Stem Cell Reports
Article

ISSCR

OPEN ACCESS

Overexpression of Nuclear Receptor 5A1 Induces and Maintains an Intermediate State of Conversion between Primed and Naive Pluripotency

Kaori Yamauchi,[1] Tatsuhiko Ikeda,[2] Mihoko Hosokawa,[3] Norio Nakatsuji,[2,3] Eihachiro Kawase,[1] Shinichiro Chuma,[3] Kouichi Hasegawa,[2,4] and Hirofumi Suemori[1,*]

[1]Laboratory of Embryonic Stem Cell Research, Department of Regeneration Science and Engineering, Institute for Frontier Life and Medical Sciences, Kyoto University, Kyoto 606-8507, Japan
[2]Institute for Integrated Cell-Material Sciences, Kyoto University, Kyoto 606-8351, Japan
[3]Laboratory of Developmental Epigenome, Department of Regeneration Science and Engineering, Institute for Frontier Life and Medical Sciences, Kyoto University, Kyoto 606-8507, Japan
[4]Institute for Stem Cell Biology and Regenerative Medicine, NCBS Campus, GKVK, Bangalore 560065, India
*Correspondence: hsuemori@infront.kyoto-u.ac.jp
https://doi.org/10.1016/j.stemcr.2020.01.012

SUMMARY

Naive and primed human pluripotent stem cells (hPSCs) have provided useful insights into the regulation of pluripotency. However, the molecular mechanisms regulating naive conversion remain elusive. Here, we report intermediate naive conversion induced by overexpressing nuclear receptor 5A1 (NR5A1) in hPSCs. The cells displayed some naive features, such as clonogenicity, glycogen synthase kinase 3β, and mitogen-activated protein kinase (MAPK) independence, expression of naive-associated genes, and two activated X chromosomes, but lacked others, such as *KLF17* expression, transforming growth factor β independence, and imprinted gene demethylation. Notably, NR5A1 negated MAPK activation by fibroblast growth factor 2, leading to cell-autonomous self-renewal independent of MAPK inhibition. These phenotypes may be associated with naive conversion, and were regulated by a DPPA2/4-dependent pathway that activates the selective expression of naive-associated genes. This study increases our understanding of the mechanisms regulating the conversion from primed to naive pluripotency.

INTRODUCTION

Human embryonic stem cells (hESCs) are derived from the inner cell mass (ICM) of blastocysts at the preimplantation stage (Thomson et al., 1998). Initially, hESCs were thought to be similar to ICM cells, such as mouse ESCs (mESCs), which are derived from the ICMs of mouse blastocysts (Evans and Kaufman, 1981). However, significant differences have been observed between hESCs and mESCs; for example, maintenance of the hESC undifferentiated state depends on fibroblast growth factor 2 (FGF2) and the transforming growth factor β (TGF-β) family (Vallier et al., 2005) but does not require leukemia inhibitory factor (LIF). In mESCs, dual inhibition of the glycogen synthase kinase 3 (GSK3) and the mitogen-activated protein kinase (MAPK)-extracellular signal-regulated kinase (ERK) pathways (termed 2i) (Ying et al., 2008) together with LIF induces mESCs into a state more closely resembling ICM cells (Boroviak et al., 2014; Huang et al., 2014). These mESCs are considered to be in a "ground" or "naive" state. In parallel with these studies, a pluripotent state of epiblast stem cells (mEpiSCs) established from post-implantation mouse embryos (Brons et al., 2007; Tesar et al., 2007) is designated as a "primed" state, which shares many common features with hESCs. Importantly, they can be converted into the naive state by reprogramming factors, including Oct3/4 (also known as Pou5f1), Nanog, Prdm14, Klf2, Klf4, Esrrb, Nr5a1, and Nr5a2, under the 2i/LIF condition (Festuccia et al., 2012; Gillich et al., 2012; Guo et al., 2009; Hall et al., 2009; Hanna et al., 2009; Silva et al., 2009). Efforts to induce a naive state of human pluripotency identified OCT3/4, KLF2, and KLF4 as important inducing factors (Hanna et al., 2010). It was also shown that simultaneous expression of NANOG and KLF2 can generate two kinds of naive cells, termed 5i/L/A and reset cells, that are considered closer to *in vivo* peri-implantation pluripotent cells in some aspects than other established naive-like cells (Chan et al., 2013; Chen et al., 2015; Gafni et al., 2013; Liu et al., 2017; Qin et al., 2016; Takashima et al., 2014; Theunissen et al., 2014; Ware et al., 2014). Comparison analyses of the cell lines unveiled significant differences between primed and naive cells in terms of cellular responses to 2i conditions, metabolism, transcriptional and epigenetic profiles, and X chromosome status. Such findings provide criteria for the definitions of human naive pluripotency, but the molecular mechanisms regulating induction of the naive state from the primed state remain unclear. In this study, we examine the effects of various transcription factors on naive state induction in hPSCs.

