

# ***LncRNA-Smad7* mediates cross-talk between Nodal/TGF- $\beta$ and BMP signaling to regulate cell fate determination of pluripotent and multipotent cells**

Xiaohui Kong<sup>1</sup>, Kun Yan<sup>2</sup>, Pujuan Deng<sup>3</sup>, Haipeng Fu<sup>1</sup>, Hongyao Sun<sup>4</sup>, Wenze Huang<sup>2,5</sup>, Shuangying Jiang<sup>6</sup>, Junbiao Dai<sup>6</sup>, Qiangfeng Cliff Zhang<sup>2,5</sup>, Jun-jie Gogo Liu<sup>3</sup> and Qiaoran Xi<sup>1,\*</sup>

<sup>1</sup>MOE Key Laboratory of Protein Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China, <sup>2</sup>Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China, <sup>3</sup>School of Life Sciences, Tsinghua-Peking Joint Center for Life Sciences, Beijing Advanced Innovation Center for Structural Biology, Tsinghua University, Beijing 100084, China, <sup>4</sup>Joint Graduate Program of Peking-Tsinghua-NIBS, Tsinghua University, Beijing 100084, China, <sup>5</sup>MOE Key Laboratory of Bioinformatics, Beijing Advanced Innovation Center for Structural Biology & Frontier Research Center for Biological Structure, Center for Synthetic and Systems Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China and <sup>6</sup>CAS Key Laboratory of Quantitative Engineering Biology, Guangdong Provincial Key Laboratory of Synthetic Genomics and Shenzhen Key Laboratory of Synthetic Genomics, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

Received February 01, 2022; Revised August 14, 2022; Editorial Decision August 16, 2022; Accepted September 01, 2022

## **ABSTRACT**

Transforming growth factor  $\beta$  (TGF- $\beta$ ) superfamily proteins are potent regulators of cellular development and differentiation. Nodal/Activin/TGF- $\beta$  and BMP ligands are both present in the intra- and extracellular milieu during early development, and cross-talk between these two branches of developmental signaling is currently the subject of intense research focus. Here, we show that the Nodal induced *lncRNA-Smad7* regulates cell fate determination via repression of BMP signaling in mouse embryonic stem cells (mESCs). Depletion of *lncRNA-Smad7* dramatically impairs cardiomyocyte differentiation in mESCs. Moreover, *lncRNA-Smad7* represses *Bmp2* expression through binding with the *Bmp2* promoter region via (CA)<sub>12</sub>-repeats that forms an R-loop. Importantly, *Bmp2* knockdown rescues defects in cardiomyocyte differentiation induced by *lncRNA-Smad7* knockdown. Hence, *lncRNA-Smad7* antagonizes BMP signaling in mESCs, and similarly regulates cell fate determination between osteocyte and myocyte formation in C2C12 mouse myoblasts. Moreover, *lncRNA-Smad7* associates with hnRNPK in mESCs and hnRNPK binds at the *Bmp2* promoter, potentially contributing to *Bmp2* expression repres-

sion. The antagonistic effects between Nodal/TGF- $\beta$  and BMP signaling via *lncRNA-Smad7* described in this work provides a framework for understanding cell fate determination in early development.

## **INTRODUCTION**

Transforming growth factor  $\beta$  (TGF- $\beta$ ) family proteins are key components of the intra- and extracellular stem cell milieu of embryonic and somatic stem cells. Nodal/Activin/TGF- $\beta$  (hereafter, TGF- $\beta$ ) and BMP signaling each play diverse roles in controlling pluripotent and multipotent cell fates (1–3). These roles are determined through transcriptional regulation mediated by the R-SMADs–SMAD4 complex coordinated by lineage-specific master regulators and chromatin associated proteins/factors (4). Cross-talk between TGF- $\beta$  and BMP signaling has been actively investigated in tissue homeostasis maintenance processes (3,5,6). For example, TGF- $\beta$  signaling suppresses BMP signaling by up-regulating genes encoding BMP signaling antagonists (e.g. *CTGF*, *Gremlin*) (7–9), and vice versa (5,10). Although cross-talk in TGF- $\beta$  and BMP signaling is inevitable in cell fate determination, the molecular mechanism responsible for this cross-talk remains elusive.

The heart is one of the earliest differentiating organs, and heart development requires stepwise regulation dependent on the precise control of genetic networks that in-

\*To whom correspondence should be addressed. Tel: +86 010 62771605; Email: xiqiaoran@mail.tsinghua.edu.cn

<sup>†</sup>Lead contact.

volve a suite of cardiac-specific transcription factors (e.g. *HAND1*, *HAND2*, *NKX2.5*, *MESPI*) (11–15), chromatin modifiers (16–18), lncRNAs (19–21), and signaling pathways (22,23). For example, Nodal and BMP signaling regulates cardiomyocyte commitment (22,24,25). While fine-scale temporal control by BMP2 signaling specifies cardiomyocyte differentiation from mouse embryonic stem cells (mESCs) (23,26), Nodal regulates asymmetric morphogenesis of heart looping by modulating cell proliferation, differentiation, and extracellular matrix composition (27). Cardiomyocyte maturation is regulated by the controlled proliferation of cardiac progenitor cells, and changes in progenitor cell proliferation rates may cause trabeculation and left ventricular noncompaction (LVNC) (28–30). It is therefore plausible that cross-talk between the two signaling branches can also function in regulating cardiomyocyte differentiation.

Numerous lncRNAs are broadly expressed during early development (18,31–33), in which the TGF- $\beta$  family plays a central role. These early development and tissue homeostasis lncRNAs are also known to be induced by the TGF- $\beta$  family, such as *DIGIT* in mESCs (34,35), *TGFB2-AS1* in HaCaT cells (36), and *lncRNA-Smad7* in mouse breast cancer and C2C12 cells (37,38). Moreover, regulatory elements embedded in lncRNA transcripts have been shown to be crucial for cellular functions through association with other factors including proteins, DNAs and RNAs (39,40). However, whether and how lncRNAs participate in cross-talk between TGF- $\beta$  and BMP signaling branches has not been documented.

In this work, we show that *lncRNA-Smad7*, which is divergently transcribed from the shared promoter of *Smad7* (37), is activated by Nodal signaling in mESCs. *lncRNA-Smad7* then regulates cardiomyocyte differentiation by repressing *Bmp2* expression, and its knockdown impairs the cardiomyocyte differentiation process. Importantly, *Bmp2* knockdown rescues these *lncRNA-Smad7* KD-related defects. In addition, *lncRNA-Smad7* represses *Bmp2* expression through binding to its promoter region via its (CA)<sub>12</sub>-repeats. Moreover, *lncRNA-Smad7* antagonizes BMP signaling in mESCs, resulting in cell fate determination between osteocyte and myocyte formation in mouse myoblast C2C12 cells. Finally, we identified hnRNPK as an interaction partner of *lncRNA-Smad7* in mESCs, and CUT&Tag assays showed that hnRNPK binds the *Bmp2* promoter, indicating that *lncRNA-Smad7* may repress *Bmp2* expression through recruitment of hnRNPK. These findings thus provide new insight into the role of lncRNAs in regulating cross-talk between Nodal/TGF- $\beta$  and BMP signaling during early development.

## MATERIALS AND METHODS

### Animal experiments

Mice were obtained from Tsinghua University Laboratory Animals Resource Center. And mice were housed at 20–22°C (12 h:12 h light:dark cycles, 50–60% humidity). All mice we used had a C57BL/6 background including adult mice (female postnatal 14 days) and pregnant mice (female, day 17.5). Embryos were isolated at the developmental

stages (E17.5). The dissected tissues were used for the qPCR to verified the gene expressions in different tissues. All animal experiments were carried out in accordance with institutional guidelines for animal welfare and approved by the Institutional Animal Care and Use Committee (IACUC) at Tsinghua University, China.

### Cell culture

Mouse ESCs E14Tg2a.IV were maintained in feeder-free condition on 0.1% gelatin-coated dishes at 37°C with 5% CO<sub>2</sub>. Basic ES cells were cultured in DMEM (Gibco, Cat. No. 8120287), 15% fetal bovine serum (Excell Bio, Cat. No. FND500), 1% penicillin/streptomycin (Biological Industries, Cat. No. 1948087), 1% non-essential amino acids (Hyclone), 1% L-glutamine (Biological Industries, Cat. No. 2008114), 1% sodium pyruvate (Sigma, Cat. No. RNB3675), 100  $\mu$ M  $\beta$ -mercaptoethanol (Sigma), 103 U/ml LIF. HEK293T (human embryonic kidney cells) and C2C12 cells are cultured in DMEM supplemented with 10% FBS (Excell Bio, Cat. No. FCS500).

### Cell differentiation

Mouse EBs formation and ectodermal differentiation were carried out as described by the supplier (ATCC) and previously reported (6).  $1.5 \times 10^5$  cells/ml in EB media were seeded in low-attached 10-cm dish without LIF to initiate the differentiation. We passaged the cells at the ratio of 1:6 on day 2 and change the media every two days. Retinoic acid (RA) was added for ectodermal differentiation on day 2. Cells were harvested on day 0 and day 2 to day 6 for total RNA extraction during the differentiation.

Cardiomyocyte differentiation was initiated via EB formation induced by hanging  $\sim 20$   $\mu$ l suspensions of drops, with roughly equivalent embryonic stem cell counts (2500 cells) (7). The drops were combined after 2 days and treated with human BMP4 (0.2 ng/ml, PEPROTECH, Cat No. 315–27), Activin A (8 ng/ml, R&D, Cat No. 338-AC), human VEGF (5 ng/ml, PEPROTECH, Cat No. 80710) and L-ascorbic Acid (1 mM, Sigma, Cat# A2218) for 32 h. EBs were then transferred to 0.1% gelatin-coated plates and incubated for 6–8 days. Cells were harvested for total RNA extraction at day 0, day 2, day 3.3, day 6, day 8 and day 10. The cells were collected for immunofluorescence staining at pH3 on day 3.3 (80 h) and cTnT on day 10.

Myogenesis differentiation of C2C12 cells was carried out in 2% fetal equine serum (Solarbio, Cat No. S9050) at the confluence of 80–90%, we changed the media every day and collected the cells on day 0, day 2, day 4 to day 6 of for total RNA extraction. Osteoblast differentiation of C2C12 cells was induced by 50 ng/ml hBMP4 at the confluence of 40–50%, we collected the cells every 6 h from hour 0 to hour 48 for total RNA extraction. Alternatively, osteoblast differentiation of C2C12 cells was initiated with 500 cells in 6-well plate by 50 ng/ml hBMP4 for up to 10 days, we change the media with hBMP4 every day and set up alkaline phosphatase staining (AP staining) on day 10.

### Rapid amplification of cDNA ends (RACE)

5' and 3' RACE was performed using the SMART RACE cDNA Amplification Kit (Clontech) according to the manufacturer's recommendations. The full length of *lncRNA-Smad7* was cloned by overlap PCR according to the sequence information of 5' and 3' of RACE. Here we used the 3' CDS primer to initiate the 3' RACE: AGCAGTGGTATCAACGCAGAGTAC(T)<sub>30</sub> V N (N = A, C, G, or T; V = A, G or C). The sequence of other primers are listed in Supplementary Table S2.

### Measurement of *lncRNA-Smad7* Copy Numbers (in E14 and C2C12 cells)

The copy number of the diluted full length of *lncRNA-Smad7* was calculated by DNA/RNA Copy Number Calculator from website (<http://endmemo.com/bio/dnacopynum.php>). The linearized full length of *lncRNA-Smad7* was cloned from a plasmid DNA (PUC19) containing *lncRNA-Smad7* using primers listed in Supplementary Table S2. A serial dilution of the linearized full length of *lncRNA-Smad7* cDNA was used to qRT-PCR to generate a standard curve for *lncRNA-Smad7*. To measure the *lncRNA-Smad7* copy number in E14 or C2C12 cells, total RNA extracted from  $1.5 \times 10^6$  and  $1.2 \times 10^6$  cells of each line was reverse transcribed into cDNAs for qPCR analysis, and the copy number could be quantitated from the standard curve. The qPCR primers were listed in Supplementary Table S1.

### PiggyBac mediated over-expression

For *Smad2* or *Smad4* reintroduction in the *Smad2* KO or *Smad4* KO mESCs: sequences harboring the *Smad2* (Gene ID: 17126) or *Smad4* (Gene ID: 17128) UTR, and CDS regions were introduced into *Smad2* KO or *Smad4* KO mESCs (PMID: 3762693) by PB-CAG system. The sequences of *Smad2* and *Smad4* were synthesized by Junbiao Dai lab.

For *lncRNA-Smad7* overexpression: The full-length transcript of *lncRNA-Smad7* were retrieved and cloned from E14 cells according to the RACE data, inserted into the PiggyBac vector resistant to hygromycin, and driven by CAG promoter. The overexpression vectors were co-transfected with pBASE by lipofectamine 2000. After drug selection of hygromycin (Amresco, Cat. No. HK547), the overexpressed cells were collected for RNA extraction and qPCR analysis.

### Generation of *lncRNA-Smad7* knock-out and 4x polyA-knock-in (KIpa) cells

The short guide RNAs (sgRNAs) were inserted into BbsI-linearized pSpCas9 (BB)-2A-GFP (PX458) vector (Addgene# 48138) respectively. The 4x polyA transcription stop cassette was cloned into the pMD19-T (TaKaRa) donor vector. We co-transfected the PX458 and donor vector containing homologous arms and 4x polyA sequence at a ratio of 1:4. Cells were transfected and GFP-positive cells were FACS sorted as single cells into 96-well plate at 48 h after transfection. The cells were cultured for 2 weeks followed

by PCR-based genotyping. One clone showing deletion of the targeted region in *lncRNA-Smad7* genomic DNA and two clones showing 4x polyA insertion were picked up for further analysis.

### Generation of shRNA mediated knock-down cells

Short hairpins (shRNAs) were designed by Public TRC Portal (the RNAi Consortium, Broad Institute) or as previously reported. Annealed oligonucleotides of shRNAs were cloned into AgeI and EcoRI restriction enzymes digested pLKO.1-puro lentiviral vector driven by U6 promoter. Scramble shRNA was used as the control shRNA. The shRNAs were transfected into HEK293T cells to package lentivirus by lipofectamine 2000 (Invitrogen, Cat# 2270695) followed by infections of mESCs or C2C12 cells. After drug selection, cells were collected for RNA extraction and qPCR analysis. Sequence of shRNAs for *lncRNA-Smad7*, *Bmp2* and *Hand1* were listed in Supplementary Table S2.

### Subcellular fractionation assay

Subcellular fractionation was performed as previously described with some subtle modifications (41). One 10-cm plate of mESCs were digested by trypsin, then spun down at 1200 rpm for 3 min to collect. We separated cytoplasmic, nucleoplasm and chromatin fractions and isolated them by TRIzol reagent (Ambion, 204409) to fully dissolve the fractions.

### Proliferation analysis

EBs of cardiomyocyte on day 3.3 (CM D3.3, a total of 80 h) were collected and fixed in 4% paraformaldehyde (PFA) at room temperature (RT). The EBs were collected to the 30% sucrose solution for 1–2 days to enable most EBs to sink to the bottom and embedded with O.C.T. compound. EB sections were permeabilized in 0.5% Triton X-100 in PBS (PBST) for 30 min at RT, incubated in primary antibody anti-pH3 overnight and AlexaFluor-488-conjugated secondary antibody for 1 h at RT. Signals were detected with A1/SIM/STORM-confocal (Nikon, 17018750). The antibodies used in this study are listed in Supplementary Table S4.

### Immunofluorescence

Immunofluorescence of cTnT: EBs were seeded on 35 mm confocal dish on day 3.3 and cultured to day 10. CMs were fixed with 4% PFA for 10 min, permeabilized with 0.5% Triton X-100 in PBS for 10 min, stained with the primary antibody cTnT overnight at 4°C and AlexaFluor-594-conjugated secondary antibody for 1 h at RT. Signals were detected with A1R confocal microscopy (A1/SIM/STORM, Nikon, 17018750; HD25, Nikon). The antibody used in this study are shown in Supplementary Table S4.

