NAR Breakthrough Article

YY1 safeguard multidimensional epigenetic landscape associated with extended pluripotency

Xiaotao Dong[1,2,3,5,†], **Rong Guo**[1,2,3,4,5,†], **Tianrong Ji**[1,2,3,5], **Jie Zhang**[1,2,3,4,5], **Jun Xu**[6], **Yaoyi Li**[1,2,3,5], **Yingliang Sheng**[1,2,3], **Yuxiang Wang**[1,2,3,4,5], **Ke Fang**[1,2,3,5], **Yulin Wen**[1,2,3,4,5], **Bei Liu**[6], **Gongcheng Hu**[1,2,3,5], **Hongkui Deng**[6] **and Hongjie Yao** ⬤[1,2,3,4,5,*]

[1]CAS Key Laboratory of Regenerative Biology, Joint School of Life Sciences, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou Medical University, Guangzhou 510530, China, [2]Bioland Laboratory (Guangzhou Regenerative Medicine and Health GuangDong Laboratory), Guangzhou 510005, China, [3]Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, GIBH-CUHK Joint Research Laboratory on Stem Cell and Regenerative Medicine, State Key Laboratory of Respiratory Disease, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China, [4]University of Chinese Academy of Sciences, Beijing 100049, China, [5]Institute of Stem Cell and Regeneration, Chinese Academy of Sciences, Beijing 100101, China and [6]School of Life Sciences, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China

ABSTRACT

Although extended pluripotent stem cells (EPSCs) have the potential to form both embryonic and extraembryonic lineages, how their transcriptional regulatory mechanism differs from that of embryonic stem cells (ESCs) remains unclear. Here, we discovered that YY1 binds to specific open chromatin regions in EPSCs. *Yy1* depletion in EPSCs leads to a gene expression pattern more similar to that of ESCs than control EPSCs. Moreover, *Yy1* depletion triggers a series of epigenetic crosstalk activities, including changes in DNA methylation, histone modifications and high-order chromatin structures. *Yy1* depletion in EPSCs disrupts the enhancer-promoter (EP) interactions of EPSC-specific genes, including *Dnmt3l*. *Yy1* loss results in DNA hypomethylation and dramatically reduces the enrichment of H3K4me3 and H3K27ac on the promoters of EPSC-specific genes by upregulating the expression of *Kdm5c* and *Hdac6* through facilitating the formation of CCCTC-binding factor (CTCF)-mediated EP interactions surrounding their loci. Furthermore, single-cell RNA sequencing (scRNA-seq) experiments revealed that YY1 is required for the derivation of extraembryonic endoderm (XEN)-like cells from EPSCs *in vitro*. Together, this study reveals that YY1 functions as a key regulator of multidimensional epigenetic crosstalk associated with extended pluripotency.

GRAPHICAL ABSTRACT
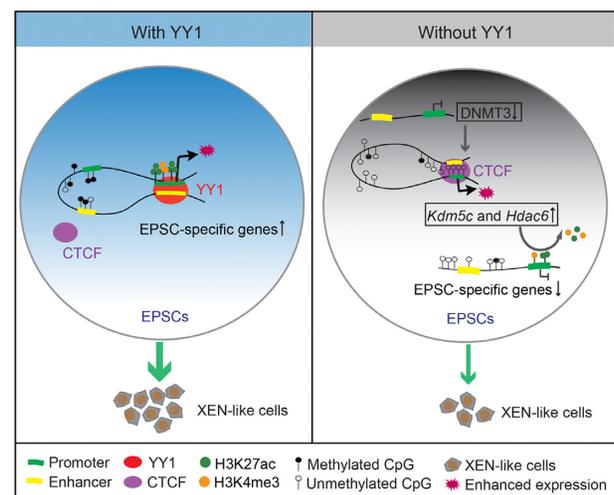
*To whom correspondence should be addressed. Tel: +86 20 32015279; Email: yao_hongjie@gibh.ac.cn
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

INTRODUCTION

Mouse blastocysts consist of the following three lineages before implantation: the epiblast (EPI), which is a precursor of all embryonic germ layers, and extraembryonic lineages, including primitive endoderm (PrE) and trophectoderm (TE) (1,2). Embryonic stem cells (ESCs) and trophoblast stem cells (TSCs) can be derived from the EPI and TE, respectively, while extraembryonic endoderm (XEN) cells are derived from PrE (3). Traditionally, ESCs are cultured in a defined medium containing two inhibitors, PD0325901 and CHIR99021 (4), which mainly contribute to embryonic lineages but have limited capacity in developing extraembryonic tissues (5,6). More recently, ESCs with expanded developmental potential designated extended pluripotent stem cells (EPSCs) or expanded potential stem cells capable of contributing to both embryonic and extraembryonic lineages in chimeras have been derived (7–10). Indeed, EPSCs could be used to derive both TSCs and XEN cells *in vitro* (7,11). Furthermore, blastocyst-like structures can be generated from EPSCs, providing a unique *in vitro* model for understanding early embryogenesis (12,13).

The stability of EPSCs at both the genetic and epigenetic levels may contribute to the robust developmental potency of these cells (14). Compared with ESCs, EPSCs show significantly lower levels of copy number variation (CNV) mutations after long-term culturing with improved genetic stability (14). EPSCs have higher global levels of 5mC modification on DNA than ESCs (7) and still retain normal genomic imprinting at the *H19* and *Snrpn* loci after long-term culturing, suggesting their epigenetic stability (14). In addition, EPSCs have more bivalent genes marked by histone H3K4me3 and H3K27me3 than ESCs (7). Except these, there is still less study on EPSC epigenetic features in the field.

YY1 is a Gli–Kruppel-type zinc finger transcription factor (15) that binds active enhancers and promoter-proximal elements and forms dimers that facilitate the EP interactions of these DNA elements (16). Moreover, YY1 participates in the regulation of imprinted genes (17,18) and X-chromosome inactivation (XCI) (19). *Yy1* knockdown during oogenesis resulted in the loss of DNA methylation on *Peg3* and *Xist* (20). YY1 is essential for early embryogenesis and adult tissue formation (21), and its deficiency causes peri-implantation lethality during mouse embryonic development (22). In addition, YY1 positively regulates transcription by targeting promoters and super-enhancers through the BAF complex in ESCs (23). However, the roles of YY1 in the regulation of the expanded developmental potential of EPSCs remain unclear.

In this study, we found that *Yy1* depletion in EPSCs triggered a series of epigenetic crosstalk among DNA methylation, histone modifications and high-order chromatin structures. *Yy1* knockdown reduced the EP interactions among placental development-specific genes. *Yy1* depletion globally reduced DNA methylation, facilitated the binding of CTCF to hypomethylated DNA regions in EPSCs and further enhanced the CTCF-mediated EP interactions of the histone H3K4me3 demethylase gene *Kdm5c* and the histone deacetylase gene *Hdac6*, thereby promoting the expression of these genes. These changes led to the reduced enrichment of H3K4me3 and H3K27ac on EPSC-specific gene promoters. Moreover, the *in vitro* F4H-induced differentiation experiments in combination with the scRNA-seq analysis suggested that YY1 plays an important role in the regulation of the differentiation of EPSCs toward an XEN-like cell phenotype.

MATERIALS AND METHODS

Mice

ICR mice were purchased from Beijing Vital River Laboratory and used to generate mouse embryonic fibroblast (MEF) feeder cells. The animals were individually housed under a 12 h light/dark cycle and provided food and water ad libitum. The animal experiments were performed according to the Guidelines for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocols were approved by the Committee on the Ethics of Animal Experiments of Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences.

Cell culture

HEK293T and MEF feeder cells were maintained in DMEM high-glucose media (HyClone, SH30022.01) with 10% FBS (Excell) under 5% CO_2 at 37°C. Mouse EPSCs and ESCs were cultured in N2B27-LCDM medium and 2i medium, respectively, under 5% CO_2 at 37°C as previously described (8). In total, 500 ml of N2B27 medium were prepared with 240 ml of DMEM/F12 (Thermo Fisher Scientific, 11330–032), 240 ml of neurobasal (Thermo Fisher Scientific, 21103-049), 2.5 ml of N2 supplement (Thermo Fisher Scientific, 17502-048), 5 ml of B27 supplement (Thermo Fisher Scientific, 12587–010), 1% GlutaMAX (Gibco, 35050061), 1% nonessential amino acids (Gibco, 11140050), 0.1 mM β-mercaptoethanol (Life Technologies, 21985023) and 1% penicillin–streptomycin (Gibco, 15070063). To prepare the N2B27-LCDM medium, small molecules and cytokines were added to the N2B27 medium at the following final concentrations: 10 ng/ml recombinant human LIF (L, Peprotech, 300-05), 3 μM CHIR99021 (C, Selleck, S2924), 2 μM (*S*)-(+)-dimethindene maleate (D, Tocris, 1425) and 2 μM minocycline hydrochloride (M, Santa Cruz Biotechnology, sc-203339). The 2i medium contains N2B27 medium with 10 ng/ml recombinant human LIF (Peprotech, 300-05), 3 μM CHIR99021 (Selleck, S2924) and 1 μM PD0325901 (Selleck, S1036). Both EPSCs and ESCs were maintained on mitomycin C-inactivated MEF feeder cells ($\sim 3 \times 10^4$ cells per cm^2)-coated dishes. The medium was changed every day, and the cells were passaged with 0.25% trypsin–EDTA (Gibco, 25200114). Both EPSCs and ESCs are clonal cell lines with the same genetic background (8).

Mouse XEN cells, which were provided by Prof. Jie Na at the School of Medicine, Tsinghua University, were cultured as previously described (24). The XEN cells were cultured in 70% MEF-conditioned and 30% fresh IDG medium without any additional supplements on gelatin-coated plates. The IDG medium contained high-glucose DMEM (HyClone, SH30022.01), 15% FBS (Gibco, 10082147), 1%

NEAA (Gibco, 11140050), 0.1 mM β-mercaptoethanol (Life Technologies, 21985023), 1 mM GlutaMAX (Gibco, 35050061), 1 mM sodium pyruvate (Gibco, 11360–070) and 20 mM HEPES (Gibco, 15630080). The XEN cells were passaged using 0.25% trypsin–EDTA (Gibco, 25200114) and replated at a 1:20 ratio once the cells reached approximately 80% confluency.

Plasmid construction

For the generation of short hairpin RNA (shRNA) vectors, the following fragments were inserted into the pLKO.1 vector (25). The shRNA sequences targeting *Yy1* were as follows: sh*Yy1*-1#: 5′-CCCTAAGCAACTGGCAGAATT-3′ and sh*Yy1*-2#: 5′-CGACGGTTGTAATAAGAAGTT-3′. For overexpression, *Yy1* cDNAs were cloned into the pSin-Flag vector (26). The plasmids used for the transfections were purified with a HiPure Plasmid EF Mini Kit (Magen, P1112-03). The sequences of the shRNA oligos used in this study are described in Supplementary Table S1. The primers used for the *Yy1* CDS amplification are listed in Supplementary Table S2.

Generation of *Yy1*-depleted EPS stable cell lines and *Yy1* stably overexpressing ES cell lines

Lentiviral supernatants for pLKO.1-Ctrl, pLKO.1-sh*Yy1*-1#, pLKO.1-sh*Yy1*-2#, pSin-Flag and pSin-Flag-*Yy1* were generated with the assistance of psPAX2, pMD2.G in HEK293T cells. EPSCs grown on MEF feeder-coated dishes were infected with lentiviruses generated from pLKO.1-Ctrl, pLKO.1-sh*Yy1*-1#, and pLKO.1-sh*Yy1*-2#. The positive cells were selected with 2 μg/ml puromycin (25,26). Then, single colonies were picked up and expanded individually on gelatin-coated dishes in N2B27-LCDM medium containing 2 μg/ml puromycin. ESCs grown on MEF feeders-coated dishes were infected with lentivirus produced from pSin-Flag or pSin-Flag-*Yy1*. The cells were selected with 2 μg/ml puromycin. Then single colonies were picked up and expanded on gelatin-coated dishes in 2i medium with 2 μg/ml puromycin.

Stable control shRNA-treated and two *Yy1* shRNA-depleted EPSC lines were grown on MEF feeder-coated dishes in N2B27-LCDM medium, and stable Flag-control and Flag-*Yy1* ESC lines were grown on MEF feeders-coated dishes in 2i medium. For the functional analysis, the cells were cultured for one passage on gelatin-coated dishes in medium with 2 μg/ml puromycin to further select positive cells and remove the MEF feeders.

Differentiation of extraembryonic cells from ESCs or EPSCs

ESCs and EPSCs were cultured on gelatin-coated plates for three passages to remove residual MEF feeders. The cells were plated into six-well plates with 1×10^4 cells per well in TS basic medium with slight modifications, which was called F4H medium (11,27,28), containing 30% MEF medium [DMEM high-glucose media (HyClone, SH30022.01) with 10% FBS (Gibco, 10082147), 2 mM L-glutamine (Gibco, 25030–81), 1% NEAA (Gibco, 11140050) and 1% penicillin–streptomycin

(Gibco, 15070063)], 70% TS basic medium [RPMI-1640 (Gibco, C11875500BT), 20% FBS (Gibco, 10082147), 1% penicillin–streptomycin (Gibco, 15070063), 2 mM L-glutamine (Gibco, 25030-81), 1 mM sodium pyruvate (Gibco, 11360-070), and 0.1 mM β-mercaptoethanol (Life Technologies, 21985023)] (29), 25 ng/ml FGF4 (Stem Cell, 78103.1) and 1 μg/ml heparin (Stem Cell, 07980).

Cell viability assay

Approximately 1000 control and *Yy1*-depleted EPSCs (without MEF feeders) were plated into 96-well plates coated with gelatin in 100 μl of N2B27-LCDM medium. The cell viability was detected at the indicated time points by using Cell Counting Kit-8 (CCK8, Beyotime, C0039) according to the manufacturer's instructions. In brief, 10 μl of CCK8 solution were added to each well and incubated for 3 h, and then, the absorbance at 450 nm was measured. Wells with no cells were used as a control. The data are presented as the mean ± SD of two independent experiments, each with five technical repeats.

Colony formation assay

In total, 1000 control and *Yy1*-depleted EPSCs (without MEF feeders) were seeded into individual wells of a six-well plate. Alkaline phosphate (AP) staining was performed with a BCIP/NBT Alkaline Phosphatase Color Development Kit (Beyotime, C3206) after 5 days of culture according to the manufacturer's instructions. The number of clones was counted. The data are presented as the mean ± SD of three independent experiments.

