Research Article

Discovery of a Novel Long Noncoding RNA Lx8-SINE B2 as a Marker of Pluripotency

Fuquan Chen, Miao Zhang, Xiao Feng, Xiaomin Li, Haotian Sun, and Xinyi Lu

State Key Laboratory of Medicinal Chemical Biology, College of Pharmacy, Nankai University, Tianjin 300307, China

Correspondence should be addressed to Xinyi Lu; luxy@nankai.edu.cn

Received 17 October 2020; Revised 31 December 2020; Accepted 22 January 2021; Published 8 February 2021

Academic Editor: Qiang Wu

Copyright © 2021 Fuquan Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Pluripotency and self-renewal of embryonic stem cells (ESCs) are marked by core transcription regulators such as Oct4, Sox2, and Nanog. Another important marker of pluripotency is the long noncoding RNA (lncRNA). Here, we found that a novel long noncoding RNA (lncRNA) Lx8-SINE B2 is a marker of pluripotency. LncRNA Lx8-SINE B2 is enriched in ESCs and downregulated during ESC differentiation. By rapid amplification of cDNA ends, we identified the full-length sequence of lncRNA Lx8-SINE B2. We further showed that transposable elements at upstream of lncRNA Lx8-SINE B2 could drive the expression of lncRNA Lx8-SINE B2. Furthermore, ESC-specific expression of lncRNA Lx8-SINE B2 was driven by Oct4 and Sox2. In summary, we identified a novel marker lncRNA of ESCs, which is driven by core pluripotency regulators.

1. Introduction

Most of the mammalian genome is composed of noncoding sequences. Among them, transposable elements (TEs) contribute to ~40% of the genome [1]. The majority of TEs are silenced, however, a small percentage of TEs are expressed during development and in diseases [2]. They play multiple roles in these processes, including function as enhancers, promoters, and long noncoding RNAs (lncRNAs) [3–6]. In vertebrates, 70% lncRNAs are composed of TEs [7]. TEs also confer tissue-specific expression on lncRNAs through the recruitment of transcription factors [3, 4, 6]. TE-derived lncRNAs actively participate in development. TE-derived lncRNA ROR functions as a sponge to miRNA and also works with hnRNPA1 to promote c-Myc expression during reprogramming [8–10]. Endogenous retrovirus HERVH-derived lncRNAs maintain pluripotency of human embryonic stem cells [3, 11–13]. Asymmetrical expression of ERV1 and ERVK-derived lncRNA LincGET in two- to four-cell mouse embryos biases cell fate toward inner cell mass [14]. These findings all suggest an important role of TE-derived lncRNA in development. Most of these findings are based on human cell lines. We are still lack of understanding of TE-derived lncRNAs in mouse embryonic stem cells (ESCs). In this study, we investigated the expression and regulation of one representative lncRNA Lx8-SINE B2 in ESCs.

2. Methods

2.1. Cell Culture. Mouse ESCs (E14) were cultured on plates coated with 0.2% gelatin (G1890, Sigma) in medium with 15% fetal bovine serum (FBS, SH30070.03, Hyclone), 2 mM L-glutamine (Gibco), 1% penicillin-streptomycin (P1400, Solarbio), 0.1 mM nonessential amino acids (Gibco), 0.1 mM β-mercaptoethanol (M3148–250, Sigma), and 10 ng/ml leukemia inhibitory factor (LIF; Z03077, GenScript). Mouse embryonic fibroblasts (MEFs) and 3-T3 cells were maintained on plates (703001, NEST Biotechnology) in Dulbecco’s modified Eagle’s medium (DMEM, Hyclone) supplemented with 10% FBS, 2 mM L-glutamine, and 1% penicillin-streptomycin. Cells were cultured at 37°C in CO₂ incubator.