RESULTS

NR5A1 Overexpression Induces Naive-Associated Gene Expression in Primed hPSCs

To understand the molecular mechanisms controlling naive conversion in hPSCs, we examined whether

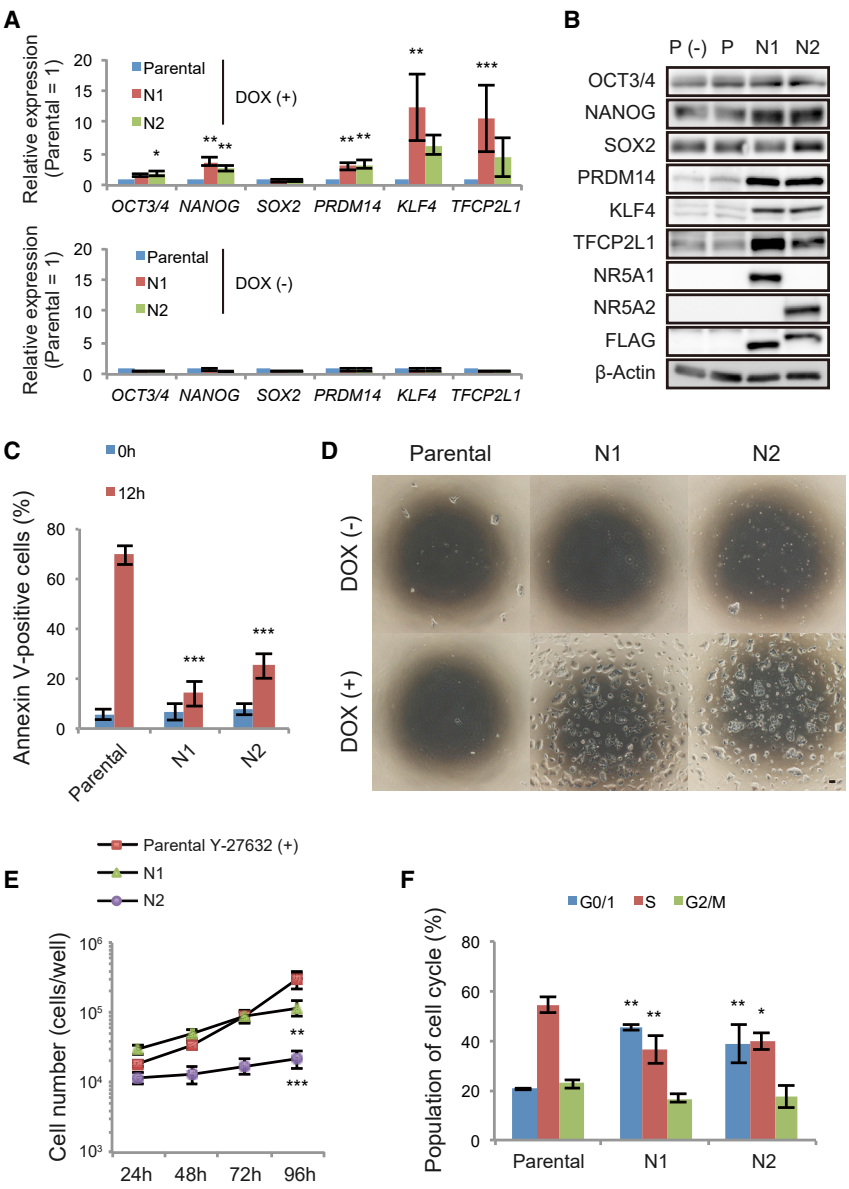


Figure 1. Overexpression of NR5A Receptors Induces Naive-Associated Gene Expression in hPSCs