Genomic DNA isolation and DNA methylation analysis

Genomic DNA (gDNA) was extracted with a TIANamp Genomic DNA Kit (Tiangen, RT121). The DNA methylation analysis was performed as previously described (30). Two micrograms of gDNA were digested overnight at 37°C with 10 U HpaII (NEB, R0171S) or McrBC (NEB, M0272L) in a 20 μl total reaction volume. Digested gDNA was loaded on a 0.8% agarose gel, and gDNA fragments were separated by electrophoresis. The quantification of the DNA signal in the gel was performed using ImageJ software.

Quantitative RT-PCR

The cells were collected and washed with PBS, and the total RNA was extracted using TRIzol® reagent (MRC, TR118-200). RNA was extracted with phenol/chloroform, precipitated with ethanol, resuspended in RNase-free water and quantified using a Nanodrop (Thermo Fisher Scientific). For the quantitative PCR, cDNAs were synthesized with HiScript® III RT SuperMix for qPCR (+ gDNA wiper) (Vazyme Biotech, R323-01). Real-time PCR was performed using SYBR Green mix (Genstar, A301-01) on a CFX96 real-time PCR system (Bio-Rad) according to the manufacturer's instructions. The data were analyzed by using the $\Delta\Delta Ct$ method (31), and *Gapdh* was used as an internal control. The primers used for the RT-qPCR assays are listed in Supplementary Table S3.

Western blot analysis

Cells were resuspended in RIPA buffer [1% Triton X-100, 0.1% sodium dodecyl sulfate (SDS), 150 mM KCl, 50 mM Tris–HCl (pH 7.4), 1 mM EDTA and 1% sodium deoxycholate], 1 mM PMSF, and 1 × protease inhibitor cocktails. The total soluble proteins were obtained by centrifugation at 15 294 × g for 10 min. The samples were separated on an SDS-PAGE gel and transferred onto a polyvinyl difluoride (PVDF) membrane (Millipore, IPVH00010). The PVDF membrane was blocked with 5% milk in TBS-T buffer. Immunoblot analysis was performed with the indicated antibodies. Then the membrane was washed with TBS-T buffer and immunoblotted. The antibodies used in this study included anti-YY1 antibody (Abcam, ab109237), anti-β-Actin antibody (Sigma-Aldrich, A2228), anti-DNMT3A antibody (Proteintech, 19366-1-AP), anti-DNMT3B antibody (Abcam, ab2851), and anti-DNMT3L antibody (Cell Signaling Technology, 13451).

RNA sequencing (RNA-seq) and bioinformatics analysis

The total RNA was extracted as described above. RNA sequencing libraries were constructed using a VAHTS mRNA-seq V3 Library Prep Kit (Vazyme Biotech, NR611). Two rounds of mRNA purification were performed to guarantee the removal of rRNA. In brief, 50 μl of mRNA capture beads were incubated with 1.5 μg of total RNA at 65°C for 5 min and then 25°C for 5 min. The supernatant was discarded, and 200 μl of beads wash buffer were added to the clean beads. Fifty microliters of Tris buffer were added to resuspend the beads, and the sample was incubated at 80°C for 5 min to release mRNA. Then, 50 μl of beads binding buffer were added to facilitate the binding of mRNA to the beads. Furthermore, ribosomal-depleted mRNAs were fragmented at 85°C for 6 min, and cDNAs were synthesized. The cDNA was purified with AMPure XP beads (Beckman Coulter, A63882), followed by end repair, adaptor-ligation, size selection of the library, and library amplification. The libraries were purified using AMPure XP beads and then sequenced on Illumina NovaSeq (Annoroad Gene Technology Co., Ltd.).

The raw reads were mapped to the mouse genome (mm10) using the STAR aligner (32). The transcript identification and counting were processed by HTSeq (33). The DESeq2 package was used to analyze the differentially expressed genes (DEGs) (34). The top Gene Ontology (GO) processes were enriched by the Metascape web-based platform (35).

scRNA-seq and bioinformatics analysis

For the 10× Genomics scRNA-seq, the cells derived from the EPSCs with or without *Yy1* depletion and the ESCs with or without *Yy1* overexpression on day 6 after the F4H treatment were dissociated with 0.25% trypsin–EDTA (Gibco, 25200114) at 37°C for 10 min. Single cell populations were resuspended in PBS with 0.04% BSA. Single-cell libraries were created using Chromium Single Cell 30 Reagent Kits and sequenced on a Novaseq 6000 sequencer. Approximately 10000 cells were captured at each time point.

The gene expression matrices were generated using Cell-Ranger software (version 6.0.2, 10 × Genomics). The following quality control steps were performed: (i) genes expressed by less than two cells were not considered; (ii) cells that expressed fewer than 400 genes and more than 8000 genes were excluded from further analysis and (iii) cells in which over 5% of unique molecular identifiers (UMIs) were derived from the mitochondrial genome were removed. The Seurat package (version 4.0.4) (36) was used to perform the single-cell analysis from whole-cell clustering. Briefly, the data normalization and scaling were performed using the SCTransform function with the default parameters. The single-cell samples were merged with the IntegrateData function with the following parameters: normalization.method = 'SCT' and dims = 1:30. A shared-nearest-neighbors (SNN) graph was constructed using the first 30 principal components before clustering the cells using the FindClusters function with a resolution of 0.4 and the default parameters. We performed a differential expression analysis to define the genes that marked each cluster using the FindconservedMarkers function with the settings min.pct = 0.1 and logfc.threshold = 0.25 using a Wilcoxon test. Each cluster for specific cell type was annotated according to the expression of the markers showing in Supplementary Table S4.

Assay for transposase accessible chromatin using sequencing (ATAC-seq) and bioinformatics analysis

The ATAC-seq experiments were performed as previously described (37). In brief, 50,000 cells were harvested, washed once with 50 μl of cold PBS and then resuspended in 50 μl of lysis buffer [10 mM Tris–HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl₂, and 0.2% (v/v) IGEPAL CA-630]. The suspension of nuclei was centrifuged at 500 × g at 4°C for 10 min. The pellet was resuspended in 50 μl of transposition reaction mix (10 μl of TD buffer, 5 μl of Tn5 transposase and 35 μl of nuclease-free H₂O) and incubated at 37°C for 30 min. Finally, DNA was isolated using a MinElute PCR Purification Kit (Qiagen, 28006). ATAC-seq libraries were constructed and purified with AMPure XP beads (Beckman Coulter, A63882) and then sequenced on HiSeq X-Ten (Annoroad Gene Technology Co., Ltd.). After trimming the adaptor sequence with Cutadapt (v.0.6.1) (38), the ATAC-seq reads were aligned to the mm10 genome by using Bowtie2 (v.2.4.1) (39) with the default parameters. Low-mapping-quality reads were filtered by SAMtools (v.1.9) (40), and duplicate reads were removed by Picard tools (v1.90). MACS2 (41) was used to call the ATAC peaks (p-value < 0.01), and the common peaks were merged by bedtools (v2.25.0) (42). ATAC-seq read counts for each sample were then calculated by the multicov function in bedtools with default settings. The DESeq2 package was used to analyze the differential ATAC peaks (34). Significantly differential ATAC peaks were defined as peaks with a q-value < 0.05 and $|\log_2FC| > 1$. Motif analysis was performed by HOMER using default settings (43). Motifs were only kept if the p-value was < 1e–20 and (the percent of target/the percent of background) was > 2.

ChIP-seq and bioinformatics analysis

The ChIP experiments were performed as previously described (44). In total, 1×10^7 cells were crosslinked with 1% formaldehyde at room temperature for 10 min. Then, the reaction was stopped by adding glycine (final concentration, 0.125 M). The crosslinked cells were lysed in ChIP SDS lysis buffer [1% SDS, 10 mM EDTA, and 50 mM Tris–HCl (pH 8.0)] containing $1 \times$ protease inhibitor cocktail and PMSF and then sonicated to achieve a chromatin size of 200–400 bp. After sonication, we collected 1% of the total supernatant for input, and the remaining supernatant was diluted with IP buffer and then co-incubated with antibody-Dynabeads protein A/G (1:1 mixed) at 4°C overnight with rotation.

ChIP for histones. The crosslinked cells were lysed in 200 μl MNase digesting buffer (0.5% NP-40, 0.5% Tween and 0.1% SDS) containing $1 \times$ protease inhibitor cocktail and PMSF. Then, 200 μl of MNase working buffer [100 mM Tris–HCl (pH 8.0) and 2 mM $CaCl_2$] and 500 U micrococcal nuclease (NEB, M0247S) were added to digest the nucleosome fragments at 37°C. The reaction was stopped by adding 100 μl of MNase stopping buffer [50 mM Tris–HCl (pH 8.0) and 25 mM EDTA]. Then, $2 \times$ RIPA buffer (1% Triton X-100, 280 mM NaCl, and 0.1% SDS) was added, and the tubes were rotated at 4°C for 10 min. After rotation, the sample was diluted with $1 \times$ RIPA buffer [10 mM Tris-HCl (pH 8.0), 140 mM NaCl, 1% Triton X-100 and 0.1% SDS], and 1% of the total sample was used as input. The remaining sample was co-incubated with antibody-Dynabeads protein A/G (1:1 mixed) at 4°C overnight with rotation. The next steps were the same as those described above.

Immune complexes were washed with the following buffers: low salt wash buffer [0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris–HCl (pH 8.0) and 150 mM NaCl], high salt wash buffer [0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris–HCl (pH 8.0) and 500 mM NaCl], LiCl wash buffer [0.25 M LiCl, 1% IGEPAL-CA630, 1% deoxycholic acid (sodium salt), 1 mM EDTA and 10 mM Tris–HCl (pH 8.0)] and TE buffer [10 mM Tris–HCl (pH 8.0) and 1 mM EDTA]. Antibody-bound chromatin was reverse-crosslinked, and the ChIPed DNA samples were purified for the ChIP-seq library generation. The antibodies included YY1 (Abcam, ab109237), CTCF (Active Motif, 61311), H3K27ac (Active Motif, 39133), H3K4me1 (Active Motif, 39297), and H3K4me3 (Millipore, 07-473).

The ChIP-seq libraries were constructed using a VAHTSTM Universal DNA Library Prep Kit for Illumina® V2 (Vazyme Biotech, ND606-01). After the PCR library amplification, size selection of adaptor-ligated DNA was performed using Agecourt AMPure XP Beads (Beckman Coulter, A63882). The libraries were sequenced on Illumina NovaSeq (Annoroad Gene Technology Co., Ltd.). After trimming the adaptor sequence with Cutadapt (v.0.6.1) (38), the ChIP-seq reads were aligned to the mm10 genome by using Bowtie2 (v.2.4.1) (39) with the default parameters. Low-mapping-quality reads were filtered by SAMtools (v.1.9), and duplicate reads were removed by Picard tools (v1.90). MACS2 (41) was used to call the peaks for YY1 ChIP-seq in EPSCs (p-value < 0.001); the other samples

had q-value < 0.01. The differential peaks were identified by the MAnorm pipeline (45), and peaks with high confidence were chosen by a p-value < 0.05 and |$\log_2$M-value| > 0.58. BigWig files were generated by deeptools (46) by the RPKM normalization method and visualized in the WashU Epigenome Browser (47).

CUT&Tag and bioinformatics analysis

The CUT&Tag experiments were performed as previously described (48). In brief, 100 000 cells were harvested and washed twice with 200 μl of wash buffer. 10 μl concanavalin A beads (Bangs Laboratories, BP531) were added per sample and incubated at room temperature for 15 min. Then, 100 μl of dig-wash buffer containing 2 mM EDTA and 1 μg of primary antibody were added. The primary antibody anti-H3K27ac (Active Motif, 39133) was added and incubated on a rotating platform at 4°C overnight. Two hundred microliters of dig-wash buffer were added to remove the unbound antibodies. Then, the reaction was incubated with pAG-Tn5 (homemade) at 4°C for 2 h. Two hundred microliters of dig-med buffer were added to remove unbound pAG-Tn5 protein. Next, the cells were resuspended in 100 μl of tagmentation buffer and incubated at 37°C for 1 h. To stop tagmentation, 2.25 μl of 0.5 M EDTA, 2.75 μl of 10% SDS and 0.5 μl of 20 mg/ml proteinase K were added and incubated at 55°C for 30 min and then at 70°C for 30 min to inactivate proteinase K. Then, DNA was extracted.

To generate the sequencing libraries, 21 μl DNA were mixed with 2 μl of a universal i5 and a uniquely barcoded i7 primer (49) using a different barcode per sample. A volume of 25 μl of NEBNext high-fidelity $2 \times$ PCR master mix (NEB, M0541L) was added and mixed. The sample was placed in a thermocycler with a heated lid using the following cycling conditions: 72°C for 5 min (gap filling); 98°C for 30 s; 14 cycles of 98°C for 10 s and 63°C for 30 s; final extension at 72°C for 1 min; and hold at 8°C. Post-PCR clean-up was performed by adding $1 \times$ volume of Ampure XP beads (Beckman Counter, A63882), and the libraries were incubated with beads at room temperature for 15 min, washed twice gently with 80% ethanol, and eluted in 25 μl of 10 mM Tris (pH 8.0). The libraries were sequenced on Illumina NovaSeq (Annoroad Gene Technology Co., Ltd.). After trimming the adaptor sequence with Cutadapt (v.0.6.1) (38), the CUT&Tag reads were aligned to the mm10 genome by using Bowtie2 (v.2.4.1) (39) with the default parameters. Low-mapping-quality reads were filtered by SAMtools (v.1.9), and duplicate reads were removed by Picard tools (v1.90). MACS2 (41) was used to call peaks with a q-value < 0.01. BigWig files were generated by deeptools (46) by the RPKM normalization method and visualized in the WashU Epigenome Browser (47).

Whole-genome bisulfite sequencing (WGBS) and bioinformatics analysis

DNA was isolated using standard phenol/chloroform extraction and ethanol precipitation. WGBS was performed as previously described (50). The C-T transformation of 1 μg genomic DNA and bisulfite treatment were performed by using an EZ DNA Methylation Gold™ Kit. Then, the

modified DNA was treated as follows: DNA fragment, end repair, A-tailing, and ligation of methylated adapter. The target DNA samples were selected and treated with bisulfite. Then, the WGBS libraries were amplified and sequenced on Illumina NovaSeq (Annoroad Gene Technology Co., Ltd.).