2i culture medium contain 50% DMEM/F12 (BasalMedia), 50% Neurobasal media (Gibco), 1% N2 supplement, 1% B27 (Gibco), 0.1 mM nonessential amino acids (Gibco), 2 mM L-glutamine(Gibco), 1% penicillin-streptomycin (P1400, Solarbio), 0.1 mM β-mercaptoethanol (M3148–250, Sigma), 1 μM MEK inhibitor PD0325901 (T6189, TargetMol), and 3 μM GSK3 inhibitor CHIR99021 (2520691, BioGems). 10 ng/ml
leukemia inhibitory factor (LIF; Z03077, GenScript) was added for 2i/LIF condition.

2.2. RNA Extraction, Reverse Transcription, and Quantitative PCR (qPCR). Total RNA was extracted with RNAiso Reagent (B9109, Takara) as described [15] and treated with DNase I to remove genomic DNA in DEPC water (B501005, Sangon Biotech). The cDNA synthesis was carried out in RNase-free tubes (401001, NEST Biotechnology) with the Transcripter First Strand cDNA Synthesis Kit (4897030001, Roche), according to the manufacturer’s instructions. Quantitative PCR (qPCR) reactions were performed using the Heff qPCR SYBR Green Master Mix (H97410, Yeasen) in a QuantStudio 6 Real-Time PCR System (Life Technologies). Primer sequences for qPCR analysis are listed in Table 1.

2.3. Depletion of Gene Expression with shRNAs. For gene knockdown, short hairpin RNAs (shRNAs) for luciferase (control) or target genes were designed by an online tool (http://sirna.wi.mit.edu/) and synthesized by GENEWIZ corporation. The shRNA plasmids were constructed using the pSuper-puro system and purified with a kit (1211-01, Biomiga). mESCs were transfected with DNA using Polyjet (SL100688, SignaGen), according to the manufacturer’s protocol. Transfected ESCs were selected with 1 μg/ml puromycin from 24 h after transfection. After four days of puromycin selection, transfected cells were harvested. The sequences of shRNAs are listed in Table 2.

2.4. 5’ and 3’ Rapid Amplification of cDNA Ends (RACE) Analysis. For 3’ RACE, first-strand cDNA synthesis is initiated at the poly(A) tail of total RNA using the anneal oligo(dt)-containing RT Adapter Primer (AP) to mRNA. Gene-specific primer pF1 was designed based on the known sequence. 3’ fragment was amplified by primer pF1 and general primer gR1, the RACE PCR products were separated on a 1.5% agarose gel and cloned into pEASY-T1 (TransGen Biotech) for Sanger sequencing. The gene-specific RACE primers used for mapping each end were from Sangon Biotech and were listed in Table 3.

2.5. Dual-Luciferase Reporter Gene Assay. Mouse ESCs were seeded at a density of 8 × 10^4 cells per well in a 24-well plate. Luciferase assay was performed as previously described [16]. The total amount of 200 ng of the various promoters of lncRNA Lx8-SINE B2 or pGL4.23 empty vector was transfected into each well of E14 ESC on a 24-well plate together with 10 ng of pCMV-Renilla. The medium was changed 12 h after transfection. After transfection of 36 h, cells were collected and lysed in 1x passive lysis buffer. The luciferase activity was determined by Dual-Luciferase Reporter Assay.
Figure 1: Mapping the full-length sequence of lncRNA Lx8-SINE B2. (a) Schematic of the 3′-rapid amplification of cDNA ends (RACE) (left) and 3′ RACE result for lncRNA Lx8-SINE B2 (right). (b) Schematic of the 5′ RACE and its result for lncRNA Lx8-SINE B2. (c) DNA sequencing of RACE using a universal primer in pEASY-T1 vector. (d) Validation of lncRNA Lx8-SINE B2 transcript size by PCR from cDNA. M, DNA marker.
2.6. Statistical Analysis. Data were analyzed with Student’s t-test (two-tailed). Significant differences were defined as ns for nonsignificant, **p < 0.01, and ***p < 0.001.