(A and B) Expression of genes associated with the naive pluripotent state. Data from parental (P), NR5A1-expressing (N1), and NR5A2-expressing (N2) cells were analyzed by the comparative ΔCt method. DOX-treated and untreated (A, top and bottom, respectively) samples are shown. In (A), expression in parental cells under each condition was set to 1. β-Actin was used as an internal control. P (−) represents parental cells cultured without DOX.

(C) Percentages of annexin V-positive cells before (0 h) and 12 h after single-cell dissociation. ***$p < 0.001$ versus parental cells at 0 or 12 h (Dunnett's test).

(D) Representative images of parental cells and NR5A transfectants in the presence or absence of DOX at day 3 after single-cell dissociation. Scale bar, 100 μm.

(E) Growth rates of parental cells and NR5A transfectants. Cells were counted every 24 h. Parental cells were treated with Y-27632 for the first 24 h of culture.

(F) Cell proliferation analysis by Click-iT 5-ethynyl-2′-deoxyuridine (EdU) assay. After EdU treatment for 1 h, the cells were co-stained with FxCycle Violet and analyzed by flow cytometry.

Data are represented as mean ± SD (A and E) or SEM (C and F) of three biological replicates. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ versus parental cells (Dunnett's test). See also Figure S1.

exogenous expression of various transcription factors could promote expression of the naive-associated genes *NANOG* and *KLF4* (Figures S1A and S1B). We selected ten genes that are either known inducers of mouse naive pluripotency or play important roles in mouse pluripotency (Festuccia et al., 2012; Gillich et al., 2012; Guo et al., 2009; Hall et al., 2009; Hanna et al., 2009; Martello et al., 2012; Niwa et al., 2009; Silva et al., 2009). A lentiviral system was used to transfect each gene into the hESC line H9, which was cultured under feeder-free conditions with mouse embryonic fibroblast (MEF)-conditioned medium to maintain a primed state (Figure S1A). The MEF-conditioned medium was used for initial screening because we routinely cultivate hESCs on MEFs with KnockOut Serum Replacement-based medium whose components are similar to those of the MEF-conditioned medium. Real-time PCR analysis revealed that expression of both *NANOG* and *KLF4* was upregulated by NR5A1 overexpression (Figure S1B). This result was unexpected as NR5A1 overexpression in mEpiSCs does not induce expression of naive-associated genes under primed conditions (Guo and Smith, 2010), prompting us to further investigate the ability of NR5A1 as an inducer of human naive pluripotency.

To facilitate our analyses, we established a doxycycline (DOX)-inducible expression system for NR5A1 (Figures S1C and S1D). An NR5A2 expression system was generated in parallel because of its high sequence similarity with NR5A1. DOX activated transgene expression in

transfectants carrying NR5A1 or NR5A2 (referred to as N1 and N2 cells unless specified otherwise) cultured in mTeSR1 (referred to as the TGF-β and FGF2 [TF] condition for primed cells; Figures S1E and S1F). Because the MEF-conditioned medium used for initial screening contains undefined factors from feeder cells, which hampers detailed molecular analysis, we used the defined medium mTeSR1 instead. Naive-associated genes, such as *PRDM14* and *TFCP2L1* (Dunn et al., 2014; Takashima et al., 2014; Theunissen et al., 2014) as well as *NANOG* and *KLF4* were upregulated at the mRNA and protein levels in the transfectants compared with parental cells cultured with DOX (referred to as parental cells unless specified otherwise; Figures 1A and 1B). We next investigated whether NR5A transfectants can survive after single-cell dissociation, because, unlike cells in the primed state, naive-state hESCs are resistant to cell death caused by dissociation. Flow cytometric analysis revealed a dramatic reduction in the proportion of annexin V-positive cells (indicating dying or dead cells) in NR5A transfectants compared with parental cells 12 h after dissociation (Figures 1C and S1G). Both N1 and N2 cells were maintained by single-cell passaging for over ten passages with sustained expression of OCT3/4, NANOG, and SOX2 (Figures 1D and S1H). However, the growth rates of both transfectants were decreased compared with parental cells (Figure 1E). Cell-cycle analysis revealed an increased G0/1 phase population and a decreased S phase population (Figure 1F). These results suggest that NR5A-induced cells lost a subset of the features of primed pluripotency. The changes in gene expression and tolerance to single-cell passaging were no longer apparent by day 10 after DOX withdrawal, indicating their dependence on NR5A expression (Figures S1I–S1K).