Subsequently, the raw reads were mapped to the computationally bisulfite-converted mouse reference genome (GRCm38/mm10) by using Bismark (51). Potential PCR duplicates were removed using the deduplicate bismark command. Approximately 2 kb tiles were calculated for all chromosomes with a 1 kb offset. UHC of 2 kb tile methylation was performed using the R hclust command with the 'ward' method. The correlation matrix was calculated by the R corrplot package.

Bridge-linker-Hi-C (BL-Hi-C) analysis

The BL-Hi-C libraries were constructed as previously described (52). Briefly, 5×10^6 cells were treated with 1% formaldehyde at room temperature for 10 min, and the crosslink was quenched by adding 2.5 M glycine to a final concentration of 0.2 M. Then, the cells were resuspended with BL-Hi-C lysis buffer 1 and incubated on ice for 15 min. After centrifugation, the cell pellet was resuspended in BL-Hi-C lysis buffer 2 and rotated at 4°C for 15 min. The cell pellet was washed once with BL-Hi-C lysis buffer 1, resuspended with 0.5% SDS and incubated at 62°C. At the end of the incubation, SDS was quenched by adding 10% Triton X-100 and ddH$_2$O incubation at 37°C for 10 min. Subsequently, the genome was digested overnight at 37°C with 100 U of HaeIII (NEB, R0108L) into blunt-end fragments. Cleaved chromatin was A-tailed by adding 10 mM dATP solution (NEB, N0440S) and Klenow Fragment (3′→5′ exo-) (NEB, M0212L) with rotation at 37°C for 40 min. Then, chromatin was treated with adenine and ligated with biotinylated bridge linker S2 (annealed by/5Phos/CGCGATATC/iBIOdT/TATCTGACT and/5Phos/GTCAGATAAGATATCGCGT) at 16°C for 4 h. After centrifugation at 3500 g at 4°C for 5 min and removal of the supernatant, the pellet was resuspended in a mixture of ddH$_2$O, Lambda Exonuclease buffer, Lambda Exonuclease (NEB, M0262L) and Exonuclease I (NEB, M0293L) and rocked at 900 rpm at 37°C for 1 h in ThermoMixer C. Next, the samples were treated with SDS and proteinase K at 55°C overnight to digest the proteins, and the DNA was purified using phenol:chloroform:isoamyl alcohol (25:24:1) extraction followed by ethanol precipitation. Then, the DNA was fragmented into 300 bp fragments on average by sonication, and the biotin-labeled DNA fragments were pulled down with Dynabeads M-280 Streptavidin beads (Thermo Fisher Scientific, 11205D). The beads were washed twice with 2 × B&W buffer and blocked with 1 × I-Block buffer at room temperature for 45 min. Then, the beads were washed twice with 1 × B&W buffer and treated with 1 mg of pre-heated Salmon Sperm-DNA (Thermo Fisher Scientific, 15632011) with rotation at room temperature for 30 min. After washing with 1 × B&W buffer, the beads were resuspended with 2 × B&W buffer, combined with sonicated DNA and rotated at room temperature for 45 min. The beads were washed five times with 2 × SSC containing 0.5% SDS, twice with 1 × B&W

buffer and once with Buffer EB (Qiagen, 19086). The DNA on the beads was end-repaired using T4 DNA polymerase (NEB, M0203L), T4 polynucleotide kinase (NEB, M0201L) and large (Klenow) fragment (NEB, M0210L) with shaking at 900 rpm at 37°C for 30 min. After washing twice with 1 × TWB at 55°C for 2 min, the DNA on the beads was A-tailed using Klenow fragment (3′→5′ exo-) (NEB, M0212L) with shaking at 900 rpm at 37°C for 30 min. The beads were washed twice with 1 × TWB at 55°C for 2 min and once with 1 × Quick ligase buffer (NEB, M2200L). The DNA on the beads was ligated with adaptor using Quick ligase (NEB, M2200L) and 20 mM Y-Adaptor (annealed by/5Phos/GATCGGAAGAGCAC ACGTCTGAACTCCAGTCAC and TACACTCTTTCC CTACACGACGCTCTTCCGATCT) at room temperature for 45 min. The beads were washed twice with 1 × TWB at 55°C for 2 min and once with buffer EB (Qiagen, 19086).

The libraries were constructed using Q5 Hot Star DNA Polymerase (NEB, M0493L) for the PCR amplification. PCR products between 300–700 bp were purified using Ampure XP beads (Beckman Coulter, A63882) and subjected to Illumina NovaSeq (Annoroad Gene Technology Co., Ltd.) for sequencing.

BL-Hi-C data analysis

The BL-Hi-C data were processed with ChIA-PET2 v0.9.2 software (53) with the parameter '-A ACGCGA TATCTTATC -B AGTCAGATAAGATAT -s 1 -m 1 -t 4 -k 2 -e 1 -l 15 -S 500' to identify chromatin interactions annotated in genome mm10. The biological replicates of each group were merged to perform the A/B compartment and topologically associated domain (TAD) analysis. ChIA-PET2 bedpe2Matrix was used to generate the contact matrices. Normalization was performed by HiCExplorer (54) hicCorrectMatrix (– correctionMethod KR). A/B compartments at 40 kb bins were identified by HiCExplorer hicPCA (–method lieberman –extraTrack H3K4me3_chip.bw –histonMarkType active). TADs and TAD boundaries were defined by HiCExplorer hicFindTADs (–correctForMultipleTesting fdr –thresholdComparisons 0.05 –minDepth 120000 – maxDepth 200000 –step 40000) at 40-kb resolution.

High confidence interactions were defined as those with a false discovery rate (FDR) < 0.05 for the downstream analysis. Promoters were defined as 2.5 kb upstream and 2.5 kb downstream of TSSs. Enhancers were defined as H3K27ac binding sites that did not overlap with promoters. To identify the loops mediated by YY1 and CTCF, we performed a combination analysis of the Hi-C data and ChIP-seq data (both YY1 and CTCF) generated in this study. The differential EP interactions mediated by YY1 or CTCF were identified using the diffloop pipeline (55) with the quickAssoc function, which was based on an overdispersed Poisson regression model. Then, differential interactions with high confidence were chosen as follows: first, interactions with p-value < 0.05 and |log$_2$FC| > 1 were selected as primary differential interactions. Then, in these primary differential interactions, for the *Yy1* shRNA-depleted EPSC versus the control shRNA-treated EPSC enhanced interactions, the interactions produced in the cells from two replicates of

Yy1 shRNA-depleted EPSC must not be zero. Similarly, for the control shRNA-treated EPSC versus the *Yy1* shRNA-depleted EPSC enhanced interactions, the interactions that either replicate of the control shRNA-treated EPSC is zero will be filtered. Finally, the filtered differential interactions were considered as differential interactions with high confidence and were used for further analysis. The data were further visualized in the WashU Epigenome Browser (47).

Transposable elements (TEs) analysis of ATAC-seq, ChIP-seq and WGBS

For the TEs analysis of ATAC-seq, ChIP-seq and WGBS, the multi-mapped reads were retained, but only the best alignment was retained for those reads; if more than one equivalent best alignment was found, one random alignment was retained. The coordinates and annotations of the TEs were downloaded from the UCSC Genome Browser (mm10) version of RepeatMasker. TE types with < 200 copies were deleted from the analysis. The average coverage signal on each TE copy was calculated using deeptools (46). The YY1 enrichment values for TEs were calculated by the average ChIP-seq signal divided the average input signal. The TEs were used for further analysis if the enrichment value was > 1.5.

Quantification and statistical analysis

The data are presented as the mean value ± SD unless otherwise indicated in the figure legend. The sample numbers and experimental repeats are indicated in the figure legends. Statistical significance in the two group comparisons was determined by Student's *t*-test (two-tailed). The differences in means were considered statistically significant at p < 0.05. The significance levels are as follows: *p < 0.05; **p < 0.01 and ***p < 0.001.

RESULTS

Transcriptome analysis of both ESCs and EPSCs

We cultured mouse EPSCs and ESCs in N2B27-LCDM medium and 2i medium, respectively, as previously described (8). Relative to ESCs, EPSCs were more homogeneous, and their colonies exhibited a clear domed morphology (Supplementary Figure S1A). To examine whether EPSCs and ESCs showed similar gene expression programs as previously described (7,8), we performed RNA-seq experiments using both EPSCs and ESCs with three biological replicates. Hierarchical cluster analysis (HCA) indicated that the gene expression patterns of the EPSCs and ESCs used in this study were highly similar to those in previously reported data (Supplementary Figure S1B). Then, we examined the DEGs (q-value < 0.05 and fold-change > 1.5) between the ESCs and EPSCs by using DESeq2. Our data indicated that there were 1270 genes with upregulated expression and 1812 genes with downregulated expression showing more than a 1.5-fold difference in expression between the EPSCs and ESCs (Supplementary Figure S1C). For example, *Lin28a*, *Gja1*, and *Spry4* showed significantly upregulated expression, while *Calcoco2*, *Arl14epl* and *Col12a1* showed downregulated expression in the EPSCs relative

to the ESCs (Supplementary Figure S1D), suggesting that these genes might be markers that can distinguish EPSCs from ESCs. The Metascape enrichment analyses of the DEGs revealed that the upregulated genes identified in EPSCs were highly enriched in the negative regulation of the Wnt and MAPK signaling pathways, *in utero* embryonic development, and placenta development (Supplementary Figure S1E, F), as previously described (7,8). Therefore, we defined the 1270 upregulated genes found in the EPSCs as EPSC-specific genes and the 1812 downregulated genes found in the EPSCs as ESC-specific genes.

Chromatin accessibility landscape between ESCs and EPSCs

To identify the differences in the chromatin accessibility landscape between ESCs and EPSCs, we performed ATAC-seq experiments and observed high reproducibility among three biological replicates ($R > 0.9$, Supplementary Figure S2A). We found that 3512 regions were more accessible in EPSCs than in ESCs; however, 6668 peaks were less accessible in EPSCs than in ESCs (Figure 1A, B). These data indicated that EPSCs have less accessible chromatin regions than ESCs. We further noticed that these hyperaccessible regions in EPSCs had relatively higher enrichments in promoters and distal intergenic regions than the hypoaccessible regions (Supplementary Figure S2B). These hyperaccessible and hypoaccessible chromatin regions were further correlated with changes in gene expression (Supplementary Figure S2C). We found that approximately 20.59% (105/510) of the genes associated with hyperaccessible regions were EPSC-specific genes and that ∼26.28% (231/879) of the genes associated with hypoaccessible regions were ESC-specific genes (Figure 1C). Histone H3K27ac has been reported as an active enhancer marker (56). We found that *Lin28a* and *Gja1,* two EPSC-specific genes, showed higher ATAC and H3K27ac signals surrounding their promoter or enhancer regions in EPSCs compared to ESCs (Supplementary Figure S2D, E). In contrast, *Calcoco2* and *Arl14epl,* two ESC-specific genes, showed lower ATAC and H3K27ac signals surrounding their promoter or enhancer regions in EPSCs compared to ESCs (Supplementary Figure S2F, G). Altogether, these results indicated that EPSCs and ESCs have different chromatin features globally.

