Published in final edited form as:
*Mol Cell*. 2014 March 20; 53(6): 1005–1019. doi:10.1016/j.molcel.2014.01.021.

# An Evolutionarily Conserved Long Noncoding RNA *TUNA* Controls Pluripotency and Neural Lineage Commitment

**Nianwei Lin**[1], **Kung-Yen Chang**[1,7], **Zhonghan Li**[1,7], **Keith Gates**[2], **Zacharia A. Rana**[1], **Jason Dang**[1], **Danhua Zhang**[2], **Tianxu Han**[1], **Chao-Shun Yang**[1], **Thomas J. Cunningham**[3], **Steven R. Head**[4], **Gregg Duester**[3], **Duc Dong**[2], and **Tariq M. Rana**[1,5,6,*]

[1]Program for RNA Biology, Sanford-Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, California 92037

[2]Genetic Disease, Sanford-Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, California 92037

[3]Development and Aging, Sanford-Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, California 92037

[4]The Scripps Research Institute, Microarray and NGS Core Facility, 10550 North Torrey Pines Road, La Jolla, California 92037

[5]Department of Pediatrics, Rady Children's Hospital San Diego and University of California San Diego, La Jolla, California 92093

[6]Biomedical Sciences Graduate Program, University of California San Diego, La Jolla, California 92093

[7]These authors contributed equally to this work.

## Abstract

**SUMMARY—**Here, we generated the first genome-scale shRNA library targeting lincRNAs in the mouse. We performed an unbiased loss-of-function study in mouse embryonic stem cells (mESCs) and identified 20 novel lincRNAs involved in the maintenance of pluripotency. Among these, *TUNA* (*Tcl1* Upstream Neuron-Associated lincRNA), was required for pluripotency and formed a complex with three RNA-binding proteins (RBPs). The *TUNA*–RBP complex was detected at the promoters of *Nanog, Sox2*, and *Fgf4*, and knockdown of *TUNA* or the individual RBPs inhibited neural differentiation of mESCs. *TUNA* showed striking evolutionary conservation of both sequence and central nervous system-restricted expression in vertebrates. Accordingly, knockdown of *tuna* in zebrafish caused impaired locomotor function, and *TUNA* expression in the brains of Huntington's patients was significantly associated with disease grade. Our results



*Corresponding author trana@sanfordburnham.orgtrana@ucsd.edu.

**Conflicts of Interest Statement**
The authors declare no conflicts of interest.

**SUPPLEMENTAL INFORMATION**
Supplemental Information includes Extended Experimental Procedures, seven figures, eight tables, and two videos can be found with this article online at http:// XX.

suggest that the lincRNA *TUNA* plays a vital role in pluripotency and neural differentiation of ESCs and is associated with neurological function of adult vertebrates.

**Keywords**

long noncoding RNAs; stem cells; pluripotency; RNA-protein complexes; neurogenesis

## INTRODUCTION

The mammalian genome encodes thousands of long noncoding RNAs (lncRNAs, >200 nucleotides), a class of RNAs increasingly recognized as playing major roles in gene regulation (Lee, 2012; Rinn and Chang, 2012). Like coding mRNAs, lncRNAs are transcribed by RNA polymerase II, 5′-capped, spliced, and polyadenylated, but they lack protein-coding potential. Recent genomics studies have identified thousands of lncRNAs in the human and mouse genomes (Derrien et al., 2012; Guttman et al., 2009), but the vast majority have no known biological function. Many lncRNAs have been identified using the transcriptional profiling approach, which presumes that a cause-and-effect relationship exists between differential gene expression and specific cellular identities or responses to stimuli (Huarte et al., 2010; Loewer et al., 2010; Wang et al., 2011). For instance, several lncRNAs have been identified as pluripotency genes based on their embryonic stem cell (ESC)-specific expression profiles (Dinger et al., 2008). In addition, dozens of lincRNAs involved in the circuitry controlling the ESC ground state were identified in a recent systematic loss-of-function screen of the majority of mouse ESC (mESC)-specific intergenic lncRNA genes (Guttman et al., 2011). Thus, transcriptional profiling has proven to be a powerful tool for discovering lncRNAs with biological functions. Nevertheless, differential expression of any gene may be a consequence of a particular biological process rather than the cause, and it is often difficult to distinguish between these possibilities. For this reason, systematic and unbiased high-throughput functional screening strategies are urgently needed to identify and characterize biologically active lncRNAs.

Here, we created the first unbiased and genome-scale high-throughput shRNA library targeting 1280 lincRNAs in the mouse genome. We identified 20 novel lincRNAs that are required for the maintenance of the pluripotency and self-renewal capacity of mESCs. One of these, which we named *Tcl1* Upstream Neuron-Associated (*TUNA*), shows remarkable sequence conservation in vertebrates and is specifically expressed in the central nervous system (CNS) of zebrafish, mice, and humans. Manipulation of *TUNA* expression in mESCs affected global gene expression, with marked changes in genes involved in cell differentiation, proliferation, cell death, and neurogenesis. *TUNA* formed an RNA–multiprotein complex that was specifically enriched at the promoters of *Sox2, Nanog*, and *Fgf4*. Consistent with its neuronal expression and function, disruption of *TUNA* expression in zebrafish caused severe behavioral defects. Finally, we noted a significant association of human *TUNA* with neurodegeneration in Huntington's disease (HD). Together, our results indicate that *TUNA* is required for the maintenance of ESC stemness and neural lineage commitment, and the association of *TUNA* with HD suggests a link between lincRNAs and the pathophysiology of neurodegenerative diseases.

# RESULTS

## A genome-scale shRNA library targeting mouse lincRNAs

To systematically analyze lincRNAs in the mouse genome, we created an unbiased genome-scale lentiviral RNAi library targeting 1280 mouse intergenic lncRNA genes annotated in the Ensembl database. We designed at least three short hairpins (shRNAs) targeting each of the 1280 lincRNAs, which generated a library of 5656 shRNAs (Table S1). Sense and antisense oligonucleotides were annealed in 96-well plates and pooled for ligation into the pLKO.1-puro lentiviral vector (Figure 1A and Figure S1A). We assessed inaccuracies and bias during library construction by two independent approaches. First, the hairpin sequences were amplified by PCR and analyzed by deep sequencing. Of the 6,991,992 reads, we found 6,379,389 (91.2%) perfect matches to the reference hairpin sequences, representing 4740 unique sequences (83.8%; Table S2). With optimized PCR cycles, we noted relatively uniform library representation (Figure S1B). Second, we sequenced 189 individual clones from a small vector pool containing 96 different hairpins (Figure S1C). Of these, we identified 166 clones (87.8%) with perfect sequence matches, of which 75 (78.1%) were unique hairpins.

To test knockdown efficiency, we evaluated three shRNAs targeting each of the 13 pluripotency-related lincRNAs previously identified in mESCs (Guttman et al., 2011). Twelve of the 13 lincRNAs were effectively depleted by at least one shRNA, and six were significantly depleted (>60%) by two or three shRNAs (Figure 1B). Taken together, the high recovery rate, relatively uniform distribution, and efficient target knockdown substantiate the quality and coverage of the shRNA library.

## Identification of lincRNA genes essential for mESC identity

Pluripotent ESCs are characterized by their ability to self-renew and to differentiate into cells of the three primary germ layers. Although many genome-wide screens have been conducted to identify protein-coding and microRNA genes involved in the maintenance of ESC identity (Chia et al., 2010; Ding et al., 2009; Ivanova et al., 2006), relatively few studies have focused on lincRNA genes. In this study, we describe the first unbiased genome-scale screen to identify lincRNAs essential for ESC stemness.

To identify lincRNAs required for ESC pluripotency, Oct4-GFP mESCs were infected with lentiviruses expressing the entire validated shRNA library together with a nontargeting control shRNA. On day 4 post-infection, the GFP+ and GFP− cell populations were purified by FACS, and extracted DNA was analyzed by deep sequencing and reference mapping (Figure 1A and Figure S2A). We identified 3265 shRNAs in three biological GFP− replicates and 3115 shRNAs in three GFP+ replicates, with very similar recovery rates for all six samples. A set of 2788 shRNAs shared by all of the experimental replicates was analyzed to assess the relative enrichment in differentiated (GFP−) compared with undifferentiated (GFP+) cells (Figure S2B).

To identify candidate lincRNAs required for maintenance of ESC identity, the average number of GFP− and GFP+ cells expressing each shRNA was calculated and expressed as a ratio. The GFP−/GFP+ ratio for the control shRNA was 1.29, reflecting the random nature

of the differentiation process. Candidate lincRNAs were selected if at least two shRNAs were present at a GFP−/GFP+ ratio of >2.5, or if one shRNA was present at a ratio >3. To more stringently filter the latter group, we surveyed their surrounding genomic regions for pluripotency-related genes, given the positive correlation between the expression of lincRNAs and their neighboring protein-coding genes (Derrien et al., 2012). Based on these criteria, a total of 21 lincRNAs were selected (Table S3) for further evaluation and functional verification by knockdown with the individual shRNAs.