NR5A1-Induced Cells Are Stably Maintained in Culture with GSK3 Inhibition

NR5A-induced cells displayed a cell proliferation defect in the primed (TF) condition, suggesting that the naive (2i/LIF) condition is more suitable for cell growth while sustaining pluripotency. We investigated whether N1 and N2 cells can be maintained in the presence of the mitogen-activated protein kinase (MEK)-ERK inhibitor PD0325901 (PD03), the GSK3 inhibitor CHIR99021 (CHIR), both inhibitors (2i), or 2i/LIF in custom mTeSR1 lacking TGF-β and FGF2, which are not required for maintaining cells in the naive state. Under 2i and 2i/LIF conditions, parental cells differentiated soon after culturing, whereas all transfectants were stably maintained over ten passages by single-cell passaging with sustained expression of OCT3/4, NANOG, and SOX2 (Figures 2A, 2B, S2A, and S2B). Interestingly, NR5A transfectants also maintained their pluripotency with PD03 or CHIR alone, with sustained expression of *NANOG* and *KLF4* (Figure 2C). However, when we examined naive pluripotency marker expression, including *DPPA3* and *ZFP42* (Takashima et al., 2014; Theunissen et al., 2014), N1 cells treated with CHIR alone (N1-CHIR cells) had significantly increased expression of *DPPA3*, but not *ZFP42* (Figure 2D). N1-CHIR cells were maintained as packed colonies, similar to the 2i and 2i/LIF conditions (Figure 2A), and expressed NANOG and KLF4 (Figure 2E). Both N1- and N2-CHIR cells formed teratomas containing the three germ layers of ectoderm, mesoderm, and endoderm when transplanted into immunodeficient mice (Figure S2C). Furthermore, the CHIR condition increased cell proliferation and enhanced entry into S phase of the cell cycle (Figures 2F and 2G). In contrast, PD03 had little or no effect, and LIF also had little effect. These results suggested that CHIR treatment alone was sufficient to maintain pluripotency and facilitate the proliferation of NR5A-induced cells.

In line with the result under the TF condition (Figures S1I–S1K), we confirmed that the pluripotency of N1-CHIR cells was dependent on NR5A1 expression, because DOX withdrawal induced differentiation even under the CHIR condition (Figure S2D). To ensure that the cellular response to CHIR was independent of the hPSC lines used, we established N1- or N2-CHIR cells derived from the hESC line KhES-1 and the human induced pluripotent stem cell (hiPSC) line 253G1, in which we constitutively overexpressed the NR5A receptors (Figure S2E). Transgene expression significantly reduced the proportions of annexin V-positive cells (Figure S2F), and the transfectants maintained their pluripotency for more than ten passages, except for N2-CHIR cells derived from KhES1 cells, which displayed >40% differentiation (Figure S2G). This result may reflect the lower NANOG, KLF4, and TFCP2L1 expression in N2-CHIR cells compared with N1-CHIR cells (Figures S2H and S2I), suggesting that NR5A2 is a weaker inducer of naive pluripotency than NR5A1.