YY1 and H3K27ac are highly enriched in hyperaccessible regions in EPSCs

To identify potential binding factors in the hyperaccessible and hypoaccessible regions of EPSCs, we investigated the transcription factor-enriched motifs within these regions. EKLF and TCFCP2L1 motifs were found in the hypoaccessible regions in EPSCs, suggesting that EKLF and TCFCP2L1 might have little effect on the maintenance of EPSCs but are important for maintaining ESCs (Figure 1D) (57). Interestingly, we found that the YY1 motif showed the greatest enrichment at the hyperaccessible sites of EPSCs (Figure 1E). We examined the YY1 expression levels in both ESCs and EPSCs by RT-qPCR and Western blot analysis and found that YY1 was highly expressed in EPSCs relative to ESCs (Figure 1F). To further confirm that YY1 could bind the hyperaccessible regions of EPSCs, we performed
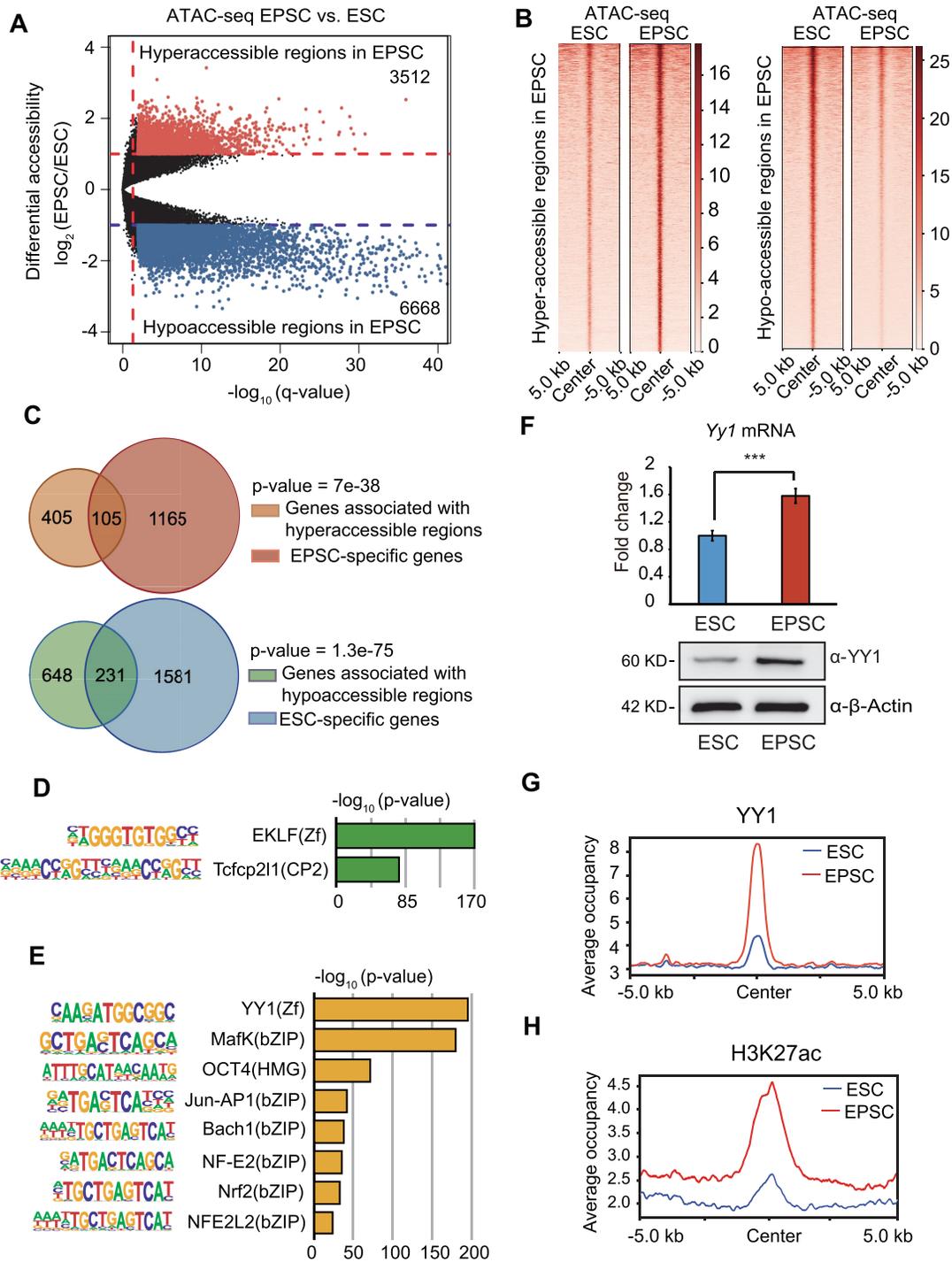
Figure 1. Chromatin accessibility features of both ESCs and EPSCs. (**A**) Volcano plot of the chromatin accessibility profiles of EPSCs and ESCs ($n = 3$ biological replicates). Red and blue dots represent significantly hyperaccessible and hypoaccessible regions, respectively, in EPSCs relative to ESCs (q-value < 0.05, fold-change > 2). (**B**) Heatmaps of ATAC-seq enrichment [normalized reads per kilobase per million (RPKM)] in hyperaccessible (left) and hypoaccessible (right) regions (center ± 5 kb) of EPSCs relative to ESCs ($n = 3$ biological replicates). (**C**) Overlap between genes with hyperaccessible regions and EPSC-specific genes (upper). Overlap between genes with hypoaccessible regions and ESC-specific genes (lower). The p-value was calculated with the hypergeometric distribution test. (**D**) Enrichment of transcription factor (TF) motifs in hypoaccessible regions of EPSCs relative to ESCs. (**E**) Enrichment of TF motifs in hyperaccessible regions of EPSCs relative to ESCs. (**F**) RT-qPCR and Western blot analysis were used to examine YY1 expression in ESCs and EPSCs. mRNA expression was tested in triplicate in three independent experiments. The data are shown as the mean ± SD. P-values were determined by two-sided Student's *t*-test. (**G**) Read count tag density pileups (in RPKM) of the ChIP-seq signals of YY1 from EPSCs and ESCs at hyperaccessible regions of EPSCs. (**H**) Read count tag density pileups (in RPKM) of the ChIP-seq signals of histone H3K27ac from EPSCs and ESCs at hyperaccessible regions of EPSCs.

YY1 ChIP-seq experiments in both ESCs and EPSCs. As expected, consistent with H3K27ac binding, YY1 was significantly enriched at the hyperaccessible regions of EPSCs relative to ESCs (Figure 1G, H). And we found that approximately 21.84% (767/3512) of the hyperaccessible regions overlapped with the YY1 binding sites in EPSCs and that approximately 16.4% (576/3512) of the hyperaccessible regions co-bound with H3K27ac in EPSCs (Supplementary Figure S2H, I). Taken together, these results showed that YY1 and H3K27ac are highly enriched in hyperaccessible regions in EPSCs.

Yy1 depletion in EPSCs results in a more similar gene expression pattern to ESCs than control EPSCs

To investigate the role of YY1 in maintaining EPSCs, we generated control shRNA and two *Yy1*-depleted stable EPSC lines (Figure 2A). Our data indicated that both *Yy1*-depleted stable EPSC lines had a cell morphology similar to that of the control shRNA-treated EPSC lines (Supplementary Figure S3A). The knockdown of *Yy1* slightly inhibited the viability and colony formation of EPSCs (Supplementary Figure S3B, C). We further performed RNA-seq experiments in EPSCs subjected to either the control shRNA treatment or *Yy1* depletion, and we analyzed the obtained data along with RNA-seq data from ESCs and wild-type EPSCs to examine the effects of *Yy1* knockdown on gene expression in EPSCs. Interestingly, the HCA indicated that the gene expression pattern of EPSCs became more similar to that of ESCs after *Yy1* depletion (Supplementary Figure S3D). We further analyzed the DEGs (q-value < 0.05 and fold-change > 1.5) and found that 1529 and 1033 genes had significantly up- and downregulated expression, respectively, after *Yy1* depletion in the EPSCs (Supplementary Figure S3E). Notably, we found that 25.6% (264/1033) of the genes with downregulated expression after *Yy1* knockdown were EPSC-specific genes (Supplementary Figure S3F). However, 33.4% (511/1529) of the genes with upregulated expression after *Yy1* knockdown were ESC-specific genes (Supplementary Figure S3G). Moreover, we found that the EPSC-specific genes related to embryonic development *in utero*, such as *Zp3*, *Epas1* and *Eno1*, showed downregulated expression after *Yy1* knockdown in both EPSC clones examined (Supplementary Figure S3H). Taken together, these results indicated that *Yy1* loss globally disrupts the gene expression pattern of EPSCs and leads to a more similar gene expression pattern to ESCs than control EPSCs.

Yy1 depletion in EPSCs disrupts the EP interactions of EPSC-specific genes

Since YY1 could mediate EP interactions as a general feature of mammalian gene control (16), we performed BL-Hi-C experiments (52) and investigated the EP interactions in EPSCs with or without *Yy1* knockdown. Our Hi-C data indicated that each library exhibited > 90% valid paired-end tags (PETs) in the mapped PETs (Supplementary Figure S4A), approximately 80% of which were found to be unique after removing PCR duplicates (Supplementary Figure S4B). Our BL-Hi-C data showed high reproducibility between different samples (Supplementary Figure S4C).

These data demonstrated that the control shRNA-treated EPSCs and the *Yy1*-depleted EPSCs showed similar distributions for both the A and B compartments (Supplementary Figure S4D). The correlation analysis revealed a minimal difference in TADs between the control shRNA-treated and *Yy1*-depleted EPSCs (Supplementary Figure S4E).

We subsequently analyzed the YY1 binding peaks using data from the ChIP-seq experiments along with the BL-Hi-C data to identify YY1-mediated EP interactions in EPSCs. Our data indicated that the interactions between the YY1-occupied enhancers and promoters significantly decreased after *Yy1* depletion (Supplementary Figure S4F, G), and 33.7% of the genes regulated by YY1-mediated EP interactions showed significant changes in their expression (Supplementary Figure S4H). We divided these genes into the following three categories after *Yy1* depletion in EPSCs: (i) genes with significantly upregulated expression, (ii) genes that were not significantly changed and (iii) genes with significantly downregulated expression. We observed that compared to both the genes with significantly upregulated expression and the genes that were not significantly changed, the genes with significantly downregulated expression had more reduced EP contact frequencies than increased contacts in response to *Yy1* depletion (Figure 2B). Furthermore, our data indicated that *Yy1* depletion resulted in a significantly decreased contact frequency of 3005 YY1-mediated EP interactions (Figure 2C). Moreover, we found that 221 genes with downregulated expression (including 39 EPSC-specific genes) had a significant decrease in EP interactions after *Yy1* depletion (Figure 2D, Supplementary Table S5). Importantly, GO analysis revealed that these 221 genes, such as *Epas1,* were involved in placental development and embryonic development *in utero* (Figure 2E, F and Supplementary Table S6). We further found that these genes with downregulated expression mediated by a reduced number of EP interactions were involved in DNA methylation and genomic imprinting (Figure 2E and Supplementary Table S6). We also found that *Yy1* depletion significantly reduced EP interactions of *Dnmt3l* (Figure 2G), which could stimulate *de novo* methylation by *Dnmt3a* and is thought to be required for the establishment of maternal genomic imprinting (58), suggesting that YY1 might regulate DNA methylation by modulating the expression of DNA methyltransferase.

Yy1 depletion induces DNA hypomethylation in EPSCs

To investigate how YY1 regulates DNA methylation in EPSCs, we measured the global DNA methylation levels in both ESCs and EPSCs with two restriction enzymes, HpaII (which cleaves unmethylated DNA) and McrBC (which cleaves methylated DNA). Our data showed that EPSCs exhibited higher levels of DNA methylation than ESCs, as previously reported (7) (Supplementary Figure S5A). Since our data indicated that YY1 positively regulates the EP interactions of *Dnmt3l*, we investigated whether YY1 regulates global DNA methylation by altering the expression of DNA methylation-associated genes. To test this hypothesis, we examined both the mRNA and protein levels of *Dnmt3a*, *Dnmt3b* and *Dnmt3l*. As expected, our data indicated that
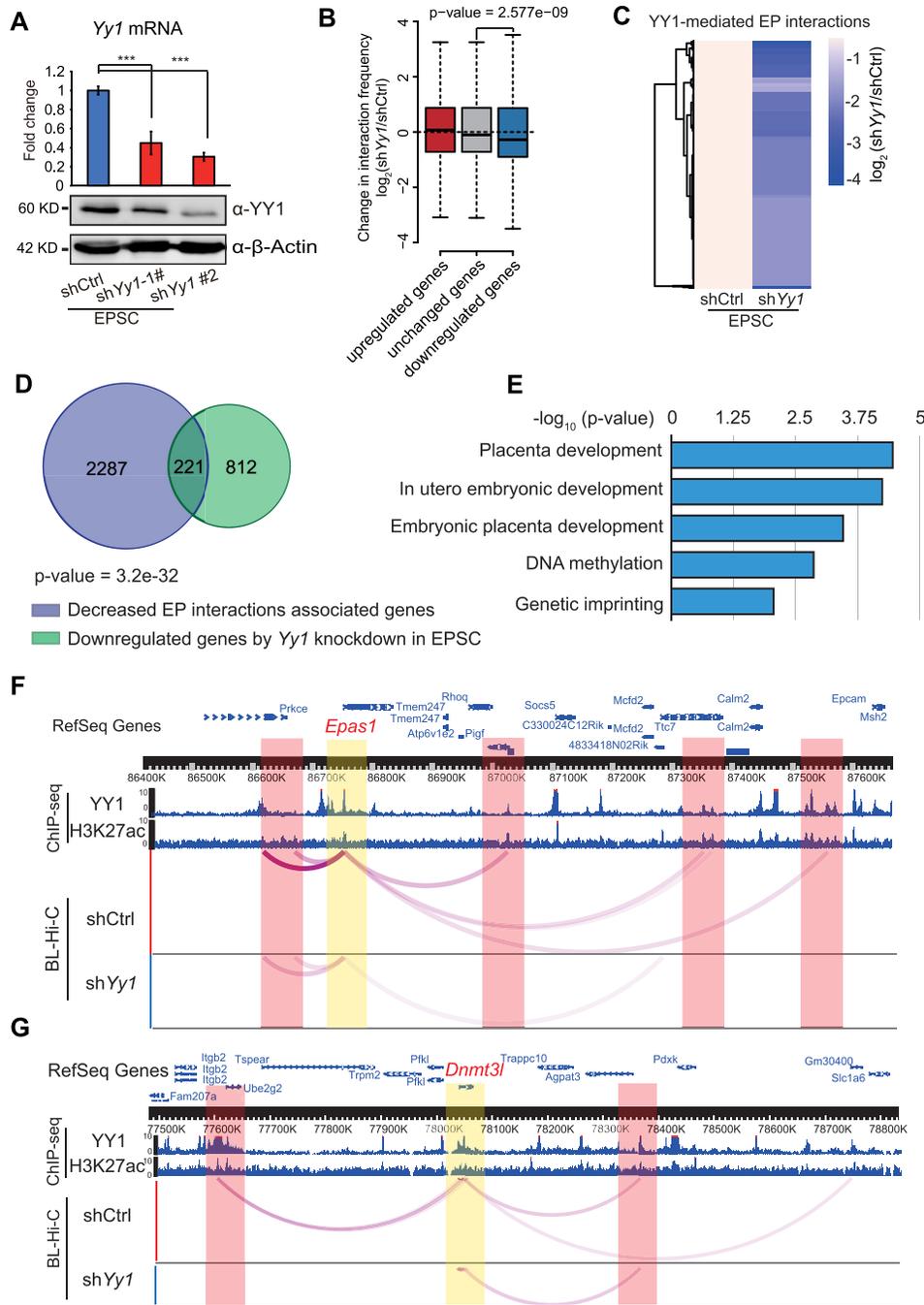
Figure 2. *Yy1* depletion disrupts the YY1-mediated EP interactions of EPSC-specific genes. (**A**) RT-qPCR and Western blot analysis were performed to determine the *Yy1* knockdown efficiency in EPSCs. mRNA expression was tested in triplicate in three independent experiments. The data are shown as the mean ± SD. P-values were determined by two-sided Student's *t*-test. (**B**) Change in the normalized interaction frequency (log$_2$ fold change) upon *Yy1* depletion in the following three different categories of EP interactions: genes with significantly upregulated expression, genes that were not significantly changed, and genes with significantly downregulated expression. (**C**) Heatmap representation of significantly decreased (blue) YY1-mediated EP interactions after *Yy1* depletion. The color key indicates the log$_2$(fold change) values of the *Yy1* shRNA-depleted EPSCs versus the control shRNA-treated EPSCs (*n* = 2 biological replicates). (**D**) Venn diagrams showing the overlap between the genes with decreased EP interactions and genes with downregulated expression after *Yy1* depletion. P-values were calculated with the hypergeometric distribution test. (**E**) GO categories of the overlapping genes shown in Figure 2D. (**F** and **G**) Representative genomic loci showing YY1 and H3K27ac binding and decreased chromatin interactions after *Yy1* knockdown in EPSCs. The promoters of *Epas1* (F) and *Dnmt3l* (G) are highlighted with yellow-shaded rectangles, and their interacting enhancers are highlighted with red-shaded rectangles.