All 44 shRNAs tested in CCE mESCs were found to deplete their target lincRNAs, albeit with differing efficiencies (Figure 2A). Notably, knockdown of 20 of the 21 lincRNA candidates induced a differentiated cell phenotype, as shown by decreased GFP expression in Oct4-GFP mESCs (Figure 2B) and reduced alkaline phosphatase (AP) activity in CCE mESCs (Figure S2C). In addition, most of the shRNAs caused a significant reduction in cell number. Consistent with the high success rate recorded during library validation, loss of ESC identity was observed with 41 of the 44 tested shRNAs, with many showing effects comparable to that of the control *Oct4* shRNA (Figure S2C). Moreover, depletion of all 20 lincRNAs, each with two independent shRNAs, caused a significant decrease in expression of the pluripotency markers *Nanog* and *Oct4* (Figures 2C and 2D). Thus, our RNAi screen identified 20 novel lincRNAs required for maintenance of ESC pluripotency.

### *Linc86023* is required for maintenance of pluripotency

We selected *linc86023* (*2810011L19Rik*) for further analysis for several reasons. First, *linc86023* is located 113.8 kb upstream of the known pluripotency-related gene *Tcl1* (Figure 3A). Second, analysis of ChIP-seq data (ENCODE/LICR Ren Laboratory, Ludwig Institute for Cancer Research, UCSD) showed enrichment of H3K4me3 at the transcription start site and of H3K36me3 across the gene body of *linc86023* in E14 and Bruce4 mESC lines (Figure 3A). Third, *linc86023* shows a remarkable degree of sequence conservation in vertebrates (Figure 3A). Indeed, *linc86023* has a higher conservation level than *Tcl1*, despite the fact that lincRNA genes are generally less conserved than protein-coding genes (Derrien et al., 2012). This exceptional degree of conservation suggests a vital function for *linc86023* in vertebrates.

*Linc86023* is located on chromosome 12 (chr12:106,574,804-106,622,141, NCBI37/mm9) and is transcribed in the opposite direction to *Tcl1* (Figure 3A). Rapid amplification of cDNA ends (RACE) and northern blot analyses identified two ~3 kb alternatively spliced forms of *linc86023* (ENSMUST00000155481/BC059025, 3281 bp; ENSMUST00000138649/AK045952, 2876 bp) (Figure 3B, Figures S3A and S3B), and both isoforms could be depleted by shRNA (Figure S3C). Insertion of EGFP into the longest predicted ORF of *linc86023* did not result in detectable protein expression, indicating that *linc86023* is a *bona fide* noncoding RNA gene (Figures S3D-F). Finally, localization of *linc86023* RNA was found in both nuclei and cytoplasm (Figures 3C and 3D).

To confirm a role for *linc86023* in regulating ESC identity, we depleted expression in CCE cells with three independent shRNAs, and in each case observed altered cell morphology, loss of AP-positive colonies, and reduced transcription of pluripotency-related genes (Figures 3E and 3F, Figure S3G). In addition, depletion of *linc86023* caused impaired cell

proliferation, whereas overexpression was associated with elevated levels of proliferation (Figures 3G and 3H), suggesting that *linc86023* plays a role in regulating the cell cycle and proliferation. Consistent with this, we found decreased levels of several positive regulators of the cell cycle and increased expression of negative regulators in cells with *linc86023* depleted (Figure S3H). *Linc86023* depletion also led to significant changes in the expression of lineage-specific differentiation markers, including downregulation of the neuroectoderm markers *Pax6* and *Sox1* (Figure S3I). Finally, *linc86023* overexpression moderately increased levels of *Nanog* and *Oct4* mRNA in CCE cells and increased the number of iPSC colonies derived from reprogrammed Oct4-GFP mouse embryonic fibroblasts (MEFS) (Figures S3J and S3K). Collectively, these data confirm the vital role of *linc86023* in maintaining ESC self-renewal and pluripotency.

## *Linc86023* (*Tcl1* Upstream Neuron-Associated; *TUNA)* is evolutionarily conserved and specifically expressed in the central nervous system

The largest exon of *linc86023* contains a highly conserved region of ~200 bp that is present in all annotated vertebrate genomes (Figure 4A). Mouse *linc86023* shows 88% and 81% sequence identity in this region to the human and zebrafish orthologs, respectively. The human ortholog, *LINC00617*, is located on chromosome 14 and is actively transcribed in H1 human ESCs (Figure S4A). To determine whether this region is essential for maintenance of ESC identity, we generated expression constructs containing a 225 bp fragment of *linc86023* (encompassing the conserved region) or the full-length gene with this region deleted. When transfected into Oct4-GFP MEFs, the conserved region was as effective as the full-length *linc86023* in generating GFP+ iPSC colonies (Figure S4B), whereas the mutant was virtually ineffective. These data therefore confirm that the ~200 bp highly conserved region of *linc86023* contains a functional motif that regulates the pluripotent state.

Interestingly, this region of *linc86023* is present in lampreys, the most primitive living vertebrate, but not in lancelets (Figure 4A), the closest invertebrate relative of vertebrates. In contrast to vertebrates, the lancelet nervous system consists of an unprotected dorsal nerve cord that extends into the head without forming a true brain. These observations therefore raised the possibility that *linc86023* might play an important role in the vertebrate CNS. To test this, we examined *linc86023* expression in 15 mouse tissues by qRT-PCR. Notably, *linc86023* was highly expressed in the brain and spinal cord, moderately expressed in the eye, and virtually absent from all other adult tissues (Figure 4B). Analysis of RNA-seq data from a broader range of tissues (Jiang et al., 2011) confirmed the CNS-restricted expression pattern of *linc86023*, with robust CNS expression evident at embryonic stage E14 (Figure S4C). This expression pattern was also observed in humans, where the highest levels were detected in the brain and moderate levels were seen in the testis (Figure S4D). Finally, the conserved CNS-specific expression of *linc86023* was confirmed by *in situ* hybridization of mouse and zebrafish embryos (Figures 4C and 4D). Because of the striking evolutionary conservation of lincRNA sequence and CNS-specific expression pattern, we named this lincRNA *TUNA*, for *Tcl1* Upstream Neuron-Associated lincRNA.

### *TUNA* is required for neural differentiation and function

Because *TUNA* displays CNS-specific expression, we asked if it might play a role in neural differentiation. To test this, we first examined *TUNA* expression in monolayer neural differentiation cultures of mESCs. Indeed, transcription of *TUNA* was greatly increased within 4 days of culture (Figure 5A) preceding the appearance of the neural stem cell marker *Nestin* on day 6. These results suggest that *TUNA* may play a crucial role in the initial phase of neural commitment of ESCs. Consistent with this, depletion of *TUNA* decreased the expression of *Nestin* and other neural progenitor cell markers such as *Sox1*, *Fgf4*, and *Zpf521* (Figure 5B). In a control experiment, *TUNA* was repressed during ESC differentiation towards the mesoderm lineage *in vitro* (Figure S5A). To investigate the global effect of *TUNA* on gene expression, we performed RNA-seq analysis at various time points after *TUNA* knockdown in mESCs. We found 990 genes with ≥3-fold difference in expression in cells treated with *TUNA* shRNA versus control shRNA, of which 530 genes were upregulated and 460 were downregulated (Figure 5C and Table S4). Notably, the upregulated clusters were enriched for genes involved in cellular development and neuronal apoptosis, and conversely, downregulated clusters were enriched for genes involved in neural tissue development, neural differentiation, cell proliferation, and neuronal recognition (Figure 5D).These results indicate that *TUNA* is induced during neural differentiation of ESCs, and accordingly, depletion of *TUNA* has a global effect on genes involved in neural lineage commitment.

