# The pluripotent stem cell-specific transcript ESRG is dispensable for human pluripotency

Kazutoshi Takahashi[1, 2, *], Michiko Nakamura[1], Megumi Narita[1], Akira Watanabe[3], Mai Ueda[1], Yasuhiro Takashima[1], Shinya Yamanaka[1, 2, 4]

[1]Department of Life Science Frontiers, Center for iPS Cell Research and Application, Kyoto University, Kyoto, Japan
[2]Gladstone Institute of Cardiovascular Disease, San Francisco, California, United States of America
[3]Graduate School of Medicine, Kyoto University, Kyoto, Japan
[4]Department of Anatomy, University of California, San Francisco, San Francisco, California, United States of America

*Corresponding Author
E-mail: kazu@cira.kyoto-u.ac.jp

## Abstract

Human pluripotent stem cells (PSCs) express human endogenous retrovirus type-H (HERV-H), which exists as more than a thousand copies on the human genome and frequently produces chimeric transcripts as long-non-coding RNAs (lncRNAs) fused with downstream neighbor genes. Previous studies showed that HERV-H expression is required for the maintenance of PSC identity, and aberrant HERV-H expression attenuates neural differentiation potentials, however, little is known about the actual of function of HERV-H. In this study, we focused on ESRG, which is known as a PSC-related HERV-H-driven lncRNA. The global transcriptome data of various tissues and cell lines and quantitative expression analysis of PSCs showed that ESRG expression is much higher than other HERV-Hs and tightly silenced after differentiation. However, the loss of function by the complete excision of the entire ESRG gene body using a CRISPR/Cas9 platform revealed that ESRG is dispensable for the maintenance of the primed and naïve pluripotent states. The loss of ESRG hardly affected the global gene expression of PSCs or the differentiation potential toward trilineage. Differentiated cells derived from ESRG-deficient PSCs retained the potential to be reprogrammed into induced PSCs (iPSCs) by the forced expression of OCT3/4, SOX2, and KLF4. In conclusion, ESRG is dispensable for the maintenance and recapturing of human pluripotency.

## Introduction

Human pluripotent stem cells (PSCs) express several types of human endogenous retroviruses (HERV) [1-3]. The HERV type-H (HERV-H) family is a primate-specific ERV element that was first integrated into new world monkeys. During further primate evolution, this family's major expansion occurred before the branch of old world monkeys [4]. The typical structure of a HERV-H consists of an interior component, HERVH-int, flanked by two long terminal repeat 7 (LTR7), which have promoter activity [5, 6]. Recent studies have demonstrated that the activity of LTR7 is highly specific in established human PSCs and relatively absent in early human embryos. In contrast, other LTR7 variants such as LTR7B, C, and Y are activated in broad types of early human embryos from the 8-cell to epiblast stages [7].

The importance of HERV-Hs in human PSCs has been shown. The knockdown of pan HERV-Hs using short hairpin RNAs (shRNAs) against conserved sequences in LTR7 regions revealed that HERV-H expression is required for the self-renewal of human PSCs [8, 9] and somatic cell reprogramming toward pluripotency [8-14]. In addition to the self-renewal, the precise expression of HERV-Hs is crucial for the neural differentiation potential of human PSCs [10, 15]. In this way, HERV-H expression contributes to PSC identity.

The transcription of HERV-H frequently produces a chimeric transcript fused with a downstream neighbor gene, which diversifies HERV-H-driven transcripts. Therefore, many HERV-H-driven RNAs contain unique sequences aside from HERV-H consensus sequences. Indeed, PSC-associated HERV-H-containing long non-coding RNAs (lncRNAs) have been reported [15-17]. One of them, ESRG (embryonic stem cell-related gene; also known as HESRG) was identified as a transcript that is predominantly expressed in undifferentiated human embryonic stem cells (ESCs) [18, 19]. ESRG is transcribed from HERV-H LTR7 promoter [8, 20] and is activated in an early stage of somatic cell reprogramming induced by the forced expression of OCT3/4, SOX2, and KLF4 (OSK) [12, 13, 20]. One previous study showed that the short hairpin RNA (shRNA)-mediated knockdown of ESRG induces the loss of PSC characters such as colony morphology and PSC markers along with the activation of differentiation markers, suggesting the indispensability of ESRG for human pluripotency [8]. However, despite these characterizations, the function of ESRG is still unknown.

In this study, we completely deleted ESRG alleles to analyze ESRG function in human PSCs with no off-target risk. Surprisingly, the loss of ESRG, which is thought to be an essential lncRNA for PSC identity, exhibited no impact on the self-renewal or differentiation potentials of both primed and naïve human PSCs. Neural progenitor cells (NPCs) derived from ESRG-deficient PSCs could be reprogrammed into induced PSC (iPSC) by OSK expression. This study revealed that ESRG is dispensable for human pluripotency.

## Results

### ESRG is robustly expressed in PSCs and tightly silenced after differentiation

The RNA sequencing (RNA-seq) and chromatin immunoprecipitation (ChIP-seq) of histone H3 modifications indicated that the ESRG locus is open and actively transcribed in human PSCs but not in differentiated cells such as human dermal fibroblasts (HDFs) (Fig. 1A). As well as other HERV-H-related genes, LTR7 elements in ESRG gene are occupied by pluripotency-associated transcription factors (TFs) such as OSK [9, 10] (Fig. 1A). Little or no ESRG expression was detected in 24 adult tissues and five fetal tissues (Fig. S1A). Compared to other PSC-associated HERV-H chimeric transcripts, ESRG expression exhibits a sharp contrast between PSCs and somatic tissues [8, 10, 15-17]. Furthermore, ESRG is expressed in PSCs, including embryonic carcinoma cell (ECC) lines, but is silenced in four cancer cell lines and ten cell lines derived from normal tissue (Fig. S1B). Quantitative reverse transcription-polymerase chain reaction (qRT-PCR) revealed that the ESRG expression is significantly higher than the expression of other HERV-H-related transcripts and is comparable to the expression of SOX2 and NANOG, which play essential roles in pluripotency (Fig. 1B). These data suggest that ESRG expression is abundant in PSCs and is tightly silenced in differentiated states.

### ESRG is dispensable for human pluripotency

The above results led us to hypothesize the important role of ESRG in human PSCs. To make a complete loss of function of the lncRNA ESRG, we employed a CRISPR/Cas9 platform and two single guide RNAs (sgRNAs) that flanked ~8,400 bp of the genomic region including the entire ESRG gene based on the human genome database and RNA-seq data (Figs. 1A and S2A). As a result, we obtained multiple independent ESRG knockout (KO) PSC lines that exhibit complete deletion of the gene body with unique minor deletion patterns in both alleles under a primed PSC culture condition (Figs. S2B and S2C). In this study, we used three clones as wild-type (WT) controls carrying intact ESRG alleles with no or minor deletions at the sgRNA recognition sites (Fig. S2D). We performed all subsequent experiments in 3 WT versus 3 KO manner. The expression of ESRG was undetectable in the KO clones by qRT-PCR (Fig. 1C). Immunocytochemistry showed that ESRG KO PSCs express the PSC core transcription factors (Fig. 1D) and PSC-specific surface antigens (Fig. 1E). The loss of ESRG made no impact on the expression of neighbor genes located within 10 Mbp of ESRG (Fig. 1F). Global transcriptome analysis revealed that the loss of ESRG altered the expression of only a few genes (Fig. 1G). Moreover, ESRG KO PSCs normally survived while maintaining the undifferentiated state as judged by alkaline phosphatase (AP) activity and the absence of any apparent genomic abnormalities (Figs. 1H and S3). These data suggest that ESRG is dispensable for the self-renewing of primed PSCs.

We also tested if ESRG is required for another state of pluripotency, the so-called naïve state, which also expresses ESRG but at a significantly lower level than the primed state (Fig. 2A). Regardless of the

ESRG expression, naïve PSCs could be established by switching the media composition and could self-renew while keeping a tightly packed colony formation (Fig. 2B) [21-23]. Furthermore, they exhibited a significantly high expression of the naïve pluripotency markers KLF4 and KLF17 and attenuated the expression of the primed PSC marker ZIC2 (Fig. 2C) [24, 25]. Microarray analysis revealed that ESRG had no effect on the global gene expression of either primed or naïve PSCs (Fig. 2D). We also differentiated ESRG WT and KO naïve PSCs to the primed pluripotent state. As a result, irrespective of the ESRG genotype, we detected the hallmarks of primed pluripotency such as flatter colony formation, the reactivation of ZIC2 and the suppression of KLF4 and KLF17, suggesting the bidirectional transition between naïve and primed pluripotency does not require ESRG (Figs. 2E and 2F). Taken together, these data demonstrate that ESRG is dispensable for the maintenance of human PSCs.

**ESRG is not involved in differentiation**

Next, we analyzed whether ESRG is required for the differentiation of PSCs by embryoid body (EB) formation. The absence of ESRG had no effect on EB formation by floating culture or differentiation into trilineage such as alpha-fetoprotein (AFP) positive (+) endoderm, smooth muscle actin (SMA) (+) mesoderm and βIII-TUBULIN (+) ectoderm (Figs. 3A and 3B). Other lineage markers such as DCN (endoderm), MSX1 (mesoderm) and MAP2 (ectoderm) were also well induced in EBs derived from either ESRG WT or KO PSCs (Fig. 3C). Global transcriptome analysis by microarray indicated the loss of ESRG caused no significant gene expression changes during EB differentiation (Fig. 3D). These data suggest that ESRG KO PSCs retained the potential to differentiate into all three germ layers.