Yy1 depletion reduced both the RNA and protein levels of *Dnmt3l*, *Dnmt3a* and *Dnmt3b* (Figure 3A, B). Next, we found that the DNA methylation level in EPSCs was dramatically reduced after *Yy1* depletion (Supplementary Figure S5B).

To further explore the effects of YY1 on genome-wide DNA methylation, we performed WGBS in the control shRNA-treated and *Yy1*-depleted EPSCs. Our WGBS data showed high reproducibility between two biological replicates (Supplementary Figure S5C). We analyzed the DNA methylation data of ESCs (59) and EPSCs subjected to either the control shRNA treatment or *Yy1* depletion. The bioinformatic analysis demonstrated that *Yy1* depletion reduced DNA methylation in EPSCs (Figure 3C). Furthermore, unsupervised hierarchical clustering (UHC) analysis revealed that the DNA methylome of the *Yy1*-depleted EPSCs showed similarities to the DNA methylation patterns of the ESCs (Figure 3D). We further found that *Yy1* depletion in EPSCs led to relatively low levels of cytosine DNA methylation on enhancers, non-CpG island (nonCGI) promoters, and transposable elements (LINE, LTR, and SINE) (Figure 3E and Supplementary Figure S5D) but had a limited effect on CGI promoters (Figure 3E).

DNA hypermethylation prevents the binding of CTCF to its target sites (60–62). CTCF regulates gene expression by forming EP interactions (63–65). To determine whether YY1 regulates CTCF binding through DNA methylation in the genome, we performed CTCF ChIP-seq experiments using both the control shRNA-treated and *Yy1*-depleted EPSCs to determine whether the reduction in DNA methylation by *Yy1* knockdown could affect CTCF binding in the genome. The integration analysis indicated that *Yy1* depletion indeed reduced the DNA methylation level at CTCF binding sites (Figure 3E). These data suggested that CTCF might be involved in the genome organization of EPSCs mediated by YY1.

Yy1 depletion facilitates CTCF-mediated EP interactions by decreasing the DNA methylation levels at CTCF binding sites in EPSCs

To investigate the change in the genome-wide binding of CTCF after *Yy1* depletion, we identified 73 325 and 74 593 CTCF peaks in the control shRNA-treated and *Yy1*-depleted EPSCs, respectively. In these sites, we observed that *Yy1* depletion only reduced 3653 binding peaks but resulted in 6274 peaks enriched relative to the control shRNA-treated EPSCs. Moreover, we found that the increased CTCF binding occupancy was accompanied by notable decreases in DNA methylation at these 6274 sites (Figure 4A, B). Furthermore, we identified 1102 CTCF-mediated EP interactions that showed an increased contact frequency upon *Yy1* depletion (Figure 4C). Our data also indicated that 131 genes (including 32 ESC-specific genes, such as *Nefm*) that showed upregulated expression upon *Yy1* depletion were associated with an increase in the EP interactions mediated by CTCF (Figure 4D, E and Supplementary Table S5). Concomitantly, the expression of *Nefm* and other ESC-specific genes, such as *Ralb*, *Sirpa*, *Zfp182*, *Serping1* and *Ube2l6*, was upregulated after *Yy1* knockdown in both EPSC clones examined (Figure 4F).

These results suggested that *Yy1* depletion promoted abnormally high expression of ESC-specific genes with decreasing DNA methylation levels in the regions mediated by CTCF-mediated EP interactions.

Yy1 depletion dramatically reduces the enrichment of H3K4me3 and H3K27ac on the promoters of EPSC-specific genes

The GO analysis revealed that the 131 genes with upregulated expression with CTCF-mediated EP interactions were involved in histone modification, covalent chromatin modification, and chromatin organization (Figure 4G). Among these genes, lysine (K)-specific demethylase 5C (*Kdm5c*) and histone deacetylase 6 (*Hdac6*), two histone-modifying enzymes associated with gene repression, attracted our attention (Supplementary Table S7), and both genes showed upregulated expression after *Yy1* knockdown in two EPSC clones (Figure 5A, B). *Yy1* depletion increased the EP interactions of both the *Kdm5c* and *Hdac6* genes (Figure 5C, D). KDM5C demethylates di- and trimethylated histone 3 lysine 4 (H3K4), and HDAC6 regulates histone deacetylation (66–68). Hence, we compared the enrichment of H3K4me3 and H3K27ac on the promoters of all genes, ESC-specific genes and EPSC-specific genes. We found that the enrichment of both H3K4me3 and H3K27ac between ESCs and EPSCs showed a limited difference among all genes (Figure 5E, F). In contrast, H3K4me3 and H3K27ac were less abundant among the ESC-specific genes but significantly enriched among the EPSC-specific genes in EPSCs compared to ESCs (Figure 5E, F). Moreover, we observed that the increased enrichment of H3K4me3 and H3K27ac among the EPSC-specific genes was more significant than the decreased enrichment of H3K4me3 and H3K27ac among the ESC-specific genes (Figure 5E, F). These data indicated that both active markers, H3K4me3 and H3K27ac, are highly associated with EPSC-specific genes.

To explore whether *Yy1* knockdown altered the enrichment of H3K4me3 and H3K27ac in EPSC-specific gene promoters, we further performed H3K4me3 ChIP-seq and H3K27ac CUT&Tag experiments using both the control shRNA-treated and *Yy1*-depleted EPSCs. Consistent with the upregulation of *Kdm5c* expression by *Yy1* knockdown (Figure 5A), we found that the enrichment of H3K4me3 in all three groups of gene promoters (all genes, ESC-specific genes, and EPSC-specific genes) was reduced (Figure 5G). Additionally, the enrichment of H3K4me3 in the EPSC-specific gene promoters was the most significantly decreased (Figure 5G). The enrichment of H3K27ac was slightly reduced in all genes and ESC-specific genes but markedly reduced in the EPSC-specific genes (Figure 5H). Taken together, we concluded that *Yy1* depletion dramatically reduces the enrichment of both active histone marks, H3K4me3 and H3K27ac, in EPSC-specific gene promoters by upregulating the expression of both *Kdm5c* and *Hdac6*.

Yy1 overexpression endows ESCs with the expression of partial EPSC specific genes

Next, we speculated whether the forced expression of *Yy1* facilitates extended pluripotency in ESCs. To test this pos-
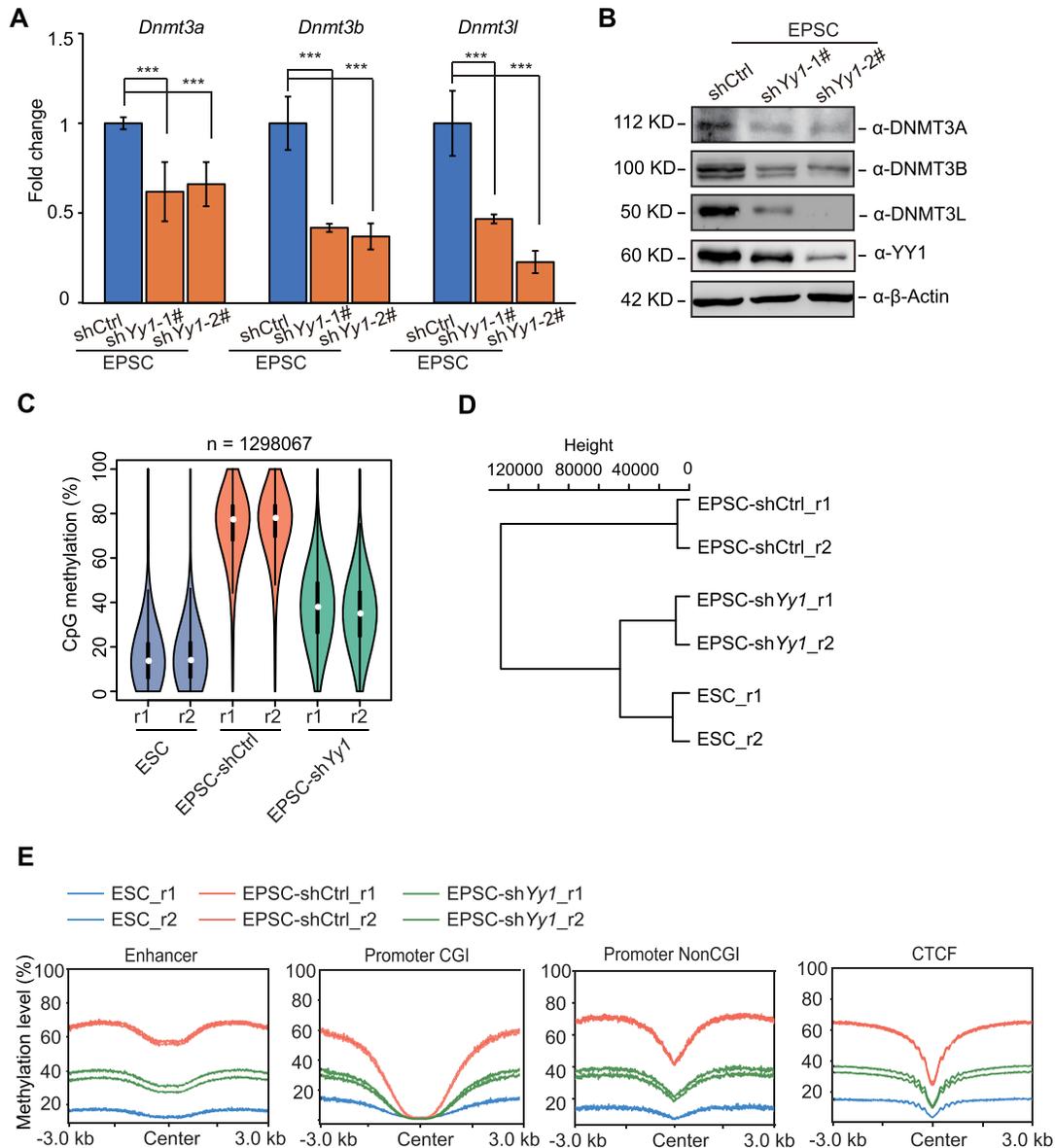
Figure 3. *Yy1* knockdown reduces DNA methylation in EPSCs. (**A**) RT-qPCR analysis of *Dnmt3a*, *Dnmt3b* and *Dnmt3l* mRNA expression in the control shRNA-treated EPSCs and the two *Yy1* shRNA-depleted EPSCs. Each gene was tested in triplicate in two independent experiments. The data are shown as the mean ± SD. P-values were determined by two-sided Student's *t*-test. (**B**) Western blot analysis of the DNMT3A, DNMT3B, DNMT3L and YY1 protein levels in the control shRNA-treated and two *Yy1* shRNA-depleted EPSCs. (**C**) Violin plots showing the CpG methylation distribution of 2 kilobase (kb) genomic tiles in ESCs and control shRNA-treated EPSCs (*n* = 2 biological replicates, indicated as r1 and r2). (**D**) UHC of the methylation levels of 2 kb genomic tiles in ESCs and control shRNA-treated and two *Yy1* shRNA-depleted EPSCs (*n* = 2 biological replicates). (**E**) Methylation profiles of enhancers, promoter CGIs, promoter nonCGIs, and CTCF binding regions in ESCs and the control shRNA-treated and two *Yy1* shRNA-depleted EPSCs (*n* = 2 biological replicates).

sibility, we overexpressed *Yy1* in ESCs under 2i/LIF conditions (Supplementary Figure S6A). Relative to the control ESCs, ESCs with *Yy1* overexpression showed much greater homogeneity and a domed cell morphology (Supplementary Figure S6B), similar to the characteristics of wild-type EPSCs. To investigate the changes in gene expression induced by *Yy1* overexpression, we performed RNA-seq experiments in both the control and *Yy1*-overexpressing ESCs. Interestingly, HCA illustrated that the gene expression pattern of the ESCs after *Yy1* overexpression became