To validate the role of *TUNA* in neural fate commitment, we performed shRNA-mediated knockdown of *TUNA* during monoculture of cells in a defined neural differentiation medium lacking serum or leukemia inhibitory factor. shRNA transduction was performed 2 days after the initiation of neural differentiation to bypass the effects on pluripotency. Consistent with the gene expression data, CCE mESCs treated with control shRNA displayed overt neuronal morphology and increased expression of Nestin (neural precursor cells) and Tuj1 (neurons) by day 7 (Figures 5E and 5F). In contrast, *TUNA*-depleted ESCs failed to differentiate and showed no expression of either Nestin or Tuj1, demonstrating that *TUNA* is required for neural differentiation of mESCs *in vitro*. To determine if *TUNA* is also functional in human neurons, we examined H3K4me3 ChIP-Seq data for cells collected from the prefrontal cortexes of a human child (4 years of age) and adult (67 years of age) [UMMS Brain Histone (Akbarian/Weng) UCSC track, GRCh37/hg19]. We found that H3K4me3 was specifically enriched at the *hTUNA* locus in neuronal, but not non-neuronal, cells from the same brain (Figure 5G). Moreover, knockdown of *hTUNA* (*LINC00617*) also blocked neural differentiation in H9 hES cells (Figure 5H). These results suggest that *hTUNA*/*LINC00617* is an important regulator during neurogenesis in human.

Finally, to determine if *TUNA* is functional in the CNS of zebrafish, we examined the effects of *tuna*-specific MOs on the locomotory response. We designed three morpholino antisense oligonucleotides (MOs) targeting the conserved region of *tuna*. MOs inhibit gene function by blocking interactions between the target site and cellular factors. For this analysis, embryos were injected with a low dose of MOs (1 ng) that did not induce developmental defects. Day 3 zebrafish larvae treated with *tuna* MOs showed greatly impaired locomotor function in touch response tests (Figures S5B and S5C and Supplemental Movies). Although

this phenotype could be due to muscular or neuronal defects (Granato et al., 1996), we did not observe obvious skeletal muscle defects in the treated embryos (Figure S5D), suggesting that the abnormal behavioral phenotype was most likely due to impaired CNS function. Future work to generate and examine *tuna* mutants will be necessary to complement our findings. Taken together, these data indicate that *TUNA* plays an essential role in neural development and function of zebrafish, mice, and human.

## *TUNA* functions by interacting with the RNA-binding proteins PTBP1, hnRNP-K, and nucleolin

We next sought to investigate the molecular mechanisms by which *TUNA* mediates its effects on ESC pluripotency. Many lincRNAs have been reported to regulate gene expression by interacting with transcription factors or chromatin-modifying complexes (Khalil et al., 2009). Because *TUNA* is enriched in the nuclear fraction of CCE cells (Figure 3D), we hypothesized that it may function through a similar mechanism.

To test this, we performed RNA pulldown experiments by incubating nuclear extracts from CCE cells with *in vitro*-synthesized biotinylated *TUNA* RNA or control *lacZ* RNA. The RNA-protein complexes were collected and resolved by SDS-PAGE. Silver staining of the gel revealed *TUNA*-specific pulldown of two bands, which were excised and analyzed by mass spectrometry (Figure 6A). Four candidate proteins were identified: polypyrimidine tract-binding protein (PTBP1/PTB/hnRNP-I), heterogeneous nuclear ribonucleoprotein K (hnRNP-K), nucleolin (NCL), and non-POU-domain-containing, octamer-binding protein (NONO). Western blot analysis identified PTBP1, hnRNP-K, and NCL as specifically pulled down with biotinylated *TUNA* RNA, but not with the control *lacZ* RNA (Figure 6B). Interestingly, murine *Ptbp1* and *Ncl* and human *hnRNP-K* have previously been identified as candidate pluripotency-associated genes (Chia et al., 2010; Ding et al., 2009).

To confirm the specificity of the *TUNA*–RBP interactions, we performed RNA immunoprecipitation (RIP) using antibodies against PTBP1 and hnRNP-K. NCL was not analyzed because antibodies suitable for RIP were not available. RNA-protein complexes were precipitated from nuclear extracts of crosslinked CCE cells and the extracted RNA was analyzed by agarose gel electrophoresis or qRT-PCR (Figures S6A and S6B). PTBP1 and hnRNP-K RIPs showed significant enrichment of *TUNA* RNA, but not *Actb* mRNA, compared to the control IgG RIPs, confirming the specificity of *TUNA* binding to these RBPs.

Next, we analyzed the function of the *TUNA*-associated RBPs in maintaining mESC pluripotency by shRNA-mediated knockdown of each protein separately. As expected, depletion of PTBP1, hnRNP-K, and NCL abolished ESC colony formation (Figure 6C) and decreased the expression of pluripotency and neural precursor markers (Figures 6D-6F). The number of cells in these cultures was also reduced, consistent with the reported roles of PTBP1, hnRNP-K, and NCL in regulation of the cell cycle, proliferation, and cell death (Moumen et al., 2005; Ohno et al., 2011; Srivastava and Pollard, 1999). Furthermore, gene expression and bioinformatics analyses identified many common genes coregulated by *TUNA* and three associated RBPs (Figure 6G). 74.2% of the genes affected by shTUNA showed altered expression levels in at least one of the three RBPs knockdown experiments,

and 21 genes showed changed expression in all 4 experiments (Table S5). Moreover, depletion of these RBPs also inhibited differentiation of CCE cells into the neural lineage (Figure S6C), suggesting that PTBP1, hnRNP-K, and NCL may mediate the effects of *TUNA* on ESC pluripotency and neurogenesis.

To determine whether PTBP1, hnRNP-K, and NCL interact with *TUNA* independently or as a multiprotein complex, we immunoprecipitated each RBP from CCE lysates treated with an RNase inhibitor or RNase, and then examined the immunoprecipitates by western blotting (Figures 6H-6J). We found that PTBP1, hnRNP-K, and NCL coimmunoprecipitated from control lysates treated with the RNase inhibitor but the interactions were abolished by treatment of lysates with RNase (Figures 6H-6J), suggesting that PTBP1, hnRNP-K, and NCL exist as multiprotein complexes with RNA, including *TUNA, in vivo*.

Recent studies have suggested that the conserved motifs in lincRNAs may serve as functional units to modulate RNA–protein or RNA–DNA interactions (Chu et al., 2011; Tsai et al., 2010). We therefore asked whether the highly conserved ~200 bp region of *TUNA* might mediate its interaction with PTBP1, hnRNP-K, and NCL. To test this, we performed RNA pulldown assays by incubating crosslinked lysates with three biotinylated constructs of *TUNA*: the full-length RNA, a fragment containing the conserved sequence alone (*TUNA-con*), or *TUNA* lacking the conserved region (*TUNA-mut*). Biotinylated *lacZ* RNA served as a control. We found that the conserved region of *TUNA* bound PTBP1, hnRNP-K, and NCL with an affinity comparable to that of the full-length RNA (Figure 6K), whereas pulldown by the construct lacking the conserved region was markedly less efficient for PTBP1 and hnRNP-K, but not NCL. Thus, we conclude that the highly conserved sequence of *TUNA* is required for the interaction with PTBP1 and hnRNP-K.

### *TUNA* mediates recruitment of PTBP1, hnRNP-K, and NCL to the *Sox2* promoter

LncRNAs are thought to modulate gene expression in part by recruiting chromatin-modifying complexes and transcription factors to target gene promoters (Chu et al., 2011; Tsai et al., 2010). To identify possible targets for the *TUNA*–RBP complex, we performed chromatin immunoprecipitation (ChIP) assays with anti-hnRNP-K antibodies and analyzed binding at the promoters of several pluripotency and neural stem cell marker genes shown to be repressed in *TUNA*-depleted cells. In extracts of cells transduced with the control shRNA, we found significant enrichment of hnRNP-K at the promoters of *Nanog*, *Sox2*, and *Fgf4*, but not at the control *GAPDH* promoter or an intergenic region (Figure 7A). Notably, the active histone mark H3K4me3 was also enriched at these promoters. In *TUNA* knockdown cells, binding of hnRNP-K and H3K4me3 at *Nanog*, *Sox2*, and *Fgf4* promoters was markedly decreased (Figure 7A), consistent with the reduced expression of these genes in *TUNA*-depleted cells (Figures 3D and S3F). These data suggest that *Nanog*, *Sox2*, and *Fgf4* are direct targets of *TUNA*.

To confirm this, we examined *TUNA* occupancy at the *Nanog*, *Sox2*, and *Fgf4* promoters by chromatin isolation by RNA purification (ChIRP) (Chu et al., 2011). Consistent with our ChIP data, the promoter of *Nanog*, *Sox2*, and *Fgf4* were significantly enriched in *TUNA*-ChIRP samples compared with the *lacZ* RNA controls (Figure 7B). These data confirm that

*TUNA* physically binds to the *Nanog*, *Sox2*, and *Fgf4* promoters and activates transcription by recruiting the multiprotein complex containing PTBP1, hnRNP-K, and NCL (Figure 7G).