Previous studies showed that HERV-H expression regulates the neural differentiation potential of human PSCs [10, 15, 26]. Thus, in addition to the random differentiation by EB formation, we tested whether ESRG contributes to the directed differentiation of PSCs into NPCs by the dual SMAD inhibition method [27, 28]. Both ESRG WT and KO PSCs were able to differentiate into expandable NPCs, which expressed the early neural lineage marker PAX6 but not OCT3/4 (Fig. 3E). Other NPC markers such as SOX1 and NES were well induced, whereas the PSC marker NANOG was silenced (Fig. 3F). These data suggest that ESRG is not responsible for HERV-H-regulated neural differentiation. Taken together, we concluded that ESRG is not required for the differentiation of human PSCs.

**ESRG is not required for somatic cell reprogramming toward pluripotency**

A previous study showed that the overexpression of ESRG improves iPSC generation [8], suggesting a positive effect on somatic cell reprogramming toward pluripotency. The activation of ESRG in the early stage of reprogramming and the high expression of ESRG during reprogramming support this hypothesis (Fig. 4A) [20]. Therefore, we reprogrammed ESRG WT and KO NPCs to iPSCs by introducing OSK. iPSCs emerged from ESRG WT and KO NPCs with comparable efficiency (Fig. 4B). This observation

suggests that ESRG is dispensable for iPSC generation. In addition, along with OSK, we transduced c-MYC, a potent enhancer of iPSC generation [29, 30], or exogenous ESRG. c-MYC but not exogenous ESRG increased the efficiency of the iPSC generation from ESRG WT and KO NPCs equally (Fig. 4B). Taken together, these data suggest that ESRG has no impact on somatic cell reprogramming toward iPSCs.

## Discussion

In this study, we completely excised the entire ESRG gene to understand its role in human PSCs while avoiding residual expression and off-target effects. Unexpectedly, ESRG KO PSCs showed no apparent phenotypes in self-renewal or differentiation potential. A previous study showed the importance of ESRG in human PSC identity by using an shRNA-mediated knockdown approach [8]. Although we used the same H9 ESC line as that study, the different strategies for the loss of function and subsequent experiments, such as knockdown and knockout, may explain the different results. For example, the speed of the loss of function may differ. Generally, RNA interference (RNAi) induces acute silencing of the target gene expression, whereas the knockout process is relatively slow and has multiple stages, including CRISPR/Cas9-mediated deletion, clonal expansion, subcloning and others. Another possibility is the off-target effect of RNAi. Similar observations have been found for the role of lncRNA Cyrano that is highly conserved in mouse and human. Knockdown by using shRNA suggested Cyrano lncRNA maintains mouse PSC identify [31], but targeted deletion of the Cyrano gene and gene silencing by CRISPR interference demonstrated no impact on mouse or human PSC identity [32-34]. Further, it has been argued that the shRNA-mediated knockdown of nuclear lncRNAs might be difficult or inefficient compared to cytoplasmic RNAs such as mRNAs [35, 36]. In addition, while small nucleotide insertions or deletions causing frameshift of the reading frames work well for the loss of function of protein-coding genes, the same is not true for non-coding RNAs. In this context, our study succeeded to generate the complete deletion of ESRG gene alleles, providing highly reliable results.

This study clearly demonstrated that ESRG is dispensable for human PSC identity. Neither primed nor naïve PSCs require ESRG for their identities, such as colony morphology or gene expression signatures, meaning ESRG is dispensable for human pluripotency, at least in an in vitro culture environment. However, since ESRG is expressed in epiblast-stage human embryos [8, 37], it might be involved in early human embryogenesis.

ESRG is stochastically activated by OSK in rare reprogrammed intermediates that have the potential to become bona fide iPSCs and is highly expressed throughout the process of reprogramming toward iPSCs [20]. In the present study, we showed that ESRG KO NPCs can be reprogrammed with the same efficiency as ESRG WT NPCs. These data suggest that ESRG is a good marker of the intermediate cells

in early stage of reprogramming rather than a functional molecule that is need for iPSC generation.

In summary, this study provides clear evidence of the dispensability of ESRG for PSC identities, such as global gene expressions and differentiation potentials, in two distinct types of pluripotent states. We also demonstrated that the function of ESRG is not required for recapturing pluripotency via somatic cell reprogramming. Finally, the tightly regulated and high expression of ESRG promises to make an excellent marker of undifferentiated PSCs both in basic research and clinical application [20, 38].

## Methods

**The culture of primed PSCs** | H9 ESC (RID:CVCL_9773) and 585A1 iPSC (RRID:CVCL_DQ06) lines were maintained in StemFiT AK02 media (Ajinomoto) supplemented with 100 ng/ml recombinant human basic fibroblast growth factor (bFGF, Peprotech) (hereafter F/A media) on a tissue culture plate coated with Laminin 511 E8 fragment (LN511E8, NIPPI). A 201B7 iPSC (RRID:CVCL_A324) line was cultured on mitomycin C (MMC)-inactivated SNL mouse feeder cells (RRID:CVCL_K227) in Primate ESC Culture medium (ReproCELL) supplemented with 4 ng/ml bFGF.

**Induction and maintenance of naïve PSCs** | The conversion of primed PSCs to the naïve state was performed as described previously [23]. Prior to naïve conversion, primed PSCs were maintained on MMC-treated primary mouse embryonic fibroblasts (PMEFs) in DFK20 media consisting of DMEM/F12 (Thermo Fisher Scientific), 20% Knockout Serum Replacement (KSR, Thermo Fisher Scientific), 1% MEM non-essential amino acids (NEAA, Thermo Fisher Scientific), 1% GlutaMax (Thermo Fisher Scientific) and 0.1 mM 2-mercaptoethanol (2-ME, Thermo Fisher Scientific)) supplemented with 4 ng/ml bFGF. The cells were harvested using CTK solution (ReproCELL) and dissociated to single cells. One hundred thousand cells were plated onto MMC-treated PMEFs in a well of a 6-well plate in DFK20 media plus bFGF and 10 μM Y-27632. Thereafter, the cells were incubated in hypoxic condition (5% $O_2$). On the next day, the media was replaced with NDiff227 (Takara) supplemented with 1 μM PD325901 (Stemgent), 10 ng/ml of recombinant human leukemia inhibitory factor (LIF, EMD Millipore) and 1 mM Valproic acid (Wako). Three days later, the media was switched to PXGL media (NDiff227 supplemented with 1 μM PD325901, 2 μM XAV939 (Wako), 2 μM Gö6983 (Sigma Aldrich) and 10 ng/ml of LIF). When round shape colonies were visible (around day 9 of the conversion), the cells were dissociated using TrypLE Express (Thermo Fisher Scientific) and plated onto a new PMEF feeder plate in PXGL media plus 10 μM Y-27632. The media was changed daily, and the cells were passaged every 4-5 days. Cells after at least 30 days of the conversion were used for the assays.

**Differentiation of naïve PSCs to the primed state** | Naïve PSCs were harvested using TrypLE Express and plated at 5 x $10^5$ cells onto a well of a LN511E8-coated 6-well plate in PXGL media supplemented with 10 μM Y-27632. On the next day, the media was replaced with F/A media. After 2 and 8 days, the cells were harvested and split to a new LN511E8-coated plate in F/A media plus 10 μM Y-27632. On day 16 of the differentiation, the cells were fixed for immunocytochemistry, and RNA samples were collected to analyze the marker gene expression.

**Induction and maintenance of NPCs** | Primed PSCs were differentiated into expandable NPCs by using the STEMdiff SMADi Neural Induction Kit (Stem Cell Technologies) as previously described [26-28]. In

brief, primed PSCs were maintained on a Matrigel (Corning)-coated plate in mTeSR1 media (Stem Cell Technologies) prior to the NPC induction. The cells were harvested using Accutase (EMD Millipore) and transferred at 3 x $10^6$ cells to a well of an AgrreWell800 plate (Stem Cell Technologies) in STEMdiff Neural Induction Medium + SMADi (Stem Cell Technologies) supplemented with 10 μM Y-27632. Five days later, uniformly sized aggregates were collected using a 37 μm Reversible Strainer (Stem Cell Technologies) and plated onto a Matrigel-coated 6-well plate in STEMdiff Neural Induction Medium + SMADi. Seven days later, neural rosette structures were selectively removed by using STEMdiff Neural Rosette Selection Reagent (Stem Cell Technologies) and plated onto a new Matrigel-coated 6-well plate in STEMdiff Neural Induction Medium + SMADi. After that, the cells were passaged every 2-3 days until day 30 post-differentiation. The established NPCs were maintained on a Matrigel-coated plate in STEMdiff Neural Progenitor Medium (Stem Cell Technologies) and passaged every 3-4 days.

**The culture of other cells** | HDFs and PLAT-GP packaging cells (RRID:CVCL_B490) were cultured in DMEM (Thermo Fisher Scientific) containing 10% fetal bovine serum (FBS, Thermo Fisher Scientific).