Immunofluorescence of MY-32 (Myosin Heavy Chain 1 Antibody): C2C12 cells were seeded on 35 mm confocal dish and induced to myogenesis differentiation with 2% fetal equine serum at the confluence of 80–90%. Myogenesis differentiated cells were fixed with 4% PFA for 10 min,

permeabilized with 0.5% Triton X-100 in PBS for 10 min, stained with the primary antibody MY-32 (Abcam, Cat# ab51263) overnight at 4°C. The following procedures were the same as we described above. The antibody used in this study are shown in Supplementary Table S4.

### ChIP

For the ChIP (Chromatin Immunoprecipitation) assay, 200 drops on day 3.3 and 1  $\mu$ l H3K27me3 antibody were enough for each ChIP assay. Cells were collected and crosslinked with 1% formaldehyde for 10 min at 37°C, then quenched by 0.125 M glycine. Cell pellets were lysed in SDS containing buffer (0.3% SDS; 50 mM Tris-HCl, pH 8.0; 20 mM EDTA) freshly supplied with protease inhibitors cocktail (Bimake, Cat# B14001) and sheared to 200–500 bp fragments by sonication. Fragmented chromatin was centrifuged at 13 000 rpm for 20 min at 4°C and the supernatants were diluted in dilution buffer (16.7 mM Tris-HCl, pH 8.0; 0.01% SDS; 1.1% Triton X-100; 1.2 mM EDTA; 167 mM NaCl). The chromatin complex was incubated for 6 h with H3K27me3 antibody (Cell Signaling Technology, Cat. No. 9733), then precipitated by 20  $\mu$ l protein A/G beads (Smart-lifescience, Cat. No. SA032025) for 2 h rotating at 4°C. The precipitated chromatin was sequentially washed 3 times with RIPA buffer (50 mM HEPES, pH 8.0; 1% NP40; 0.7% DOC; 0.5 M LiCl) and then twice with TE buffer (50 mM Tris-HCl, pH 8.0; 1 mM EDTA). Precipitated chromatin complexes were eluted in elution buffer (50 mM Tris-HCl, pH 8.0; 1 mM EDTA, 1% SDS) for 1 h at 65°C, and the supernatant was incubated at 65°C for 6 h on a thermos shaker for de-crosslinking.

RNA and proteins were digested using RNase A or proteinase K, respectively. And the DNA was purified by phenol chloroform extraction and ethanol precipitation. We eluted the samples with 30  $\mu$ l water for ChIP-seq and qPCR analysis. The enrichment for ChIP-qPCR and ChIP-seq was normalized to the sample incubated without antibody. The primer sequences of qPCR and antibody are shown in Supplementary Table S1 and S4.

### Cleavage under targets & tagmentation (CUT & Tag)

Cleavage under targets & tagmentation (CUT & Tag) (42) were conducted in E14 mESCs to profile the chromatin association of hnRNPK. 1.0 million cells were collected for CUT & Tag assay of hnRNPK according to the manufacturer's protocol (Yeasen, Cat. No. 12598ES12). 2  $\mu$ g hnRNPK antibody (Invitrogen, MA5-36291) was used for each CUT & Tag assay. The DNA library was prepared through PCR according to the manufacturer's protocol (Yeasen).

### ChIRP

ChIRP (Chromatin isolation by RNA purification) was performed as previously described with modifications (43), the 50–59-nt DNA biotinylated probes (BGI tech), more stringent crosslinking, hybridization and wash conditions were employed. About  $5 \times 10^7$  mESCs were crosslinked in 10-cm dish by 3% formaldehyde for 10 min at 37°C, quenched by

0.375 M glycine and harvested at 1500 rpm for 10 min at 4°C. Cells were resuspended in 800  $\mu$ l lysis buffer (50 mM Tris-Cl, pH 7.0; 10 mM EDTA; 1% SDS) with addition of protease inhibitor cocktail, PMSF and RNase inhibitor. The lysate was sonicated to 200–500 bp, then centrifuged at 13 000 rpm for 20 min at 4°C. We used 100 pmol probe mix for 800  $\mu$ l lysis, and incubated the chromatin complex for 6 h at 37°C rotating all the time. The chromatin complex was then immunoprecipitated by streptavidin beads (Thermo, Cat. No. 20349) for 2 h at 37°C. The precipitated chromatin was sequentially washed 4 times by wash buffer (2x SSC; 0.5% SDS). DNA was eluted sequentially by RNase elution buffer (50 mM NaHCO<sub>3</sub>, 1% SDS, 100  $\mu$ g RNaseA, 100 U RNase H) at 37°C for 1 h. De-crosslinked chromatin was subjected to proteinase treatment and the DNA was prepared as described in ChIP-seq method. The enrichment of ChIRP-seq signals were normalized to the *IncRNA-Smad7* knock-out cells. The sequences of probes used in this study are listed in Supplementary Table S3.

### ChIRP-MS

Mass spectrometry was performed followed ChIRP experiment. Equal amount of E14 cells were used for the ChIRP and control group. Proteins was eluted in 300  $\mu$ l RIPA lysis buffer (Solarbio) with 10  $\mu$ l RNaseA (20 mg/ml) and 10  $\mu$ l RNaseH1 (10 U/ $\mu$ l) and incubated 2 h at 37°C. The beads were discarded by centrifugation, and the supernatant were recovered for the following experiment. 5 $\times$  SDS loading buffer was added to the supernatant and the samples were prepared for 10 min at 100°C. Protein samples were separated in SDS-PAGE gel, and stained with Coomassie brilliant blue R-250. Each lane was cut into 2 pieces for the following mass spectrometry.

The gel bands of interest were excised from the gel, reduced with 5 mM of DTT and alkylated with 11 mM iodoacetamide which was followed by in-gel digestion with sequencing grade modified trypsin in 50 mM ammonium bicarbonate at 37°C overnight. The peptides were extracted twice with 0.1% trifluoroacetic acid in 50% acetonitrile aqueous solution for 1 h and then dried in a speedVac. Peptides were redissolved in 25  $\mu$ l 0.1% trifluoroacetic acid, and 6  $\mu$ l of the extracted peptides were analyzed by Orbitrap Fusion mass spectrometer.

For LC-MS/MS analysis, the peptides were separated by a 120 min gradient elution at a flow rate 0.30  $\mu$ l/min with a Thermo-Dionex Ultimate 3000 HPLC system, which was directly interfaced with an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). The analytical column was a home-made fused silica capillary column (75  $\mu$ m ID, 150 mm length; Upchurch, Oak Harbor, WA) packed with C-18 resin (300 Å, 5  $\mu$ m, Varian, Lexington, MA). Mobile phase A consisted of 0.1% formic acid, and mobile phase B consisted of 100% acetonitrile and 0.1% formic acid. The Q Exactive mass spectrometer was operated in the data-dependent acquisition mode using Xcalibur 2.1.2 software and there was a single full-scan mass spectrum in the orbitrap (300–1800  $m/z$ , 70 000 resolution) followed by 20 data-dependent MS/MS scans at 27% normalized collision energy (HCD). The MS/MS spectra from each LC-MS/MS run were searched using

an in-house Proteome Discoverer (Version PD1.4, Thermo-Fisher Scientific, USA). Static peptide modification included carbamidomethylation (C), dynamic oxidation (M). Two trypsin missed cleavage was allowed. Precursor tolerance and ion fragment tolerance were set at 20 ppm and 0.02 Da, respectively. Confidence levels were set to 1% FDR (high confidence) and 5% FDR (middle confidence). The top ranked proteins of ChIRP-MS were listed in Supplementary Table S6.

### Dual-luciferase reporter assay

*Bmp2* promoter region (chr2: 133551243–133551936) with or without GT-repeats and *Pou5f1* promoter (chr17: 35503947–35504243) region were cloned into pGL3-promoter vector. mESCs E14 ( $1 \times 10^5$  in each well of 12-well plate) were co-transfected by 900 ng luciferase reporter vector and 100 ng renilla vector (internal control) with lipofectamine 2000. Cells were harvested 48 h after transfection. Luciferase activities were examined using the Dual-Luciferase Reporter Assay System (Promega Cat# E1910) and measured by enzyme-labeled instrument (PerkinElmer).

### ELISA

The autocrine BMP2 protein levels were quantified using a commercially available ELISA kit (Abcam, Cat. No. ab119582). All samples were collected from the media culturing cells for 48 h. The media were centrifuged 1000 rpm for 5 min and measured immediately or stored at  $-80^\circ\text{C}$  less than one month. All samples were assayed according to the manufacturer's instructions and tested in duplicates by personnel blinded for each group. The optical density of each well was determined using a microplate reader at an absorbance of 450 nm. No interference and no cross-reactivity were expected based on the manufacturer's instructions. The minimum detectable dose (MDD) of BMP2 ELISA ranged from 50 to 200 pg/ml. The dynamic range of BMP2 ELISA ranged from 15.625 to 1000 pg/ml.

### Protein production and purification

The recombinant Flag-tagged hnRNPK of mouse was constructed in PCI backbone. The plasmid was transfected into the HEK293T cells when the optical density of the culture reached 0.7 (20  $\mu\text{g}$  plasmid for each T75 flask cells). Each T75 flask cells were passaged to one 15 cm dish 12 h post-transfection and harvested 36 h post-transfection. The cell pellet of each dish was suspended in 1 ml cold lysis buffer (400 mM KCl; 20 mM  $\text{MgCl}_2$ ; 40 mM HEPES-KOH, PH 7.0; 0.5% NP40), containing 1 mM DTT and EDTA-free protease inhibitor cocktail. Cells were lysed by sonication on ice for 30 s intermittently at 30% power. The resulting lysates were centrifuged at  $4^\circ\text{C}$  for 15 min at 13 000 rpm, then the supernatants were saved for purification experiments. The 50  $\mu\text{l}$  anti-Flag resins were added to the supernatant of each sample and incubated for 6 h at  $4^\circ\text{C}$ . Then, the resins were washed with lysis buffer for four times at  $4^\circ\text{C}$  followed with incubation with the elution buffer (20 mM Tris-HCl, PH 7.6; 500 mM NaCl) containing 300  $\mu\text{g}/\text{ml}$

flag peptide for twice (125  $\mu\text{l}$  each time) at  $4^\circ\text{C}$ . The hnRNPK protein was further purified with heparin column and gel-filtration column (HiLoad 16/60 Superdex200) by AKTA system (GE Healthcare). The protein purity was monitored by SDS-PAGE.

### IVT (*in vitro* transcription)

Full length, (CA)<sub>12</sub>, and (UG)<sub>25</sub> containing segments of *lncRNA-Smad7* were transcribed *in vitro*. Briefly, the purified DNA templates of *lncRNA-Smad7* fragment were transcribed *in vitro* by T7 polymerase (0.2 mg/ml) for 3 h at  $37^\circ\text{C}$ . DNase I was then added to the transcribed RNA for 30 min at  $37^\circ\text{C}$ . The mixture was centrifuged to remove any precipitate. The digested RNA was purified by phenol-chloroform (PH 5.2) at the ration 1:1 followed with centrifugation for 30 min at 13 000 rpm at  $4^\circ\text{C}$ . The supernatant was concentrated by ultrafiltration spin columns (Millipore) with DEPC water for 100 000 folds at least. Finally, the RNA was concentrated to  $\sim 0.5\text{--}2.0$   $\mu\text{g}/\mu\text{l}$ .

### EMSA (electrophoretic mobility shift assay)

For RNA-protein EMSA: Increasing amounts of purified hnRNPK was incubated with 1.5 pmol *in vitro* transcribed *lncRNA-Smad7* fragments for 30 min at RT (10 mM Hepes PH 7.5; 400 mM NaCl; 5 mM  $\text{MgCl}_2$ ; 0.01% Triton X-100; 10% glycine; 2 mM DTT).

For RNA-DNA EMSA: The Cyanine5 (Cy5) labeled single strand oligonucleotides were generated by incubating 5'NH<sub>2</sub> labeled oligonucleotides (Sangon Biotech) and the reactive dye Cyanine5 NHS ester (LumiProbe) at 1:1 molar ratio and kept at RT for 4 h. Labeled nucleic acid were then separated from excess free fluorophores through ethanol precipitation. Increasing amounts of the *in vitro* transcribed *lncRNA-Smad7* fragments were incubated with 0.15 pmol Cy5 labeled single strand oligonucleotide [(TG)<sub>17</sub>], [(CA)<sub>17</sub>], and [TCTAGTGA-(TG)<sub>17</sub>-TCCATGTG] for 30 min at RT. Double strands 50nt oligonucleotides (CA: TG)<sub>17</sub> were annealed by Cy5 labeled [TCTAGTGA-(TG)<sub>17</sub>-TCCATGTG] and [CACATGGA-(CA)<sub>17</sub>-TCACTAGA] oligonucleotide.

For RNA-DNA and R-loop antibody S9.6 EMSA: the RNA/DNA hybrids containing 0.15 pmol 5'-cy5-labeled DNA and 1.8 pmol RNA (12 folds to the DNA), and increasing amount of S9.6 antibody were incubated in the buffer [10 mM Tris-HCl (pH 7.5), 2.5mM  $\text{MgCl}_2$ , 2.5M Glycerol, 50 mM NaCl and 1 mM DTT] at RT for 30 min. Then, 2.5 U and 5.0 U RNase H1 (NEB), 0.01mg RNase A (Transgene) or vehicle was added to the DNA/RNA/S9.6 complex.

The complexes were separated in a 1.2% agarose gel in 0.5 $\times$  TAE buffer at 100 V for 50 min on ice. Finally, fluorescent signal was capture by Amersham Typhoon (GE Healthcare Bio-Sciences AB, 67130069) and nucleic acids stained with Superstain (CWBIO).

### Negative staining EM of RNA

The structure of *lncRNA-Smad7* was verified by negative staining electron microscopy (EM). All RNA samples were diluted at a final concentration of 2.5  $\mu\text{M}$  in the

DEPC water and negatively stained in 2% (w/v) uranyl acetate solution following the standard deep stain procedure on holey-carbon coated EM copper grids covered with a thin layer of continuous carbon (44). And all the negatively stained specimens were examined on an FEI Tecnai-F20 electron microscope operated at 200 kV acceleration voltage at 50 000 nominal magnification with a range of defocus from 3 to 3.5  $\mu\text{m}$ . The electron micrographs were taken on a Gatan Ultrascan4000 4k  $\times$  4k CCD camera.

### DRIP-qPCR

DNA–RNA immunoprecipitation (DRIP) was performed by using S9.6 antibody recognizing DNA–RNA hybrid along chromosomes as described previously (45–47) with several modifications. In brief,  $2 \times 10^7$  cells were collected and the genome was extracted by the genome extraction kit (TIANGEN).

RNase A and RNase H treatment: half of the genomic DNA (gDNA) was treated by 5  $\mu\text{l}$  RNase A for 2 h at 37°C, then 1/10 volume of 10 $\times$  RNase H buffer and RNase H (NEB) was added, followed by incubation at 37°C overnight. RNase-treated and -untreated gDNAs were digested with Mse I, Dde I, Alu I, and Mbo I (NEB; final concentration: 75 U/ml for each enzyme) at 37°C. Fragmented gDNA was purified by phenol:chloroform (25:24), resuspended in TE buffer and quantified by Qubit 3.0 (Invitrogen). Then, each sample was immunoprecipitated with 1 $\times$  DRIP binding buffer [10 mM NaPO<sub>4</sub> (pH 7.0), 140 mM NaCl and 0.05% Triton X-100] and S9.6 antibody (American Type Culture Collection, HB-8730, 10  $\mu\text{g}$  for 1  $\mu\text{g}$  gDNA) in a shaker at 4°C overnight. After adding Protein A/G beads and incubating at 4°C for 3 h, the beads were washed four times with 1 $\times$  DRIP binding buffer. Next, the S9.6 associated complexes were eluted by elution buffer [50 mM Tris–HCl (pH 8.0), 0.5% SDS and 10 mM EDTA] containing proteinase K at 55°C for 1 h. The complex was purified with phenol/chloroform extraction, and the supernatant was transferred into a new tube. 1/10 volume of 3 M NaAc, 1  $\mu\text{l}$  Acryl carrier (Solarbio), and 2.5 volume ethanol was added to precipitated the complex at –20°C at least 2 h. The DRIPed DNA pellet was washed with 70% ethanol, air-dried, resuspended in DEPC water and used for qPCR. The sequence of the primers are listed in Supplementary Table S1.