Among the *TUNA*-targeted genes, *Sox2* is particular interesting. *Sox2* is highly expressed in pluripotent ESCs and neural precursor cells, where it plays a critical role in establishing and maintaining pluripotency (Takahashi and Yamanaka, 2006) and in neurogenesis (Bergsland et al., 2011). Moreover, microarray analysis of 13 human brain regions (Human Brain Atlas Microarrays, Sestan Lab, Yale University) showed coexpression of *TUNA* and *Sox2* in the hippocampus, striatum, thalamus, and cerebellum, but not in neocortical areas (Figure 7C). Both genes were expressed most highly in the striatum and thalamus, which consist primarily of neural cell bodies. This is consistent with the specific expression of *TUNA* in neuronal cells in human brains (Figure 5G).

Such highly coordinated expression of *TUNA* and *Sox2* suggests that they may target a common set of genes. To test this, we analyzed gene expression in *TUNA* shRNA-treated mESCs and compared the results with a previous analysis of gene expression in *Sox2* knockdown mESCs (Hutchins et al., 2013). Remarkably, 562 genes were found to be modulated by both *TUNA* and *Sox2* (Figure 7D). This gene set showed highly significant enrichment of genes involved in development, differentiation, neurogenesis, proliferation, and neuronal cell death (Figure 7D). These data suggest that *TUNA* and *Sox2* may control the ESC state and neurogenesis by regulating a common set of genes. Remarkably, overexpression of Sox2 was sufficient to partially rescue the shTUNA-mediated neural differentiation phenotype (Figure 7E and Figure S7A).

### *TUNA* is associated with Huntington's disease

The strong spatial and cell-type restricted pattern of *TUNA* expression in the brain prompted us to ask whether *TUNA* might play a role in neurodegenerative diseases, which characteristically affect neurons in discrete brain regions. Indeed, analysis of the genes affected by *TUNA* depletion in CCE cells identified marked changes in numerous genes linked to human neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, and HD (Figure S7B). Some of the most marked changes were observed in genes associated with HD, an autosomal dominant disorder that typically causes death within 20 years of the onset of motor, cognitive, and psychiatric symptoms.

The earliest and most severe neuronal damage in HD occurs in the striatum, a component of the basal ganglia, which functions as a relay station for communication between the limbic system and the frontal lobe (Walker, 2007). Of note, we found that *TUNA* was most highly expressed in the thalamus and striatum (Figure 7C), supporting a possible association with this disease. To examine this further, we retrieved data from a gene expression study of four regions of the brains of 44 HD patients and 36 unaffected subjects (Hodges et al., 2006). Neuropathological staging of HD was rated from Grade 0 to Grade 4 based on the macroscopic appearance of the brain and loss of neurons in the head of the caudate nucleus, the most affected area in the striatum (Vonsattel et al., 1985). Intriguingly, *hTUNA* expression was significantly associated with pathological disease severity, decreasing significantly as the disease grade increased (Figure 7F). In contrast, *hTUNA* expression in

the motor cortex, prefrontal association cortex, and cerebellum was not affected by the disease grade (Figure S7C). To ensure that the decreased expression of *hTUNA* was not simply due to neuronal cell death, we analyzed the expression of the neural marker gene *Neurod1* in the same HD and control brains. In contrast to *hTUNA*, we found no disease stage-related change in *Neurod1* expression, indicating that the same number of neurons was evaluated at each stage (Figure S7D). Taken together, these results suggest that deregulation of *hTUNA* in the caudate nucleus may be involved in the pathophysiology of HD.

## Discussion

Advances in high resolution microarray and next-generation sequencing technology led to the discovery of thousands of short and long ncRNAs. Current data from the ENCODE consortium suggest that as much as 75% of the human genome may be transcribed, and >9640 lncRNA loci have been identified to date. Nevertheless, the biological roles of only ~100 lncRNAs have been characterized, and it remains unclear whether some or all of the remaining lncRNAs are biologically active (Derrien et al., 2012; Djebali et al., 2012). Most of the lncRNAs with known functional roles were identified by transcriptional profiling of different cell types. One drawback to this approach is that it does not distinguish between causative and consequential changes in gene expression. To overcome this, we generated the first unbiased high-throughput shRNA library targeting 1280 lincRNAs in the mouse genome. We achieved a high recovery rate (~80%), relatively uniform distribution, and effective knockdown with our shRNA library. We believe this library will allow genome-wide RNAi screens to be performed in various biological systems and disease models and will thus greatly improve our understanding of the roles of lincRNAs in an array of cell- and behavior-specific regulatory networks.

Understanding the molecular events required for ESCs to maintain a balance between pluripotency and lineage commitment is crucial to advance the use of stem cell-based therapies in regenerative medicine. Although many genome-wide screens have been conducted to identify protein-coding and microRNA genes that maintain the self-renewal and differentiative capacity of ESCs (Chia et al., 2010; Ding et al., 2009; Ivanova et al., 2006), the search for lincRNAs with similar functions is still in its infancy (Dinger et al., 2008; Guttman et al., 2011; Loewer et al., 2010). To avoid possible bias introduced by library construction, ESC transduction, and PCR amplification, we identified enriched shRNAs in differentiated versus undifferentiated cells. Of the 21 lincRNA candidates that satisfied the selection criteria, 20 were functionally verified by demonstrating characteristic changes in cell morphology, AP activity, and marker gene expression. This screening method showed remarkable efficacy and a very low false positive rate. We found that knockdown of *TUNA* resulted in loss of pluripotency and disruption of global gene expression in mESCs, and many of the affected genes are involved in controlling the cell cycle and proliferation. Consistent with this, mESC proliferation was decreased by depletion of *TUNA* whereas *TUNA* overexpression promoted proliferation. These findings suggest that *TUNA* may influence the cell cycle regulatory network of mESCs, a possibility consistent with the known involvement of the cell cycle machinery in the establishment or/and maintenance of the stem cell state (White and Dalton, 2005).

Many lncRNAs contribute to the epigenetic regulation of gene expression by serving as modular scaffolds for histone modification complexes (Tsai et al., 2010). We found that *TUNA* interacts with three previously identified multifunctional proteins; PTBP1, hnRNP-K, and NCL, each of which has been implicated in the maintenance of ESC pluripotency (Chia et al., 2010; Ding et al., 2009). One of the functions ascribed to hnRNP-K is the temporal control of neural differentiation through post-transcriptional regulation of *p21* mRNA (Yano et al., 2005). The nucleolar phosphoprotein NCL is highly expressed in proliferating cells, where it functions in chromatin remodeling (Angelov et al., 2006) and transcription (Dempsey et al., 1999; Yang et al., 1994). Of note, there is evidence that NCL regulates the cell cycle, apoptosis, and maintenance of stemness in ESCs (Srivastava and Pollard, 1999; Yang et al., 2011). Finally, PTBP1 has been implicated in cell cycle regulation and neural differentiation, predominantly through post-transcriptional mechanisms (Ohno et al., 2011; Zheng et al., 2012). The diverse roles of these proteins suggest that the *TUNA*–RBP complex may regulate gene expression through multiple mechanisms.

Collectively, our demonstration of the association of *TUNA* with ESC pluripotency, neural differentiation, and HD progression suggests a new layer of complexity in the networks controlling stem cell biology and disease pathophysiology.

## EXPERIMENTAL PROCEDURES

### shRNA library construction and mapping

At least three short hairpins were designed for each of the 1280 annotated lincRNAs in the mouse genome (Ensembl release 61, February 2011). A total of 5656 shRNAs (Table S1) were cloned into the vector (Moffat et al., 2006). Deep sequencing of the 107 bp amplicons was performed using the Illumina HiSeq 2000 system. For details, see Supplementary Experimental Procedures.

### Lentiviral library preparation

The lentiviral library was prepared as previously described with some modifications (Moffat et al., 2006). For details, see Supplementary Experimental Procedures.

### Identification of putative pluripotency-associated lincRNAs

Three biological replicates of Oct4-GFP mESCs were transduced with the shRNA library (including a control shRNA) and cultured for 4 days. The cells were then harvested for fluorescence-activated cell sorting (FACS). Genomic DNA was extracted from GFP+ and GFP− cells, and sequences of integrated shRNAs were analyzed. For details, see Supplementary Experimental Procedures.

### *In situ* hybridization (ISH) of mouse embryos

Whole mount ISH of mouse embryos was performed as previously described, with some modifications (Wilkinson, 1992). Wild-type mouse embryos at E13.5 were fixed overnight in 4% PFA at 4°C. For details, see Supplementary Experimental Procedures.

### RNA pulldown assay and mass spectrometry

RNA pulldown experiments were performed as described previously (Rinn et al., 2007) with some modifications. The complexes were eluted, resolved by SDS-PAGE gel, and silver stained with Silver Stain Plus (Bio-Rad). Bands specifically pulled down by biotinylated *TUNA* were excised from the gel, digested, and analyzed by mass spectrometry. For details, see Supplementary Experimental Procedures.