**Embryoid body (EB) differentiation** | PSCs were cultured on a Matrigel-coated plate in mTeSR1 media until reaching confluency prior to EB formation. The cells were harvested using CTK solution (ReproCELL), and cell clumps were transferred onto an ultra-low binding plate (Corning) in DFK20 media. For the first 2 days, 10 μM Y-27362 was added to the media to improve cell survival. The media was changed every other day. After 8 days of floating culture, the EBs were we transferred onto a tissue culture plate coated with 0.1% gelatin (EMD Millipore) and maintained in DFK20 media for another 8 days.

**Plasmid** | Full-length ESRG complementary DNA (cDNA) was amplified using ESRG-S and ESRG-AS primers and inserted into the BamHI/NotI site of a pMXs retroviral vector [39] using In-Fusion technology (Clontech). The primer sequences for the cloning are available in S1 Table.

**Reprogramming** | Retroviral transduction of the reprogramming factors was performed as described previously [12, 20]. A pMXs retroviral vector encoding human OCT3/4 (RRID:Addgene_17217), human SOX2 (RRID:Addgene_17218), human KLF4 (RRID:Addgene_17219), human c-MYC (RRID:Addgene_17220) and ESRG (6 μg each) along with 3 μg of pMD2.G (gift from Dr. D. Trono; RRID:Addgene_12259) was transfected into PLAT-GP packaging cells, which were plated at 3.6 x $10^6$ cells per 100 mm dish the day before transfection, using FuGENE6 transfection reagent (Promega). Two days after the transfection, virus-containing supernatant was collected and filtered through a 0.45 μm-pore size cellulose acetate filter to remove the cell debris. Viral particles were precipitated using Retro-X

Concentrator (Clontech) and resuspended in STEMdiff Neural Progenitor Medium containing 8 µg/ml Polybrene (EMD Millipore). Then, appropriate combinations of viruses were mixed and used for the transduction to NPCs. This point was designated day 0. The cells were harvested on day 3 post-transduction and replated at 5 x $10^4$ cells per well of a LN511E8-coated 6-well plate in STEMdiff Neural Progenitor Medium. The following day (day 4), the medium was replaced with F/A media, and the medium was changed every other day. The iPSC colonies were counted on day 24 post-transduction. Bona fide iPSC colonies were distinguished from non-iPSC colonies by their morphological differences and/or alkaline phosphatase activity.

**Deletion of ESRG gene** | A ribonucleoprotein complex consisting of 40 pmol of Alt-R S.p. HiFi Cas9 Nuclease V3 (Integrated DNA Technologies) and two single guide RNAs (sgRNAs): sgESRG-U (5'-AGAGAAUACGAAGCUAAGUG-3') and sgESRG-L (5'-AUUGCAGUUGUCACAUGACA-3'), 150 pmol each; SYNTHEGO) was introduced into 5 x $10^5$ H9 ESCs (passage number 49) from a subconfluent culture using a 4D-Nucleofector System with X Unit (Lonza) and P3 Primary Cell 4D-Nucleofector Kit S (Lonza) with the CA173 program. Three days after the nucleofection, the cells were harvested and replated at 500 cells onto a LN511E8-coated 100 mm dish in F/A media supplemented with 10 µM Y-27632. The cells were maintained until the colonies grew big enough for subcloning. The colonies were mechanically picked up, dissociated using TrypLE select and plated onto a LN511E8-coated 12-well plate in F/A media supplemented with 10 µM Y-27632.

The genomic DNA of the expanded clones was purified using the DNeasy Blood & Tissue Kit (QIAGEN). Fifty nanograms of purified DNA was used for quantitative polymerase chain reaction (PCR) using TaqMan Genotyping Master Mix (Thermo Fisher Scientific) on an ABI7900HT Real Time PCR System (Applied Biosystems). TaqMan Assays (Thermo Fisher Scientific) such as ESRG_cn1 (Hs05898393_cn) and ESRG_cn2 (Hs06675423_cn) detected the ESRG locus and TaqMan Copy Number Reference Assay human RNase P (4403326, Thermo Fisher Scientific) was used as an endogenous control. To verify the indel patterns in wild-type clones, fragment around the sgESRG-U and sgESRG-L recognition sites were amplified with ESRG-U-S/ESRG-U-AS and ESRG-L-S/ESRG-L-AS primer sets, respectively. The amplicons were purified using the QIAquick PCR Purification Kit (QIAGEN) and subjected to sequencing. To check the deleted sequences in the knockout clones, a fragment with ESRG-U-S/ESRG-L-AS primers was amplified. Conventional PCR was performed using KOD Xtreme Hot Start DNA Polymerase (EMD Millipore). The fragments were cloned into pCR-Blunt II TOPO using the Zero Blunt TOPO PCR Cloning Kit (Thermo Fisher Scientific), and the sequencing was verified using M13 forward and M13 reverse universal primers. The sequence data was analyzed using SnapGene software (GSL Biotech LLC). The primer sequences are provided in S1 Table.

**RNA isolation and reverse-transcription polymerase chain reaction** | The cells were lysed with QIAzol reagent (QIAGEN), and the total RNA was purified using a miReasy Mini Kit (QIAGEN) according to the manufacturer's protocol. The reverse transcription (RT) of 1 μg of purified RNA was done by using SuperScript III First-Strand Synthesis SuperMix (Thermo Fisher Scientific). Quantitative RT-PCR was performed using TaqMan Assays with TaqMan Universal Master Mix II, no UNG (Applied Biosystems) on an ABI7900HT or a QuantoStudio 5 Real Time PCR System (Applied Biosystems). The $C_t$ values of the undetermined signals caused by too low expression was set at 40. The levels of mRNA were normalized to the GAPDH expression, and the relative expression was calculated as the fold-change from the control. Information about the TaqMan Assays is shown in S2 Table.

**Gene expression analysis by microarray** | The total RNA samples were purified using the miReasy Mini Kit, and the quality was evaluated using a 2100 Bioanalyzer (Agilent Technologies). Two hundred nanograms of total RNA was labeled with Cyanine 3-CTP and used for hybridization with SurePrint G3 Human GE 8x60K (version 1 (G4851A) and version 3 (G4851C), Agilent Technologies) and the one-color protocol. The hybridized arrays were scanned with a Microarray Scanner System (G2565BA, Agilent Technologies), and the extracted signals were analyzed using the GeneSpring version 14.6 software program (Agilent Technologies). Gene expression values were normalized by 75th percentile shifts. Differentially expressed genes between ESRG WT and KO ESCs were extracted by t-tests with Benjamini and Hochberg corrections [fold change (FC) > 2.0, false-discovery rate (FDR) < 0.05].

**Immunocytochemistry** | The cells were washed once with PBS, fixed with fixation buffer (BioLegend) for 15 min at room temperature and blocked in PBS containing 1% bovine serum albumin (BSA, Thermo Fisher Scientific), 2% normal donkey serum (Sigma-Aldrich) for 45 min at room temperature. For the staining of intracellular proteins, the fixed cells were permeabilized by adding 0.2% TritonX-100 (Teknova) during the blocking process. Then the cells were incubated with primary antibodies diluted in PBS containing 1% BSA at 4°C overnight. After washing with PBS, the cells were incubated with secondary antibodies diluted in PBS containing 1% BSA and 1 μg/ml Hoechst 33342 (Thermo Fisher Scientific) for 45 min at room temperature in the dark. The fluorescent signals were detected using a BZ-X710 imaging system (KEYENCE). The antibodies and dilution rate were as follows: anti-OCT3/4 (1:250, 611203, BD Biosciences), anti-SOX2 (1:100, ab97959, Abcam), anti-NANOG (1:100, ab21624, Abcam), anti-KLF17 (1:100, HPA024629, Atlas Antibodies), anti-PAX6 (1:1,000, 901301, BioLegend), SSEA3 (1:100, 09-0044, Stemgent), SSEA4 (1:100, 09-0006, Stemgent), SSEA5 (1:100, 355201, BioLegend), TRA-1-60 (1:100, MAB4360, EMD Millipore), TRA-2-49/6E (1:100, 358702, BioLegend), anti-AFP (1:200, GTX15650, GeneTex), anti-SMA (1:200, CBL171-I, EMD Millipore), anti-βIII-TUBULIN (1:1,000, XMAB1637, EMD

Millipore), Alexa 488 Plus anti-mouse IgG (1:500, A32766, Thermo Fisher Scientific), Alexa 647 Plus anti-rabbit IgG (1:500, A32795, Thermo Fisher Scientific), Alexa 594 anti-rat IgM (1:500, A21213, Thermo Fisher Scientific) and Alexa 555 anti-mouse IgM (1:500, A21426, Thermo Fisher Scientific).

**Quantification and statistical analysis** | Data are presented as the mean ± standard deviation unless otherwise noted. Sample number (n) indicates the number of replicates in each experiment. The number of experimental repeats is indicated in the figure legends. To determine statistical significance, we used the unpaired t-test for comparisons between two groups using Excel Microsoft 365 (Microsoft). Statistical significance was set at $p < 0.05$. All graphs and heatmaps were generated using GraphPad Prism 8 software (GraphPad).

