY, Johnson R, Stanton LW (2012) Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. EMBO J 31(3):522–533

52. Guttman M et al (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature 477(7364):295–300

53. Ulitsky I et al (2011) Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell 147(7):1537–1550

54. Lin N et al (2014) An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. Mol Cell 53(6):1005–1019

55. Sheik Mohamed J et al (2010) Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells. RNA 16(2):324–337

56. Zhang A et al (2013) The human long non-coding RNA-RoR is a p53 repressor in response to DNA damage. Cell Res 23(3):340–350

57. Xu N et al (2009) MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells. Cell 137(4):647–658

58. Wang Y et al (2013) Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. Dev Cell 25(1):69–80

59. Zhao J et al (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell 40(6):939–953

60. Davidovich C et al (2013) Promiscuous RNA binding by Polycomb repressive complex 2. Nat Struct Mol Biol 20(11):1250–1257

61. Takahashi N et al (2009) Deletion of Gtl2, imprinted non-coding RNA, with its differentially methylated region induces lethal parent-origin-dependent defects in mice. Hum Mol Genet 18(10):1879–1888

62. Zhou Y et al (2010) Activation of paternally expressed genes and perinatal death caused by deletion of the Gtl2 gene. Development 137(16):2643–2652

63. Stadtfeld M et al (2010) Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature 465(7295):175–181

64. Shen X et al (2009) Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell 139(7):1303–1314

65. Kaneko S et al (2014) Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. Mol Cell 53(2):290–300

66. Wysocka J et al (2005) WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. Cell 121(6):859–872

67. Migliori V et al (2012) Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. Nat Struct Mol Biol 19(2):136–144

68. Gomez JA et al (2013) The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus. Cell 152(4):743–754

69. Ang YS et al (2011) Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. Cell 145(2):183–197

70. Li X et al (2012) The histone acetyltransferase MOF is a key regulator of the embryonic stem cell core transcriptional network. Cell Stem Cell 11(2):163–178

71. Yang YW et al (2014) Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. Elife 3:e02046

72. Tan Y et al (2013) Acetylated histone H3K56 interacts with Oct4 to promote mouse embryonic stem cell pluripotency. Proc Natl Acad Sci USA 110(28):11493–11498

73. Xie CQ et al (2009) Expression profiling of nuclear receptors in human and mouse embryonic stem cells. Mol Endocrinol 23(5):724–733

74. Jain AK et al (2016) LncPRESS1 Is a p53-regulated LncRNA that safeguards pluripotency by disrupting SIRT6-mediated de-acetylation of histone H3K56. Mol Cell 64(5):967–981

75. Butler JT et al (2009) Changing nuclear landscape and unique PML structures during early epigenetic transitions of human embryonic stem cells. J Cell Biochem 107(4):609–621

76. Peric-Hupkes D et al (2010) Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol Cell 38(4):603–613

77. Chen Q et al (2018) Long noncoding RNA NEAT1, regulated by the EGFR pathway, contributes to glioblastoma progression through the WNT/beta-catenin pathway by scaffolding EZH2. Clin Cancer Res 24(3):684–695

78. Zhao H, Jin Y (2017) Signaling networks in the control of pluripotency. Curr Opin Genet Dev 46:141–148

79. Ying QL, Smith A (2017) The art of capturing pluripotency: creating the right culture. Stem Cell Rep 8(6):1457–1464

80. Huang G et al (2015) Molecular basis of embryonic stem cell self-renewal: from signaling pathways to pluripotency network. Cell Mol Life Sci 72(9):1741–1757

81. Tu J et al (2018) Gas5 is an essential lncRNA regulator for self-renewal and pluripotency of mouse embryonic stem cells and induced pluripotent stem cells. Stem Cell Res Ther 9(1):71

82. Xu C et al (2016) Long non-coding RNA GAS5 controls human embryonic stem cell self-renewal by maintaining NODAL signalling. Nat Commun 7:13287

83. Daneshvar K et al (2016) DIGIT Is a conserved long noncoding rna that regulates GSC expression to control definitive endoderm differentiation of embryonic stem cells. Cell Rep 17(2):353–365

84. Sun Z et al (2018) The long noncoding RNA Lncenc1 maintains naive states of mouse ESCs by promoting the glycolysis pathway. Stem Cell Rep 11:741–755

85. Boffelli D, Nobrega MA, Rubin EM (2004) Comparative genomics at the vertebrate extremes. Nat Rev Genet 5(6):456–465

86. Diederichs S (2014) The four dimensions of noncoding RNA conservation. Trends Genet 30(4):121–123

87. Li L et al (2013) Targeted disruption of Hotair leads to homeotic transformation and gene derepression. Cell Rep 5(1):3–12

88. Nitsche A, Stadler PF (2017) Evolutionary clues in lncRNAs. Wiley Interdiscip Rev RNA 8(1)

89. Bejerano G et al (2004) Ultraconserved elements in the human genome. Science 304(5675):1321–1325

90. Rosenbloom KR et al (2015) The UCSC Genome Browser database: 2015 update. Nucleic Acids Res 43(Database issue):D670–D681

91. Casper J et al (2018) The UCSC Genome Browser database: 2018 update. Nucleic Acids Res 46(D1):D762–D769

92. Ryu T, Seridi L, Ravasi T (2012) The evolution of ultraconserved elements with different phylogenetic origins. BMC Evol Biol 12:236