### RNA-seq analysis

Cells were collected by trypsin digestion and the RNA was isolated by RNA isolation kit (Dakewe, Cat. No. 8034111). High-throughput sequencing was performed by Novogene on a HiSeq X. We mapped the RNA-seq data to the mouse reference genome (mm10) using HISAT2 (version 2.1.0). The levels of gene expression were calculated by Cufflinks (version 2.2.1) based on mm10 annotations. Differential testing and log<sub>2</sub> fold change calculations were performed using Cuffdiff (version 2.2.1), with the implementation of two biological replicates. Heatmap was performed by Rstudio and Gene Ontology analysis was performed using DAVID (<https://david.ncifcrf.gov/>).

### Statistical analysis

The intensity of the western blotting bands was quantified by Image J and presented after normalization to the loading control.  $\beta$ -Actin (Actb) was used as the internal control for qPCR analysis. Quantitative data in this study are presented as mean  $\pm$  S.D. from three independent experiments at least ( $n \geq 3$ ) unless otherwise stated, and compared statistically by unpaired Student's *t* test. Statistical significance was indicated as follows: \* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$ . Statistical parameters for each experiment, including values of *n* and statistical significance, can be found in the figure legends. Graphs were generated using GraphPad Prism (6.01).

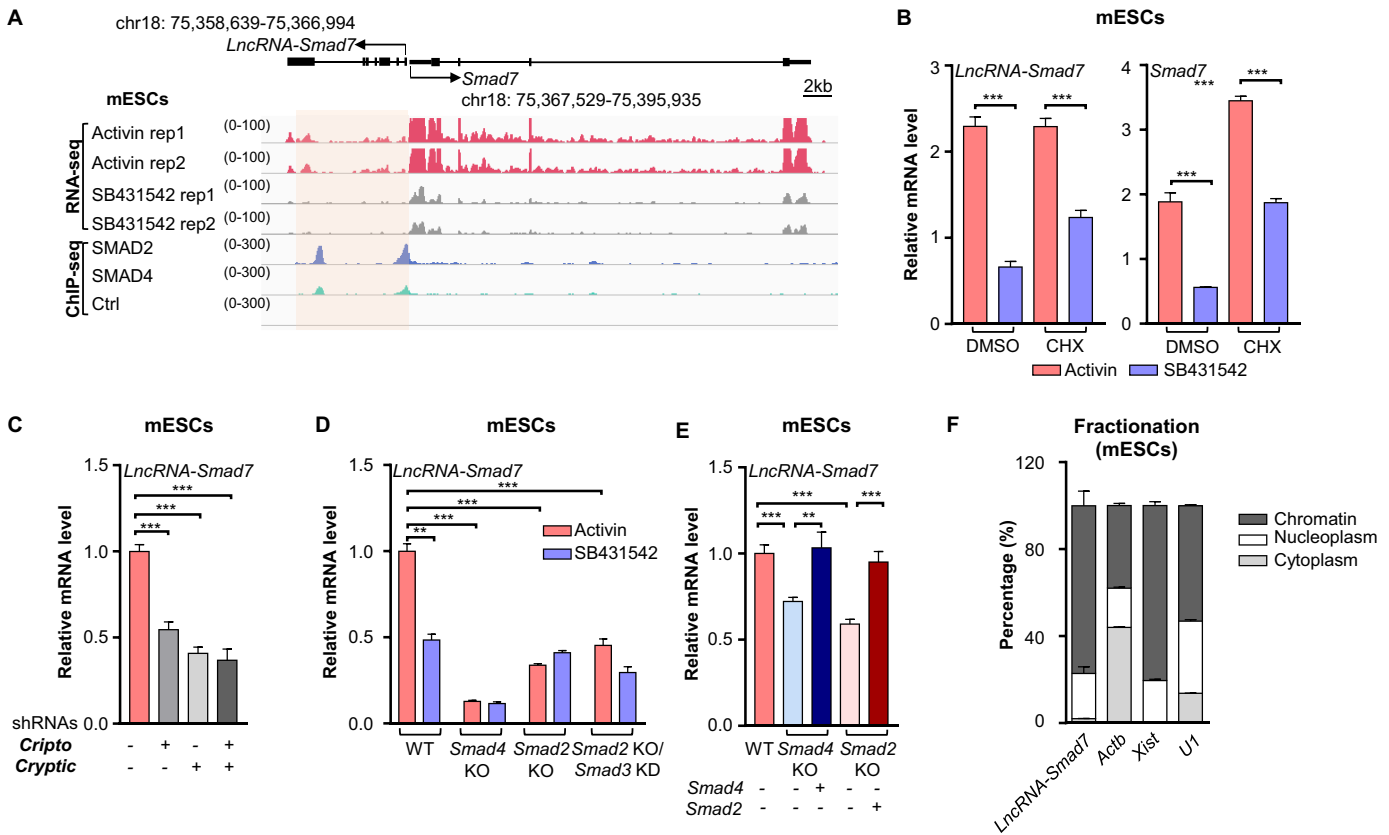
## RESULTS

### Nodal-driven SMAD signaling activates *lncRNA-Smad7* transcription in mESCs

We set out to explore *lncRNAs* responsive to Nodal signaling in mESCs (Supplementary Figure S1A). Activin A (hereafter, Activin) is used as a substitute for Nodal in this study because it is more readily available from mammalian sources than Nodal and they act through the same receptors, except that Nodal also requires the Cripto and Cryptic co-receptors (48,49). We used Activin to activate Nodal signaling and the ALK4/5/7 inhibitor SB431542 to block stimulation by Nodal-like factors in mESCs. RNA-seq analysis showed that transcription of both *lncRNA-Smad7* (Supplementary Figure S1B) and *Smad7* are induced by Activin (Figure 1A, B). Copy number quantification indicated that *lncRNA-Smad7* transcripts were present at about two copies per cell in both E14 and C2C12 cells (Supplementary Figure S1C). ChIP-seq analysis of SMAD2 and SMAD4 indicated that these transcription factors bind at the promoter and gene body of *lncRNA-Smad7* in mESCs (Figure 1A). In addition, *lncRNA-Smad7* expression was significantly decreased in mESCs with knockdown for the Nodal co-receptors Cripto and Cryptic (Figure 1C and Supplementary Figure S1D), which suggested that *lncRNA-Smad7* is indeed responsive to Nodal signaling.

To verify that *lncRNA-Smad7* is a primary target of Nodal signaling, mESCs were pretreated with the protein synthesis inhibitor cycloheximide (hereafter, CHX) and *lncRNA-Smad7* and *Smad7* expression were quantified by qPCR. The results indicated that both *lncRNA-Smad7* and *Smad7* are transcriptionally activated in the presence of Activin, suggesting that their expression did not require nascent synthesis of protein factors, and that *lncRNA-Smad7* is a primary target of Nodal signaling (Figure 1B). Moreover, no induction of *lncRNA-Smad7* was observed upon Activin treatment in mESCs with *Smad2* KO, *Smad2* KO/*Smad3* KD or *Smad4* KO (Figure 1D), while the induction of *lncRNA-Smad7* was rescued by reconstituting *Smad2* or *Smad4* expression (Figure 1E). Taken together, these results demonstrated that *lncRNA-Smad7* is directly activated by Nodal-driven SMAD signaling in mESCs.

Fractionation assays to determine *lncRNA-Smad7* localization indicated that it mainly resides in the nucleus of mESCs, supporting its function in associating with chromatin (Figure 1F). Moreover, qPCR assays indicated that



**Figure 1.** Nodal-driven SMAD signaling activates the transcription of *lncRNA-Smad7* in mESCs. (A) Gene tracks of RNA-seq data for Activin A (hereafter, Activin)- and SB431542-treated mESCs (GSE115169); ChIP-seq of SMAD2 and SMAD4 (GSE125116) in mESCs at *lncRNA-Smad7* and *Smad7* loci (mm10: mouse reference genome 10). (B) qPCR analysis of *lncRNA-Smad7* and *Smad7* expressions in Activin (50 ng/ml)- and SB431542 (10  $\mu$ M)- treated mESCs for 2 h pretreated with cycloheximide (hereafter, CHX, 0.1 mg/ml) or DMSO for 1 hour. DMSO was used as the negative control for CHX. (C) qPCR analysis of *lncRNA-Smad7* expression in *Cripto* KD, *Cryptic* KD, and *Cripto/Cryptic* double KD mESCs by shRNAs. KD, knock-down. (D) qPCR analysis of *lncRNA-Smad7* expression in the indicated mESCs with Activin (50 ng/ml) and SB431542 (10  $\mu$ M) treatment. KO, knock-out. (E) qPCR analysis of *lncRNA-Smad7* expression in *Smad2* KO and *Smad4* KO mESCs with or without expression of *Smad2* or *Smad4* respectively. (F) Fractionation assay showed the subcellular localization of *lncRNA-Smad7* and the other indicated transcripts in mESCs. The *Actb* ( $\beta$ -actin), *U1*, and *Xist* mRNAs were used as cytoplasm, nucleoplasm, and chromatin markers for the fractions respectively. All RNA abundance values are absolutely quantified, the data are representative of three independent experiments. (B–E) All data are representative of three independent experiments. Data are presented as mean  $\pm$  S.D. (Standard Deviation),  $n = 3$ , asterisks indicate a difference from control, \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (two-tailed Student's  $t$  test). See also Supplementary Figure S1.

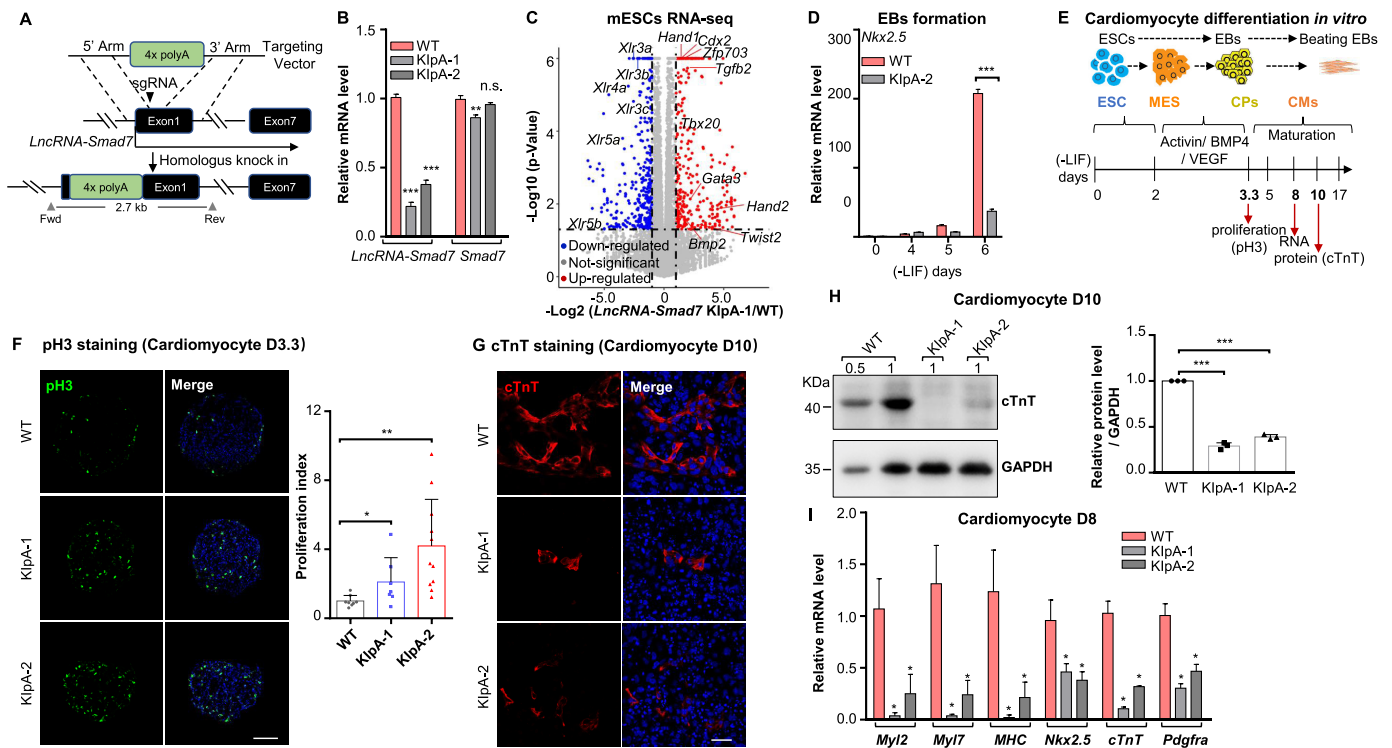
*lncRNA-Smad7* transcripts were highly enriched in skeletal muscle and heart on postnatal day 14 or week 8 in adult mice (Supplementary Figure S1E and S1F), and was expressed at lower levels in the liver, heart, and brain of embryos (Supplementary Figure S1E and S1F). It should be noted that no obvious homolog of murine *lncRNA-Smad7* was identified in other mammals (Supplementary Figure S1G). We then cloned *lncRNA-Smad7* transcripts from mESCs by RACE (rapid amplification of cDNA ends), which revealed an alternative isoform that was 81 bp shorter at the 5' end of exon1 compared with other previously reported transcript isoforms (37,38) (Supplementary Figure S1H).

#### *lncRNA-Smad7* knockdown impairs cardiomyocyte differentiation in mESCs

We next used CRISPR-Cas9 to generate *lncRNA-Smad7* knockdown (KD) cell lines (*lncRNA-Smad7* KIPa) by inserting a 4 $\times$  polyA transcription stop cassette in exon1

(Figure 2A and Supplementary Figure S2A). QPCR analysis showed a  $\sim$ 70% reduction in *lncRNA-Smad7* expression in the *lncRNA-Smad7* KIPa cells, whereas no major effects were observed on *Smad7* expression (Figure 2B). Further RNA-seq analysis in *lncRNA-Smad7* KIPa and WT mESCs with gene ontology (GO) analysis showed that the differentially expressed genes (DEGs) were mainly involved in development (Supplementary Figure S2B). However, *lncRNA-Smad7* KIPa mESCs exhibited morphology and alkaline phosphatase (AP) activity similar to WT mESCs (Supplementary Figure S2C). Using a 2-fold cutoff to identify up- and down-regulated DEGs in the *lncRNA-Smad7* KIPa mESCs revealed a set of up-regulated DEGs well-known to participate in heart development and trophoblast lineage determination, including *Bmp2*, *Cdx2* and *Hand1*, etc. (Figure 2C and Supplementary Figure S2D). In addition, BMP2 is a known factor regulating *Hand1* and *Cdx2* expression during early development (50–52).