In primed cells, suppression of MEK-ERK signaling by PD03 decreased *OCT3/4* and *NANOG* expression soon after treatment (Figure S2J). In naive cells, reactivation of MEK-ERK signaling by PD03 withdrawal induces downregulation of *NANOG* and *KLF4* (Theunissen et al., 2014). Conversely, our results revealed that the presence or absence of PD03 had little impact on pluripotency in NR5A-induced cells. To understand this, we investigated whether PD03 suppressed ERK1/2 phosphorylation by MEK (Figure 2H). Phosphorylated ERK1/2 was not detected in PD03-treated control or N1 cells, indicating that PD03 acted as expected. PD03 treatment induced phosphorylation of MEK1/2 in both parental and N1 cells, probably because the suppression of ERK phosphorylation by PD03 reduced negative feedback regulation, as reported previously (Sturm et al., 2010). Interestingly, activation of MEK1/2 and ERK1/2 by FGF2, a known activator of

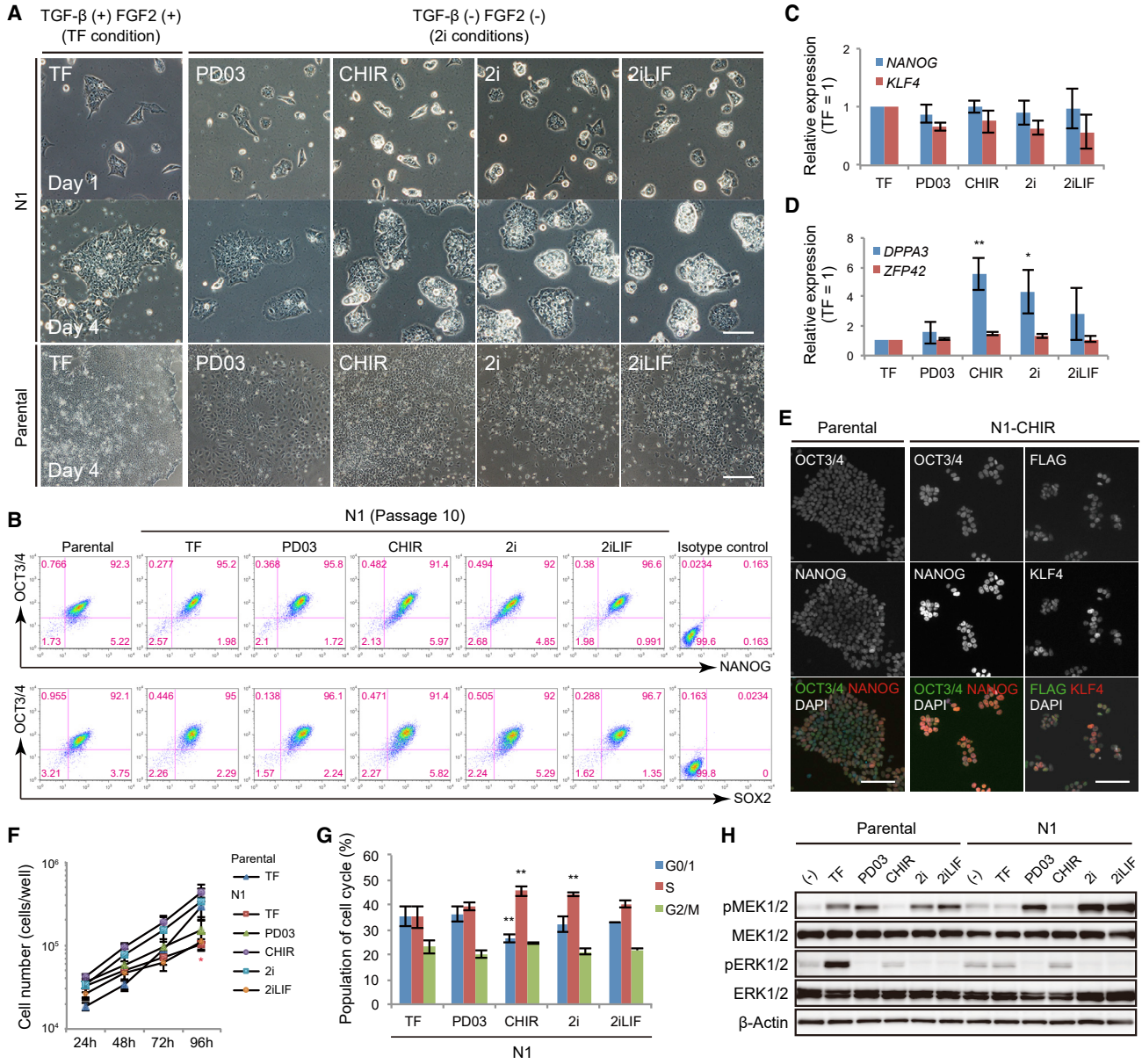


Figure 2. NR5A1-Induced Cells Are Stably Maintained in Culture with a GSK3 Inhibitor

(A) Representative images of parental at day 4 and N1 cells at passage 10 after culturing with combined GSK3 and MEK-ERK inhibition.