3. Results

3.1. Mapping the Full-Length Sequence of lncRNA Lx8-SINE B2. Through mining the previous publication [17], it was shown that lincRNA-1282 was expressed in ESCs and its depletion leads to downregulation of c-Myc [17], which is an important reprogramming factor. Therefore, we set out to perform RACE to identify the full-length of lincRNA-1282 [17], which is a partial sequence of lncRNA Lx8-SINE B2. To identify the full length of Lx8-SINE B2, we performed 3' RACE and 5' RACE with primers as designed (Figures 1(a) and 1(b)). Our amplicons for both 5' and 3' RACE were visible as a single DNA band without multiple or unspecific bands (Figures 1(a) and 1(b)). Next, we sequenced the amplicons and identified the sequences of lincRNA Lx8-SINE B2 found, we designed primers to amplify the full length of lncRNA Lx8-SINE B2 and subcloned the lncRNA into TA cloning vector (Figure 1(d)). The lncRNA Lx8-SINE B2 was revealed to be a 734 bp lncRNA.

3.2. Expression Pattern of lncRNA Lx8-SINE B2. We searched the sequences of lncRNA Lx8-SINE B2 against the mouse...
genome (mm10) and discovered that lncRNA Lx8-SINE B2 contained 3 exons, which are located between Adgrv1 and Lysmd3 gene (Figure 2(a)). Exon 1 of IncRNA Lx8-SINE B2 overlapped with LINE1 family Lx8 and its third exon overlapped with SINE B2 (Figure 2(a)); therefore, we named this lncRNA as Lx8-SINE B2. We designed primers on the nonrepeat region of exon 2 and 3 to detect the expression of IncRNA Lx8-SINE B2. Interestingly, it was noticed that lncRNA Lx8-SINE B2 was downregulated during ESC differentiation, similar to the pluripotency gene Oct4, Sox2, and Esrrb, according to qPCR results (Figure 2(b)). We also found that IncRNA Lx8-SINE B2 was also expressed in ESCs instead of differentiated cells such as MEF (Figure 2(c)). Furthermore, we demonstrated that the expression of lncRNA Lx8-SINE B2 was not affected by the alternation of ESC culture condition. Its expression was slightly upregulated in the presence of 2i/LIF or 2i condition in contrast to the serum/LIF culture condition (Figure 2(d)). These suggest IncRNA Lx8-SINE B2 as a marker of ESC.

3.3. Promoter Structure of IncRNA Lx8-SINE B2. After that, we examined how the specific expression of IncRNA Lx8-SINE B2 was achieved. The upstream 1 kb promoter region of IncRNA Lx8-SINE B2 contains ORR1D2 and SINE B1 (Figure 3(a)). To study how Lx8-SINE B2 is regulated in ESCs, we cloned -623 bp to +327 bp of IncRNA Lx8-SINE B2 gene into luciferase reporter (Figures 3(a) and 3(b)). We also created various truncation versions of this region to identify the core promoter of Lx8-SINE B2 (Figure 3(b)). The region corresponding to ERV, origin-region repeat 1 type D2 (ORR1D2, -157 bp to +3 bp) carried the strongest promoter activity in contrast to those of other truncations (Figure 3(c)). The promoter activity of ORR1D2 was specific to ESCs but inactivated in 3T3 fibroblasts (Figure 3(d)). These results support that IncRNA Lx8-SINE B2 is driven by ERV ORR1D2, implicating that TEs not only contribute to the exons of IncRNAs but also the promoter of IncRNAs.

3.4. Transcriptional Regulation of IncRNA Lx8-SINE B2 by Oct4 and Sox2. To identify which transcription factor activates IncRNA Lx8-SINE B2, we depleted three core pluripotency transcription factors (Oct4, Sox2, and Nanog) (Figures 4(a)–4(c)). Depletion of Oct4 or Sox2, but not Nanog, strongly suppressed IncRNA Lx8-SINE B2 expression (Figures 4(a)–4(c)). We also examined the expression of IncRNA Lx8-SINE B2 after the depletion of Oct4, Sox2, and Nanog (Figures 4(a)–4(c)). However, depletion of either Sox2 or Oct4, but not Nanog, affected the promoter activity of ORR1D2 (Figures 4(d)–4(f)). Sox2 depletion imposed stronger inhibition on ORR1D2 than Oct4 and Nanog (Figures 4(d)–4(f)). Furthermore, we examined the binding of Oct4, Sox2, and Nanog on the promoter of IncRNA Lx8-SINE B2. Consistent with results from luciferase assay, only Oct4 and Sox2 bound to the promoter according to our analysis of published ChIP-seq data (Figure 4(g)). These results suggest that Sox2 and Oct4 directly bind to ORR1D2 to activate Lx8-SINE B2 in ESCs (Figure 4(g)).