We then induced embryoid body (EB) formation for mesendoderm and ectoderm differentiation assays in WT



**Figure 2.** Knock-down of *lncRNA-Smad7* impairs cardiomyocyte differentiation of mESCs. (A) Scheme of the strategy to generate 4× polyA stop cassette knock-in mESCs at 5' end of *lncRNA-Smad7*. (B) qPCR analysis of *lncRNA-Smad7* and *Smad7* expressions of the *lncRNA-Smad7* KIpA and wide type (WT) mESCs. (C) Volcano plot of the DEGs (2-fold cutoff) in *lncRNA-Smad7* KIpA cells compared with WT mESCs. The data are representative of two independent experiments. (D) qPCR analysis of *Nkx2.5* expression in *lncRNA-Smad7* KIpA and WT EBs on the indicated day. (E) Schematic drawing showed in vitro differentiation of mESCs to cardiomyocytes. MES, mesoderm stage; CPs, cardiac progenitors; CMs, cardiomyocytes. (F) (Left) Immunofluorescence staining showed the pH3 (green) in *lncRNA-Smad7* KIpA and WT cells on CM day 3.3 (CM D3.3). Nuclei were stained with DAPI (blue); scale bar, 100  $\mu$ m; (right) quantification of pH3 positive ratios showed in (F, left). Scatter plots represent the individual views (normalized to the WT sample). Data are presented as mean  $\pm$  S.D., two-tailed Student's *t* test,  $n \geq 7$ , \**P* < 0.05, \*\**P* < 0.01. (G) Immunofluorescence staining of cTnT (red) for *lncRNA-Smad7* KIpA and WT CMs on day 10. Nuclei were stained with DAPI (blue); scale bar, 25  $\mu$ m. (H, left) Western blot analysis of the total cTnT protein levels in *lncRNA-Smad7* KIpA and WT CMs on day 10 (D10), GAPDH was used as the loading control. 0.5 and 1 indicated the percentage of loading for each sample; (right) quantification of the cTnT protein levels showed in (H, left) (normalized to WT samples of each independent experiments). Scatter plots represent the three biologically independent experiments. The intensities for each samples were measured by ImageJ. (I) qPCR analysis of cardiac-specific genes on day 8 (D8) in *lncRNA-Smad7* KIpA and WT CMs. (B, D, H, I) All data are representative of three independent experiments. Data are presented as mean  $\pm$  S.D.,  $n = 3$ , \**P* < 0.05, \*\**P* < 0.01 (two-tailed Student's *t* test). See also Supplementary Figures S2 and S3.

and *lncRNA-Smad7* KIpA mESCs to investigate the role(s) of *lncRNA-Smad7* in these differentiation processes (Supplementary Figure S2E) (53,54). Compared to WT cells at day 0, the expression levels of both *lncRNA-Smad7* and *Smad7* were transiently reduced (i.e. lower on EB day 4 and ectodermal differentiation day 3) (Supplementary Figure S2F, G) during the mesoderm differentiation. While depletion of *lncRNA-Smad7* during differentiation did not lead to consistent effects on *Smad7* expression (Supplementary Figure S2H). No obvious different effect was detected during the ectoderm-specific gene expression in *lncRNA-Smad7* KD cells (Supplementary Figure S2I), but early mesoderm differentiation was significantly impacted (Supplementary Figure S2J) compared to wild type cells. Furthermore, expression of the cardiomyocyte marker *Nkx2.5* was dramatically decreased in *lncRNA-Smad7* KIpA EBs (Figure 2D), suggesting that *lncRNA-Smad7* might participate in cardiomyocyte differentiation.

We next investigated the functional impacts of *lncRNA-Smad7* in cardiomyocyte differentiation (Figure 2E)

(19,55). Successful cardiomyocyte differentiation was validated through classic immunofluorescence staining assays for cTnT (cardiac troponin T) (Supplementary Figure S3A). The cardiomyocyte markers *Nkx2.5* and *cTnT* were strongly induced in cardiomyocytes (CMs), whereas *lncRNA-Smad7* expression was reduced on day 3.3 (80 h) of cardiac progenitor cell (CPs) differentiation (Supplementary Figure S3B). Changes in cardiac progenitor cell proliferation rates are known to lead to defects in cardiomyocyte maturation (13,23). It is thus notable that CP proliferation rates in *lncRNA-Smad7* KIpA EBs increased significantly, as indicated by phospho-Histone H3 (pH3) levels at this time point (Figure 2F). However, no obvious difference was observed in EB size between WT and *lncRNA-Smad7* KIpA cells on cardiomyocyte day 3.3 (CM D3.3) (Supplementary Figure S3C).

Interestingly, immunofluorescence staining showed that the proportion of cTnT positive CMs was significantly reduced among *lncRNA-Smad7* KIpA cells compared to WT cells (Figure 2G). Consistent with this finding, cTnT protein levels were significantly decreased in *lncRNA-Smad7* KIpA



CMs on day 10 (Figure 2H). Moreover, the expression of cardiac-specific genes (e.g. *Myl2*, *Myl7*, *MHC*, *Nkx2.5* and *cTnT*) was markedly reduced in *lncRNA-Smad7* KD cells on day 8 of cardiomyocyte differentiation (Figure 2I). Thus, reducing *lncRNA-Smad7* expression can stimulate the proliferation of cardiac progenitor cells but impair cardiomyocyte maturation in vitro.

#### ***lncRNA-Smad7* represses *Bmp2* expression and associates at its promoter**

Next, we performed RNA-seq analysis in both *lncRNA-Smad7* KIPa and WT cardiac progenitor cells. GO analysis showed that the up-regulated DEGs in *lncRNA-Smad7* KIPa cells ( $n = 101$ ) were mainly associated with organismal development, including heart development (Supplementary Figure S4A). In line with this finding, a large suite of heart development-related genes were significantly up-regulated in *lncRNA-Smad7* KIPa cells on CM D3.3 (Figure 3A). In particular, the expression of cardiovascular-specific genes such as *Bmp2*, *Hand1* and *Hand2* were also dramatically increased in *lncRNA-Smad7* KIPa cells on CM D3.3 (Figure 3A), which were also significantly up-regulated in the *lncRNA-Smad7* KIPa mESCs (Figure 2C).

We then introduced a *lncRNA-Smad7* expression vector into *lncRNA-Smad7* KIPa mESCs to overexpress *lncRNA-Smad7*. The elevated expression of *Bmp2* was repressed under *lncRNA-Smad7* overexpression in *lncRNA-Smad7* KIPa cells (Figure 3B), suggesting that *Bmp2* was directly suppressed by *lncRNA-Smad7*. We further generated two *lncRNA-Smad7* KD mESC lines mediated by shRNAs and initiated differentiation in both WT and KD ESCs. Consistent with the KIPa phenotype, *Bmp2* expression was up-regulated in *lncRNA-Smad7* KD cells (Figure 3C). In addition, the highest levels of *lncRNA-Smad7* expression were detected in embryonic mouse hearts at E10.5 to E16.5, whereas *Bmp2* expression concomitantly declined during this period (Supplementary Figure S4B). Conversely, in the early stages of cardiomyocyte differentiation in vitro, *Bmp2* expression increased while *lncRNA-Smad7* expression decreased (Supplementary Figure S3B).

In addition, we quantified autocrine BMP2 levels in culture media by ELISA, which showed that secreted BMP2 levels were higher in *lncRNA-Smad7* KIPa cells than that in WT cells from 71.4 to 134.1 ng/ml (Figure 3D). In line with this result, expression of the downstream BMP target, *Id1*, was increased in *lncRNA-Smad7* KIPa cells compared to that in WT mESCs (Figure 3E). We also detected a slight but steady  $\sim 1.1$  fold increase in phospho-SMAD1/5/8 at SMAD1/5 residues Ser463/465 and Smad8 sites Ser465/467 in *lncRNA-Smad7* KIPa mESCs, further illustrating that *lncRNA-Smad7* KD leads to upregulation of BMP signaling (Figure 3F).

We then performed ChIRP-seq (chromatin isolation by RNA purification and sequencing) (43) for endogenous *lncRNA-Smad7* using antisense DNA oligos tiling along the entire transcript of *lncRNA-Smad7* to obtain genome-wide binding sites of *lncRNA-Smad7*. To this end, we used *lncRNA-Smad7* truncation knock-out mutant (KO) mESCs as the negative control. We found that *lncRNA-Smad7*

strongly associated with its own gene locus in WT mESCs but not in KO cells, validating the specificity of DNA affinity capture (Supplementary Figure S4C). Interestingly, motif analysis of ChIRP DNA-seq data indicated that *lncRNA-Smad7* preferentially binds at CA- or GT-rich genomic regions (Figure 3G). Moreover, the ChIRP DNA-seq results showed that *lncRNA-Smad7* could bind a *Bmp2* promoter region containing (CA:TG)<sub>17</sub>-repeats (Figure 3H).

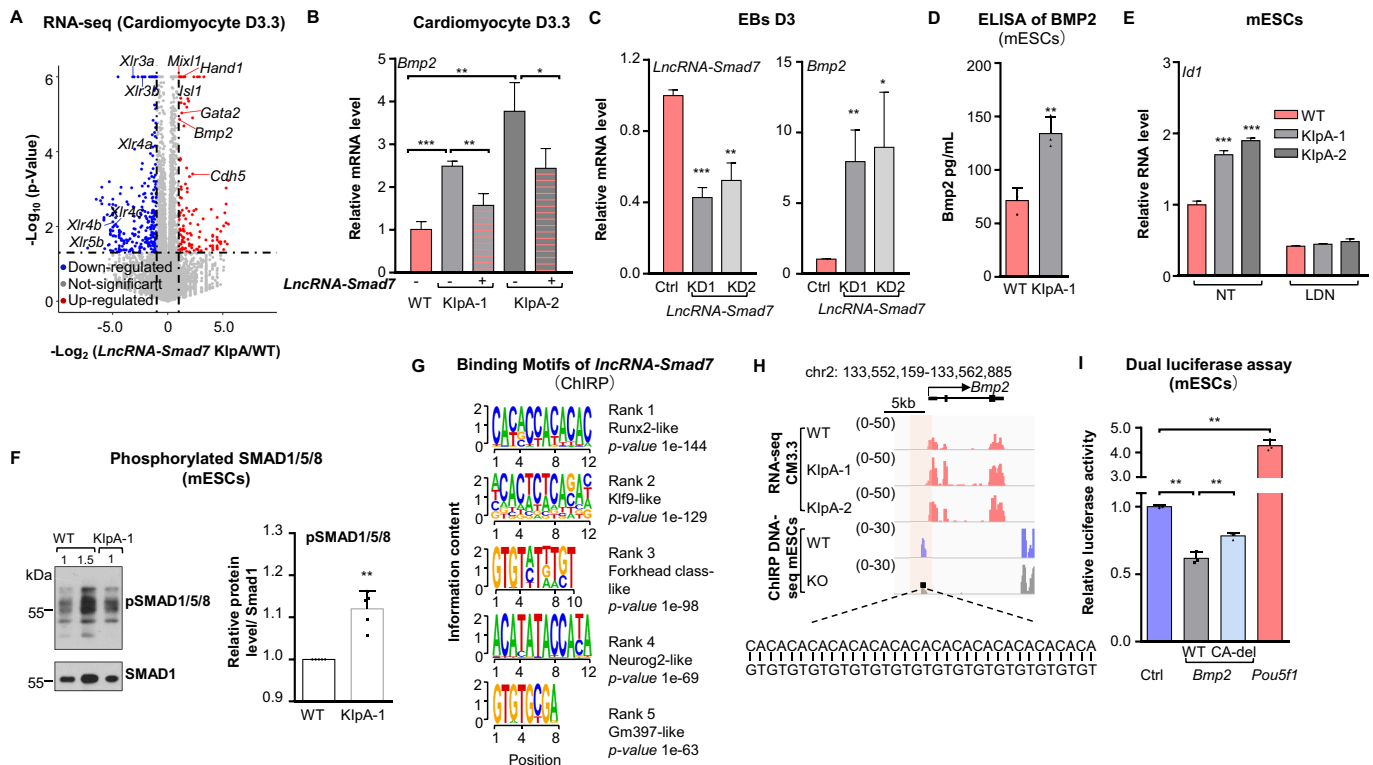
We then cloned the *Bmp2* promoter region harboring the (CA:TG)<sub>17</sub>-repeats into the pGL3-promoter vector to examine its transcriptional regulatory activity. The *Pou5f1* promoter was used as a positive regulatory element control. Luciferase reporter assays in mESCs showed that the *Bmp2* promoter region exerted repressive regulatory effects in a manner partially dependent on (CA:TG)<sub>17</sub>-repeats, since deletion of the (CA:TG)<sub>17</sub>-repeat region led to a significant reduction in transcriptional repression of luciferase signal (Figure 3I).

Collectively, these results indicated that *lncRNA-Smad7* represses *Bmp2* expression in both mESCs and cardiac progenitor cells (CM D3.3) and binds to a promoter region containing (CA:TG)<sub>17</sub>-repeats.

#### ***lncRNA-Smad7* directly associates with *Bmp2* promoter via (CA)<sub>12</sub>-repeats in vitro**

To further characterize the binding interactions between *lncRNA-Smad7* and the promoter region of *Bmp2*, we conducted gel-shift assays using variants of *lncRNA-Smad7* transcribed *in vitro* (Figure 4A and Supplementary Figure S4D) with Cyanine 5 (Cy5)-labeled double stranded (ds) or single stranded (ss) DNA fragments containing (CA:TG)<sub>17</sub>-repeats. The results showed that full length *lncRNA-Smad7* associates with both ds and ss DNAs containing (TG)<sub>17</sub>-repeats, and that *lncRNA-Smad7* exhibited a greater shift with ss DNA than ds DNA (Figure 4B). In addition, *lncRNA-Smad7* could bind with ss DNAs containing (TG)<sub>17</sub> repeats, but not with (CA)<sub>17</sub>-containing ss DNAs (Figure 4C). It should also be noted that both (CA)<sub>12</sub>- and (UG)<sub>25</sub>-repeats are present in *lncRNA-Smad7* transcripts (Figure 4A).