### Chromatin isolation by RNA purification (ChIRP)

ChIRP experiments were performed using a protocol described previously (Chu et al., 2011). For details, see Supplementary Experimental Procedures.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank Dr. Howard Chang and Ci Chu for help and guidance with the ChIRP experiments. We are grateful to David Corey and Jeanne Lawrence for advice in FISH experiments and Alysson R. Muotri for help in neuronal differentiation of hES cells. We are thankful to Sanford-Burnham Medical Research Institute Genomics and Informatics and Data Management shared resource for RNA array experiments and RNA deep sequencing data analysis. We also thank Khatereh Motamedchaboki of the Proteomics Core facility for identification of RNA-associated proteins. This work was supported in part by grants from the National Institutes of Health.

## REFERENCES

Angelov D, Bondarenko VA, Almagro S, Menoni H, Mongelard F, Hans F, Mietton F, Studitsky VM, Hamiche A, Dimitrov S, et al. Nucleolin is a histone chaperone with FACT-like activity and assists remodeling of nucleosomes. EMBO J. 2006; 25:1669–1679. [PubMed: 16601700]

Bergsland M, Ramskold D, Zaouter C, Klum S, Sandberg R, Muhr J. Sequentially acting Sox transcription factors in neural lineage development. Genes Dev. 2011; 25:2453–2464. [PubMed: 22085726]

Buckley NJ, Johnson R, Zuccato C, Bithell A, Cattaneo E. The role of REST in transcriptional and epigenetic dysregulation in Huntington's disease. Neurobiol Dis. 2010; 39:28–39. [PubMed: 20170730]

Chia NY, Chan YS, Feng B, Lu X, Orlov YL, Moreau D, Kumar P, Yang L, Jiang J, Lau MS, et al. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature. 2010; 468:316–320. [PubMed: 20953172]

Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. Mol Cell. 2011; 44:667–678. [PubMed: 21963238]

Dee CT, Hirst CS, Shih YH, Tripathi VB, Patient RK, Scotting PJ. Sox3 regulates both neural fate and differentiation in the zebrafish ectoderm. Dev Biol. 2008; 320:289–301. [PubMed: 18572157]

Dempsey LA, Sun H, Hanakahi LA, Maizels N. G4 DNA binding by LR1 and its subunits, nucleolin and hnRNP D, A role for G-G pairing in immunoglobulin switch recombination. J Biol Chem. 1999; 274:1066–1071. [PubMed: 9873052]

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 2012; 22:1775–1789. [PubMed: 22955988]

Ding L, Paszkowski-Rogacz M, Nitzsche A, Slabicki MM, Heninger AK, de Vries I, Kittler R, Junqueira M, Shevchenko A, Schulz H, et al. A genome-scale RNAi screen for Oct4 modulators

defines a role of the Paf1 complex for embryonic stem cell identity. Cell Stem Cell. 2009; 4:403–415. [PubMed: 19345177]

Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, Askarian-Amiri ME, Ru K, Solda G, Simons C, et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res. 2008; 18:1433–1445. [PubMed: 18562676]

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al. Landscape of transcription in human cells. Nature. 2012; 489:101–108. [PubMed: 22955620]

Granato M, van Eeden FJ, Schach U, Trowe T, Brand M, Furutani-Seiki M, Haffter P, Hammerschmidt M, Heisenberg CP, Jiang YJ, et al. Genes controlling and mediating locomotion behavior of the zebrafish embryo and larva. Development. 1996; 123:399–413. [PubMed: 9007258]

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature. 2009; 458:223–227. [PubMed: 19182780]

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature. 2011; 477:295–300. [PubMed: 21874018]

Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C, et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol. 2010; 28:503–510. [PubMed: 20436462]

Hodges A, Strand AD, Aragaki AK, Kuhn A, Sengstag T, Hughes G, Elliston LA, Hartog C, Goldstein DR, Thu D, et al. Regional and cellular gene expression changes in human Huntington's disease brain. Hum Mol Genet. 2006; 15:965–977. [PubMed: 16467349]

Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell. 2010; 142:409–419. [PubMed: 20673990]

Hutchins AP, Choo SH, Mistri TK, Rahmani M, Woon CT, Keow Leng Ng C, Jauch R, Robson P. Co-Motif Discovery Identifies an Esrrb-Sox2-DNA Ternary Complex as a Mediator of Transcriptional Differences Between Mouse Embryonic and Epiblast Stem Cells. Stem Cells. 2013; 31:269–281. [PubMed: 23169531]

Ivanova N, Dobrin R, Lu R, Kotenko I, Levorse J, DeCoste C, Schafer X, Lun Y, Lemischka IR. Dissecting self-renewal in stem cells with RNA interference. Nature. 2006; 442:533–538. [PubMed: 16767105]

Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M, Gingeras TR, Oliver B. Synthetic spike-in standards for RNA-seq experiments. Genome Res. 2011; 21:1543–1551. [PubMed: 21816910]

Keppetipola N, Sharma S, Li Q, Black DL. Neuronal regulation of pre-mRNA splicing by polypyrimidine tract binding proteins, PTBP1 and PTBP2. Crit Rev Biochem Mol Biol. 2012; 47:360–378. [PubMed: 22655688]

Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc Natl Acad Sci U S A. 2009; 106:11667–11672. [PubMed: 19571010]

Kim K, Choi J, Heo K, Kim H, Levens D, Kohno K, Johnson EM, Brock HW, An W. Isolation and characterization of a novel H1.2 complex that acts as a repressor of p53-mediated transcription. J Biol Chem. 2008; 283:9113–9126. [PubMed: 18258596]

Lee JT. Epigenetic regulation by long noncoding RNAs. Science. 2012; 338:1435–1439. [PubMed: 23239728]

Loewer S, Cabili MN, Guttman M, Loh YH, Thomas K, Park IH, Garber M, Curran M, Onder T, Agarwal S, et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat Genet. 2010; 42:1113–1117. [PubMed: 21057500]

Matsui M, Prakash TP, Corey DR. Transcriptional silencing by single-stranded RNAs targeting a noncoding RNA that overlaps a gene promoter. ACS chemical biology. 2013; 8:122–126. [PubMed: 23082936]

Moumen A, Masterson P, O'Connor MJ, Jackson SP. hnRNP K: an HDM2 target and transcriptional coactivator of p53 in response to DNA damage. Cell. 2005; 123:1065–1078. [PubMed: 16360036]

Ohno S, Shibayama M, Sato M, Tokunaga A, Yoshida N. Polypyrimidine tract-binding protein regulates the cell cycle through IRES-dependent translation of CDK11(p58) in mouse embryonic stem cells. Cell Cycle. 2011; 10:3706–3713. [PubMed: 22037210]

Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annu Rev Biochem. 2012; 81:145–166. [PubMed: 22663078]

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell. 2007; 129:1311–1323. [PubMed: 17604720]

Srivastava M, Pollard HB. Molecular dissection of nucleolin's role in growth and cell proliferation: new insights. FASEB J. 1999; 13:1911–1922. [PubMed: 10544174]

Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006; 126:663–676. [PubMed: 16904174]

Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY. Long noncoding RNA as modular scaffold of histone modification complexes. Science. 2010; 329:689–693. [PubMed: 20616235]

Vonsattel JP, Myers RH, Stevens TJ, Ferrante RJ, Bird ED, Richardson EP Jr. Neuropathological classification of Huntington's disease. J Neuropathol Exp Neurol. 1985; 44:559–577. [PubMed: 2932539]

Walker FO. Huntington's disease. Lancet. 2007; 369:218–228. [PubMed: 17240289]

Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature. 2011; 472:120–124. [PubMed: 21423168]

White J, Dalton S. Cell cycle control of embryonic stem cells. Stem Cell Rev. 2005; 1:131–138. [PubMed: 17142847]

Wilkinson, DG. In situ hybridization: A Practical Approach. IRL Press; Oxford: 1992. Whole mount in situ hybridization of vertebrate embryos.