To exclude the possibility that Oct4 and Sox2 activate neighboring genes of IncRNA Lx8-SINE B2 together with it, we examined the expression of Lysmd3 and Adgrv1 during
**Figure 4:** Continued.
ESC differentiation. Different from lncRNA Lx8-SINE B2, both Lysmd3 and Adgrv1 were unaffected by LIF withdrawal (Figure 5(a)). Furthermore, the expression of Lysmd3 and Adgrv1 were activated by depletion of Oct4 or Sox2, suggesting they are regulated differently from Lx8-SINE B2 (Figures 5(b) and 5(c)). Moreover, the expression of LINE1 and SINE B2 were not affected by Oct4 or Sox2 depletion (Figure 5(d)), confirming the specificity of Oct4 and Sox2 in activating the expression of lncRNA Lx8-SINE B2.

4. Discussion

In summary, we identified a novel pluripotency marker lncRNA Lx8-SINE B2, whose expression is driven by the binding of Oct4 and Sox2 on ORR1D2. Oct4 and Sox2 are the core pluripotency regulators in ESCs [18, 19]. Oct4 and Sox2 can drive the expression of lncRNAs in cancer cells and ESCs [20–22]. The binding profiles of OCT4 are different in human and mouse ESCs [23], which can be explained by its binding differences on species-specific TEs [23]. Here, we found that Oct4 and Sox2 targeted mouse TE ORR1D2 to drive ESC-specific lncRNA expression (Figure 4), further supporting the important role of TEs in driving the expression of species-specific lncRNAs. There are many pluripotency markers; however, we provide Lx8-SINE B2 as an additional novel marker of pluripotency. It lies at the downstream of key pluripotency genes Oct4 and Sox2 (Figure 4). It composes of TEs and is distinct from traditional markers of pluripotency. In comparison to other ESC markers, Lx8-SINE B2 is unique as an ORR1D2-driven pluripotency marker, which demonstrates that transposable elements can function as cell type-specific lncRNA and promoter, similarly...
to protein-coding genes. Finally, its depletion is associated with the downregulation of Myc in ESCs [17]; therefore, Lx8-SINE B2 expression also reflects Myc expression status of ESCs. Myc represses primitive endoderm differentiation [24]. Myc also maintains ESC pluripotency and self-renewal [25]. Therefore, we speculate that the depletion of lncRNA Lx8-SINE B2 may cause a phenotype similar to that of Myc downregulation.

Our study demonstrates that different types of TEs combine to form lncRNA and drive lncRNA expression (Figures 2 and 3), implicating TEs as important components of lncRNA. TEs in lncRNAs work as an important RNA domain [26, 27]. TEs within lncRNAs regulate the tissue-specific expression of lncRNAs [4, 28]. In human, lncRNAs containing HERVH are specifically expressed in human ESCs [3, 4, 7]. TEs within lncRNAs also contribute to their functions. For example, SINE B2 in antisense lncRNA of Uchl1 interacts with Uchl1 mRNA and promotes the translation of Uchl1 through enhancing the association of mRNA with polysome [29]. These studies demonstrate that TEs are
critical to the expression and function of lncRNAs. Given that lncRNA Lx8-SINE B2 is composed of TE Lx8 and SINE B2, it will be interesting to investigate whether ORR1D2 drives the expression of other lncRNAs and the function of Lx8 and SINE B2 within lncRNAs in the future study.

5. Conclusion
In conclusion, we mapped the full-length sequence of lncRNA Lx8-SINE B2 and found it as an ESC-specific lncRNA. We also found that it was driven by ORR1D2 which was bound by Sox2 and Oct4 to drive its transcription. These findings support TEs as important compositions and promoter of lncRNA.