**Data availability** | RNA-seq and ChIP-seq (GSE56569 and GSE89976) and Gene expression microarray (GSE54848, GSE156834 and GSE159101) results are accessible in the Gene Expression Omnibus database of the National Center for Biotechnology Information website.

## Acknowledgements

We would like to thank M. Iwasaki, M. Koyanagi-Aoi, A. Kunitomi, K. Okita, M. Ohnuki and D. Trono for sharing materials and data, MA. Khurram, SD. Perli, S. Wang and K. Tomoda for discussions, and Y. Kawahara, M. Lancero and R. Hirohata for technical assistance. We are also grateful to K. Essex, K. Higashi, K. Kamegawa, M. Otsuki, M. Saito and S. Takeshima for administrative support, and P. Karagiannis for crucial reading of the manuscript.

## Funding

This work was supported by Grants-in-Aid for Scientific Research (20K20585) from the Japanese Society for the Promotion of Science (JSPS); a grant from the Core Center for iPS Cell Research (19bm0104001h0007), Research Center Network for Realization of Regenerative Medicine from Japan Agency for Medical Research and Development (AMED); a grant from the Japan Foundation for Applied Enzymology; a grant from the Fujiwara Memorial Foundation; a grant from the Takeda Science Foundation; and the iPS Cell Research Fund from Center for iPS Cell Research and Application, Kyoto University. The study was also supported by funding from Mr. H. Mikitani, Mr. M. Benioff, and the L.K. Whittier Foundation.

## Author contributions

Conceptualization, K.T. and S.Y.; Methodology, K.T., M.U., and Y.T.; Investigation, K.T., M.Nakamura, M.Narita, and A.W.; Formal Analysis, K.T and A.W.; Writing – Original Draft, K.T.; Funding Acquisition, K.T. and S.Y.; Resources, K.T., M.U., and Y.T.; Supervision, K.T. and S.Y.

## Competing interests

K.T. is on the scientific advisory board of I Peace, Inc. without salary. S.Y. is a scientific advisor (without salary) of iPS Academia Japan. All other authors have no conflict of interest.

## Figure legends

**Figure 1 | ESRG is dispensable for primed pluripotency**

(A) Epigenetic status of the ESRG locus. Shown are RNA-seq and ChIP-seq data for histone modifications and PSC core transcription factor (TF) binding on the ESRG locus in HDFs and iPSCs on human genome assembly hg19. The green arrowheads at the bottom indicate the location of the LTR7 elements. (B) Expression of PSC-associated mRNAs and HERV-H chimeric RNAs. Shown are the averaged expressions of the indicated transcripts in H9 ESCs, 585A1 iPSCs and 201B7 iPSCs. Error bars and white lines indicate min. to max. and the mean of each gene expression, respectively. Values are compared to GAPDH. n=3. (C) Expression of ESRG in ESRG WT and KO PSC clones. Values are normalized by GAPDH and compared with primed H9 ESCs. n=3. (D) Expression of PSC core transcription factors. Bars,100 μm. (E) Expression of PSC-specific surface antigens. Bars, 100 μm. (F) Expression of neighbor genes <10 Mbp apart from ESRG gene. Values are normalized by GAPDH and compared with parental primed H9 ESCs. n=3. (G) Global gene expression. Scatter plots compare the microarray data of ESRG WT and KO primed PSCs. The colored plots indicate differentially expressed genes (DEGs) with statistical significance (FC>2.0, FDR, 0.05). Number of DEGs are shown in the figure. n=3. (H) Plating efficiency. Shown are the number of AP (+) colonies raised from 100 or 200 ESRG WT and KO PSCs. n=3. Numerical values for B, C, F, and H are available in S1 Data.

**Figure 2 | No impact of ESRG on naïve pluripotency**

(A) The ESRG expression. Shown are relative expressions of ESRG in primed PSCs, naïve PSCs, NPCs and HDFs. Values are normalized by GAPDH and compared with the primed 585A1 iPSC line. *P<0.05 vs. primed PSCs by unpaired t-test. n=3. (B) Conversion to naïve pluripotency. Shown are representative images of ESRG WT and KO primed and naïve PSCs under phase contrast and of immunocytochemistry for KLF17 (red) and OCT3/4 (green). Bars, 200 μm. (C) The expression of primed and naïve PSC markers. Shown are the relative expressions of common PSC markers (POU5F1 and NANOG), a primed PSC marker (ZIC2) and naïve PSC markers (KLF4 and KLF17). Values are normalized by GAPDH and compared with primed H9 ESCs. n=3. (D) Global transcriptome. Scatter plots comparing the microarray data of ESRG WT and KO naïve PSCs. The colored plot indicates DEG with statistical significance (FC>2.0, FDR,0.05). Number of DEGs are shown in the figure. n=3. (E) Differentiation to primed pluripotency. Representative images of ESRG WT and KO naïve PSCs before and after conversion to the primed pluripotent state are shown. Bars, 200 μm. (F) The expression of primed and naïve PSC markers. Shown are the relative expressions of the marker genes in (C) in ESRG WT and KO naïve PSCs before and after the differentiation to the primed pluripotent state. Values are normalized by GAPDH and compared with primed H9 ESCs. n=3. Numerical values for A, C, and F are available in S1 Data.

**Figure 3 | ESRG-deficient PSCs are capable of differentiating.**
(A) Differentiation by EB formation. Bars, 500 μm. (B) Trilineage differentiation. Bars, 200 μm. (C) The expression of differentiation markers. Shown are the relative expressions of PSC markers (POU5F1 and NANOG) and differentiation markers (DCN, MSX1 and MAP2) on days 8 and 16 of EB differentiation. Values are normalized by GAPDH and compared with primed H9 ESCs. n=3. (D) Global gene expression of differentiation derivatives. Scatter plots compare the microarray data of ESRG WT and KO PSC-derived EBs on days 8 and 16. Number of DEGs (FC>2.0, FDR,0.05) are shown in the figure. n=3. (E) NPC differentiation. Representative images of ESRG WT and KO PSCs and NPCs under phase contrast and of immunocytochemistry for PAX6 (red) and OCT3/4 (green) are shown. Bars, 200 μm. (F) The expression of NSC markers. Shown are the relative expressions of PSC markers (POU5F1 and NANOG) and NPC markers (PAX6, SOX1 and NES) in ESRG WT and KO PSCs and NPCs. Values are normalized by GAPDH and compared with primed H9 ESCs. n=3. Numerical values for C and F are available in S1 Data.

**Figure 4 | ESRG is dispensable for iPSC reprogramming.**
(A) The expression of ESRG during reprogramming. The heatmap shows the normalized intensities of ESRG, POU5F1 (endogenous), SOX2 (endogenous) and NANOG expression from microarray data in the time course of iPSC reprogramming (days 0-49) and established iPSCs (far right). n=3. (B) The effect of ESRG on iPSC generation. Shown are the numbers of AP (+) iPSC colonies 24 days after the transduction of OSK along with Mock (n=4), ESRG (n=4) and c-MYC (n=5). Numerical values for A and B are available in S1 Data.

## Supporting information

**S1 Fig. ESRG expression profiles.** Expression of ESRG in human tissues. (A) Shown are the normalized intensities of ESRG expression from the microarray data of PSCs (H9 ESC), 24 adult tissues and five fetal tissues. (B) Expression of ESRG in human cell lines. The normalized intensities of ESRG expression from the microarray data of several PSC lines including H9 ESC, 201B7 iPSC, 585A1 iPSC, 2102Ep embryonic carcinoma cells (ECC) and NTERA-2 ECC, cancer cell lines such as MCF7, HepG2, HeLa and Jurkat, and normal tissue-derived cells such as adipose tissue-derived mesenchymal stem cells (AdMSC), dental pulp-derived MSCs (DpMSC), human dermal fibroblasts (HDF), peripheral blood mononuclear cells (PBMC), bronchial epithelial cells (BrEC), prostate epithelial cells (PrEC), hepatocytes (Hep), epidermal keratinocytes (EKc), neural progenitor cells (NPC) and astrocytes (Astrocyte) are shown. Numerical values for A and B are available in S1 Data.

**S2 Fig. Deletion of ESRG locus.** (A) The scheme of ESRG targeting. The locations of sgRNAs for targeting (sgESRG-U and -L), primers for genotyping (U-S/AS and L-S/AS) and TaqMan Assays for copy number analyses (cn1 and cn2) are shown. The sequences of sgRNAs and primers are provided in the Methods section and S1Table. (B) The copy number of ESRG gene. The copy number of ESRG gene in ESRG WT (clones 1, 21 and 28), a heterozygous clone (Het) that lacks one ESRG allele and KO (clones 10, 18 and 23) were quantified by qPCR using TaqMan Copy Number Assays (cn1 and 2). Values are normalized by RNase P and compared with parental H9 ESCs. n=3. (C) The sequences around the deletion sites in ESRG KO ESC clones verified by Sanger sequencing. (D) The sequences around the sgRNA recognition sites upstream (sgESRG-U) and downstream (sgESRG-L) of the ESRG locus in ESRG WT ESC clones verified by Sanger sequencing. Numerical values for B are available in S1 Data.

**S3 Fig. Karyotypes of PSC clones used in the study.** Representative images of G-band staining show that all clones used in the study maintained normal female karyotypes (46XX).