Yang A, Shi G, Zhou C, Lu R, Li H, Sun L, Jin Y. Nucleolin maintains embryonic stem cell self-renewal by suppression of p53 protein-dependent pathway. J Biol Chem. 2011; 286:43370–43382. [PubMed: 22013067]

Yang TH, Tsai WH, Lee YM, Lei HY, Lai MY, Chen DS, Yeh NH, Lee SC. Purification and characterization of nucleolin and its identification as a transcription repressor. Mol Cell Biol. 1994; 14:6068–6074. [PubMed: 8065340]

Yano M, Okano HJ, Okano H. Involvement of Hu and heterogeneous nuclear ribonucleoprotein K in neuronal differentiation through p21 mRNA post-transcriptional regulation. J Biol Chem. 2005; 280:12690–12699. [PubMed: 15671036]

Zheng S, Gray EE, Chawla G, Porse BT, O'Dell TJ, Black DL. PSD-95 is post-transcriptionally repressed during early neural development by PTBP1 and PTBP2. Nat Neurosci. 2012; 15:381–388. S381. [PubMed: 22246437]

## Highlights

- Genome-wide RNAi screen identified 20 novel lincRNAs controlling pluripotency.
- LincRNA *TUNA* is required for pluripotency and neural differentiation.
- *TUNA* interacts with RNA-binding proteins through a conserved sequence.
- *TUNA* expression in Huntington's patients was associated with disease grade.

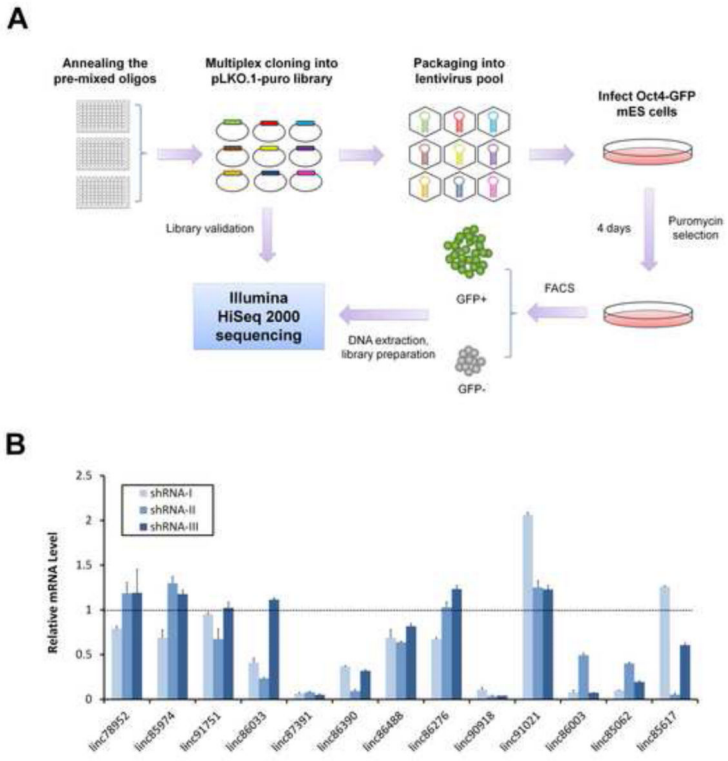


**Figure 1. Construction and validation of a genome-scale shRNA library for mouse lincRNAs**

(A) Pairs of pre-mixed DNA oligonucleotides were annealed in 96-well plates, pooled, and cloned into pLKO.1-puro lentiviral vector. Pooled shRNA plasmids were validated by both Sanger sequencing and deep sequencing. Oct4-GFP mESCs were transduced with the lentiviral library for 24 h and then selected with puromycin for an additional 4 days. Undifferentiated (GFP +) and differentiated (GFP−) cells were sorted by FACS and DNA was extracted for deep sequencing analysis.

(B) Knockdown efficiency was determined by depletion of 13 lincRNAs known to be expressed in mESCs. CCE mESCs were transduced with three lincRNA-specific shRNAs, and lincRNA expression was analyzed by qRT-PCR. Transcript levels were normalized to *Actb* mRNA. Data are the mean ± SD of triplicates. See also Figure S1 and Table S2.

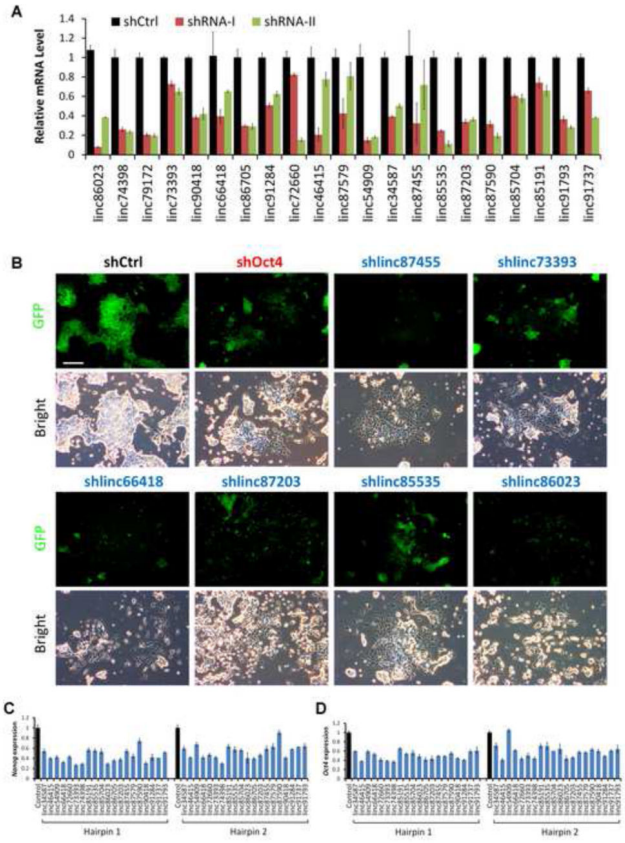


**Figure 2. Identification of lincRNAs involved in maintenance of mESC self-renewal and pluripotency**

(A) Knockdown of lincRNA transcription in CCE mESCs 4 days after infection. Transcript levels were measured by qRT-PCR and normalized to *Actb* mRNA levels. Data are the mean ± SD of triplicates.

(B) Fluorescence (upper panels) and bright-field (lower panels) micrographs of Oct4-ESCs after infection with a nontargeting control shRNA or shRNAs targeting *Oct4* and six novel lincRNAs. Scale bars, 100 μm.

(C and D) Relative expression of *Nanog* (C) and *Oct4* (D) mRNA after knockdown of 20 selected lincRNAs, each with two shRNAs (Hairpin 1 and Hairpin 2). Expression levels were normalized to *Actb*. Data are the mean ± SD of triplicates. See also Figure S2 and Table S3.

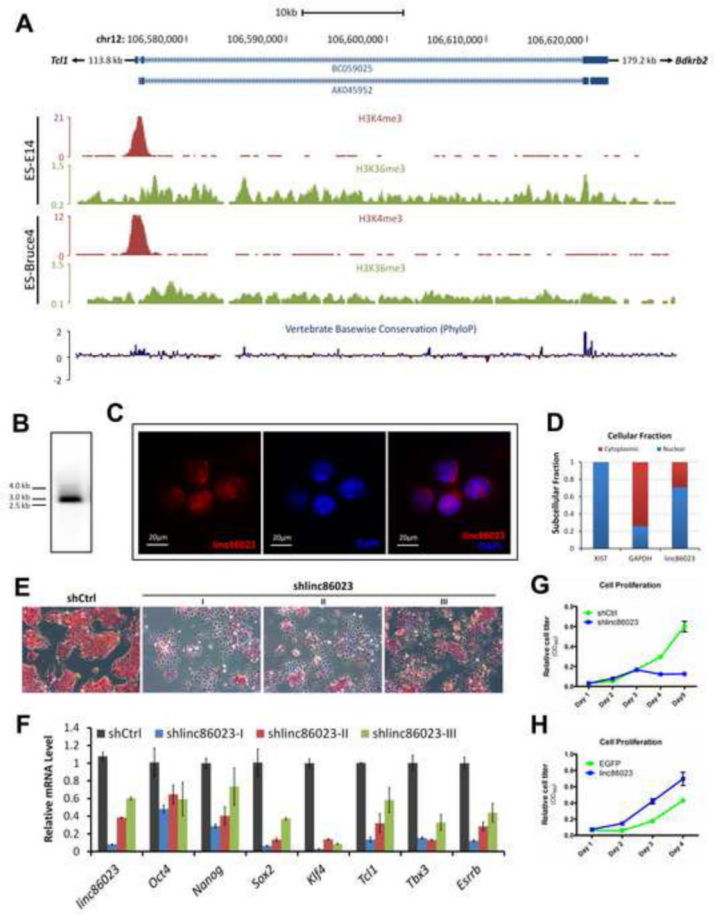


**Figure 3. A highly conserved lincRNA, *linc86023*, is required for maintenance of mESC pluripotency**

(A) Schematic of the mouse *linc86023* locus on chromosome 12 (UCSC genome version NCBI37/mm9). BC059025 (3281 bp) and AK045952 (2876 bp) are alternatively transcribed forms. Blue rectangles represent flanking exons, and blue arrowheads indicate the direction of transcription. Middle panels show ChIP-seq signals of active histone marks H3K4me3 (red) and H3K36me3 (green) in E14 and Bruce mESC lines (data from ENCODE/LICR Histone). The bottom profile shows the level of *linc86023* sequence conservation in vertebrates.