Data Availability
Published ChIP data analyzed by Cistrome [30] in this study are GSE54103 for Sox2 [31], GSE78073 for Oct4 [32], and GSE56312 for Nanog [33].

Conflicts of Interest
We declare that there is no conflict of interest present for this study.

Authors’ Contributions
X. Lu conceived and designed the study. F.C., M.Z., and X. Li performed most experiments. X.F. and H.S. did bioinformatics analysis. F.C., M.Z., and X. Lu wrote the manuscript.

Acknowledgments
This work is supported by grants from the National Natural Science Foundation of China (31871488, 32070858), the National Key Research and Development Program of China (2018YFA0107000), and Fundamental Research Funds for the Central Universities (63151113, 63161150). We thank Ruiqing Chen for his technical assistance.

References
[1] R. A. Weiss, “Exchange of genetic sequences between viruses and hosts,” Current Topics in Microbiology and Immunology, vol. 407, pp. 1–29, 2017.
[2] K. A. O’Donnell and K. H. Burns, “Mobilizing diversity: transposable element insertions in genetic variation and disease,” Mobile DNA, vol. 1, no. 1, p. 21, 2010.
[3] X. Lu, F. Sachs, L. A. Ramsay et al., “The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity,” Nature Structural & Molecular Biology, vol. 21, no. 4, pp. 423–425, 2014.
[4] D. Kelley and J. Rinn, “Transposable elements reveal a stem cell-specific class of long noncoding RNAs,” Genome Biology, vol. 13, no. 11, p. R107, 2012.
[5] C. D. Todd, O. Deniz, D. Taylor, and M. R. Branco, “Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells,” eLife, vol. 8, 2019.
[6] B. Miao, S. Fu, C. Lyu, P. Gontarz, T. Wang, and B. Zhang, “Tissue-specific usage of transposable element-derived promoters in mouse development,” Genome Biology, vol. 21, no. 1, p. 255, 2020.
[7] A. Kapusta, Z. Kronenberg, V. J. Lynch et al., “Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs,” PLoS Genetics, vol. 9, no. 4, article e1003470, 2013.
[8] S. Loewer, M. N. Cabili, M. Gottman et al., “Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells,” Nature Genetics, vol. 42, no. 12, pp. 1113–1117, 2010.
[9] J. Huang, A. Zhang, T. T. Ho et al., “Linc-RoR promotes c-Myc expression through hnRNP I and AUF1,” Nucleic Acids Research, vol. 44, no. 7, pp. 3059–3069, 2016.
[10] Y. Wang, Z. Xu, J. Jiang et al., “Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal,” Developmental Cell, vol. 25, no. 1, pp. 69–80, 2013.
[11] J. Durruthy-Durruthy, V. Sebastiani, M. Wossidlo et al., “The primate-specific noncoding RNA HPAT5 regulates pluripotency during human preimplantation development and nuclear reprogramming,” Nature Genetics, vol. 48, no. 1, pp. 44–52, 2016.
[12] J. He, X. Fu, M. Zhang et al., “Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells,” Nature Communications, vol. 10, no. 1, p. 34, 2019.
[13] J. Wang, G. Xie, M. Singh et al., “Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells,” Nature, vol. 516, no. 7531, pp. 405–409, 2014.
[14] J. Wang, L. Wang, G. Feng et al., “Asymmetric expression of LincGET biases cell fate in two-cell mouse embryos,” Cell, vol. 175, no. 7, pp. 1887–1901.e18, 2018, e18.
[15] E. Fu, J. Shen, Z. Dong et al., “Histone demethylase Kdm2a regulates germ cell genes and endogenous retroviruses in embryonic stem cells,” Epigenomics, vol. 11, no. 7, pp. 751–766, 2019.
[16] F. Chen, W. Zhang, X. Xie, T. Gao, Z. Dong, and X. Lu, “Histone chaperone FACT represses retrotransposon MERVL and MERVL-derived cryptic promoters,” Nucleic Acids Research, vol. 48, no. 18, pp. 10211–10225, 2020.
[17] M. Gottman, J. Donaghey, B. W. Carey et al., “lincRNAs act in the circuitry controlling pluripotency and differentiation,” Nature, vol. 477, no. 7364, pp. 295–300, 2011.
[18] L. A. Boyer, T. I. Lee, M. F. Cole et al., “Core transcriptional regulatory circuitry in human embryonic stem cells,” Cell, vol. 122, no. 6, pp. 947–956, 2005.
[19] X. Chen, H. Xu, P. Yuan et al., “Integration of external signaling pathways with the core transcriptional network in embryonic stem cells,” Cell, vol. 133, no. 6, pp. 1106–1117, 2008.
[20] J. Jen, Y. A. Tang, Y. H. Lu, C. C. Lin, W. W. Lai, and Y. C. Wang, “Oct4 transcriptionally regulates the expression of long non-coding RNAs NEAT1 and MALAT1 to promote lung cancer progression,” Molecular Cancer, vol. 16, no. 1, p. 104, 2017.
[21] J. Sheik Mohamed, P. M. Gaughwin, B. Lim, P. Robson, and L. Lipovich, “Conserved long noncoding RNAs transcriptionally regulated by Oct4 and Nanog modulate pluripotency in mouse embryonic stem cells,” RNA, vol. 16, no. 2, pp. 324–337, 2010.
[22] W. Gao, C. Q. Qi, M. G. Feng, P. Yang, L. Liu, and S. H. Sun, "SOX2-induced upregulation of IncRNA LINC01561 promotes non-small-cell lung carcinoma progression by sponging miR-760 to modulate SHCBP1 expression," *Journal of Cellular Physiology*, vol. 235, no. 10, pp. 6684–6696, 2020.