(B) Flow cytometric analysis of pluripotency markers in parental and N1 cells at passage 10 cultured under each condition.

(C and D) Expression of the indicated genes in N1 cells. RNA was extracted from cells cultured for >10 passages under each condition. Data were analyzed by the comparative ΔCt method. The TF condition was set to 1.

(E) Immunocytochemical staining of pluripotency markers and FLAG in parental and CHIR-treated N1 (N1-CHIR) cells at passage 6.

(F) Growth rates of N1 cells. Cells after switching from TF to each condition were counted every 24 h. Parental cell data from Figure 1E are shown here for comparison.

(G) Cell proliferation analysis of N1 cells cultured for >3 passages under inhibitor conditions by EdU assay.

(H) Phosphorylated MEK1/2 and ERK1/2 in parental and N1 cells. The cells were transferred from the TF condition (mTeSR1) to custom mTeSR1 lacking TGF-β and FGF2. After 23 h of treatment, inhibitors were added to each culture for 1 h. Under TF conditions, custom mTeSR1 was changed to the TF condition. Under (−) conditions, the cells were treated with DMSO. β-Actin was used as a loading control. Scale bars, 100 μm. Data are represented as mean ± SD (C, D, and F) or SEM (G) of three biological replicates. *$p < 0.05$; **$p < 0.01$ versus TF condition (C, D, and G) or versus parental cells (F); Dunnett's test. See also Figures S2 and S3.

MEK-ERK signaling, was abrogated in N1 cells. We, therefore, examined whether NR5A1 overexpression decreased the expression of FGFR1, a major receptor for FGF2 (Ornitz et al., 1996). FGFR1 was significantly downregulated at the mRNA and protein levels after NR5A1 overexpression, while the low levels of *FGFR2*, *3*, and *4* remained unchanged (Figures S2K and S2L). These results suggest that decreased FGFR1 expression diminished the response to FGF2, suppressing MEK/ERK activation in N1 cells. Taken together, we can conclude that the pluripotent state induced by NR5A1 is independent of MEK-ERK signaling, making treatment with PD03 unnecessary. Therefore, we focused on investigating the characteristics of N1 cells cultured with CHIR.

As the self-renewal of primed hESCs is dependent on TGF-β/activin/nodal signaling (Weinberger et al., 2016), we next cultured N1 cells with the activin receptor-like kinase inhibitors A83-01 (A83) and SB431542 (SB43). Their dependency on TGF-β/activin/nodal signaling became pronounced as the passage number increased (Figure S3A), indicating that NR5A1-induced cells were dependent on these pathways.

In the primed state, hESCs maintain a glycolytic metabolic state with a low mitochondrial respiration capacity, which changes to a low glycolysis, highly aerobic respiration-dependent state in the naive state (Guo et al., 2016; Sperber et al., 2015; Takashima et al., 2014; Zhou et al., 2012). Therefore, we analyzed the metabolic state of N1-CHIR cells in comparison with parental cells, N1-DOX (−), and N1-DOX (+) (−) cells, which were transferred from the CHIR condition with DOX to the TF condition without DOX (Figures S3B and S3C). Although the basal oxygen consumption rate revealed minimal changes between the cell lines, maximal respiration was significantly increased in N1-CHIR cells compared with the other lines.

Previous reports have demonstrated a higher frequency of genomic instability in naive cells subjected to continuous passaging by single-cell dissociation (Pastor et al., 2016; Theunissen et al., 2014). In this study, while NR5A transfectants derived from H9 cells displayed normal karyotypes for >10 passages, chromosomal abnormalities were observed at 25 and 19 passages in N1 cells derived from KhES-1 and 253G1 cells, respectively (Figures S3D and S3E). This indicates that the risk of genomic instability increased by single-cell dissociation, even though the cells retained their pluripotent state.