**S1 Table. Primers used in this study.**

**S2 Table. TaqMan Assays used in this study.**

**S1 Data. In separate sheets, the excel spreadsheet contains the numerical values for Figs. 1B, 1C, 1F, 1H, 2A, 2C, 2F, 3C, 3F, 4A, and 4B; S1A, S1B and S2B Figs.**

## References

1. Santoni FA, Guerra J, Luban J. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. Retrovirology. 2012;9:111. Epub 2012/12/21. doi: 10.1186/1742-4690-9-111. PubMed PMID: 23253934; PubMed Central PMCID: PMC3558390.

2. Kelley D, Rinn J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol. 2012;13(11):R107. Epub 2012/11/28. doi: 10.1186/gb-2012-13-11-r107. PubMed PMID: 23181609; PubMed Central PMCID: PMC3580499.

3. Fuchs NV, Loewer S, Daley GQ, Izsvak Z, Lower J, Lower R. Human endogenous retrovirus K (HML-2) RNA and protein expression is a marker for human embryonic and induced pluripotent stem cells. Retrovirology. 2013;10:115. Epub 2013/10/26. doi: 10.1186/1742-4690-10-115. PubMed PMID: 24156636; PubMed Central PMCID: PMCPMC3819666.

4. Mager DL, Freeman JD. HERV-H endogenous retroviruses: presence in the New World branch but amplification in the Old World primate lineage. Virology. 1995;213(2):395-404. Epub 1995/11/10. doi: 10.1006/viro.1995.0012. PubMed PMID: 7491764.

5. Jern P, Sperber GO, Ahlsen G, Blomberg J. Sequence variability, gene structure, and expression of full-length human endogenous retrovirus H. Journal of virology. 2005;79(10):6325-37. Epub 2005/04/29. doi: 10.1128/JVI.79.10.6325-6337.2005. PubMed PMID: 15858016; PubMed Central PMCID: PMC1091717.

6. Jern P, Sperber GO, Blomberg J. Definition and variation of human endogenous retrovirus H. Virology. 2004;327(1):93-110. Epub 2004/08/26. doi: 10.1016/j.virol.2004.06.023. PubMed PMID: 15327901.

7. Goke J, Lu X, Chan YS, Ng HH, Ly LH, Sachs F, et al. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. Cell stem cell. 2015;16(2):135-41. Epub 2015/02/07. doi: 10.1016/j.stem.2015.01.005. PubMed PMID: 25658370.

8. Wang J, Xie G, Singh M, Ghanbarian AT, Rasko T, Szvetnik A, et al. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature. 2014. Epub 2014/10/16. doi: 10.1038/nature13804. PubMed PMID: 25317556.

9. Lu X, Sachs F, Ramsay L, Jacques PE, Goke J, Bourque G, et al. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. Nat Struct Mol Biol. 2014;21(4):423-5. Epub 2014/04/01. doi: 10.1038/nsmb.2799. PubMed PMID: 24681886.

10. Ohnuki M, Tanabe K, Sutou K, Teramoto I, Sawamura Y, Narita M, et al. Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. Proceedings of the National Academy of Sciences of the United States of America. 2014. Epub 2014/08/07. doi: 10.1073/pnas.1413299111. PubMed PMID: 25097266.

11. Friedli M, Turelli P, Kapopoulou A, Rauwel B, Castro-Diaz N, Rowe HM, et al. Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. Genome research. 2014. Epub 2014/06/01. doi: 10.1101/gr.172809.114. PubMed PMID: 24879558.

12. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007;131(5):861-72. PubMed PMID: 18035408.

13. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;126(4):663-76. PubMed PMID: 16904174.

14. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, et al. Induced pluripotent stem cell lines derived from human somatic cells. Science (New York, NY). 2007;318(5858):1917-20. PubMed PMID: 18029452.

15. Koyanagi-Aoi M, Ohnuki M, Takahashi K, Okita K, Noma H, Sawamura Y, et al. Differentiation-defective phenotypes revealed by large-scale analyses of human pluripotent stem cells. Proceedings of the National Academy of Sciences of the United States of America. 2013. Epub 2013/11/22. doi: 10.1073/pnas.1319061110. PubMed PMID: 24259714.

16. Loewer S, Cabili MN, Guttman M, Loh YH, Thomas K, Park IH, et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nature genetics. 2010;42(12):1113-7. Epub 2010/11/09. doi: 10.1038/ng.710. PubMed PMID: 21057500; PubMed Central PMCID: PMC3040650.

17. Ng SY, Johnson R, Stanton LW. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. The EMBO journal. 2012;31(3):522-33. Epub 2011/12/24. doi: 10.1038/emboj.2011.459. PubMed PMID: 22193719; PubMed Central PMCID: PMCPMC3273385.

18. Zhao M, Ren C, Yang H, Feng X, Jiang X, Zhu B, et al. Transcriptional profiling of human embryonic stem cells and embryoid bodies identifies HESRG, a novel stem cell gene. Biochemical and biophysical research communications. 2007;362(4):916-22. Epub 2007/09/07. doi: 10.1016/j.bbrc.2007.08.081. PubMed PMID: 17803967.

19. Li G, Ren C, Shi J, Huang W, Liu H, Feng X, et al. Identification, expression and subcellular localization of ESRG. Biochemical and biophysical research communications. 2013;435(1):160-4. Epub 2013/05/01. doi: 10.1016/j.bbrc.2013.04.062. PubMed PMID: 23628413.

20. Rand TA, Sutou K, Tanabe K, Jeong D, Nomura M, Kitaoka F, et al. MYC Releases Early Reprogrammed Human Cells from Proliferation Pause via Retinoblastoma Protein Inhibition. Cell reports. 2018;23(2):361-75. Epub 2018/04/12. doi: 10.1016/j.celrep.2018.03.057. PubMed PMID: 29641997.

21. Takashima Y, Guo G, Loos R, Nichols J, Ficz G, Krueger F, et al. Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell. 2014;158(6):1254-69. Epub 2014/09/13. doi: 10.1016/j.cell.2014.08.029. PubMed PMID: 25215486; PubMed Central PMCID: PMC4162745.

22. Theunissen TW, Powell BE, Wang H, Mitalipova M, Faddah DA, Reddy J, et al. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell stem cell. 2014;15(4):471-87. Epub 2014/08/05. doi: 10.1016/j.stem.2014.07.002. PubMed PMID: 25090446; PubMed Central PMCID: PMC4184977.

23. Guo G, von Meyenn F, Rostovskaya M, Clarke J, Dietmann S, Baker D, et al. Epigenetic resetting of human pluripotency. Development (Cambridge, England). 2017;144(15):2748-63. Epub 2017/08/03. doi: 10.1242/dev.146811. PubMed PMID: 28765214; PubMed Central PMCID: PMCPMC5560041.

24. Di Stefano B, Ueda M, Sabri S, Brumbaugh J, Huebner AJ, Sahakyan A, et al. Reduced MEK inhibition preserves genomic stability in naive human embryonic stem cells. Nature methods. 2018;15(9):732-40. Epub 2018/08/22. doi: 10.1038/s41592-018-0104-1. PubMed PMID: 30127506; PubMed Central PMCID: PMCPMC6127858.

25. Collier AJ, Panula SP, Schell JP, Chovanec P, Plaza Reyes A, Petropoulos S, et al. Comprehensive Cell Surface Protein Profiling Identifies Specific Markers of Human Naive and Primed Pluripotent States. Cell stem cell. 2017;20(6):874-90.e7. Epub 2017/03/28. doi: 10.1016/j.stem.2017.02.014. PubMed PMID: 28343983; PubMed Central PMCID: PMCPMC5459756.

26. Takahashi K, Jeong D, Wang S, Narita M, Jin X, Iwasaki M, et al. Critical Roles of Translation Initiation and RNA Uridylation in Endogenous Retroviral Expression and Neural Differentiation in Pluripotent Stem Cells. Cell reports. 2020;31(9):107715. Epub 2020/06/04. doi: 10.1016/j.celrep.2020.107715. PubMed PMID: 32492424.

27. Chambers SM, Fasano1 CA, Papapetrou EP, Tomishima M, Sadelain M, Studer L. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nature biotechnology. 2009;27(3):275-80.

28. Doi D, Samata B, Katsukawa M, Kikuchi T, Morizane A, Ono Y, et al. Isolation of human induced pluripotent stem cell-derived dopaminergic progenitors by cell sorting for successful transplantation. Stem cell reports. 2014;2(3):337-50. Epub 2014/03/29. doi: 10.1016/j.stemcr.2014.01.013. PubMed PMID: 24672756; PubMed Central PMCID: PMC3964289.

29. Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, et al. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nature biotechnology. 2008;26(1):101-6. PubMed PMID: 18059259.

30. Wernig M, Meissner A, Cassady JP, Jaenisch R. c-Myc is dispensable for direct reprogramming of mouse fibroblasts. Cell stem cell. 2008;2(1):10-2. Epub 2008/03/29. doi: S1934-5909(07)00315-3 [pii] 10.1016/j.stem.2007.12.001. PubMed PMID: 18371415.