(B) Northern blot of RNA from CCE mESCs indicating the *linc86023* transcript size (~3000 nt).

(C) RNA FISH for *linc86023* in CCE cells.

(D) *Linc86023* was found in both nuclear and cytoplasmic fractions. Cellular fractionation was performed in CCE cells followed by RNA isolation, and mRNA levels of *XIST*, *GAPDH* and *linc86023* were measured by qRT-PCR. The relative subcellular fraction of each gene was shown.

(E) Alkaline phosphatase staining of CCE mESCs on day 4 following transduction with a control shRNA or three independent shRNAs targeting *linc86023*. Scale bars, 50 μm.

(F) Decreased expression of *linc86023* and seven pluripotency genes after knockdown of *linc86023* by three shRNAs. qRT-PCR was performed 4 days after transduction. Gene expression was normalized to *Actb* mRNA levels. Data are the mean ± SD of triplicates.

(G and H) CCE mESC cell proliferation after knockdown (G) and overexpression (H) of *linc86023*. Cell proliferation (measured as the absorbance at 490 nm) was measured 4 days after shRNA treatment (G) or 5 days after transfection with pcDNA3-*linc86023* (H). See also Figure S3.

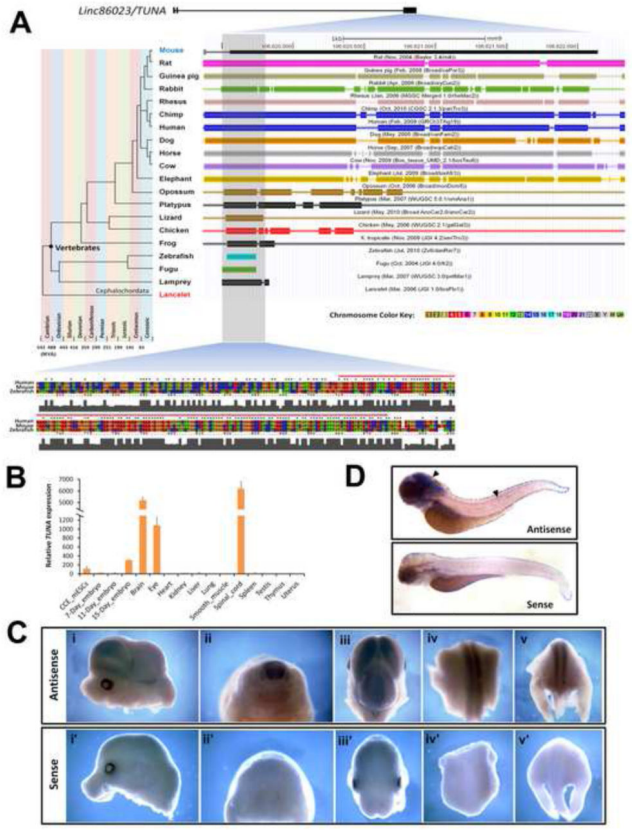


**Figure 4. *Linc86023* (*TUNA*) is evolutionarily conserved and expressed specifically in the central nervous system**

(A) Comparative genomic alignment of 19 species to the mouse genome (mm9) at the 5′ end of the largest exon of *linc86023*/*TUNA*. Chromosome numbers are indicated by the color key. Bottom panel shows alignment of the human, mouse, and zebrafish sequence around the highly conserved region (~200 bp, red line).

(B) Expression of *TUNA* in 15 mouse tissues was measured by qRT-PCR and normalized to *Actb* mRNA levels.

(C) *In situ* hybridization of *TUNA* RNA in E13.5 mouse embryos. Panels show: (i and i') side view of the head, (ii and ii') transverse plane of the body, (iii and iii') overhead view, (iv and iv') dorsal view of the middle body, and (v and v') dorsal view of the lower body. Upper panels show the embryo hybridized with the anti-sense probe, while lower panels with the sense probe.

(D) Whole mount *in situ* hybridization of *tuna* in zebrafish embryo (72 h postfertilization) showing *tuna* expression in the brain and spinal cord (arrowhead). The lower panel shows the embryo hybridized with the sense probe. See also Figure S4.

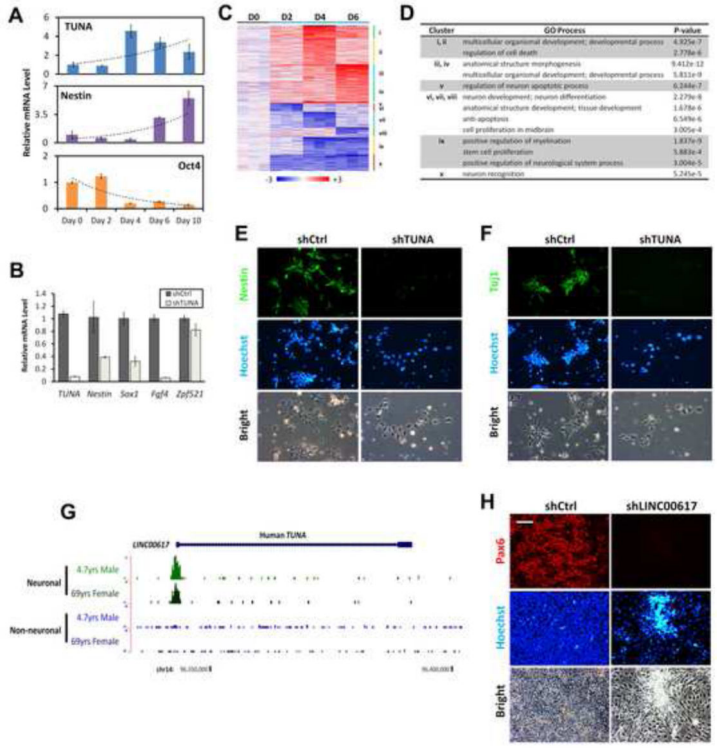


**Figure 5. LincRNA *TUNA* is required for neuronal differentiation of mESCs**

(A) qRT-PCR analysis of *TUNA*, *Nestin*, and *Oct4* expression during neuronal differentiation of CCE mESCs. Total RNA was extracted on the indicated days of monolayer neural differentiation cultures, and relative mRNA levels were normalized to 18S rRNA. Data are the mean ± SD of triplicates.

(B) qRT-PCR analysis of neuronal lineage genes in CCE mESCs following shRNA-mediated silencing of *TUNA*. RNA was extracted four days after transduction, and mRNA levels were normalized to *Actb*.

(C) Heat map showing hierarchical clustering of differentially expressed genes following *TUNA* knockdown. Shown are genes with ≥3-fold difference in expression in *TUNA*-depleted versus control CCE mESCs. RNA was extracted on days 0, 2, 4, and 6 of neural differentiation culture.

(D) Enriched gene ontology (GO) processes of the ten gene clusters identified in (C).

(E and F) Fluorescence and bright-field micrographs of *in vitro*-differentiated CCE mESCs transduced with control or *TUNA*-specific shRNA. CCE cells were transduced after two days of differentiation and analyzed at day 7. Upper panels show cells immunostained for Nestin (E) and Tuj1 (F). Middle panels show Hoechst nuclear staining, and lower panels show bright-field images. Scale bars, 100 μm.

(G) Brain Histone H3K4me3 ChIP-Seq analysis (UMMS Brain Histone (Akbarian/Weng) UCSC track, GRCh37/hg19) of the human *TUNA* ortholog (*LINC00617*) in neuronal and non-neuronal nuclei collected from the prefrontal cortex of a 4.7-year-old male and a 69-year-old female.

(H) Knockdown of human *TUNA* (*LINC00617*) abolished neural differentiation in human. The monolayer culture method was performed in H9 hES cells. Lentiviral infection was performed on day 3 upon neural induction. Neural progenitor cells were immunostained for Pax6 at day 8. Scale bars, 50 μm. See also Figure S5 and Table S4.