Next, we examined the DNA binding ability of *lncRNA-Smad7* RNA segments containing (CA)<sub>12</sub>- and (UG)<sub>25</sub>-repeats (Figure 4A). These assays indicated that the (CA)<sub>12</sub>-containing RNA segment (534 bp) could bind with either (TG)<sub>17</sub>-containing ss DNAs or (CA:TG)<sub>17</sub>-containing ds DNAs in a similar manner to full length *lncRNA-Smad7* (Figure 4D). In addition, its association with ss (TG)<sub>17</sub> DNA was dependent on the (CA)<sub>12</sub> sequence in the RNA (Figure 4E). In the contrast, the (CA)<sub>12</sub>-containing RNA segment showed negligible binding with ss (CA)<sub>17</sub> DNA (Figure 4E). Negative staining EM (electron microscopy) of RNA revealed that deletion of the (CA)<sub>12</sub> repeats dramatically impacted the *in vitro* folding conformation of the (CA)<sub>12</sub>-containing *lncRNA-Smad7* segment (Supplementary Figure S4E). Although the full length RNA did not associate with ss (CA)<sub>17</sub> DNA, we detected interactions between the (UG)<sub>25</sub> RNA segment with ss (CA)<sub>17</sub> DNA (Supplementary Figure S4F), although it remains unclear why UG-repeats in full length *lncRNA-Smad7* cannot

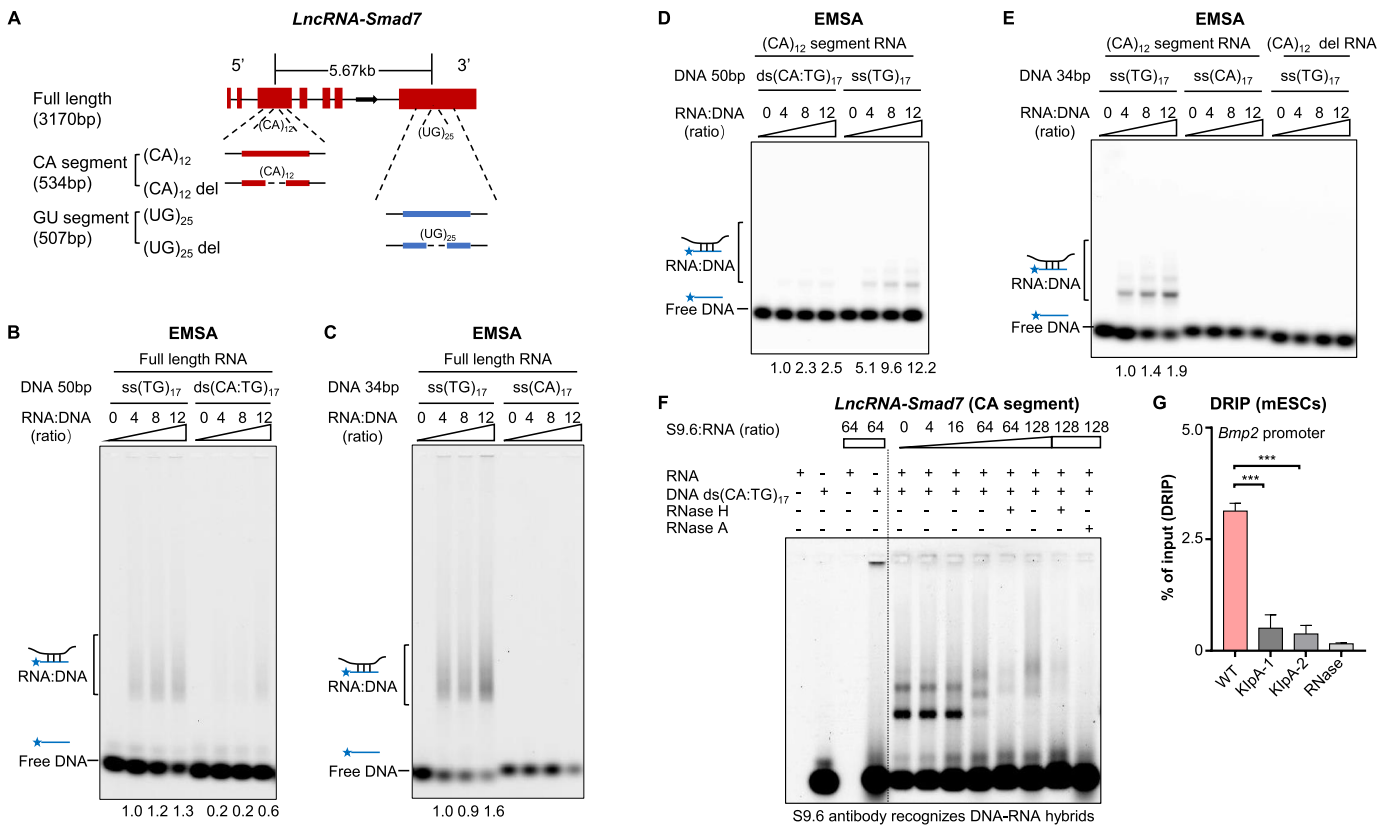


**Figure 3.** *LncRNA-Smad7* represses *Bmp2* transcription during cardiomyocyte differentiation and binds to *Bmp2* promoter. (A) Volcano plot of the DEGs (2-fold cutoff) in *lncRNA-Smad7* KIpA cells compared with WT cells in the stage of cardiac progenitor cells (cardiomyocyte D3.3). The data are representative of two independent experiments. (B) qPCR analysis of *Bmp2* expression in the indicated cells on cardiomyocyte D3.3. (C) qPCR analysis of *Bmp2* and *lncRNA-Smad7* expression in *lncRNA-Smad7* KD and control EBs on day 3 (D3). (D) ELISA assay showed autocrine *Bmp2* protein levels in *lncRNA-Smad7* KIpA mESCs compared to WT. The data are representative of three biologically independent experiments. (E) qPCR analysis of *Id1* expression in *lncRNA-Smad7* KIpA and WT mESCs. BMP type I receptor inhibitor LDN193189 (hereafter, LDN, 1  $\mu$ M) was used to block BMP signaling. (F) Western blot analysis (left panels) and quantification (right panel) of phospho-Smad1/5/8 in *lncRNA-Smad7* KIpA and WT mESCs (normalized to WT samples of each independent experiments). Smad1 was used as the loading control. Data are presented as mean  $\pm$  S.D. Scatter plots represent the biologically independent experiments ( $n = 5$ ). 1 and 1.5 indicated the amount of loading for each sample. (G) Motif analysis of the ChIP DNA-seq of *lncRNA-Smad7*. (H) IGV tracks showed: RNA-seq of WT and *lncRNA-Smad7* KIpA cells at CM3.3; DNA-seq of *lncRNA-Smad7* ChIP in WT and *lncRNA-Smad7* whole locus knock-out (KO) mESCs at *Bmp2* locus (mm10). (I) Dual-luciferase reporter assay showed the activities of *Bmp2* and *Pou5f1* promoter regions in mESCs. Ctrl was the empty vector; CA-del, *Bmp2* promoter region without the (CA:TG)<sub>17</sub>-repeats. (B–E, I) All data are representative of three independent experiments. Data are presented as mean  $\pm$  S.D.,  $n = 3$ , \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (two-tailed Student's  $t$  test). See also Supplementary Figure S4.

base pair with CA-repeats in ss DNA. We speculated that full length *lncRNA-Smad7* might form tertiary structures that allow RNA association with certain sequences (i.e. (TG)<sub>17</sub> repeats) while inhibiting RNA binding with other sequences (i.e. (CA)<sub>17</sub> repeats), which could be critical for its function suppressing *Bmp2* expression. We further analyzed the CA-repeats containing lncRNAs responsive to nodal/TGF- $\beta$  signaling in mESCs (56) and hESCs (57), indicating that there may be CA-repeats containing lncRNAs which function similarly to mouse *lncRNA-Smad7* in both *Mus Musculus* and *Homo Sapiens* (Supplementary Table S5).

Previous reports have shown that DNA and RNA can form R-loops or triplexes to regulate gene expression (58–62). Since *lncRNA-Smad7* preferentially associates with the ss (TG)<sub>17</sub> repeats over interactions with ds (CA:TG)<sub>17</sub> of the *Bmp2* promoter in vitro, we hypothesized that *lncRNA-Smad7* may form an R-loop with DNA fragments containing (TG) repeats at the *Bmp2* promoter. Exposure to the

R-loop antibody S9.6 in EMSAs resulted in a so-called supershift in DNA/RNA complex, while RNaseH1 treatment abolished the supershift band (Figure 4F), suggesting that the (CA)<sub>12</sub>-containing *lncRNA-Smad7* segment indeed forms an R-loop with DNA. Moreover, DRIP (DNA/RNA hybrid immunoprecipitation)-seq analysis in iPSCs (47) facilitated identification of an R-loop in the *Bmp2* promoter region that overlapped with the ChIP signal of *lncRNA-Smad7* in mESCs (Supplementary Figure S4G). In light of this evidence, we performed DRIP-qPCR experiments using S9.6 antibody to validate the RNA/DNA hybrids (R-loop) at the *Bmp2* promoter in WT and *lncRNA-Smad7* KD mESCs. The DRIP-qPCR results showed that the R-loop signal was markedly stronger in WT mESCs than in *lncRNA-Smad7* KIpA cells (Figure 4G), indicating that *lncRNA-Smad7* forms the R-loop at the *Bmp2* promoter region. Taken together, these findings implied that *Bmp2* promoter R-loop could potentially contribute to the transcriptional repression of *Bmp2* by *lncRNA-Smad7*.



**Figure 4.** *LncRNA-Smad7* directly binds at the promoter region of *Bmp2* via CA repeats *in vitro*. (A) Schematic drawing showed *in vitro* transcription of full length, (CA)<sub>12</sub> and (UG)<sub>25</sub>-containing segments of *LncRNA-Smad7*. (B) EMSA showed the binding affinity of 50 bp single strand TG-repeats ss(TG)<sub>17</sub> and double strands (CA:TG) repeats ds(CA:TG)<sub>17</sub> with the increasing amounts of the full length of *LncRNA-Smad7*. (C) EMSA showed the binding affinity of 34 bp ss(TG)<sub>17</sub> and ss(CA)<sub>17</sub> with the increasing amounts of the full length of *LncRNA-Smad7*. (D) EMSA showed binding affinity of 50 bp ds(CA:TG)<sub>17</sub> and ss(TG)<sub>17</sub> with the increasing amounts of the *in vitro* transcribed (CA)<sub>12</sub> containing segment of *LncRNA-Smad7*. (E) EMSA showed binding affinity of 34 bp ss(TG)<sub>17</sub> or ss(CA)<sub>17</sub> repeats with the increasing amounts of (CA)<sub>12</sub> containing segment of *LncRNA-Smad7* and (CA)<sub>12</sub> containing segment without (CA)<sub>12</sub> repeats. 0.15 pmol Cy5-labeled DNA was used as a fixed amount in (B–E). (F) EMSA showed binding affinity of R-loop antibody S9.6 and DNA/RNA complex, 2.5 U and 5 U of RNaseH1 or 0.1 mg RNaseA were added in the indicated samples. 0.15 pmol 5'-cy5-labeled DNA and 1.8 pmol RNA (12-fold to the DNA) were used as a fixed amount in this experiment. (G) DRIP-qPCR showed the R-loop signal at the *Bmp2* promoter region in WT and *LncRNA-Smad7* KIpA mESCs. RNase indicated RNase A and RNase H1. (G) All data are representative of three independent experiments. Data are presented as mean ± S.D.,  $n = 3$ , \*\*\* $P < 0.001$  (two-tailed Student's  $t$  test). See also Supplementary Figure S4.

### **Bmp2** KD rescues defects in cardiomyocyte differentiation in *LncRNA-Smad7* KD cells

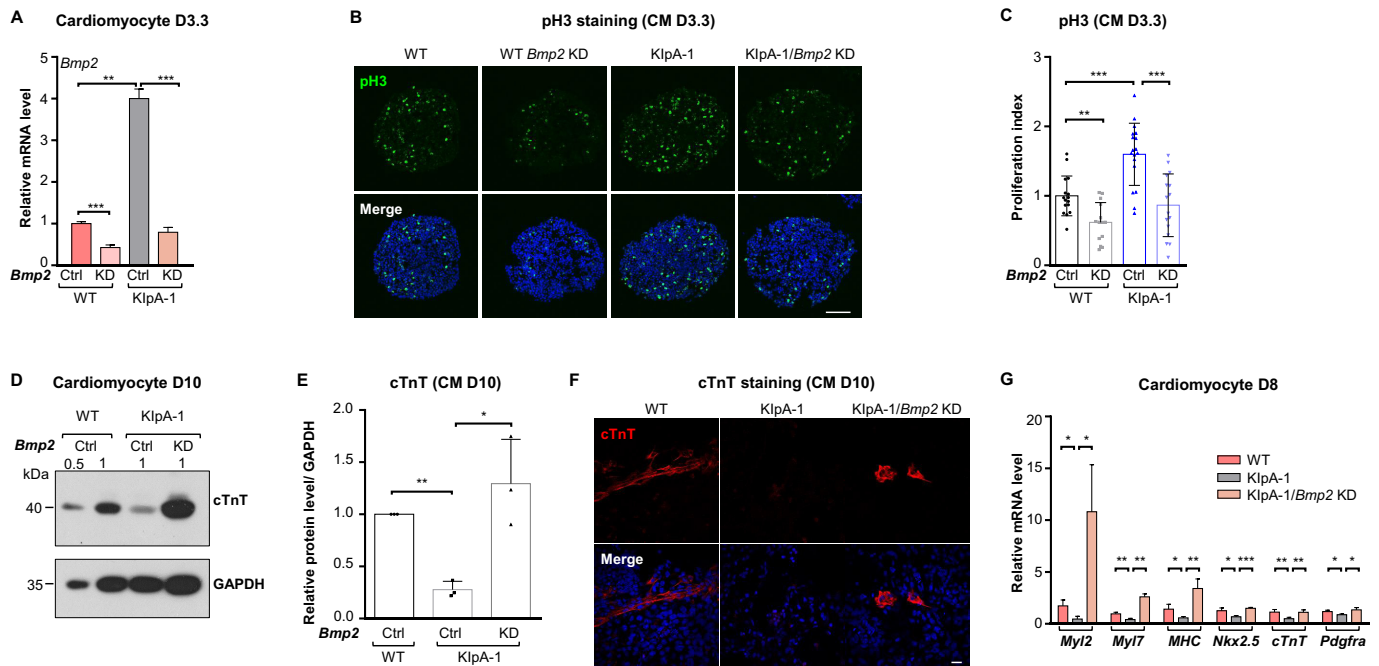
Previous studies have shown that *Bmp2* or *Hand1* overexpression increases the proliferation rate of cardiac progenitor cells and consequently impairs cardiomyocyte maturation (13,23). Based on the results above, we assumed that up-regulation of cardiac-specific genes such as *Bmp2* and *Hand1* on CM D3.3 could impair cardiomyocyte differentiation. To test this possibility, we generated *Bmp2* KD mESCs by shRNA to verify whether defects in cardiomyocyte differentiation caused by *LncRNA-Smad7* depletion could be rescued by *Bmp2* KD (Figure 5A).

Cardiomyocyte differentiation assays using *LncRNA-Smad7* KIpA, *Bmp2* KD/*LncRNA-Smad7* KIpA, *Bmp2* KD or WT mESCs revealed that CP proliferation rates were inhibited under *Bmp2* depletion in both WT and *LncRNA-Smad7* KIpA cells, as indicated by immunofluorescence staining for pH3 on CM D3.3 (Figure 5B, C).

Subsequent quantification of total protein accumulation and immunofluorescence staining for cTnT indicated

that *Bmp2* KD could rescue defects in CM maturation in *LncRNA-Smad7* KIpA cells (Figure 5D–F). This result was supported by restoration of increased cardiac-specific marker expression in *LncRNA-Smad7* KIpA/*Bmp2* KD cells (Figure 5G). Collectively, these findings indicated that *Bmp2* KD rescues impaired cardiomyocyte differentiation in *LncRNA-Smad7*-depleted cells. Thus, *Bmp2* is directly repressed by *LncRNA-Smad7* transcriptionally.

Previous studies have reported that *Hand1* is regulated by BMP signaling (23,50). We found that *Hand1* expression was significantly down-regulated in *Bmp2* KD EBs on CM D3.3 (Supplementary Figure S5A) and up-regulated by BMP2 treatment (Supplementary Figure S5B), indicating that *Hand1* might be regulated by a *LncRNA-Smad7*-*Bmp2* axis. We thus established *Hand1* KD cell lines in both *LncRNA-Smad7* KIpA and WT cells (Supplementary Figure S5C). *Hand1* KD led to the enhanced CM maturation, indicated by cTnT protein levels at CM D10 (Supplementary Figure S5D). In addition, cardiac-specific marker gene expression was rescued in *LncRNA-Smad7* KIpA/*Hand1* KD cells (Supplementary Figure S5E). These results



**Figure 5.** *Bmp2* KD rescues cardiomyocyte differentiation defects in *lncRNA-Smad7* KD cells. (A) qPCR analysis of *Bmp2* expression on CM D3.3. (B) Immunofluorescence staining showed the pH3 levels (green) in *Bmp2* KD, *lncRNA-Smad7* KIpA, *lncRNA-Smad7* KIpA/*Bmp2* KD and control EBs on CM D3.3. Nuclei were stained with DAPI (blue); scale bar, 100  $\mu$ m. (C) Quantification of pH3 positive ratios showed in (B). Scatter plots represent the individual views (normalized to the WT sample). Data are presented as mean  $\pm$  S.D., two-tailed Student's *t* test,  $n \geq 14$ ,  $**P < 0.01$ ,  $***P < 0.001$ . (D) Western blot showed the total protein level of cTnT in *lncRNA-Smad7* KIpA, *lncRNA-Smad7* KIpA/*Bmp2* KD and control CMs on D10. GAPDH was used as the loading control. 0.5 and 1 indicated the percentage of loading for each sample. (E) Quantification of the cTnT protein intensities showed in (D). Scatter plots represent the three biologically independent experiments. The intensities for each sample were measured by ImageJ. (F) Immunofluorescence staining of cTnT (red) for *lncRNA-Smad7* KIpA, *lncRNA-Smad7* KIpA/*Bmp2* KD, and WT CMs on D10. Nuclei were stained with DAPI (blue); scale bar, 25  $\mu$ m. (G) qPCR analysis of key cardiac-specific genes in the indicated CMs on D10. (A, E, G) All data are representative of three independent experiments. Data are presented as mean  $\pm$  S.D.,  $n = 3$ ,  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$  (two-tailed Student's *t* test). See also Supplementary Figure S5.

together implied that a *lncRNA-Smad7-Bmp2* axis may regulate cardiomyocyte differentiation via *Hand1*. However, we observed no association between *lncRNA-Smad7* and the *Hand1* genomic locus using ChIRP-seq analysis, suggesting that *Hand1* was not directly regulated by *lncRNA-Smad7* (Supplementary Figure S5F).