31. Smith KN, Starmer J, Miller SC, Sethupathy P, Magnuson T. Long Noncoding RNA Moderates MicroRNA Activity to Maintain Self-Renewal in Embryonic Stem Cells. Stem cell reports. 2017;9(1):108-21. Epub 2017/06/06. doi: 10.1016/j.stemcr.2017.05.005. PubMed PMID: 28579393; PubMed Central PMCID: PMCPMC5511051.

32. Hunkler HJ, Hoepfner J, Huang CK, Chatterjee S, Jara-Avaca M, Gruh I, et al. The Long Non-coding RNA Cyrano Is Dispensable for Pluripotency of Murine and Human Pluripotent Stem Cells. Stem cell reports. 2020;15(1):13-21. Epub 2020/06/13. doi: 10.1016/j.stemcr.2020.05.011. PubMed PMID: 32531193; PubMed Central PMCID: PMCPMC7363876.

33. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, et al. CRISPR-mediated modular RNA-guided

regulation of transcription in eukaryotes. Cell. 2013;154(2):442-51. Epub 2013/07/16. doi: 10.1016/j.cell.2013.06.044. PubMed PMID: 23849981; PubMed Central PMCID: PMCPMC3770145.

34. Mandegar MA, Huebsch N, Frolov EB, Shin E, Truong A, Olvera MP, et al. CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs. Cell stem cell. 2016;18(4):541-53. Epub 2016/03/15. doi: 10.1016/j.stem.2016.01.022. PubMed PMID: 26971820; PubMed Central PMCID: PMCPMC4830697.

35. Lennox KA, Behlke MA. Cellular localization of long non-coding RNAs affects silencing by RNAi more than by antisense oligonucleotides. Nucleic acids research. 2016;44(2):863-77. Epub 2015/11/19. doi: 10.1093/nar/gkv1206. PubMed PMID: 26578588; PubMed Central PMCID: PMCPMC4737147.

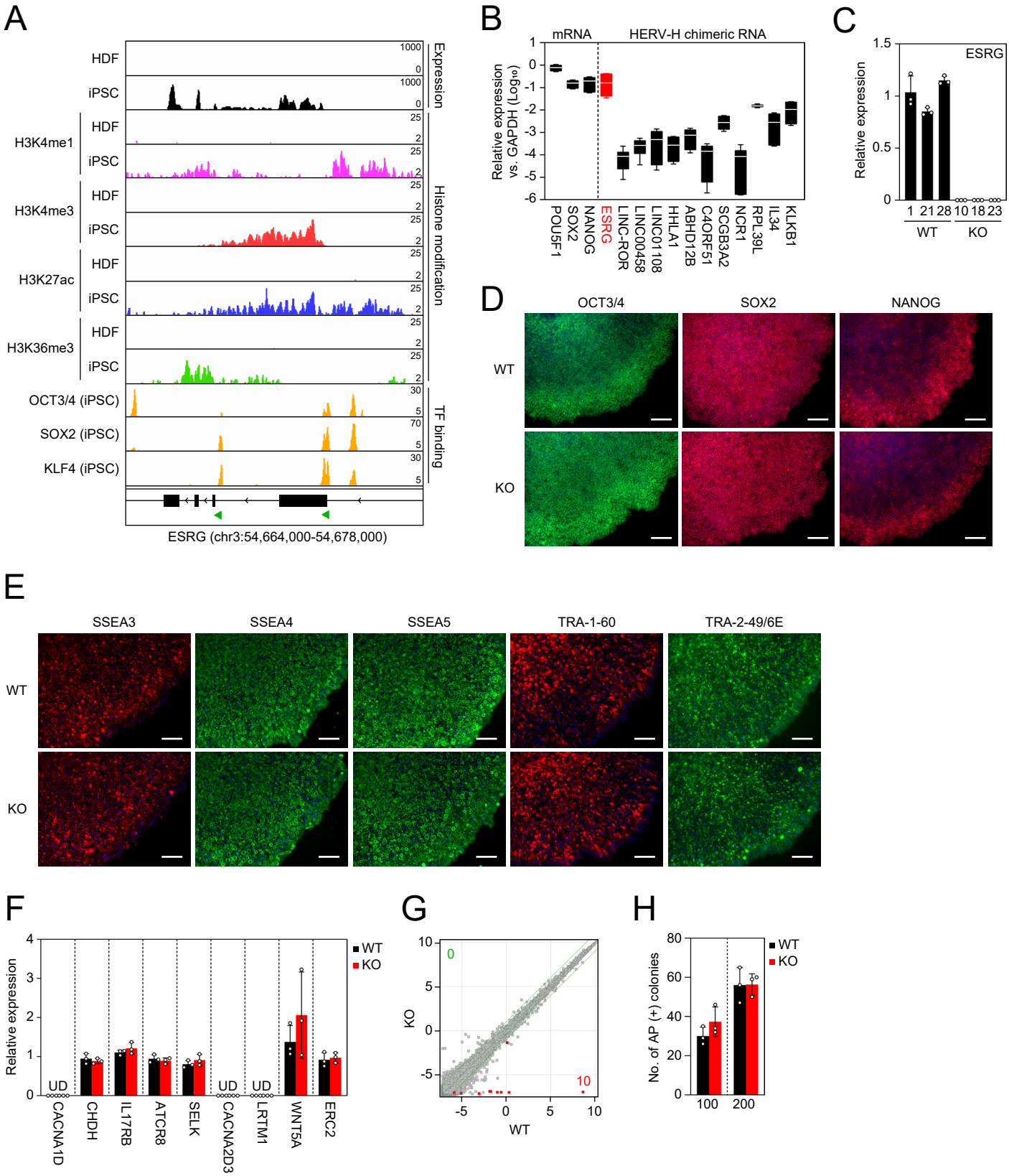
36. Liu SJ, Lim DA. Modulating the expression of long non-coding RNAs for functional studies. EMBO reports. 2018;19(12). Epub 2018/11/24. doi: 10.15252/embr.201846955. PubMed PMID: 30467236; PubMed Central PMCID: PMCPMC6280642.

37. Izsvák Z, Wang J, Singh M, Mager DL, Hurst LD. Pluripotency and the endogenous retrovirus HERVH: Conflict or serendipity? BioEssays : news and reviews in molecular, cellular and developmental biology. 2016;38(1):109-17. Epub 2016/01/07. doi: 10.1002/bies.201500096. PubMed PMID: 26735931.

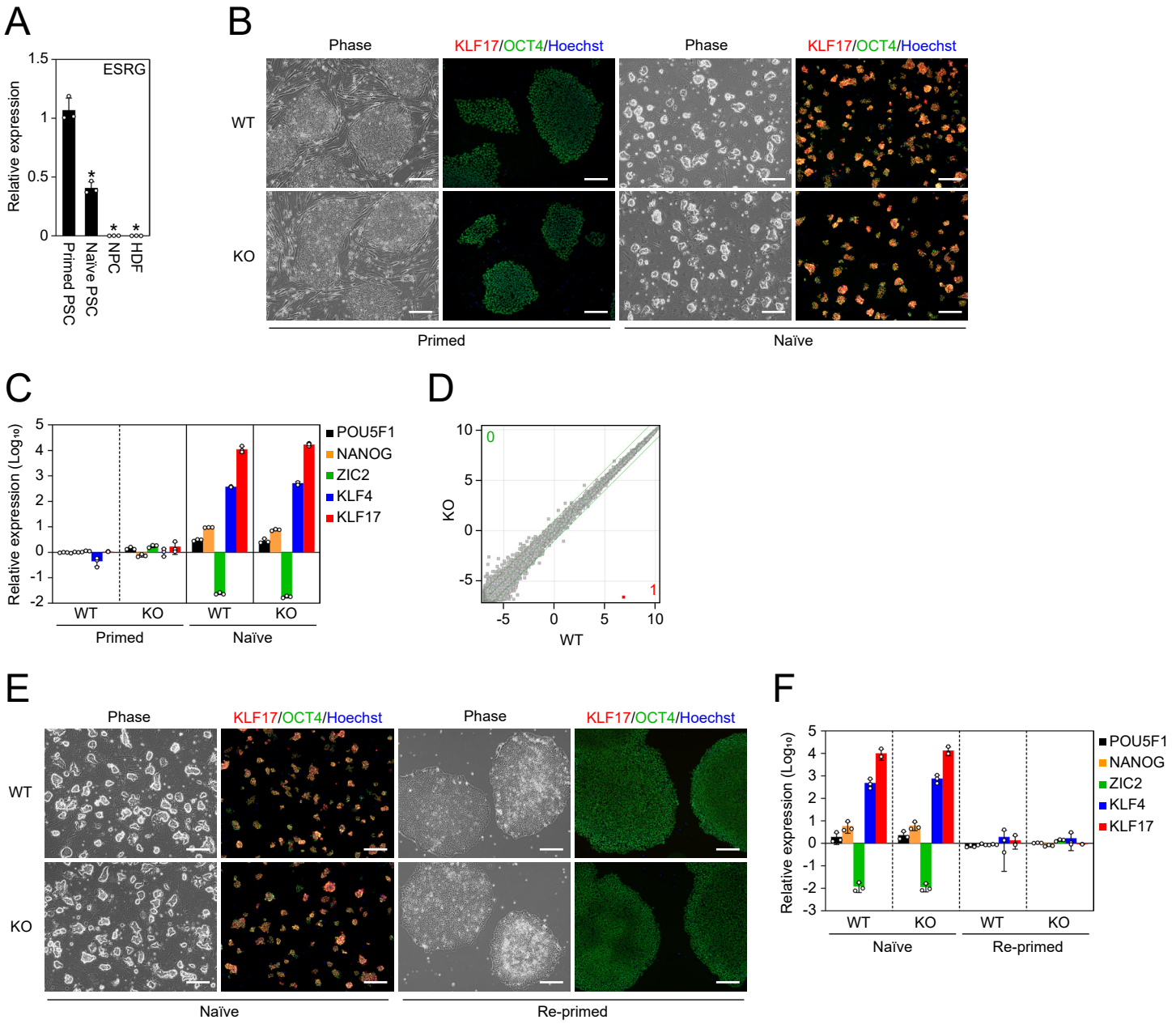
38. Sekine K, Tsuzuki S, Yasui R, Kobayashi T, Ikeda K, Hamada Y, et al. Robust detection of undifferentiated iPSC among differentiated cells. Scientific reports. 2020;10(1):10293. Epub 2020/06/26. doi: 10.1038/s41598-020-66845-6. PubMed PMID: 32581272; PubMed Central PMCID: PMCPMC7314783 conflict of interest.