NR5A1-Induced Cells Have Transcriptional Features of Both Primed and Naive Pluripotency

To examine the global transcriptional state of NR5A1-induced cells grown under the CHIR condition, we first performed microarray analysis of N1-CHIR cells, comparing them with parental cells cultured under TF or CHIR conditions (parental-TF and parental-CHIR cells, respectively) and N1 cells cultured under the TF condition (N1-TF cells; Figures 3A–3C). Principal-component analysis and heatmaps showed similar transcriptional states for N1 cells under both culture conditions, but these patterns were completely different from those of parental-TF or parental-CHIR cells (Figures 3A and 3B). However, a set of genes associated with naive pluripotency, including *ZFP57*, *FGF4*, *KHDC3L*, and *CTSF* was upregulated in N1-CHIR cells compared with N1-TF cells (Figure 3B; Table S3), suggesting that the CHIR condition is more suitable for the promotion of naive-associated gene expression than the TF condition. CHIR treatment of parental cells induced a marked increase in the expression of gene sets associated with ectoderm differentiation and Wnt signaling (Figures 3C and S4A), whereas upregulation of these genes was not observed in N1-CHIR cells, indicating that the response of NR5A1-induced cells to GSK3 blockade was altered at the transcriptional level. We next performed RNA sequencing (RNA-seq) analysis of parental-TF and N1-CHIR cells (Figures 3D–3I); comparative analysis of RNA-seq datasets from this study and of human naive-like cell lines (Chan et al., 2013; Gafni et al., 2013; Takashima et al., 2014; Theunissen et al., 2014) indicated that N1-CHIR cells were closer to 3iL and NHSM cells than to 4i/L/A, 5i/L/A, WIBR3-DOX, or reset cells (Figure 3D; Table S4). Among the naive-associated genes examined, *KLF2* and *KLF17* expression was lower in N1-CHIR cells than in 4i/L/A, 5i/L/A, WIBR3-DOX, or reset cells (Figures 3E and S4B). Overexpression of NR5A1 markedly increased DNA methyltransferase 3 beta (DNMT3B) expression at the mRNA and protein levels (Figures S4B and S4C). We next examined the expression of transposable elements (TEs), such as human endogenous retroviruses (HERVs), HERV-associated long terminal repeats (LTRs), and SINE-VNTR-*Alu* (SVA) elements, as diminished transcription of LTR7 and HERVH-int elements and elevated transcription of LTR5_Hs, HERVK-int, and SVA elements was reported in 5i/L/A and reset cells (Guo et al., 2017; Theunissen et al., 2016). LTR5_Hs and HERVK-int elements were listed in the top 100 TEs ($\log_2$ fold change > 1.5, adjusted $p < 0.05$; Table S5). The heatmap revealed significantly increased expression of these transcripts in N1 cells (Figures 3F and 3G). However, while expression of some LTR7 and HERVH-int elements was reduced, a large number of elements showed increased expression (Figures 3H and 3I). SVA elements were low and unchanged between parental-TF and N1-CHIR cells (Figure 4D, right panel). These results demonstrate that NR5A1-induced cells have features of both primed and naive pluripotency, suggesting that they represent an intermediate state of conversion between these states.