Taken together, *lncRNA-Smad7* directly represses *Bmp2* transcription, thereby prevents expression of *Hand1* which encodes the main factor regulating cardiomyocyte differentiation.

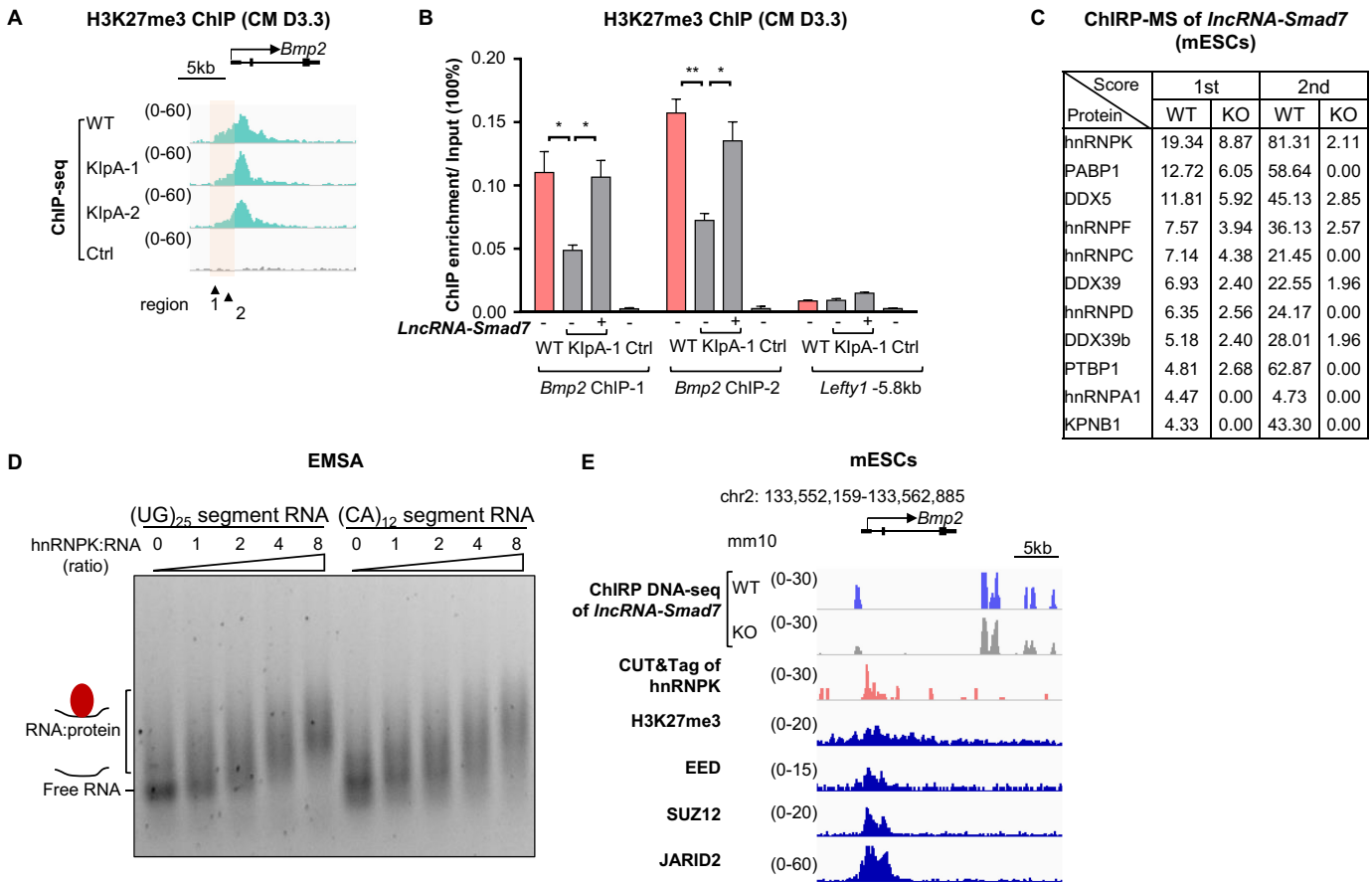
#### ***lncRNA-Smad7* interacts with hnRNP and potentially regulates H3K27me3 modification at the *Bmp2* promoter locus**

In mESCs, developmental genes are silenced which often carry bivalent mark H3K4me3 and H3K27me3, such as those in the *Bmp2* promoter (63), although they remain poised for activation during differentiation (Supplementary Figure S6A). H3K27me3 modifications, catalyzed by polycomb repressive complex 2 (PRC2), are enriched at the transcriptionally repressed loci of several developmental genes (64), but are removed during early mammalian development (65). Given that *Bmp2* expression was repressed by *lncRNA-Smad7*, we next investigated whether H3K27me3 modifications in the *Bmp2* promoter were maintained through *lncRNA-Smad7*. Consistent with this possibility, ChIP-seq analysis showed that

the H3K27me3 signal in the bivalent chromatin domain of *Bmp2* was decreased in *lncRNA-Smad7* KIpA cells compared to that in WT cells on CM D3.3 (Figure 6A). Moreover, the attenuated H3K27me3 signal at the *Bmp2* promoter region was significantly increased by overexpression of *lncRNA-Smad7* in *lncRNA-Smad7* KD cells (Figure 6B). Interestingly, the *lncRNA-Smad7* associates at *Bmp2* promoter (Figure 3H), suggesting that *lncRNA-Smad7* may facilitate H3K27me3 modification at *Bmp2* promoter.

To identify any additional binding partners of *lncRNA-Smad7* that may participate in transcriptional regulation by *lncRNA-Smad7*, we conducted another ChIRP assay followed by mass spectrometry analysis (LC-MS/MS). This experiment revealed that the candidate *lncRNA-Smad7* interacting proteins were RNA-binding proteins (Figure 6C and Supplementary Table S6), among which hnRNP was the top ranked potential interaction partner based on the score of LC-MS/MS. In addition, RNA immunoprecipitation (RIP) assays published by another group indicated that hnRNP could indeed interact with *lncRNA-Smad7* in mESCs (Supplementary Figure S6B) (66).

hnRNP has been described as the principal binding factor responsible for bridging PRC1 with *Xist* and is necessary for PRC1 and PRC2 recruitment (67,68). Moreover, hnRNP is also required to mediate H3K27me3 modifications across targeted chromatin regions (66). To

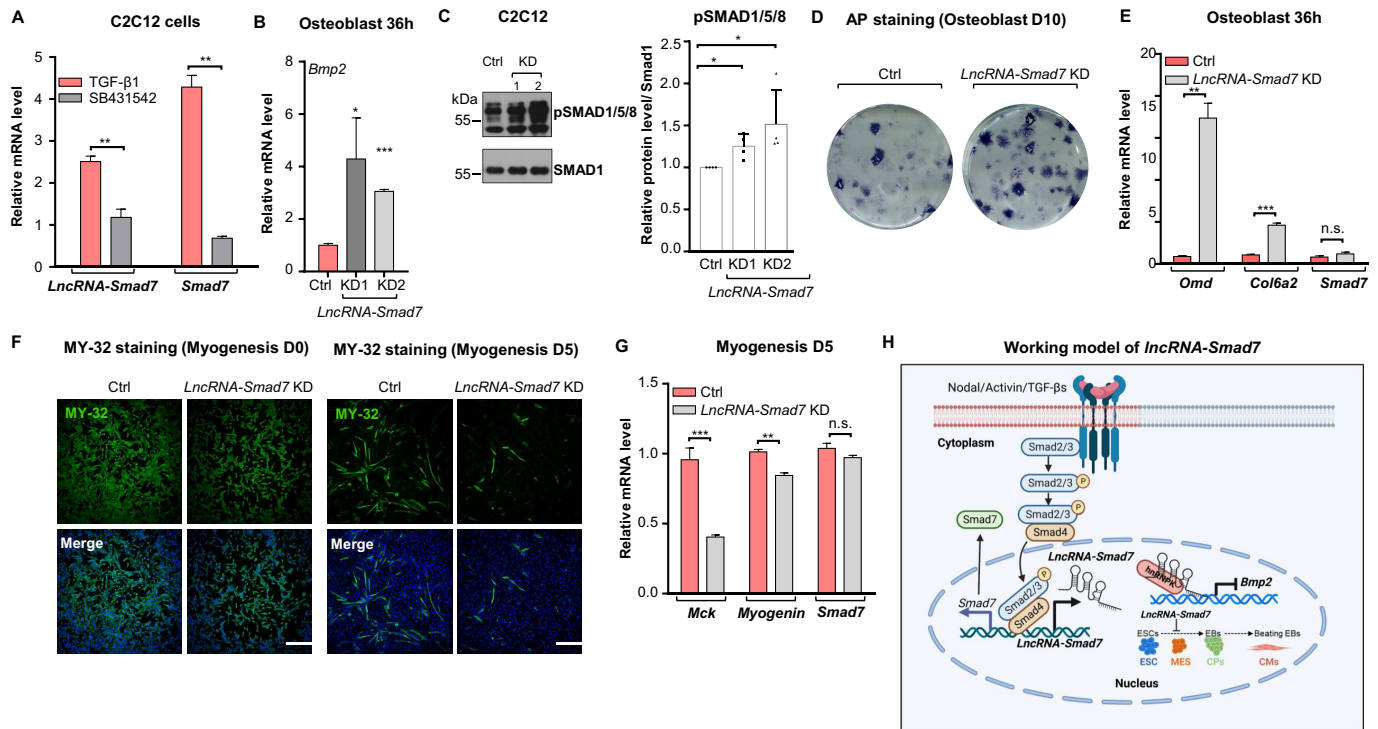


**Figure 6.** *LncRNA-Smad7* KD leads to decreased H3K27me3 modification at the *Bmp2* promoter region and *lncRNA-Smad7* interacts with hnRNPK. (A) IGV tracks showed H3K27me3 ChIP-seq of WT and *lncRNA-Smad7* KlpA progenitor cells on CM D3.3 (mm10). (B) ChIP-qPCR analysis of H3K27me3 signals at the promoter regions of *Bmp2* on CM D3.3. The IgG was set for the control group, and  $-5.8$  kb region of *Lefty1* was used as a negative control for H3K27me3 enrichment. (C) Combination of mass spectrometry (MS) and ChIRP indicated the interacted proteins of *lncRNA-Smad7* in mESCs with two independent experiments (related to Supplementary Table S6). The KO cells are set as the negative control. (D) EMSA showed the association of (UG)<sub>25</sub> containing and (CA)<sub>12</sub> containing segments of *lncRNA-Smad7* with hnRNPK *in vitro*. 1.6 pmol RNA were used as a fixed amount in this experiment. (E) IGV tracks showed ChIRP-DNA seq of *lncRNA-Smad7*, CUT & Tag of hnRNPK, and ChIP-seq of H3K27me3 and PRC2 complex in mESCs at *Bmp2* loci (GSE183465, GSE103258, GSE127117) (mm10). (B) All data are representative of three independent experiments. Data are presented as mean  $\pm$  S.D.,  $n = 3$ , \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (two-tailed Student's *t* test). See also Supplementary Figure S6.

further examine the direct interaction between hnRNPK and *lncRNA-Smad7*, we incubated purified hnRNPK (Supplementary Figure S6C) with *in vitro* transcribed segments of *lncRNA-Smad7* containing (CA)<sub>12</sub> or (UG)<sub>25</sub> repeats (Supplementary Figure S4D) for electrophoretic mobility shift assays (EMSA). The results showed that *lncRNA-Smad7* segments containing either repeat could directly bind to hnRNPK in a dose-dependent manner (Figure 6D and Supplementary Figure S6D), suggesting that *lncRNA-Smad7* interacts with hnRNPK *in vitro*. In addition, hnRNPK CUT & Tag assays indicated that hnRNPK associates with the *Bmp2* promoter region (Figure 6E). Interestingly, analysis of published datasets showed that PRC2 and PRC1 complex also associate with promoter region of *Bmp2* that is enriched with *lncRNA-Smad7* and hnRNPK (Figure 6E and Supplementary Figure S6E) (69,70). Thus, hnRNPK may facilitate repression of *Bmp2* by *lncRNA-Smad7* through association with the *Bmp2* promoter, most likely via maintenance of H3K27me3 marks.

### *LncRNA-Smad7* regulates cell fate determination by repressing *Bmp2* in C2C12 cells

The above data led us to hypothesize that this cross-talk mediated by *lncRNA-Smad7* may represent a general regulatory mechanism that is conserved among cells which recognize both TGF- $\beta$ /Nodal and BMP ligands. C2C12 cells are progenitor mesenchymal cells with the potential to differentiate into osteoblasts or myoblasts. In these cells, BMP signaling is known to reverse cell fate determination from myocytes to osteocytes, while TGF- $\beta$  signaling can negatively regulate osteoblast formation (63). Thus, using C2C12 cells, we first verified that *lncRNA-Smad7* is responsive to TGF- $\beta$  signaling (Figure 7A), then we conducted differentiation assays in which C2C12 cells were separately induced to differentiate into either myoblasts or osteoblasts (64,65). QPCR-based analysis indicated that *lncRNA-Smad7* expression levels were significantly elevated during myogenesis compared to that in uninduced C2C12 cells, whereas its expression initially increased then substantially decreased during osteocyte differentiation (Supplementary Figures S7A, B).



**Figure 7.** *LncRNA-Smad7* regulates cell fate determination of C2C12 cells. (A) qPCR analysis of *LncRNA-Smad7* and *Smad7* expressions in C2C12 cells treated with TGF- $\beta$ 1 (10 ng/ml) or SB431542 (25  $\mu$ M). (B) qPCR analysis of *Bmp2* expression in *LncRNA-Smad7* KD (shRNA) and control C2C12 cells at osteoblast differentiation 36 h. Data are presented as mean  $\pm$  S.D.,  $n = 8$ , \* $P < 0.05$ , \*\* $P < 0.01$  (two-tailed Student's  $t$  test). (C) Western blot showed the phospho-Smad1/5/8 levels in the two *LncRNA-Smad7* KD cell lines (1 and 2) and control C2C12 cells. Smad1 was used as the loading control; quantification of the phospho-Smad1/5/8 intensities showed in Figure 7C. Scatter plots represent the four biologically independent experiments. Data are presented as mean  $\pm$  S.D., \* $P < 0.05$  (two-tailed Student's  $t$  test). (D) AP staining of *LncRNA-Smad7* KD and control C2C12 cells on osteoblast differentiation day 10. The data are representative of four independent experiments. (E) qPCR analysis of *Smad7* and marker genes (*Omd*, *Col6a2*) in the indicated C2C12 cells on osteoblast differentiation 36 h. (F) Immunofluorescence staining showed the MY-32 levels (green) of *LncRNA-Smad7* KD and control C2C12 cells on myogenesis day 0 and day 5. Nuclei were stained with DAPI (blue); scale bar, 100  $\mu$ m. (G) qPCR analysis showed the expressions of *Smad7* and marker genes (*Mck*, *Myogenin*) in the indicated C2C12 cells on myogenesis differentiation day 5. (H) Working model of *LncRNA-Smad7* in cross-talk regulation of TGF- $\beta$  and BMP signalings. Figure was created in BioRender.com. (A, E, G) All data are representative of three independent experiments. Data are presented as mean  $\pm$  S.D., \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (two-tailed Student's  $t$  test). See also Supplementary Figure S7.