more similar to that of EPSCs than control ESCs (Supplementary Figure S6C). We further examined the DEGs (q-value < 0.05 and fold-change > 2) and found that 1644 and 2055 genes showed significantly up- and downregulated expression, respectively, in the *Yy1*-overexpressing ESCs. Notably, we found that 21.2% (436/2055) of the genes with downregulated expression after *Yy1* overexpression were ESC-specific genes (Supplementary Figure S6D). However, 15.7% (258/1644) of the genes with upregulated expression after *Yy1* overexpression were EPSC-specific genes (Sup-
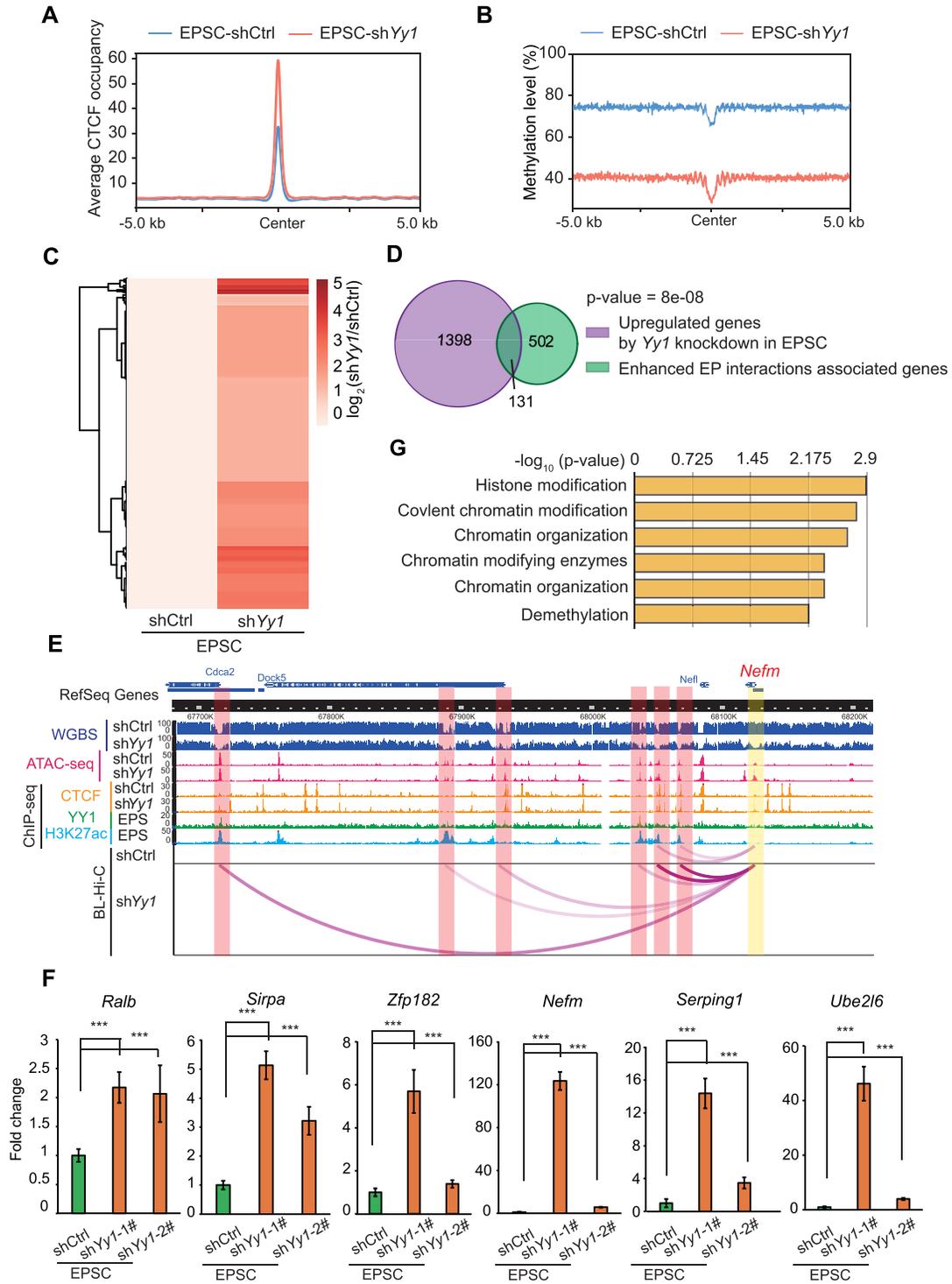
Figure 4. *Yy1* knockdown increases CTCF-mediated EP interactions by decreasing the methylation levels at CTCF binding sites in EPSCs. (**A**) Tag density profile of significantly increased CTCF occupancy after *Yy1* knockdown. These tags were from RPKM normalized ChIP-seq signals. (**B**) Methylation profiles of CTCF binding sites showing significant increases after *Yy1* knockdown. (**C**) Heatmap representation of significantly enhanced EP interactions mediated by CTCF after *Yy1* knockdown. (**D**) Venn diagrams showing the overlap between the genes associated with enhanced EP interactions and the genes with upregulated expression after *Yy1* knockdown. P-values were calculated with the hypergeometric distribution test. (**E**) Representative genomic loci showing YY1 and H3K27ac binding, DNA methylation, accessible chromatin, and enhanced chromatin interactions after *Yy1* knockdown in EPSCs. The promoter of *Nefm* is indicated with yellow-shaded rectangles, and its interacting enhancers are highlighted with red-shaded rectangles. (**F**) RT-qPCR analysis of genes with upregulated expression related to CTCF-mediated EP interactions after *Yy1* depletion. Each gene was tested in triplicate in two independent experiments. The data are shown as the mean ± SD. P-values were determined by two-sided Student's *t*-test. (**G**) GO categories of the overlapping genes shown in Figure 4D.
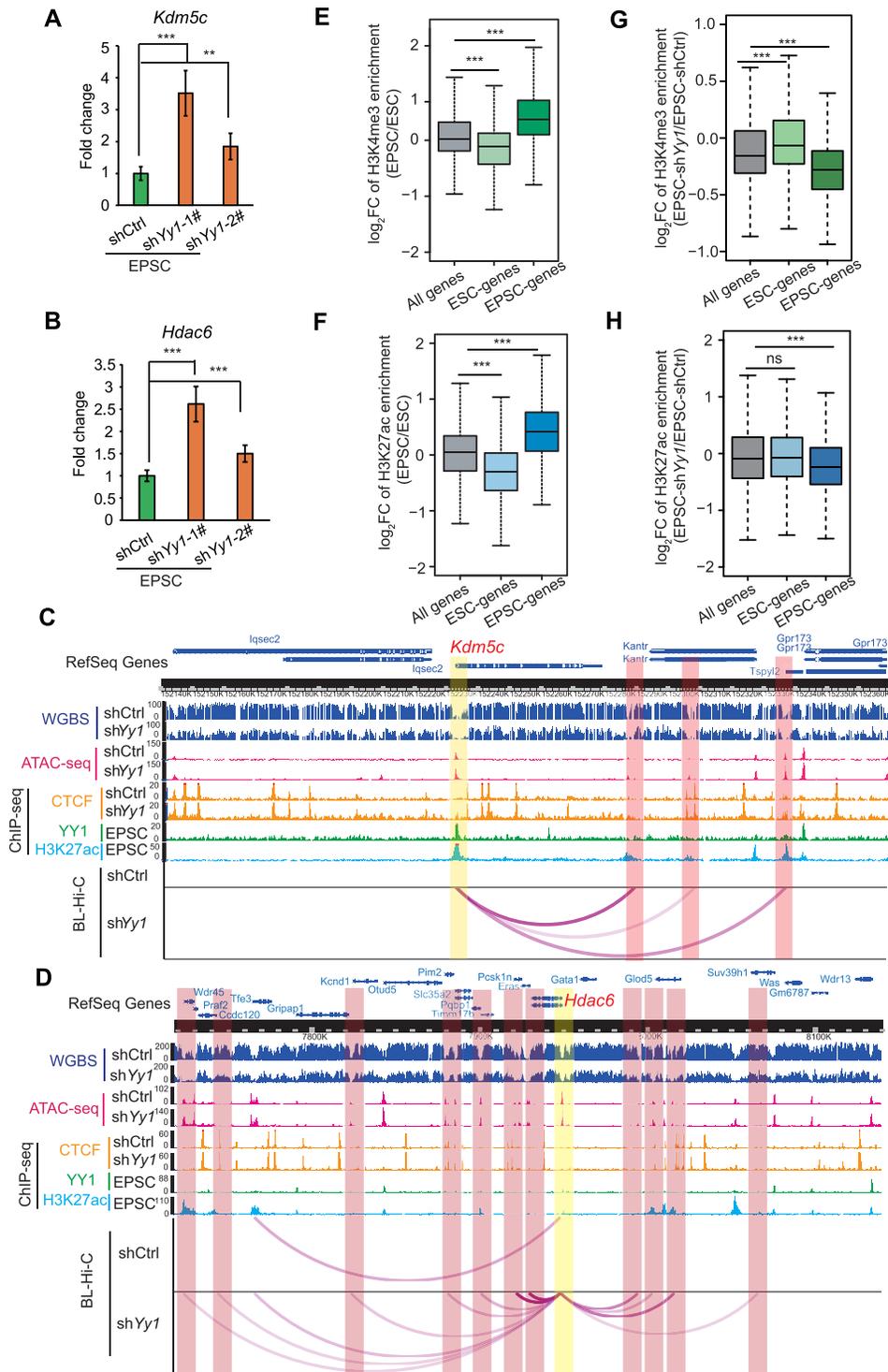
Figure 5. Effects of *Yy1* depletion on the H3K4me3 and H3K27ac levels on EPSC-specific genes. (**A** and **B**) RT-qPCR analysis of the expression of *Kdm5c* (A) and *Hdac6* (B) in the control shRNA-treated and two *Yy1* shRNA-depleted EPSCs. The gene was tested in triplicate in two independent experiments. The data are shown as the mean ± SD. P-values were determined by two-sided Student's *t*-test. (**C** and **D**) Representative genomic loci showing YY1 and H3K27ac binding, DNA methylation, accessible chromatin, and enhanced chromatin interactions after *Yy1* knockdown in EPSCs. The promoters of *Kdm5c* (C) and *Hdac6* (D) are highlighted with yellow-shaded rectangles, and their interacting enhancers are highlighted with red-shaded rectangles. (**E** and **F**) Boxplot showing enrichment fold change (FC) of H3K4me3 (E) and H3K27ac (F) in EPSCs versus ESCs in regions of all genes, ESC-specific genes and EPSC-specific genes. The centerline indicates the median value, while the box and whiskers represent the interquartile range (IQR) and $1.5 \times$ IQR, respectively. ***p-value < 0.001 by the Mann–Whitney *U* test versus the FC of all genes. (**G** and **H**) Boxplot showing enrichment FC of H3K4me3 (G) and H3K27ac (H) in *Yy1* shRNA-depleted EPSCs versus the control shRNA-treated EPSCs in regions of all genes, ESC-specific genes and EPSC-specific genes. The centerline indicates the median value, while the box and whiskers represent the IQR and $1.5 \times$ IQR, respectively. ***p-value < 0.001 by the Mann–Whitney *U* test versus the FC of all genes.

plementary Figure S6E). Additionally, *Yy1* overexpression in ESCs upregulated the expression of DNA methylation-associated genes (*Dnmt3a*, *Dnmt3b* and *Dnmt3l*) and extraembryonic-associated genes (*Eno1, Phlda2*, *Zfp36l1, Zp3*, *Erf* and *sh3pxd2a*), although the increase in expression was not comparable to that of EPSCs (Supplementary Figure S6F). Taken together, these results indicated that the overexpression of *Yy1* could induce relative low, but detectable expression of partial EPSC specific genes.

YY1 balances *in vitro* extraembryonic derivatives between ESCs and EPSCs

To determine whether YY1 could promote the formation of extraembryonic cells, we derived extraembryonic cells from either ESCs with/without *Yy1* overexpression or EPSCs with/without *Yy1* knockdown for six days using F4H medium (Figure 6A). We observed significant differences in the colony morphology between ESCs and EPSCs starting on day 2 (Supplementary Figure S7A). Subsequently, the ESC-derived cells formed a flat monolayer (green circled region), and the EPSC-derived cells were domed and formed a dense multilayer (red circled region) on day 6 (Supplementary Figure S7A). These data revealed that different derivation capacities exist within these cell types. Based on this system, our data showed that the *Yy1*-depleted EPSCs formed a flatter cell monolayer similar to the ESC-derived cells (Supplementary Figure S7B). Moreover, the *Yy1*-overexpressing ESC-derived cells displayed domed cells and formed a denser cell multilayer similar to the EPSC-derived cells on day 6 after differentiation (Supplementary Figure S7B). The bulk RNA-seq experiments indicated that the gene expression pattern of *Yy1*-depleted EPSC-derived cells was somewhat similar to that of the ESC-derived cells, while the derivatives of the *Yy1*-overexpressing ESCs were somewhat similar to the EPSC derivatives (Supplementary Figure S7C).

To examine whether the derivatives of ESCs or EPSCs express TSC- or XEN-specific marker genes during extraembryonic differentiation, we analyzed gene expression in the cells derived from ESCs and EPSCs on day 0 and 6 after F4H treatment, respectively. Our data showed that the expression of pluripotency genes was significantly decreased and that the TSC marker genes showed slightly upregulated expression in both the ESC- and EPSC-derived cells on day 6 (Supplementary Figure S7D). Interestingly, we noted that the expression of XEN marker genes in the EPSC-derived cells was significantly upregulated relative to that in the ESC-derived cells, indicating that EPSCs have a stronger XEN-derivatization ability than ESCs in this system (Supplementary Figure S7D). Meanwhile, we found that the expression of some XEN marker genes was not highly induced upon *Yy1* depletion (Supplementary Figure S7D). To further assess whether YY1 affects the expression of XEN marker genes after F4H treatment, RT-qPCR analysis indicated that the induction of *Gata4*, *Gata6* and *Sox17*, three XEN marker genes, was lower in the derivatives of both *Yy1*-depleted EPSCs on day 6 (Supplementary Figure S7E, left). However, the *Yy1* overexpression did not upregulate the expression of all three marker genes we tested (Supple-

mentary Figure S7E, right). These data suggested that YY1 might regulate the derivation of XEN-like cells from EPSCs.

YY1 is required for the derivation of XEN-like cells from EPSCs

Although the expression level of XEN marker genes was significantly upregulated in the F4H-treated EPSCs on day 6, the expression level was much lower than that in XEN cells (Supplementary Figure S7E). To further dissect the cellular heterogeneity and explore the regulatory function of YY1, we performed scRNA-seq of the cells derived from EPSCs with or without *Yy1* depletion and ESCs with or without *Yy1* overexpression on day 6 after F4H treatment (Figure 6B). In total, 11 102, 6072, 8634 and 8814 single-cell transcriptomes were obtained from these four samples after the quality control (Supplementary Figure S8A). The Pearson correlations between the mRNA numbers and reads in these four samples were > 0.9 (Supplementary Figure S8B). Dimensionality reduction and visualization via uniform manifold approximation and projection (UMAP) allowed us to divide these cells into eleven distinct clusters designated and annotated as clusters 0 to 10 based on known marker genes as follows (Supplementary Table S4) (28,69,70): anterior primitive streak-like cells (cluster 1*, Eomes* and *Mixl1*), mesenchyme-like cells (cluster 4, *Acta2* and *Pmp22*), primordial germ cell (PGC)-like cells (cluster 6, *Dnd1* and *Dppa3*), XEN-like cells (cluster 7, *Gata6* and *Sox17*) and unknown cells (remaining clusters that could not be annotated) (Figure 6C–E and Supplementary Figure S8C, D and Supplementary Table S8). As expected, the extraembryonic derivatives displayed a high degree of heterogeneity with a low percentage of XEN-like cells in the differentiated cell populations (Figure 6E). Interestingly, the ratio of XEN-like cells among the EPSC derivatives was approximately 5-fold higher than that among the ESC derivatives (Figure 6E). Furthermore, *Yy1* depletion reduced the ratio of XEN-like cells among the EPSC derivatives (cluster 7, chi-square test p-value < 0.001) and caused an abnormal increase in anterior primitive streak-like cells (cluster 1, chi-square test p-value < 0.001) and mesenchyme-like cells (cluster 4, chi-square test p-value < 0.001) (Figure 6E and Supplementary Figure S8E). Consistent with the *in vitro* differentiation phenotype of the ESCs overexpressing *Yy1* (Supplementary Figure S7E, right), *Yy1* overexpression did not enhance the ratio of XEN-like cells among the ESC derivatives (Figure 6E and Supplementary Figure S8E). Taken together, these results indicated that *in vitro* extraembryonic derivatives show high heterogeneity and that YY1 is necessary but not sufficient for the derivation of XEN-like cells from EPSCs.