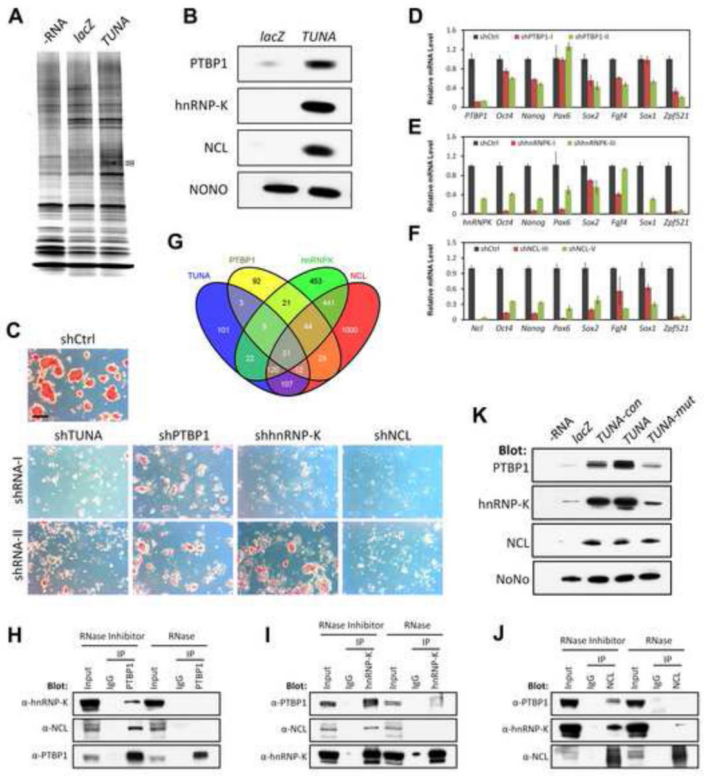


**Figure 6. LincRNA *TUNA* physically interacts with PTBP1, hnRNP-K, and NCL in mESCs**

(A) RNA pulldown of *TUNA*-associated proteins from CCE mESCs. Biotinylated *TUNA* RNA or a control *lacZ* RNA were incubated with nuclear extracts and collected with streptavidin beads. Isolated proteins were resolved by SDS-PAGE and silver stained. Two *TUNA*-specific bands (arrowheads) were excised and subjected to mass spectrometry.

(B) Western blotting of *TUNA* and *lacZ* RNA-associated proteins with antibodies to PTBP1, hnRNP-K, and NCL. A nonspecifically associated protein (NONO) served as the loading control.

(C) AP staining of CCE mESCs transduced with PTBP1-, hnRNP-K-, and NCL-specific shRNAs for 4 days. Two independent shRNAs were analyzed for each protein. Scale bars, 200 μm.

(D–F) qRT-PCR of pluripotency and neural lineage marker genes in CCE mESCs four days after transduction with shRNAs specific for PTBP1 (D), hnRNP-K (E), and NCL (F). Two independent shRNAs were analyzed for each protein. mRNA levels were normalized to *Actb*. Data are the mean ± SD of triplicates.

(G) Many genes are coregulated by *TUNA* and its associated RBPs. shRNA mediated knockdown was performed in CCE cells, and RNA samples were collected on day 4 for microarray analysis. The venn diagram shows the genes with altered transcription (fold change > 1.5, $p < 0.05$) among four knockdown experiments.

(H–J) Coimmunoprecipitation of *TUNA*-associated proteins. CCE mESC lysates were immunoprecipitated with antibodies to PTBP1 (G), hnRNP-K (H), and NCL (I). Immunoprecipitates were treated with RNase or an RNase inhibitor and then analyzed by western blotting with the indicated antibodies.

(K) RNA pulldown assays of *TUNA*-associated proteins from CCE mESC extracts. Cell lysates were incubated with biotinylated full-length wild-type *TUNA* RNA (*TUNA*), the ~200 bp conserved sequence (*TUNA-con*), *TUNA* lacking the conserved region (*TUNA-mut*), or control *lacZ* RNA. RNA-associated proteins were analyzed by western blotting with the indicated antibodies.

See also Figure S6 and Table S5.

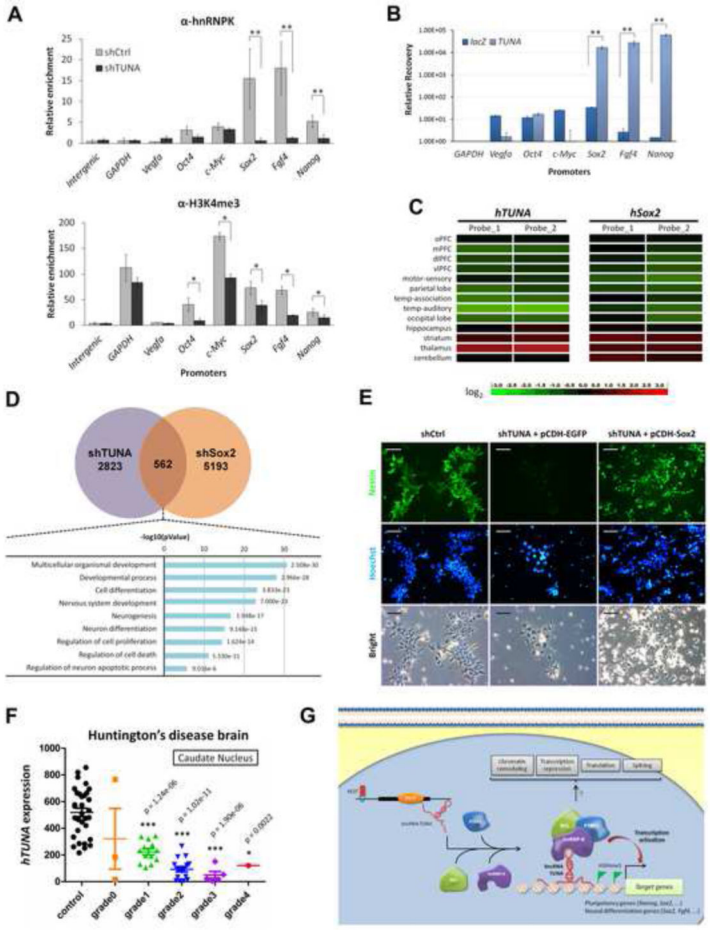


**Figure 7. LincRNA *TUNA* mediates hnRNP-K binding to the *Sox2*, *Nanog*, and *Fgf4* promoters**

(A) ChIP analysis of CCE nuclear extracts four days after transduction with a control or *TUNA*-specific shRNA. Binding of H3K4me3 (left) or hnRNP-K (right) at the promoter regions was quantified by real-time PCR, and is shown as the relative enrichment compared with IgG. *GAPDH* promoter and an intergenic region served as control chromatin loci. Results are the means ± SD of three independent experiments. *p < 0.05, **p < 0.01 by two-tailed Student's t-test.

(B) ChIRP analysis of chromatin occupancy of *TUNA* RNA. CCE cells were crosslinked and sonicated to obtain chromatin DNA of ~100–500 bp. Samples were incubated with 29 DNA probes against *TUNA* RNAs. DNA from the input and precipitated chromatin was analyzed by real-time PCR. Probes against *lacZ* RNA were used as a negative control. ChIRP signals at the *Vegfa*, *Oct4*, *c-Myc*, *Sox2*, *Fgf4*, and *Nanog* promoters were normalized to those at the *GAPDH* promoter. **p < 0.01 by two-tailed Student's t-test.

(C) Microarray data (Sestan Lab Human Brain Atlas) showing coexpression of *TUNA* and *Sox2* in areas of the human brain. RNA isolated from 13 brain regions was analyzed with two probes per gene. Tissues included the orbital (oPFC), medial (mPFC), dorsolateral (dlPFC), and ventrolateral (vlPFC) regions of the prefrontal cortex.

(D) Genes coregulated by *TUNA* and *Sox2*. Venn diagram of genes showing ≥2-fold change in expression following knockdown of *TUNA* or *Sox2* (microarray data from Hutchins et al., 2013). GO analysis of 562 common genes is shown below with p values for each biological function category.

(E) Overexpression of *Sox2* partially rescued the shTUNA-mediated neural differentiation phenotype. The CCE clones stably expressing *hSox2* or *EGFP* were verified by western blot, and subject to *in vitro* neural differentiation using the same procedure as Figure 5E. Scale bars, 100 μm.

(F) Expression of human *TUNA* in the caudate nucleus of brains from Huntington's disease patients according to disease severity. Gene expression microarray analysis of 44 HD brains and 36 normal brains (Hodges et al., 2006). Grade 0 to 4 staging was based on the macroscopic appearance of the brain and neuronal loss in the caudate nucleus. *$p < 0.005$, ***$p < 0.000001$ by two-tailed Student's test.

(G) Model for lincRNA *TUNA* function. *TUNA* recruits a protein complex containing PTBP1, hnRNP-K, and NCL to the promoters of multiple target genes involved in pluripotency maintenance and neuronal differentiation. The *TUNA*–RBP complex functions as a transcriptional activator of *Nanog*, *Sox2*, and *Fgf4*, and may be involved in other regulatory mechanisms such as transcription repression, chromatin remodeling, translation, and splicing under distinct chromatin contexts or in different cells. See also Figure S7.