In addition to the above trends, relative expression analysis further indicated that *Bmp2* was significantly increased in C2C12 cells with *LncRNA-Smad7* KD (using the same shRNAs as in mESCs) (Figure 7B). Consistent with this finding, phospho-SMAD1/5/8 levels were also substantially increased in *LncRNA-Smad7* KD C2C12 cells (Figure 7C). Moreover, alkaline phosphatase (AP) activity, an osteoblast-specific differentiation marker, was significantly up-regulated in *LncRNA-Smad7* KD cells (Figure 7D). In line with the known functions of BMP signaling in these differentiation processes, the osteoblast differentiation markers *Omd* and *Col6a2* were significantly increased (Figure 7E), while expression levels of the myogenesis differentiation markers, *Mck* and *Myogenin*, were dramatically lower under *LncRNA-Smad7* KD in C2C12 cells (Figure 7F). Additionally, immunofluorescence staining with antibody targeting fast skeletal myosin (MY32) indicated that less myotubes were fused in *LncRNA-Smad7* KD cells compared with WT control cells (Figure 7G and Supplementary Figure S7C). These cumulative results thus expanded the functional relevance of *LncRNA-Smad7* and cross-talk between TGF- $\beta$ /Nodal and BMP signaling in the cell fate determination of multi-potent progenitor cells.

Collectively, these findings indicated that *LncRNA-Smad7*, regulated by Nodal/TGF- $\beta$ , transcriptionally represses *Bmp2* to control cell fate determination during early developmental stages. This involvement of *LncRNA-Smad7* in the cross-talk between the Nodal/TGF- $\beta$  and BMP signaling pathways thus defines a layer of regulation beyond that of the known coding genes in these pathways (Figure 7H).

## DISCUSSION

TGF- $\beta$  (Nodal/Activin/TGF- $\beta$ ) and BMP ligands are present in niches of cells during early development (23,71). Here, we identified the Nodal activated *LncRNA-Smad7*, which associates at the *Bmp2* promoter and represses *Bmp2* expression in mESCs. Hence, Nodal induced *LncRNA-Smad7* mediates cross-talk between Nodal and BMP signaling in mESCs. Moreover, we demonstrated that *LncRNA-Smad7* regulates cardiomyocyte differentiation by repressing *Bmp2*. Interestingly, this cross-talk also applies to the cell fate determination of C2C12 cells. In sum, our study illustrates a *LncRNA-Smad7* mediated cross-talk mechanism

between Nodal/TGF- $\beta$  and BMP signaling in pluripotent and multipotent cells (Figure 7E).

TGF- $\beta$  family members play essential roles in early development (2,4,72). It is well-established that TGF- $\beta$  and BMP signaling are antagonistic to each other at different levels (5) and such cross-talk has been reported in various biological processes (73–75). Multiple studies have shown that TGF- $\beta$  and BMP signaling restricts each other by competing for the limited cellular supply of SMAD4 or R-SMADs (76–79), and TGF- $\beta$  signaling can suppress BMP signaling by promoting genes expression of BMP signaling antagonists (5,7,8,80). Our data support a model wherein these two pathways are converged via *lncRNA-Smad7* (Figure 7E). This type of the cross-talk represents general phenomena in early development maintaining the balances between different signaling pathways to determine cell fate.

Chromatin associated lncRNAs have been reported to regulate developmental and cellular functions by bridging transcriptional regulators and chromatin (40,81,82). *lncRNA-Smad7* associating with *Bmp2* promoter region and up-regulation of *Bmp2* expression in *lncRNA-Smad7* KD cells suggested that *lncRNA-Smad7* transcriptionally represses *Bmp2* expression. Motif analysis of ChIRP-seq showed that *lncRNA-Smad7* tends to bind at CA- or TG-rich regions of the genome. Interestingly, the *Bmp2* promoter region has (CA:TG)<sub>17</sub> repeats motif which is overlapped with *lncRNA-Smad7* binding site and the *lncRNA-Smad7* transcript contains both a (CA)<sub>12</sub>- and a (UG)<sub>25</sub>-repeat regions. The in vitro binding assays further confirmed that the (CA)<sub>12</sub> containing RNA segment of *lncRNA-Smad7* associates with ss (TG)<sub>17</sub>-repeats at the *Bmp2* promoter. Studies showed that RNA could associate with DNA to form the R-loop (58,59) or triplex (60–62). Thus, it is worth investigating how exactly *lncRNA-Smad7* associates with this chromatin region through structural biology to elucidate the complex of *lncRNA-Smad7* with the *Bmp2* promoter.

Polycomb complex is required for maintaining pluripotency in embryonic stem cells (83,84). Many lncRNAs have been shown to function as transcriptional repressive factors (66,85,86) by recruiting the polycomb complex to the target genes (36). Here, we have demonstrated that *lncRNA-Smad7* is essential to maintain the H3K27me3 modification in the *Bmp2* promoter region. And hnRNPK was validated to directly bind with *lncRNA-Smad7* in vitro. Since hnRNPK was reported to function as a bridge between lncRNA and polycomb complex by associating with both factors (66,67), it is likely that *lncRNA-Smad7* represses *Bmp2* expression by recruiting hnRNPK at the *Bmp2* promoter region, consequently bringing transcriptional repressive factors (e.g. polycomb complex) to the locus. In the future, elucidating the assembling of *lncRNA-Smad7*, the hnRNPK, and polycomb complex at *Bmp2* promoter region is worth the investigation.

Mouse *lncRNA-Smad7* is not conserved among species based on the primary sequences. However, RNA molecules are folded into complex of three-dimensional structures which is more conserved than its primary sequences, and RNA structures often play essential roles in RNAs' function (39,40). It is possible that there are RNAs in other species with similar function as mouse *lncRNA-Smad7*

which might share the conserved structure but not primary sequences. In addition, our data showed that (CA)<sub>12</sub>-repeats in *lncRNA-Smad7* is essential for the binding of *lncRNA-Smad7* to the *Bmp2* promoter region. There are some CA-repeats containing lncRNAs in human, and we speculate that there may be CA-repeats containing lncRNAs which function similarly to mouse *lncRNA-Smad7* in Homo sapiens (Supplementary Table S5). Recent studies showed more techniques have been developed to probe the RNA structure and its RNA binding proteins (87,88), especially for those highly abundant RNA. If these techniques allow to detect the structure of the low abundant RNA like *lncRNA-Smad7*, it is worth taking advantage of these techniques to reveal the RNA structure in the future.

Given that *lncRNA-Smad7* and *Smad7* share the same promoter and there are potential regulatory elements of *Smad7* embedded in the *lncRNA-Smad7* locus, it is not feasible to generate *lncRNA-Smad7* KO mice by truncating the whole locus of *lncRNA-Smad7* since the truncation will affect *Smad7* expression. Once the functional domain of *lncRNA-Smad7* is identified, it would be a plausible way for us to generate KO mice by deleting the functional domain of *lncRNA-Smad7*, which will elucidate the function of *lncRNA-Smad7* in early development.

## DATA AVAILABILITY

Raw data have been deposited in the GEO database with the series accession GSE178136. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD036394.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We would like to thank Drs X. Shen and Y. Yin for the plasmid PB-CAG-BGHpA and helpful discussions; Drs J. Na and R. Huang for the RNA-seq analysis of mouse embryo hearts; Drs Q. Shen and Q. Bai for the plasmid pGL3; Drs Q. Sun and Q. Li for the R-loop antibody S9.6; Drs H. Deng and C. Zhao at Protein Chemistry Facility for mass spectrometry analysis; Y. Chen at Imaging Core Facility for assistance of using Nikon A1R HD25; P. Jiao and J. Ji at Center of Biomedical Analysis for flow-cytometry analysis.

## FUNDING

National Natural Science Foundation of China (NSFC) [91540108]; Tsinghua-Peking Center for Life Sciences (to Q.X.); Chunfeng Fund [2021Z99CFY020]; Tsinghua University, Beijing (to J.J.G.L.).

*Conflict of interest statement.* None declared.

## REFERENCES

- David, C.J. and Massague, J. (2018) Contextual determinants of TGF $\beta$  action in development, immunity and cancer. *Nat. Rev. Mol. Cell Biol.*, **19**, 419–435.

2. Senft, A.D., Costello, I., King, H.W., Mould, A.W., Bikoff, E.K. and Robertson, E.J. (2018) Combinatorial smad2/3 activities downstream of nodal signaling maintain embryonic/extra-embryonic cell identities during lineage priming. *Cell Rep.*, **24**, 1977–1985.
3. Wang, W., Rigueur, D. and Lyons, K.M. (2020) TGFbeta as a gatekeeper of BMP action in the developing growth plate. *Bone*, **137**, 115439.
4. Bai, J. and Xi, Q. (2018) Crosstalk between TGF-beta signaling and epigenome. *Acta Biochim. Biophys. Sin. (Shanghai)*, **50**, 322.
5. Ning, J., Zhao, Y., Ye, Y. and Yu, J. (2019) Opposing roles and potential antagonistic mechanism between TGF-beta and BMP pathways: implications for cancer progression. *EBioMedicine*, **41**, 702–710.
6. Yu, Y., Lin, Y., Yang, G. and Tian, L. (2018) The interplay between TGF-beta/SMAD and BMP/SMAD signaling pathways in the epithelial mesenchymal transition of A549 cells induced by silica. *Toxicol. Mech. Methods*, **28**, 286–292.
7. Mundy, C., Gannon, M. and Popoff, S.N. (2014) Connective tissue growth factor (CTGF/CCN2) negatively regulates BMP-2 induced osteoblast differentiation and signaling. *J. Cell Physiol.*, **229**, 672–681.
8. Sanders, L.N., Schoenhard, J.A., Saleh, M.A., Mukherjee, A., Ryzhov, S., McMaster, W.G. Jr, Nolan, K., Gumina, R.J., Thompson, T.B., Magnuson, M.A. et al. (2016) BMP antagonist gremlin 2 limits inflammation after myocardial infarction. *Circ. Res.*, **119**, 434–449.
9. Abreu, J.G., Ketpura, N.I., Reversade, B. and De Robertis, E.M. (2002) Connective-tissue growth factor (CTGF) modulates cell signalling by BMP and TGF-beta. *Nat. Cell Biol.*, **4**, 599–604.
10. Meurer, S.K., Esser, M., Tihaa, L. and Weiskirchen, R. (2012) BMP-7/TGF-beta1 signalling in myoblasts: components involved in signalling and BMP-7-dependent blockage of TGF-beta-mediated CTGF expression. *Eur. J. Cell Biol.*, **91**, 450–463.
11. George, R.M., Maldonado-Velez, G. and Firulli, A.B. (2020) The heart of the neural crest: cardiac neural crest cells in development and regeneration. *Development*, **147**, dev188706.
12. Zhao, Y., Londono, P., Cao, Y., Sharpe, E.J., Proenza, C., O'Rourke, R., Jones, K.L., Jeong, M.Y., Walker, L.A., Buttrick, P.M. et al. (2015) High-efficiency reprogramming of fibroblasts into cardiomyocytes requires suppression of pro-fibrotic signalling. *Nat. Commun.*, **6**, 8243.
13. Risebro, C.A., Smart, N., Dupays, L., Breckenridge, R., Mohun, T.J. and Riley, P.R. (2006) Hand1 regulates cardiomyocyte proliferation versus differentiation in the developing heart. *Development*, **133**, 4595–4606.
14. Bondue, A., Lapouge, G., Paulissen, C., Semeraro, C., Iacovino, M., Kyba, M. and Blanpain, C. (2008) Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. *Cell Stem. Cell*, **3**, 69–84.
15. Lin, X., Swedlund, B., Ton, M.N., Ghazanfar, S., Guibentif, C., Paulissen, C., Baudelet, E., Plaindoux, E., Achouri, Y., Calonne, E. et al. (2022) Mesp1 controls the chromatin and enhancer landscapes essential for spatiotemporal patterning of early cardiovascular progenitors. *Nat. Cell Biol.*, **24**, 1114–1128.
16. Liu, Z., Chen, O., Zheng, M., Wang, L., Zhou, Y., Yin, C., Liu, J. and Qian, L. (2016) Re-patterning of H3K27me3, H3K4me3 and DNA methylation during fibroblast conversion into induced cardiomyocytes. *Stem. Cell Res.*, **16**, 507–518.
17. Ai, S., Peng, Y., Li, C., Gu, F., Yu, X., Yue, Y., Ma, Q., Chen, J., Lin, Z., Zhou, P. et al. (2017) EED orchestration of heart maturation through interaction with HDACs is H3K27me3-independent. *Elife*, **6**, e24570.
18. Liu, J., Liu, S., Han, L., Sheng, Y., Zhang, Y., Kim, I.M., Wan, J. and Yang, L. (2021) LncRNA HBL1 is required for genome-wide PRC2 occupancy and function in cardiogenesis from human pluripotent stem cells. *Development*, **148**, dev199628.
19. Guo, X., Xu, Y., Wang, Z., Wu, Y., Chen, J., Wang, G., Lu, C., Jia, W., Xi, J., Zhu, S. et al. (2018) A linc1405/eomes complex promotes cardiac mesoderm specification and cardiogenesis. *Cell Stem Cell*, **22**, 893–908.
20. Klattenhoff, C.A., Scheuermann, J.C., Surface, L.E., Bradley, R.K., Fields, P.A., Steinhauer, M.L., Ding, H., Butty, V.L., Torrey, L., Haas, S. et al. (2013) Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell*, **152**, 570–583.
21. George, M.R., Duan, Q., Nagle, A., Kathiriyai, I.S., Huang, Y., Rao, K., Haldar, S.M. and Bruneau, B.G. (2019) Minimal in vivo requirements for developmentally regulated cardiac long intergenic non-coding RNAs. *Development*, **146**, dev185314.
22. Sorensen, D.W. and van Berlo, J.H. (2020) The role of TGF-beta signaling in cardiomyocyte proliferation. *Curr. Heart Fail Rep.*, **17**, 225–233.
23. Prados, B., Gomez-Apinaniz, P., Papoutsis, T., Luxan, G., Zaffran, S., Perez-Pomares, J.M. and de la Pompa, J.L. (2018) Myocardial bmp2 gain causes ectopic EMT and promotes cardiomyocyte proliferation and immaturity. *Cell Death Dis.*, **9**, 399.
24. Dogra, D., Ahuja, S., Kim, H.T., Rasouli, S.J., Stainier, D.Y.R. and Reischauer, S. (2017) Opposite effects of activin type 2 receptor ligands on cardiomyocyte proliferation during development and repair. *Nat. Commun.*, **8**, 1902.
25. Cai, W., Albini, S., Wei, K., Willems, E., Guzzo, R.M., Tsuda, M., Giordani, L., Spiering, S., Kurian, L., Yeo, G.W. et al. (2013) Coordinate nodal and BMP inhibition directs Baf60c-dependent cardiomyocyte commitment. *Genes. Dev.*, **27**, 2332–2344.
26. Yuasa, S., Itabashi, Y., Koshimizu, U., Tanaka, T., Sugimura, K., Kinoshita, M., Hattori, F., Fukami, S., Shimazaki, T., Ogawa, S. et al. (2005) Transient inhibition of BMP signaling by noggin induces cardiomyocyte differentiation of mouse embryonic stem cells. *Nat. Biotechnol.*, **23**, 607–611.
27. Desgrange, A., Le Garrec, J.F., Bernheim, S., Bonnelykke, T.H. and Meilhac, S.M. (2020) Transient nodal signaling in left precursors coordinates opposed asymmetries shaping the heart loop. *Dev. Cell*, **55**, 413–431.
28. Chen, X., Qin, L., Liu, Z., Liao, L., Martin, J.F. and Xu, J. (2015) Knockout of SRC-1 and SRC-3 in mice decreases cardiomyocyte proliferation and causes a noncompaction cardiomyopathy phenotype. *Int. J. Biol. Sci.*, **11**, 1056–1072.
29. Chen, H., Shi, S., Acosta, L., Li, W., Lu, J., Bao, S., Chen, Z., Yang, Z., Schneider, M.D., Chien, K.R. et al. (2004) BMP10 is essential for maintaining cardiac growth during murine cardiogenesis. *Development*, **131**, 2219–2231.
30. Wengrofsky, P., Armenia, C., Oleszak, F., Kupferstein, E., Rednam, C., Mitre, C.A. and McFarlane, S.I. (2019) Left ventricular trabeculation and noncompaction cardiomyopathy: a review. *EC Clin. Exp. Anat.*, **2**, 267–283.
31. Pauli, A., Rinn, J.L. and Schier, A.F. (2011) Non-coding RNAs as regulators of embryogenesis. *Nat. Rev. Genet.*, **12**, 136–149.
32. Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A. and Rinn, J.L. (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.*, **25**, 1915–1927.
33. Perry, R.B. and Ulitsky, I. (2016) The functions of long noncoding RNAs in development and stem cells. *Development*, **143**, 3882–3894.
34. Daneshvar, K., Pondick, J.V., Kim, B.M., Zhou, C., York, S.R., Macklin, J.A., Abualteen, A., Tan, B., Sigova, A.A., Marcho, C. 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**, 353–365.
35. Daneshvar, K., Ardehali, M.B., Klein, I.A., Hsieh, F.K., Kratkiewicz, A.J., Mahpour, A., Cancelliere, S.O.L., Zhou, C., Cook, B.M., Li, W. et al. (2020) lncRNA DIGIT and BRD3 protein form phase-separated condensates to regulate endoderm differentiation. *Nat. Cell Biol.*, **22**, 1211–1222.
36. Papoutsoglou, P., Tsubakihara, Y., Caja, L., Moren, A., Pallis, P., Ameer, A., Heldin, C.H. and Moustakas, A. (2019) The TGFB2-AS1 lncRNA regulates TGF-beta signaling by modulating corepressor activity. *Cell Rep.*, **28**, 3182–3198.
37. Arase, M., Horiguchi, K., Ehata, S., Morikawa, M., Tsutsumi, S., Aburatani, H., Miyazono, K. and Koinuma, D. (2014) Transforming growth factor-beta-induced lncRNA-Smad7 inhibits apoptosis of mouse breast cancer jygmc(A) cells. *Cancer Sci.*, **105**, 974–982.
38. Song, C., Wang, J., Ma, Y., Yang, Z., Dong, D., Li, H., Yang, J., Huang, Y., Plath, M., Ma, Y. et al. (2018) Linc-smad7 promotes myoblast differentiation and muscle regeneration via sponging miR-125b. *Epigenetics*, **13**, 591–604.
39. Dominguez, D., Freese, P., Alexis, M.S., Su, A., Hochman, M., Palden, T., Bazile, C., Lambert, N.J., Van Nostrand, E.L., Pratt, G.A. et al. (2018) Sequence, structure, and context preferences of human RNA binding proteins. *Mol. Cell*, **70**, 854–867.