39. Morita S, Kojima T, Kitamura T. Plat-E: an efficient and stable system for transient packaging of retroviruses. Gene Ther. 2000;7(12):1063-6. PubMed PMID: 10871756.

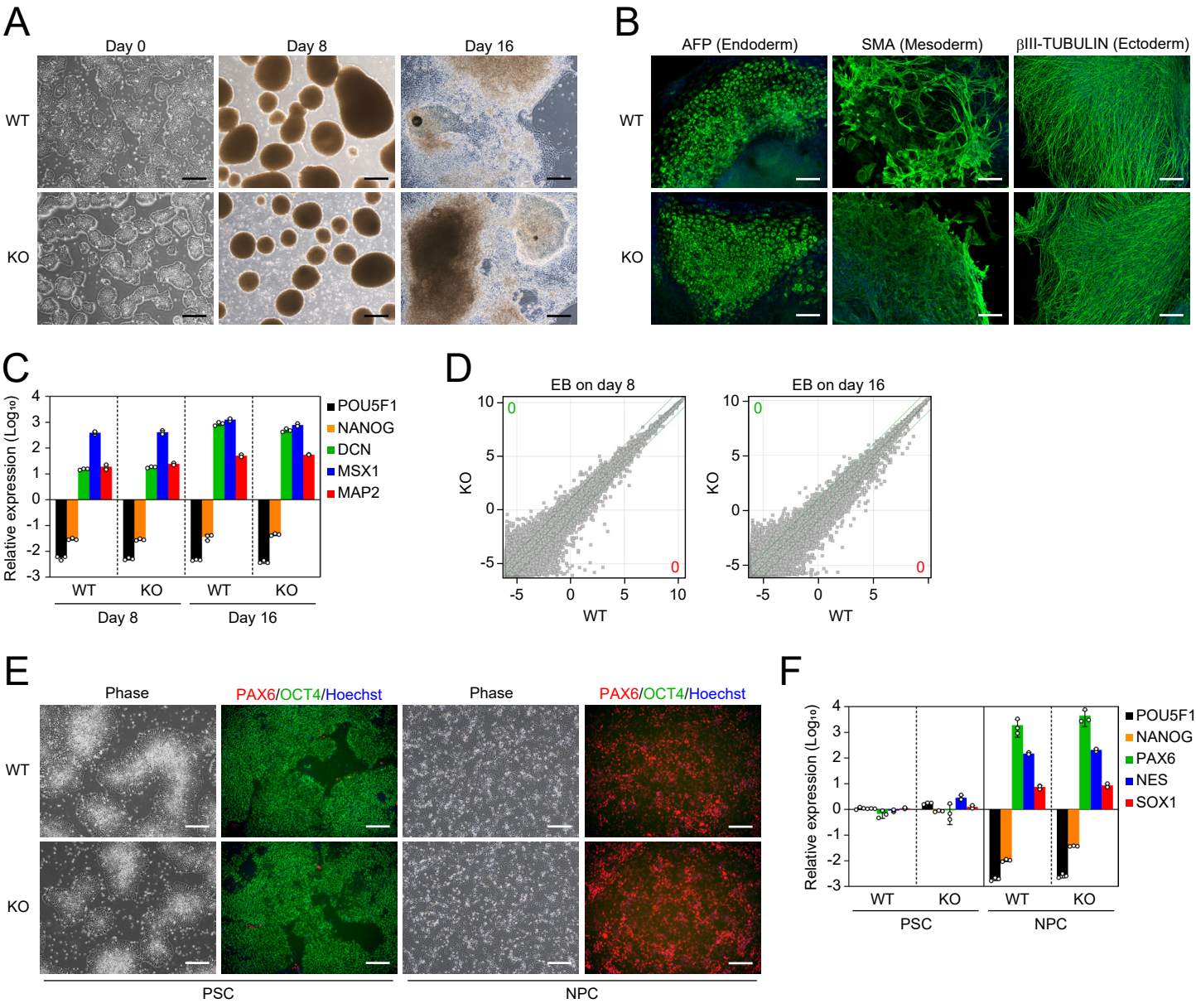
# Fig. 1



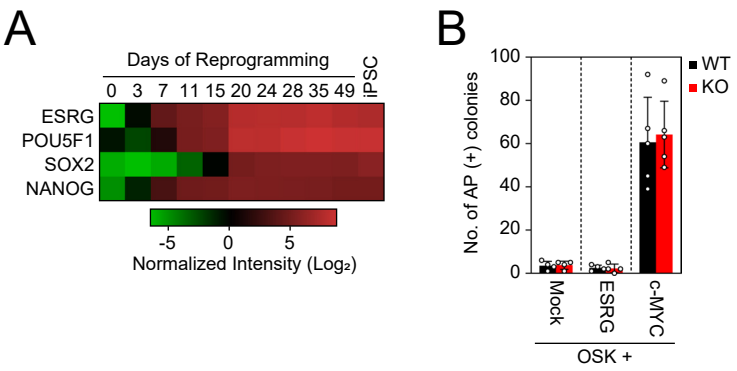
# Fig. 2



# Fig. 3

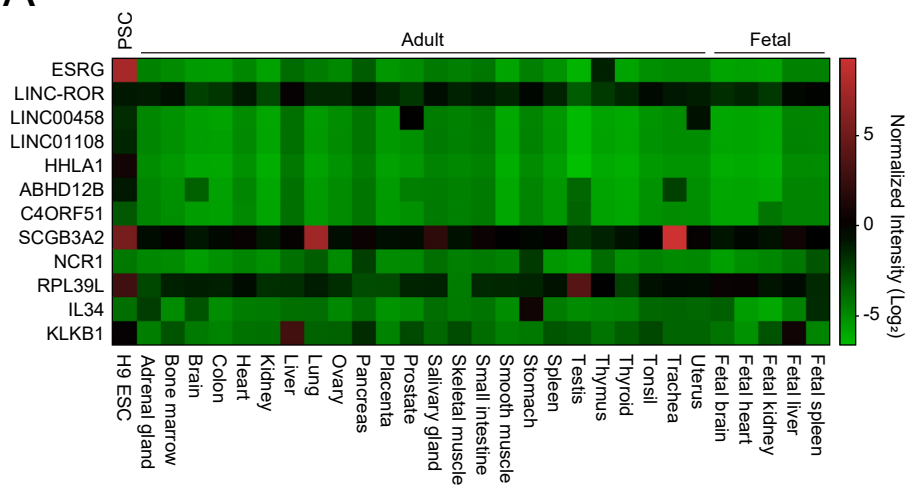


# Fig. 4

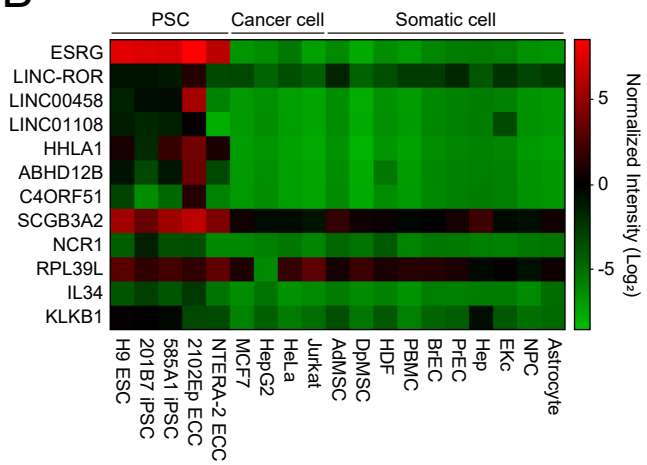


# S1 Fig.

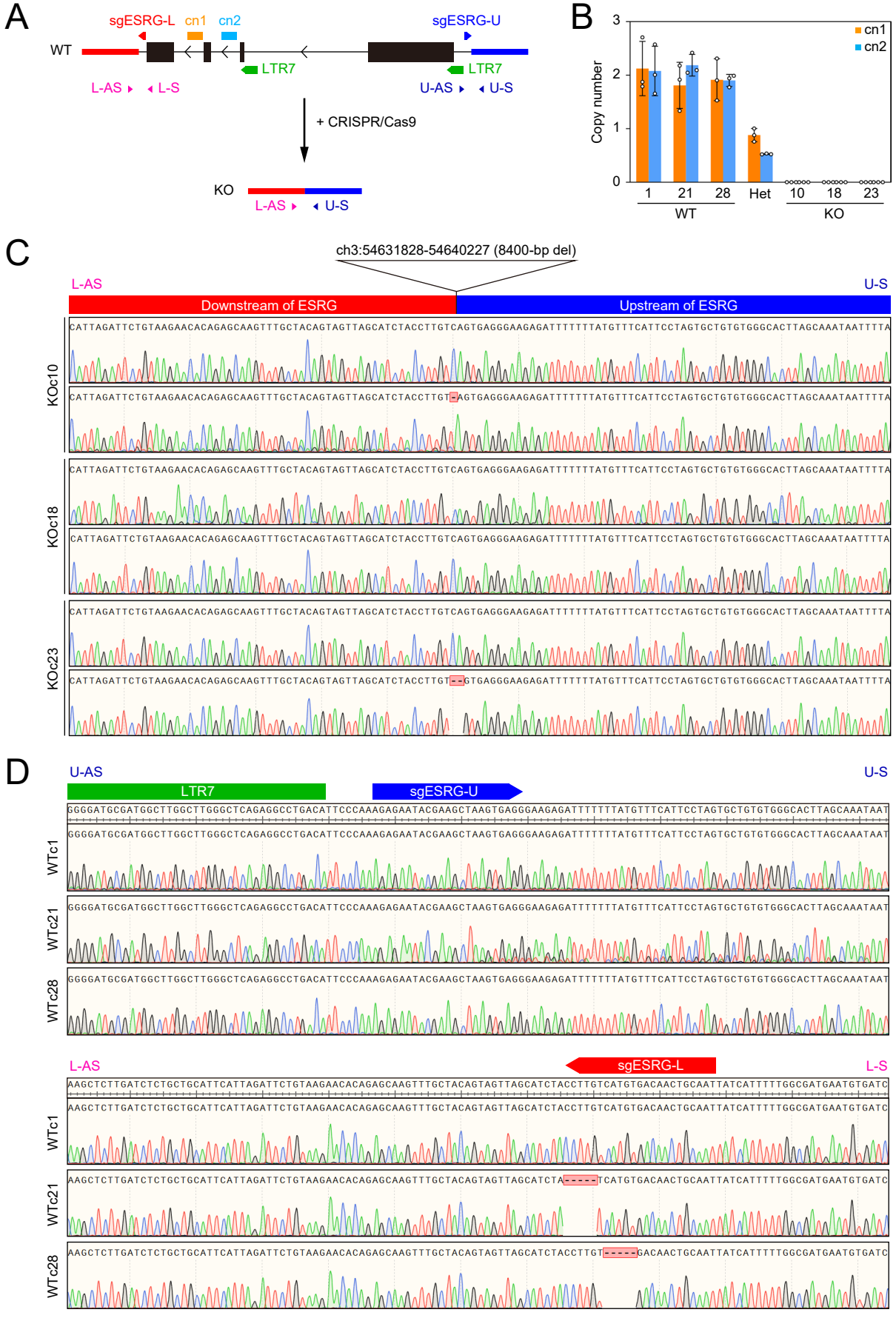
A



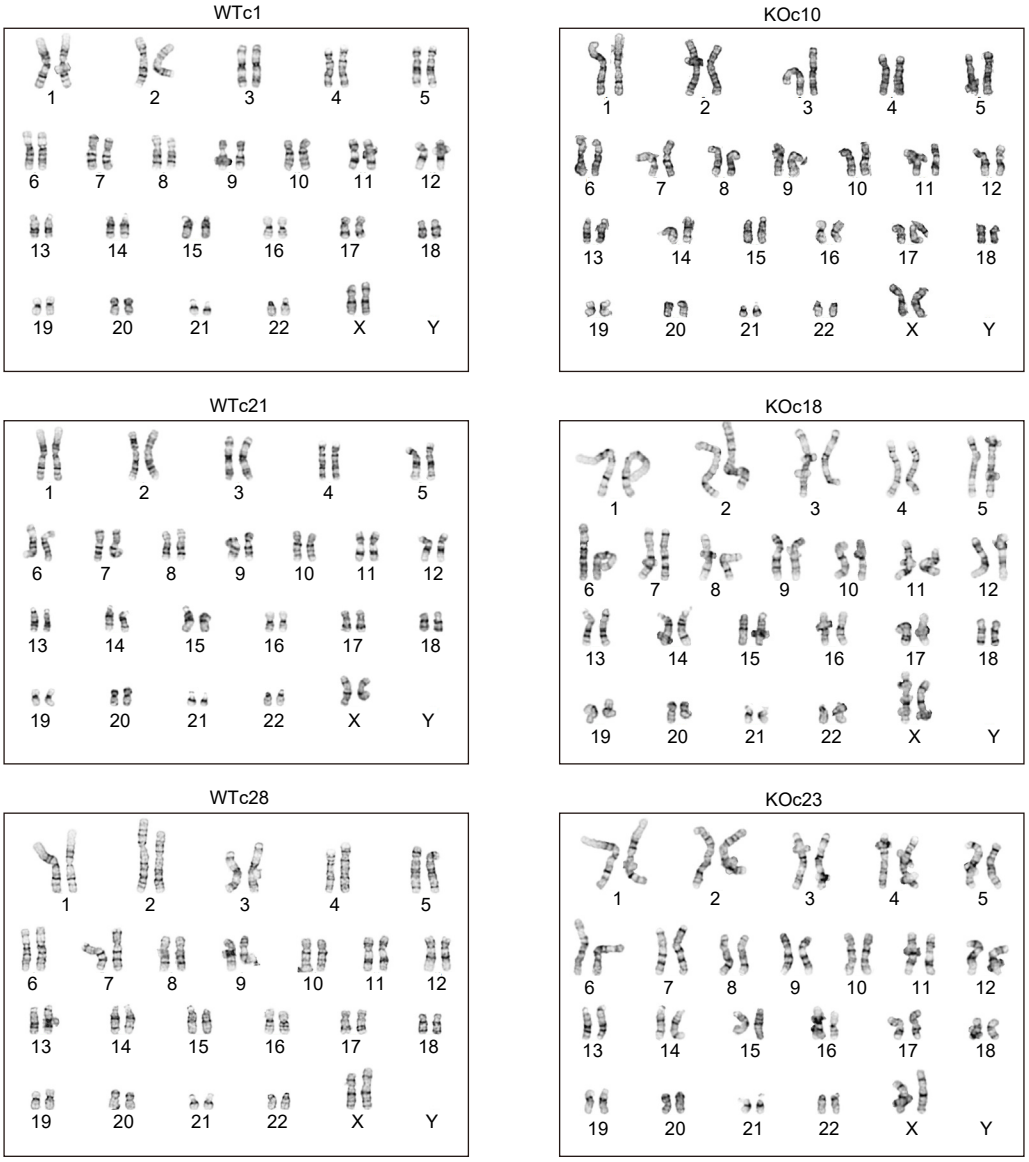

B



# S2 Fig.



# S3 Fig.



## S1 Table. Primers used in this study

| Name | Sequence (5' to 3') |
|---|---|
| ESRG-U-S | ACAATTACATGTCAATGCTGCGAG |
| ESRG-U-AS | GTGGAATGTCATCAGTTAAGGTGG |
| ESRG-L-S | CACTGCAAAGATCACATTCATCGC |
| ESRG-L-AS | GGCTGGAGCTTAGAATGGTTCAG |