[23] G. Kunarso, N. Y. Chia, J. Jeyakani et al., "Transposable elements have rewired the core regulatory network of human embryonic stem cells," *Nature Genetics*, vol. 42, no. 7, pp. 631–634, 2010.

[24] K. N. Smith, A. M. Singh, and S. Dalton, "Myc represses primitive endoderm differentiation in pluripotent stem cells," *Cell Stem Cell*, vol. 7, no. 3, pp. 343–354, 2010.

[25] N. V. Varlakhanova, R. F. Cotterman, W. N. deVries et al., "Myc maintains embryonic stem cell pluripotency and self-renewal," *Differentiation; Research in Biological Diversity*, vol. 80, no. 1, pp. 9–19, 2010.

[26] V. Fort, G. Khelifi, and S. M. I. Hussein, "Long non-coding RNAs and transposable elements: A functional relationship," *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, vol. 1868, no. 1, article 118837, 2021.

[27] A. P. Hutchins and D. Pei, "Transposable elements at the center of the crossroads between embryogenesis, embryonic stem cells, reprogramming, and long non-coding RNAs," *Science Bulletin*, vol. 60, no. 20, pp. 1722–1733, 2015.

[28] T. Chishima, J. Iwakiri, and M. Hamada, "Identification of transposable elements contributing to tissue-specific expression of long non-coding RNAs," *Genes*, vol. 9, no. 1, 2018.

[29] C. Carrieri, L. Cimatti, M. Biagioli et al., "Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat," *Nature*, vol. 491, no. 7424, pp. 454–457, 2012.

[30] T. Liu, J. A. Ortiz, L. Taing et al., "Cistrome: an integrative platform for transcriptional regulation studies," *Genome Biology*, vol. 12, no. 8, p. R83, 2011.

[31] J. Chen, Z. Zhang, L. Li et al., "Single-molecule dynamics of enhancosome assembly in embryonic stem cells," *Cell*, vol. 156, no. 6, pp. 1274–1285, 2014.

[32] J. Shin, T. W. Kim, H. Kim et al., "Aurkb/PP1-mediated resetting of Oct4 during the cell cycle determines the identity of embryonic stem cells," *eLife*, vol. 5, article e10877, 2016.

[33] C. Galonska, M. J. Ziller, R. Karnik, and A. Meissner, "Ground state conditions induce rapid reorganization of core pluripotency factor binding before global epigenetic reprogramming," *Cell Stem Cell*, vol. 17, no. 4, pp. 462–470, 2015.