40. Constanty, F. and Shkumatava, A. (2021) lncRNAs in development and differentiation: from sequence motifs to functional characterization. *Development*, **148**, dev182741.
41. Yin, Y., Yan, P., Lu, J., Song, G., Zhu, Y., Li, Z., Zhao, Y., Shen, B., Huang, X., Zhu, H. *et al.* (2015) Opposing roles for the lncRNA haunt and its genomic locus in regulating HOXA gene activation during embryonic stem cell differentiation. *Cell Stem Cell*, **16**, 504–516.
42. Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K. and Henikoff, S. (2019) CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.*, **10**, 1930.
43. Chu, C., Quinn, J. and Chang, H.Y. (2012) Chromatin isolation by RNA purification (ChIRP). *J. Vis. Exp.*, <https://doi.org/10.3791/3912>.
44. Liu, X. and Wang, H.W. (2011) Single particle electron microscopy reconstruction of the exosome complex using the random conical tilt method. *J. Vis. Exp.*, <https://doi.org/10.3791/2574>.
45. Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N.J. and Dean, C. (2013) R-loop stabilization represses antisense transcription at the arabidopsis FLC locus. *Science*, **340**, 619–621.
46. Xu, W., Xu, H., Li, K., Fan, Y., Liu, Y., Yang, X. and Sun, Q. (2017) The R-loop is a common chromatin feature of the arabidopsis genome. *Nat. Plants*, **3**, 704–714.
47. Li, Y., Song, Y., Xu, W., Li, Q., Wang, X., Li, K., Wang, J., Liu, Z., Velychko, S., Ye, R. *et al.* (2020) R-loops coordinate with SOX2 in regulating reprogramming to pluripotency. *Sci. Adv.*, **6**, eaba0777.
48. Reissmann, E., Jornvall, H., Blokzijl, A., Andersson, O., Chang, C., Minchiotti, G., Persico, M.G., Ibanez, C.F. and Brivanlou, A.H. (2001) The orphan receptor ALK7 and the activin receptor ALK4 mediate signaling by nodal proteins during vertebrate development. *Genes Dev.*, **15**, 2010–2022.
49. Yeo, C. and Whitman, M. (2001) Nodal signals to smads through Cripto-dependent and Cripto-independent mechanisms. *Mol. Cell*, **7**, 949–957.
50. Bonilla-Claudio, M., Wang, J., Bai, Y., Klysiak, E., Selever, J. and Martin, J.F. (2012) Bmp signaling regulates a dose-dependent transcriptional program to control facial skeletal development. *Development*, **139**, 709–719.
51. Bernardo, A.S., Faial, T., Gardner, L., Niakan, K.K., Ortmann, D., Senner, C.E., Callery, E.M., Trotter, M.W., Hemberger, M., Smith, J.C. *et al.* (2011) BRACHYURY and CDX2 mediate BMP-induced differentiation of human and mouse pluripotent stem cells into embryonic and extraembryonic lineages. *Cell Stem Cell*, **9**, 144–155.
52. Valera, E., Isaacs, M.J., Kawakami, Y., Izpisua Belmonte, J.C. and Choe, S. (2010) BMP-2/6 heterodimer is more effective than BMP-2 or BMP-6 homodimers as inducer of differentiation of human embryonic stem cells. *PLoS One*, **5**, e11167.
53. Xi, Q., Wang, Z., Zaromytidou, A.I., Zhang, X.H., Chow-Tsang, L.F., Liu, J.X., Kim, H., Barlas, A., Manova-Todorova, K., Kaartinen, V. *et al.* (2011) A poised chromatin platform for TGF-beta access to master regulators. *Cell*, **147**, 1511–1524.
54. Murry, C.E. and Keller, G. (2008) Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell*, **132**, 661–680.
55. Kattman, S.J., Witty, A.D., Gagliardi, M., Dubois, N.C., Niapour, M., Hotta, A., Ellis, J. and Keller, G. (2011) Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. *Cell Stem Cell*, **8**, 228–240.
56. Luo, M., Bai, J., Liu, B., Yan, P., Zuo, F., Sun, H., Sun, Y., Xu, X., Song, Z., Yang, Y. *et al.* (2019) H3K18ac primes mesendodermal differentiation upon nodal signaling. *Stem Cell Rep.*, **13**, 642–656.
57. Zorzan, I., Pellegrini, M., Arboit, M., Incarnato, D., Maldotti, M., Forcato, M., Tagliazucchi, G.M., Carbognin, E., Montagner, M., Oliviero, S. *et al.* (2020) The transcriptional regulator ZNF398 mediates pluripotency and epithelial character downstream of TGF-beta in human PSCs. *Nat. Commun.*, **11**, 2364.
58. Cloutier, S.C., Wang, S., Ma, W.K., Al Husini, N., Dhoondia, Z., Ansari, A., Pascuzzi, P.E. and Tran, E.J. (2016) Regulated formation of lncRNA-DNA hybrids enables faster transcriptional induction and environmental adaptation. *Mol. Cell*, **62**, 148.
59. Chedin, F. (2016) Nascent connections: R-Loops and chromatin patterning. *Trends Genet.*, **32**, 828–838.
60. Li, Y., Syed, J. and Sugiyama, H. (2016) RNA-DNA triplex formation by long noncoding RNAs. *Cell Chem. Biol.*, **23**, 1325–1333.
61. Postepska-Igielska, A., Giwojna, A., Gasri-Plotnitsky, L., Schmitt, N., Dold, A., Ginsberg, D. and Grummt, I. (2015) lncRNA khps1 regulates expression of the Proto-oncogene SPHK1 via triplex-mediated changes in chromatin structure. *Mol. Cell*, **60**, 626–636.
62. Kalwa, M., Hanzelmann, S., Otto, S., Kuo, C.C., Franzen, J., Jousens, S., Fernandez-Rebollo, E., Rath, B., Koch, C., Hofmann, A. *et al.* (2016) The lncRNA HOTAIR impacts on mesenchymal stem cells via triple helix formation. *Nucleic Acids Res.*, **44**, 10631–10643.
63. Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823–837.
64. Harikumar, A. and Meshorer, E. (2015) Chromatin remodeling and bivalent histone modifications in embryonic stem cells. *EMBO Rep.*, **16**, 1609–1619.
65. Bogliotti, Y.S. and Ross, P.J. (2012) Mechanisms of histone H3 lysine 27 trimethylation remodeling during early mammalian development. *Epigenetics*, **7**, 976–981.
66. Schertzer, M.D., Bracer, K.C.A., Starmer, J., Cherney, R.E., Lee, D.M., Salazar, G., Justice, M., Bischoff, S.R., Cowley, D.O., Ariel, P. *et al.* (2019) lncRNA-Induced Spread of polycomb controlled by genome architecture, RNA abundance, and CpG island DNA. *Mol. Cell*, **75**, 523–537.
67. Pintacuda, G., Wei, G., Roustan, C., Kirmizitas, B.A., Solcan, N., Cerase, A., Castello, A., Mohammed, S., Moindrot, B., Nesterova, T.B. *et al.* (2017) hnRNPK recruits PGC3/5-PRC1 to the xist RNA B-Repeat to establish polycomb-mediated chromosomal silencing. *Mol. Cell*, **68**, 955–969.
68. Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E. and Chang, H.Y. (2015) Systematic discovery of xist RNA binding proteins. *Cell*, **161**, 404–416.
69. Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J.T., Sadreyev, R.I., Dekker, J. and Kingston, R.E. (2017) Polycomb repressive complex 1 generates discrete compacted domains that change during differentiation. *Mol. Cell*, **65**, 432–446.
70. Oksuz, O., Narendra, V., Lee, C.H., Descostes, N., LeRoy, G., Raviram, R., Blumenberg, L., Karch, K., Rocha, P.P., Garcia, B.A. *et al.* (2018) Capturing the onset of PRC2-Mediated repressive domain formation. *Mol. Cell*, **70**, 1149–1162.
71. Zhang, S., Chen, T., Chen, N., Gao, D., Shi, B., Kong, S., West, R.C., Yuan, Y., Zhi, M., Wei, Q. *et al.* (2019) Implantation initiation of self-assembled embryo-like structures generated using three types of mouse blastocyst-derived stem cells. *Nat. Commun.*, **10**, 496.
72. Senft, A.D., Bikoff, E.K., Robertson, E.J. and Costello, I. (2019) Genetic dissection of nodal and bmp signalling requirements during primordial germ cell development in mouse. *Nat Commun*, **10**, 1089.
73. Meng, X.M., Chung, A.C. and Lan, H.Y. (2013) Role of the TGF-beta/BMP-7/Smad pathways in renal diseases. *Clin. Sci. (Lond.)*, **124**, 243–254.
74. Wu, M., Chen, G. and Li, Y.P. (2016) TGF-beta and BMP signaling in osteoblast, skeletal development, and bone formation, homeostasis and disease. *Bone Res.*, **4**, 16009.
75. Massague, J. (2008) TGFbeta in cancer. *Cell*, **134**, 215–230.
76. Sartori, R., Schirwis, E., Blaauw, B., Bortolanza, S., Zhao, J., Enzo, E., Stantzou, A., Mousel, E., Toniolo, L., Ferry, A. *et al.* (2013) BMP signaling controls muscle mass. *Nat. Genet.*, **45**, 1309–1318.
77. Candia, A.F., Watabe, T., Hawley, S.H., Onichtchouk, D., Zhang, Y., Derynck, R., Niehrs, C. and Cho, K.W. (1997) Cellular interpretation of multiple TGF-beta signals: intracellular antagonism between activin/BVg1 and BMP-2/4 signaling mediated by smads. *Development*, **124**, 4467–4480.
78. Yuan, G., Zhan, Y., Gou, X., Chen, Y. and Yang, G. (2018) TGF-beta signaling inhibits canonical BMP signaling pathway during palate development. *Cell Tissue Res.*, **371**, 283–291.
79. Gronroos, E., Kingston, I.J., Ramachandran, A., Randall, R.A., Vizan, P. and Hill, C.S. (2012) Transforming growth factor beta inhibits bone morphogenetic protein-induced transcription through novel phosphorylated smad1/5-smad3 complexes. *Mol. Cell Biol.*, **32**, 2904–2916.
80. Michos, O., Panman, L., Vintersten, K., Beier, K., Zeller, R. and Zuniga, A. (2004) Gremlin-mediated BMP antagonism induces the

- epithelial-mesenchymal feedback signaling controlling metanephric kidney and limb organogenesis. *Development*, **131**, 3401–3410.
81. Graindorge, A., Pinheiro, I., Nawrocka, A., Mallory, A.C., Tsvetkov, P., Gil, N., Carolis, C., Buchholz, F., Ulitsky, I., Heard, E. *et al.* (2019) In-cell identification and measurement of RNA-protein interactions. *Nat. Commun.*, **10**, 5317.
  82. Dossin, F., Pinheiro, I., Zylicz, J.J., Roensch, J., Collombet, S., Le Saux, A., Chelmicki, T., Attia, M., Kapoor, V., Zhan, Y. *et al.* (2020) SPEN integrates transcriptional and epigenetic control of X-inactivation. *Nature*, **578**, 455–460.
  83. Shan, Y., Liang, Z., Xing, Q., Zhang, T., Wang, B., Tian, S., Huang, W., Zhang, Y., Yao, J., Zhu, Y. *et al.* (2017) PRC2 specifies ectoderm lineages and maintains pluripotency in primed but not naive ESCs. *Nat. Commun.*, **8**, 672.
  84. Miller, S.A., Damle, M., Kim, J. and Kingston, R.E. (2021) Full methylation of H3K27 by PRC2 is dispensable for initial embryoid body formation but required to maintain differentiated cell identity. *Development*, **148**, dev196329.
  85. Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A. *et al.* (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 11667–11672.
  86. Fang, H., Bonora, G., Lewandowski, J.P., Thakur, J., Filippova, G.N., Henikoff, S., Shendure, J., Duan, Z., Rinn, J.L., Deng, X. *et al.* (2020) Trans- and cis-acting effects of fibre on epigenetic features of the inactive x chromosome. *Nat. Commun.*, **11**, 6053.
  87. Cai, Z., Cao, C., Ji, L., Ye, R., Wang, D., Xia, C., Wang, S., Du, Z., Hu, N., Yu, X. *et al.* (2020) RIC-seq for global in situ profiling of RNA-RNA spatial interactions. *Nature*, **582**, 432–437.
  88. Corley, M., Burns, M.C. and Yeo, G.W. (2020) How RNA-binding proteins interact with RNA: molecules and mechanisms. *Mol. Cell*, **78**, 9–29.