| ESRG-BamH1-S | AGTTAATTAAGGATCCGCTGACTCTCTTTTCGGACTCAGC |
| ESRG-Not1-AS | ACTGTGCTGGCGGCCGCACTCATCAAACCATTTGAATTTAATTGC |

## S2 Table. TaqMan Assays used in this study

| Target | Assay ID |
|---|---|
| ESRG | Hs03666618_s1 |
| LINC-ROR | Hs04332550_m1 |
| LINC00458 | Hs05005988_m1 |
| LINC01108 | Hs04402672_m1 |
| HHLA1 | Hs00903176_g1 |
| ABHD12B | Hs00997975_g1 |
| C4ORF51 | Hs03037752_m1 |
| SCGB3A2 | Hs00369678_m1 |
| NCR1 | Hs00950813_m1 |
| RPL39L | Hs01027925_m1 |
| IL34 | Hs01050928_g1 |
| KLKB1 | Hs00168478_m1 |
| POU5F1 | Hs04260367_gH |
| SOX2 | Hs01053409_s1 |
| NANOG | Hs02387400_g1 |
| ZIC2 | Hs00600845_m1 |
| KLF4 | Hs00358836_m1 |
| KLF17 | Hs00702999_m1 |
| PAX6 | Hs00240871_m1 |
| NES | Hs04187831_g1 |
| SOX1 | Hs01057642_s1 |
| GAPDH | Hs02786624_g1 |