93. Johnsson P et al (2014) Evolutionary conservation of long non-coding RNAs; sequence, structure, function. Biochim Biophys Acta 1840(3):1063–1071

94. Visel A et al (2008) Ultraconservation identifies a small subset of extremely constrained developmental enhancers. Nat Genet 40(2):158–160

95. Hezroni H et al (2017) A subset of conserved mammalian long non-coding RNAs are fossils of ancestral protein-coding genes. Genome Biol 18(1):162

96. Harmston N, Baresic A, Lenhard B (2013) The mystery of extreme non-coding conservation. Philos Trans R Soc Lond B Biol Sci 368(1632):20130021

97. Polychronopoulos D et al (2017) Conserved non-coding elements: developmental gene regulation meets genome organization. Nucleic Acids Res 45(22):12611–12624

98. Ahituv N et al (2007) Deletion of ultraconserved elements yields viable mice. PLoS Biol 5(9):e234

99. Dickel DE et al (2018) Ultraconserved enhancers are required for normal development. Cell 172(3):491–499 e15

100. Pennacchio LA et al (2013) Enhancers: five essential questions. Nat Rev Genet 14(4):288–295

101. Licastro D et al (2010) Promiscuity of enhancer, coding and non-coding transcription functions in ultraconserved elements. BMC Genom 11:151

102. Pennacchio LA et al (2006) In vivo enhancer analysis of human conserved non-coding sequences. Nature 444(7118):499–502

103. Braconi C et al (2011) Expression and functional role of a transcribed noncoding RNA with an ultraconserved element in hepatocellular carcinoma. Proc Natl Acad Sci USA 108(2):786–791

104. Calin GA et al (2007) Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. Cancer Cell 12(3):215–229

105. Fabbri M et al (2008) MicroRNAs and noncoding RNAs in hematological malignancies: molecular, clinical and therapeutic implications. Leukemia 22(6):1095–1105

106. Fassan M et al (2014) Transcribed ultraconserved noncoding RNAs (T-UCR) are involved in Barrett's esophagus carcinogenesis. Oncotarget 5(16):7162–7171

107. Galasso M et al (2014) A large scale expression study associates uc.283-plus lncRNA with pluripotent stem cells and human glioma. Genome Med 6(10):76

108. Goto K et al (2016) The transcribed-ultraconserved regions in prostate and gastric cancer: DNA hypermethylation and micro-RNA-associated regulation. Oncogene 35:3598–3606

109. Nan A et al (2016) A transcribed ultraconserved noncoding RNA, Uc.173, is a key molecule for the inhibition of lead-induced neuronal apoptosis. Oncotarget 7(1):112–124

110. Olivieri M et al (2016) Long non-coding RNA containing ultra-conserved genomic region 8 promotes bladder cancer tumorigenesis. Oncotarget 7:20636–20654

111. Lujambio A et al (2010) CpG island hypermethylation-associated silencing of non-coding RNAs transcribed from ultraconserved regions in human cancer. Oncogene 29(48):6390–6401

112. Magistri M et al (2012) Regulation of chromatin structure by long noncoding RNAs: focus on natural antisense transcripts. Trends Genet 28(8):389–396

113. Katzman S et al (2007) Human genome ultraconserved elements are ultraselected. Science 317(5840):915

114. Mattick JS, Makunin IV (2005) Small regulatory RNAs in mammals. Hum Mol Genet 14 Spec No 1:R121–R132

115. Nelson AC, Wardle FC (2013) Conserved non-coding elements and cis regulation: actions speak louder than words. Development 140(7):1385–1395

116. Sandelin A et al (2004) Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. BMC Genom 5(1):99

117. Zhou J et al (2017) Conserved expression of ultra-conserved non-coding RNA in mammalian nervous system. Biochim Biophys Acta 1860(12):1159–1168

118. Colasante G et al (2008) Arx is a direct target of Dlx2 and thereby contributes to the tangential migration of GABAergic interneurons. J Neurosci 28(42):10674–10686

119. Mehta SL, Dharap A, Vemuganti R (2014) Expression of transcribed ultraconserved regions of genome in rat cerebral cortex. Neurochem Int 77:86–93

120. Feng J et al (2006) 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(11):1470–1484

121. Berghoff EG et al (2013) Evf2 (Dlx6as) lncRNA regulates ultraconserved enhancer methylation and the differential transcriptional control of adjacent genes. Development 140(21):4407–4416

122. Mercer TR, Dinger ME, Mattick JS (2009) Long non-coding RNAs: insights into functions. Nat Rev Genet 10(3):155–159

123. Ramos AD et al (2013) Integration of genome-wide approaches identifies lncRNAs of adult neural stem cells and their progeny in vivo. Cell Stem Cell 12(5):616–628

124. Bernstein BE et al (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125(2):315–326

125. Lee TI et al (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. Cell 125(2):301–313

126. Fiorenzano A et al (2018) An ultraconserved element containing lncRNA preserves transcriptional dynamics and maintains ESC self-renewal. Stem Cell Rep 10(3):1102–1114

127. Radhakrishnan B, Alwin Prem Anand A (2016) Role of miRNA-9 in brain development. J Exp Neurosci 10:101–120

128. Flynn RA, Chang HY (2014) Long noncoding RNAs in cell-fate programming and reprogramming. Cell Stem Cell 14(6):752–761