DISCUSSION

The organization of accessible chromatin across the genome provides an opportunity for permissible physical interactions between chromatin-binding factors and DNA *cis* elements to cooperatively regulate gene expression (71). Pluripotent stem cells contain more open chromatin than differentiated cells, resulting in a more complex regulatory network (72). This study showed that YY1 acts as a key regulator in the modulation of genome organization through

Figure 6. YY1 promotes the formation of XEN-like cells *in vitro*. (**A**) Strategy for the derivation of extraembryonic cells from ESCs and EPSCs. RNA was collected, and the expression of genes was tested at day 0 and day 6. (**B**) Dimensionality reduction and clustering of the scRNA-seq data of the control shRNA-EPSC-derived cells ($n = 11\,102$), *Yy1*-depleted EPSC-derived cells ($n = 6072$), Flag-control-overexpressing ESC-derived cells ($n = 8634$), and Flag-*Yy1*-overexpressing ESC-derived cells (n = 8814). Different colors represent different clustered cell groups. The cluster names are from the cell annotation file in Supplementary Table S4. (**C**) Dot plot showing average scaled expression and the percentage of cells within each cluster for selected marker genes. Color intensity of each dot represents the average expression level of marker genes in the cells per cluster. And dot size reflects the percentage of cells expressing the marker genes per cluster (the cutoff value was 0.1). (**D**) Violin plots showing the expression levels of marker genes. (**E**) Fractions of different clusters in the control shRNA-EPSC-derived cells, *Yy1*-depleted EPSC-derived cells, and Flag-control or Flag-*Yy1*-overexpressing ESC-derived cells. Asterisks denote chi-square test p-value < 0.001.

multiple layers of epigenetic crosstalk to maintain the EPSC identity. First, *Yy1* depletion decreased the expression of *Phlda2*, *Dnmt3l* and *Zfp36l1*, which are important for the development of the mouse placenta (73–76), and *Spry2* and *Spry4*, which participate in a negative feedback loop of MAPK activation (77), by disrupting the formation of EP interactions in EPSCs (Supplementary Table S5). Second, *Yy1* knockdown altered DNA methylation by regulating the expression of *Dnmt3a, Dnmt3b* and *Dnmt3l* and CTCF global occupation. CTCF is a DNA methylation-sensitive binding factor (62), and mediates long-range chromatin interactions (63,64). The decrease in DNA methylation led to an increase in CTCF-mediated EP interactions, which could explain the abnormally high expression of ESC-specific genes in the EPSCs after *Yy1* depletion. Third, the epigenetic crosstalk between DNA methylation and histone H3K4me3 (78), between RNA methylation and histone H3K9me2 (79), and between DNA methylation and CTCF-mediated chromatin looping (80) play important roles in gene regulation and cell fate determination. *Yy1* knockdown reduced the binding of both H3K4me3 and H3K27ac to EPSC-specific genes by increasing the CTCF-mediated EP interactions of the *kdm5c* gene encoding histone H3K4me3 demethylase and the *Hdac6* gene required for histone deacetylation.

The LINE, IAP, and MERVL TEs contain the YY1 motif (81). YY1 can bind IAP in ESCs (82,83). Indeed, we also observed that YY1 could bind RLTR10D2, RLTR10A and IAPEY_LTR in EPSCs (Supplementary Figure S9A, B). YY1 also plays a functional role in the regulation of chromatin accessibility and DNA methylation at these TEs since *Yy1* depletion increased chromatin accessibility (Supplementary Figure S9C, D) but reduced the DNA methylation (Supplementary Figure S9E, F) of RLTR10D2, RLTR10A and IAPEY_LTR bound by YY1. Taken together, our data also suggested that YY1 can regulate the accessibility and DNA methylation of specific TEs by binding them.

Whether ESCs can be used to derive extraembryonic cells is still controversial. Conventional studies have reported that ESCs cannot differentiate into TE and PrE derivatives (7,84). However, several studies have indicated that ESCs can be derived from extraembryonic cells, such as TSCs or XEN cells, by adjusting the culture conditions (85–87). *In vitro* EPSC derivation is also controversial. Yang *et al.* successfully induced TSC-like cells that could differentiate into mature trophoblasts using improved TX medium with feeder cells as the substrate (7). In contrast, Posfai *et al.* could not obtain cells expressing TSC marker genes from EPSCs derived with TS medium (28). These results demonstrated that the culture conditions play critical roles in regulating the *in vitro* developmental potential of cells. FGF4 is sufficient for XEN cell derivation (87); moreover, FGF4 and heparin could be used to establish XEN cell lines from ESCs and blastocysts (27,88). We developed an F4H culture medium containing FGF4 and heparin and derived some XEN-like cells from both EPSCs and ESCs, although the *in vitro*-derived cell population showed high heterogeneity. Through this strategy, we found that EPSCs produced more XEN-like cells than ESCs (Figure 6E), which is consistent with previous conclusions (7).

In addition, *Yy1* depletion reduced the XEN-derivatization ability of EPSCs (Figure 6E). However, *Yy1* overexpression in ESCs did not increase the ratio of XEN-like cells among the ESC derivatives (Figure 6E and Supplementary Figure S8E), which might also be due to the limitations of the *in vitro* culture system we used. Although the YY1 level in the ESCs with *Yy1* overexpression was approximately 10-fold higher than that in the wild-type ESCs, the expression of DNMT genes and extraembryonic-associated genes was still relatively low compared with the expression in EPSCs (Supplementary Figure S6F), indicating that YY1 alone is not sufficient to switch epigenetic environments during cell fate transition from ESCs.

YY2 binds to a consensus sequence similar to YY1 (89,90), although both shared and mutually exclusive sites have been identified in ESCs (91). YY2 is also expressed in the embryonic stages preceding blastocysts and TSCs (89,92). Although we found that *Yy2* was highly expressed in EPSCs relative to ESCs, the relative expression level of *Yy2* was lower than that of *Yy1* (data not shown). Therefore, we focused on the role of YY1 in the regulation of the fate of EPSCs; however, the synergistic regulation of EPSC identity by YY2 and/or YY1 should be explored in future studies.

Our scRNA-seq data showed that, under the right culture conditions, both ESCs and EPSCs could turn into cells expressing combinations of marker genes representative of anterior primitive streak cells, mesenchyme cells and PGCs, and these cell groups appeared during the establishment of the antero-posterior (A-P) axis (E5.5–E6.5) before the formation of the gastrula during mouse embryonic development (1,69,93). This phenomenon might be due to *in vitro* FGF4-induced *Bmp4* expression (Supplementary Figure S7D), which has been shown to facilitate the establishment of the A-P axis during early embryonic development (94). Additionally, our data indicated that *Yy1* promotes the production of PGC-like cells from ESCs, which may provide ideas for improving ESC-to-PGC transformation efficiency by modulating *Yy1* levels.

Overall, this study revealed that YY1 functions as an important epigenetic factor in regulating different epigenetic crosstalk, which in turn affect the maintenance and differentiation potential of EPSCs.

DATA AVAILABILITY

The NCBI GEO number of the RNA-seq, ATAC-seq, ChIP-seq, CUT&Tag, scRNA-seq, WGBS and BL-Hi-C data described in this paper is GSE157720 and the accession number of the Genome Sequence Archive of the Beijing Institute of Genomics (BIG) Data Center is CRA004888. Published RNA-seq data were from ERP005641 (7) and GSE89303 (8). The WGBS data of ESCs were from GSE99488 (59). The sequencing data, including the sample names, cell lines, replicate information and number of reads, are listed in Supplementary Table S9.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Rossant,J. and Tam,P. (2009) Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. *Development*, **136**, 701–713.
2. Harrison,S., Sozen,B., Christodoulou,N., Kyprianou,C. and Zernicka-Goetz,M. (2017) Assembly of embryonic and extraembryonic stem cells to mimic embryogenesis in vitro. *Science*, **356**, eaal1810.
3. Zylicz,J.J. (2020) Defined stem cell culture conditions to model mouse blastocyst development. *Curr. Protoc. Stem Cell Biol.*, **52**, e105.
4. Ying,Q.L., Wray,J., Nichols,J., Batlle-Morera,L., Doble,B., Woodgett,J., Cohen,P. and Smith,A. (2008) The ground state of embryonic stem cell self-renewal. *Nature*, **453**, 519–523.
5. Beddington,R. and Robertson,E. (1989) An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development*, **105**, 733–737.
6. De Los Angeles,A., Ferrari,F., Xi,R., Fujiwara,Y., Benvenisty,N., Deng,H., Hochedlinger,K., Jaenisch,R., Lee,S., Leitch,H.G. *et al.* (2015) Hallmarks of pluripotency. *Nature*, **525**, 469–478.
7. Yang,J., Ryan,D.J., Wang,W., Tsang,J.C., Lan,G., Masaki,H., Gao,X., Antunes,L., Yu,Y., Zhu,Z. *et al.* (2017) Establishment of mouse expanded potential stem cells. *Nature*, **550**, 393–397.
8. Yang,Y., Liu,B., Xu,J., Wang,J., Wu,J., Shi,C., Xu,Y., Dong,J., Wang,C., Lai,W. *et al.* (2017) Derivation of pluripotent stem cells with in vivo embryonic and extraembryonic potency. *Cell*, **169**, 243–257.
9. Gao,X., Nowak-Imialek,M., Chen,X., Chen,D., Herrmann,D., Ruan,D., Chen,A.C.H., Eckersley-Maslin,M.A., Ahmad,S., Lee,Y.L. *et al.* (2019) Establishment of porcine and human expanded potential stem cells. *Nat. Cell Biol.*, **21**, 687–699.
10. Yang,J., Ryan,D.J., Lan,G., Zou,X. and Liu,P. (2019) In vitro establishment of expanded-potential stem cells from mouse pre-implantation embryos or embryonic stem cells. *Nat. Protoc.*, **14**, 350–378.
11. Zhu,K., Liu,Y., Fan,C., Zhang,M., Cao,H., He,X., Li,N., Chu,D., Li,F., Zou,M. *et al.* (2020) Etv5 safeguards trophoblast stem cells differentiation from mouse EPSCs by regulating fibroblast growth factor receptor 2. *Mol. Biol. Rep.*, **47**, 9259–9269.
12. Li,R., Zhong,C., Yu,Y., Liu,H., Sakurai,M., Yu,L., Min,Z., Shi,L., Wei,Y., Takahashi,Y. *et al.* (2019) Generation of blastocyst-like structures from mouse embryonic and adult cell cultures. *Cell*, **179**, 687–702.
13. Sozen,B., Cox,A.L., De Jonghe,J., Bao,M., Hollfelder,F., Glover,D.M. and Zernicka-Goetz,M. (2019) Self-Organization of mouse stem cells into an extended potential blastoid. *Dev. Cell*, **51**, 698–712.
14. Li,H., Zhao,C., Xu,J., Xu,Y., Cheng,C., Liu,Y., Wang,T., Du,Y., Xie,L., Zhao,J. *et al.* (2019) Rapid generation of gene-targeted EPS-derived mouse models through tetraploid complementation. *Protein Cell*, **10**, 20–30.
15. Shi,Y., Lee,J.S. and Galvin,K.M. (1997) Everything you have ever wanted to know about Yin Yang 1. *Biochim. Biophys. Acta*, **1332**, F49–F66.
16. Weintraub,A.S., Li,C.H., Zamudio,A.V., Sigova,A.A., Hannett,N.M., Day,D.S., Abraham,B.J., Cohen,M.A., Nabet,B., Buckley,D.L. *et al.* (2017) YY1 is a structural regulator of enhancer-promoter loops. *Cell*, **171**, 1573–1588.
17. He,H., Ye,A., Perera,B.P.U. and Kim,J. (2017) YY1's role in the Peg3 imprinted domain. *Sci. Rep.*, **7**, 6427.
18. Thiaville,M.M. and Kim,J. (2011) Oncogenic potential of yin yang 1 mediated through control of imprinted genes. *Crit. Rev. Oncog.*, **16**, 199–209.
19. Makhlouf,M., Ouimette,J.F., Oldfield,A., Navarro,P., Neuillet,D. and Rougeulle,C. (2014) A prominent and conserved role for YY1 in Xist transcriptional activation. *Nat. Commun.*, **5**, 4878.
20. Kim,J.D., Kang,K. and Kim,J. (2009) YY1's role in DNA methylation of Peg3 and Xist. *Nucleic Acids Res.*, **37**, 5656–5664.
21. Vella,P., Barozzi,I., Cuomo,A., Bonaldi,T. and Pasini,D. (2012) Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells. *Nucleic Acids Res.*, **40**, 3403–3418.
22. Donohoe,M., Zhang,X., McGinnis,L., Biggers,J., Li,E. and Shi,Y. (1999) Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality. *Mol. Cell Biol.*, **19**, 7237–7244.
23. Wang,J., Wu,X., Wei,C., Huang,X., Ma,Q., Huang,X., Faiola,F., Guallar,D., Fidalgo,M., Huang,T. *et al.* (2018) YY1 positively regulates transcription by targeting promoters and super-enhancers through the BAF complex in embryonic stem cells. *Stem Cell Rep.*, **10**, 1324–1339.
24. Guo,J., Wang,P., Sozen,B., Qiu,H., Zhu,Y., Zhang,X., Ming,J., Zernicka-Goetz,M. and Na,J. (2021) Machine learning-assisted high-content analysis of pluripotent stem cell-derived embryos in vitro. *Stem Cell Rep.*, **16**, 1331–1346.
25. Moffat,J., Grueneberg,D.A., Yang,X., Kim,S.Y., Kloepfer,A.M., Hinkle,G., Piqani,B., Eisenhaure,T.M., Luo,B., Grenier,J.K. *et al.* (2006) A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell*, **124**, 1283–1298.
26. Li,J., Huang,K., Hu,G., Babarinde,I.A., Li,Y., Dong,X., Chen,Y.S., Shang,L., Guo,W., Wang,J. *et al.* (2019) An alternative CTCF isoform antagonizes canonical CTCF occupancy and changes chromatin architecture to promote apoptosis. *Nat. Commun.*, **10**, 1535.
27. Niakan,K.K., Schrode,N., Cho,L.T. and Hadjantonakis,A.K. (2013) Derivation of extraembryonic endoderm stem (XEN) cells from mouse embryos and embryonic stem cells. *Nat. Protoc.*, **8**, 1028–1041.
28. Posfai,E., Schell,J.P., Janiszewski,A., Rovic,I., Murray,A., Bradshaw,B., Yamakawa,T., Pardon,T., El Bakkali,M., Talon,I. *et al.* (2021) Evaluating totipotency using criteria of increasing stringency. *Nat. Cell. Biol.*, **23**, 49–60.
29. Kubaczka,C. and Schorle,H. (2016) Protocol for the direct conversion of murine embryonic fibroblasts into trophoblast stem cells. *J. Vis. Exp.*, **113**, 54277.
30. Graf,U., Casanova,E.A., Wyck,S., Dalcher,D., Gatti,M., Vollenweider,E., Okoniewski,M.J., Weber,F.A., Patel,S.S., Schmid,M.W. *et al.* (2017) Pramel7 mediates ground-state pluripotency through proteasomal-epigenetic combined pathways. *Nat. Cell. Biol.*, **19**, 763–773.
31. Livak,K.J. and Schmittgen,T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta delta C(T)) method. *Methods*, **25**, 402–408.
32. Dobin,A., Davis,C., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
33. Anders,S., Pyl,P. and Huber,W. (2015) HTSeq–a python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.
34. Love,M., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
35. Zhou,Y., Zhou,B., Pache,L., Chang,M., Khodabakhshi,A.H., Tanaseichuk,O., Benner,C. and Chanda,S.K. (2019) Metascape

provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.*, **10**, 1523.

36. Butler,A., Hoffman,P., Smibert,P., Papalexi,E. and Satija,R. (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.*, **36**, 411–420.

37. Buenrostro,J.D., Wu,B., Chang,H.Y. and Greenleaf,W.J. (2015) ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.*, **109**, 21–29.

38. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.*, **17**, 10–12.

39. Langdon,W.B. (2015) Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. *BioData Min.*, **8**, 1.

40. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and Processing,GenomeProjectData, S. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

41. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.

42. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

43. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.

44. Li,H., Lai,P., Jia,J., Song,Y., Xia,Q., Huang,K., He,N., Ping,W., Chen,J., Yang,Z. *et al.* (2017) RNA helicase DDX5 inhibits reprogramming to pluripotency by miRNA-based repression of RYBP and its PRC1-dependent and -independent functions. *Cell Stem Cell*, **20**, 462–477.

45. Shao,Z., Zhang,Y., Yuan,G., Orkin,S. and Waxman,D. (2012) MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. *Genome Biol.*, **13**, R16.

46. Ramirez,F., Dundar,F., Diehl,S., Gruning,B.A. and Manke,T. (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.*, **42**, W187–W191.

47. Zhou,X., Maricque,B., Xie,M., Li,D., Sundaram,V., Martin,E.A., Koebbe,B.C., Nielsen,C., Hirst,M., Farnham,P. *et al.* (2011) The human epigenome browser at washington university. *Nat. Methods*, **8**, 989–990.

48. Kaya-Okur,H., Wu,S., Codomo,C., Pledger,E., Bryson,T., Henikoff,J., Ahmad,K. and Henikoff,S. (2019) CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.*, **10**, 1930.

49. Buenrostro,J., Wu,B., Litzenburger,U., Ruff,D., Gonzales,M., Snyder,M., Chang,H. and Greenleaf,W. (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, **523**, 486–490.

50. Kulis,M., Heath,S., Bibikova,M., Queirós,A., Navarro,A., Clot,G., Martínez-Trillos,A., Castellano,G., Brun-Heath,I., Pinyol,M. *et al.* (2012) Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat. Genet.*, **44**, 1236–1242.

51. Krueger,F. and Andrews,S. (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, **27**, 1571–1572.

52. Liang,Z., Li,G., Wang,Z., Djekidel,M.N., Li,Y., Qian,M.P., Zhang,M.Q. and Chen,Y. (2017) BL-Hi-C is an efficient and sensitive approach for capturing structural and regulatory chromatin interactions. *Nat. Commun.*, **8**, 1622.

53. Li,G., Chen,Y., Snyder,M.P. and Zhang,M.Q. (2017) ChIA-PET2: a versatile and flexible pipeline for ChIA-PET data analysis. *Nucleic Acids Res.*, **45**, e4.

54. Wolff,J., Bhardwaj,V., Nothjunge,S., Richard,G., Renschler,G., Gilsbach,R., Manke,T., Backofen,R., Ramírez,F. and Grüning,B. (2018) Galaxy hicexplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.*, **46**, W11–W16.

55. Lareau,C. and Aryee,M. (2018) diffloop: a computational framework for identifying and analyzing differential DNA loops from sequencing data. *Bioinformatics*, **34**, 672–674.

56. Creyghton,M.P., Cheng,A.W., Welstead,G.G., Kooistra,T., Carey,B.W., Steine,E.J., Hanna,J., Lodato,M.A., Frampton,G.M., Sharp,P.A. *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 21931–21936.

57. Atlasi,Y., Megchelenbrink,W., Peng,T., Habibi,E., Joshi,O., Wang,S.Y., Wang,C., Logie,C., Poser,I., Marks,H. *et al.* (2019) Epigenetic modulation of a hardwired 3D chromatin landscape in two naive states of pluripotency. *Nat. Cell Biol.*, **21**, 568–578.

58. Jia,D., Jurkowska,R.Z., Zhang,X., Jeltsch,A. and Cheng,X. (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature*, **449**, 248–251.

59. Bao,S., Tang,W.W., Wu,B., Kim,S., Li,J., Li,L., Kobayashi,T., Lee,C., Chen,Y., Wei,M. *et al.* (2018) Derivation of hypermethylated pluripotent embryonic stem cells with high potency. *Cell Res.*, **28**, 22–34.

60. Ayala-Ortega,E., Arzate-Mejía,R., Pérez-Molina,R., González-Buendía,E., Meier,K., Guerrero,G. and Recillas-Targa,F. (2016) Epigenetic silencing of miR-181c by DNA methylation in glioblastoma cell lines. *BMC Cancer*, **16**, 226.

61. Hark,A., Schoenherr,C., Katz,D., Ingram,R., Levorse,J. and Tilghman,S. (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature*, **405**, 486–489.

62. Bell,A.C. and Felsenfeld,G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, **405**, 482–485.

63. Gomez-Diaz,E. and Corces,V.G. (2014) Architectural proteins: regulators of 3D genome organization in cell fate. *Trends Cell Biol.*, **24**, 703–711.

64. Hu,G., Dong,X., Gong,S., Song,Y., Hutchins,A.P. and Yao,H. (2020) Systematic screening of CTCF binding partners identifies that BHLHE40 regulates CTCF genome-wide distribution and long-range chromatin interactions. *Nucleic Acids Res.*, **48**, 9606–9620.

65. Arzate-Mejía,R., Recillas-Targa,F. and Corces,V. (2018) Developing in 3D: the role of CTCF in cell differentiation. *Development*, **145**, dev137729.

66. Christensen,J., Agger,K., Cloos,P., Pasini,D., Rose,S., Sennels,L., Rappsilber,J., Hansen,K., Salcini,A. and Helin,K. (2007) RBP2 belongs to a family of demethylases, specific for tri-and dimethylated lysine 4 on histone 3. *Cell*, **128**, 1063–1076.

67. Iwase,S., Lan,F., Bayliss,P., Torre-Ubieta,L, Huarte,M., Qi,H., Whetstine,J., Bonni,A., Roberts,T. and Shi,Y. (2007) The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. *Cell*, **128**, 1077–1088.

68. Seidel,C., Schnekenburger,M., Dicato,M. and Diederich,M. (2015) Histone deacetylase 6 in health and disease. *Epigenomics*, **7**, 103–118.

69. Pijuan-Sala,B., Griffiths,J.A., Guibentif,C., Hiscock,T.W., Jawaid,W., Calero-Nieto,F.J., Mulas,C., Ibarra-Soria,X., Tyser,R.C.V., Ho,D.L.L. *et al.* (2019) A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature*, **566**, 490–495.

70. Zhong,Y., Choi,T., Kim,M., Jung,K.H., Chai,Y.G. and Binas,B. (2018) Isolation of primitive mouse extraembryonic endoderm (pXEN) stem cell lines. *Stem Cell Res.*, **30**, 100–112.

71. Klemm,S., Shipony,Z. and Greenleaf,W. (2019) Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.*, **20**, 207–220.

72. Schlesinger,S. and Meshorer,E. (2019) Open chromatin, epigenetic plasticity, and nuclear organization in pluripotency. *Dev. Cell*, **48**, 135–150.

73. Aapola,U., Kawasaki,K., Scott,H.S., Ollila,J., Vihinen,M., Heino,M., Shintani,A., Kawasaki,K., Minoshima,S., Krohn,K. *et al.* (2000) Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. *Genomics*, **65**, 293–298.

74. Bourc'his,D., Xu,G., Lin,C., Bollman,B. and Bestor,T. (2001) Dnmt3L and the establishment of maternal genomic imprints. *Science*, **294**, 2536–2539.

75. Tunster,S.J., Tycko,B. and John,R.M. (2010) The imprinted Phlda2 gene regulates extraembryonic energy stores. *Mol. Cell Biol.*, **30**, 295–306.

76. Stumpo,D.J., Byrd,N.A., Phillips,R.S., Ghosh,S., Maronpot,R.R., Castranio,T., Meyers,E.N., Mishina,Y. and Blackshear,P.J. (2004) Chorioallantoic fusion defects and embryonic lethality resulting from disruption of Zfp36l1, a gene encoding a CCCH tandem zinc finger protein of the tristetraprolin family. *Mol. Cell Biol.*, **24**, 6445–6455.

77. Cabrita,M.A. and Christofori,G. (2008) Sprouty proteins, masterminds of receptor tyrosine kinase signaling. *Angiogenesis*, **11**, 53–62.

78. Fu,K., Bonora,G. and Pellegrini,M. (2020) Interactions between core histone marks and DNA methyltransferases predict DNA methylation patterns observed in human cells and tissues. *Epigenetics*, **15**, 272–282.

79. Li,Y., Xia,L., Tan,K., Ye,X., Zuo,Z., Li,M., Xiao,R., Wang,Z., Liu,X., Deng,M. *et al.* (2020) N^6-Methyladenosine co-transcriptionally directs the demethylation of histone H3K9me2. *Nat. Genet.*, **52**, 870–877.

80. Flavahan,W.A., Drier,Y., Johnstone,S.E., Hemming,M.L., Tarjan,D.R., Hegazi,E., Shareef,S.J., Javed,N.M., Raut,C.P., Eschle,B.K. *et al.* (2019) Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs. *Nature*, **575**, 229–233.

81. He,J., Fu,X., Zhang,M., He,F., Li,W., Abdul,M.M., Zhou,J., Sun,L., Chang,C., Li,Y. *et al.* (2019) Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. *Nat. Commun.*, **10**, 34.

82. Guallar,D., Perez-Palacios,R., Climent,M., Martinez-Abadia,I., Larraga,A., Fernandez-Juan,M., Vallejo,C., Muniesa,P. and Schoorlemmer,J. (2012) Expression of endogenous retroviruses is negatively regulated by the pluripotency marker Rex1/Zfp42. *Nucleic Acids Res.*, **40**, 8993–9007.

83. Schoorlemmer,J., Perez-Palacios,R., Climent,M., Guallar,D. and Muniesa,P. (2014) Regulation of mouse retroelement MuERV-L/MERVL expression by REX1 and epigenetic control of stem cell potency. *Front. Oncol.*, **4**, 14.

84. Yamanaka,Y., Ralston,A., Stephenson,R.O. and Rossant,J. (2006) Cell and molecular regulation of the mouse blastocyst. *Dev. Dyn.*, **235**, 2301–2314.

85. Schenke-Layland,K., Angelis,E., Rhodes,K., Heydarkhan-Hagvall,S., Mikkola,H. and Maclellan,W. (2007) Collagen IV induces trophoectoderm differentiation of mouse embryonic stem cells. *Stem Cells*, **25**, 1529–1538.

86. Prudhomme,J. and Morey,C. (2016) Epigenesis and plasticity of mouse trophoblast stem cells. *Cell Mol. Life Sci.*, **73**, 757–774.

87. Cho,L., Wamaitha,S., Tsai,I., Artus,J., Sherwood,R., Pedersen,R., Hadjantonakis,A. and Niakan,K. (2012) Conversion from mouse embryonic to extra-embryonic endoderm stem cells reveals distinct differentiation capacities of pluripotent stem cell states. *Development*, **139**, 2866–2877.

88. Kunath,T., Arnaud,D., Uy,G., Okamoto,I., Chureau,C., Yamanaka,Y., Heard,E., Gardner,R., Avner,P. and Rossant,J. (2005) Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts. *Development*, **132**, 1649–1661.

89. Perez-Palacios,R., Macias-Redondo,S., Climent,M., Contreras-Moreira,B., Muniesa,P. and Schoorlemmer,J. (2016) In vivo chromatin targets of the transcription factor Yin Yang 2 in trophoblast stem cells. *PLoS One*, **11**, e0154268.

90. Kim,J.D., Faulk,C. and Kim,J. (2007) Retroposition and evolution of the DNA-binding motifs of YY1, YY2 and REX1. *Nucleic Acids Res.*, **35**, 3442–3452.

91. Tahmasebi,S., Jafarnejad,S.M., Tam,I.S., Gonatopoulos-Pournatzis,T., Matta-Camacho,E., Tsukumo,Y., Yanagiya,A., Li,W., Atlasi,Y., Caron,M. *et al.* (2016) Control of embryonic stem cell self-renewal and differentiation via coordinated alternative splicing and translation of YY2. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 12360–12367.

92. Pérez-Palacios,R., Climent,M., Santiago-Arcos,J., Macías-Redondo,S., Klar,M., Muniesa,P. and Schoorlemmer,J. (2021) YY2 in mouse preimplantation embryos and in embryonic stem cells. *Cells*, **10**, 1123.

93. Lawson,K.A., Dunn,N.R., Roelen,B.A., Zeinstra,L.M., Davis,A.M., Wright,C.V., Korving,J.P. and Hogan,B.L. (1999) Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev.*, **13**, 424–436.

94. Gotoh,N., Manova,K., Tanaka,S., Murohashi,M., Hadari,Y., Lee,A., Hamada,Y., Hiroe,T., Ito,M., Kurihara,T. *et al.* (2005) The docking protein FRS2alpha is an essential component of multiple fibroblast growth factor responses during early mouse development. *Mol. Cell Biol.*, **25**, 4105–4116.