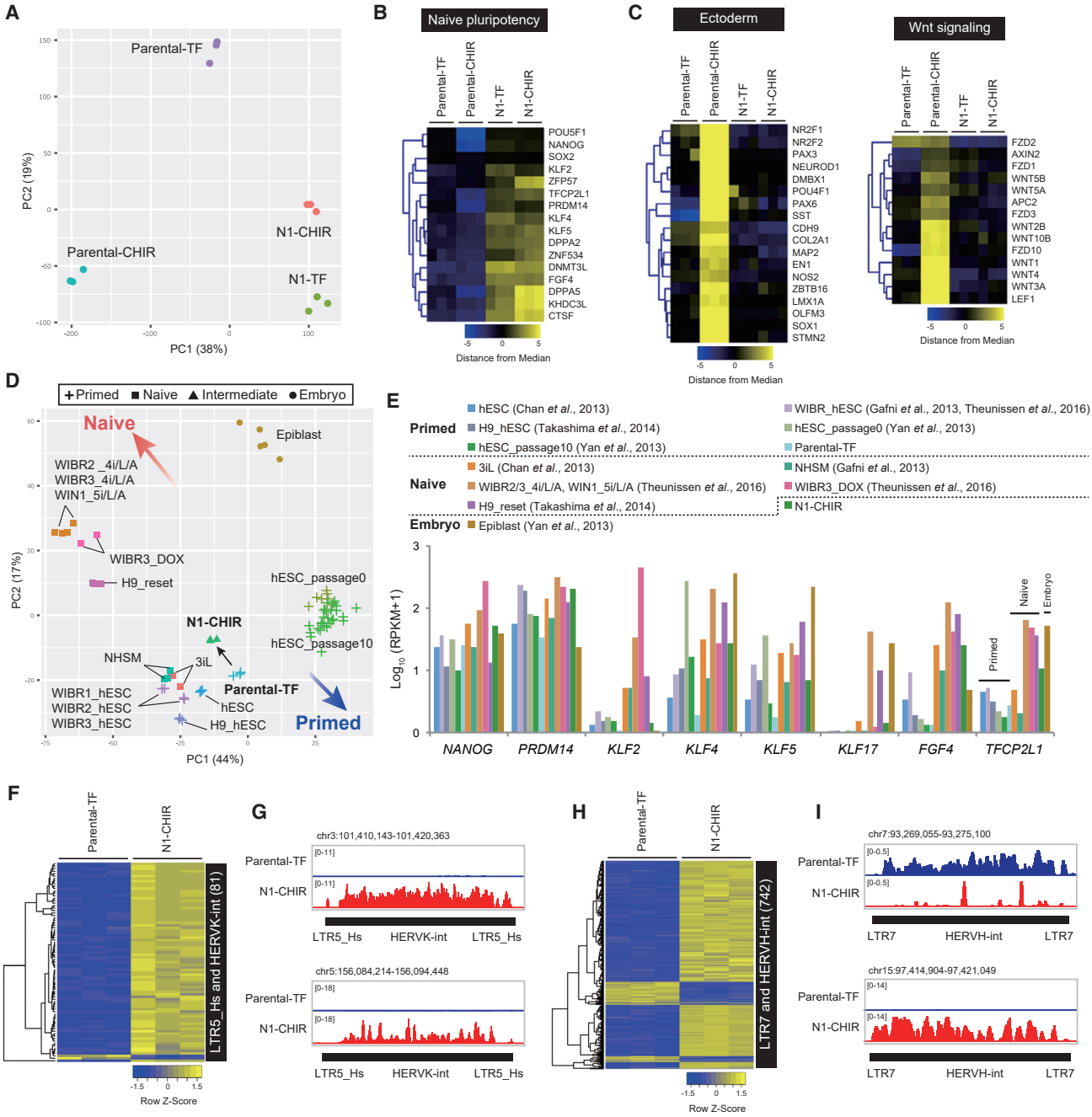


Figure 3. NR5A1-Induced Cells Bear Transcriptional Features of Both Primed and Naive Pluripotency

(A) Principal-component analysis (PCA) of microarray data from parental and N1 cells cultured under TF or CHIR conditions.

(B and C) Heatmaps of genes selected from microarray data under each condition.

(D) PCA of RNA-seq data from this study and previously published human naive-like cell lines (3iL, NHSM, WIBR2_4i/L/A, WIBR3_4i/L/A, WIN1_5i/L/A, and WIBR3_DOX, H9_reset), human primed cell lines (hESC_passage0 and 10), and human embryos.

(E) Expression of individual genes in the RNA-seq data in (D). Values are shown as the $\log_{10}$ (RPKM+1). RPKM, reads per kilobase of exon per million mapped reads.

(F–I) Heatmaps (adjusted $p < 0.05$, $\log_2$ FC > 1.5; F and H) and expression patterns (G and I) of selected TEs under each condition.

All microarray and RNA-seq data in this study were obtained from three biological replicates. See also Figure S4 and Tables S3, S4, and S5.

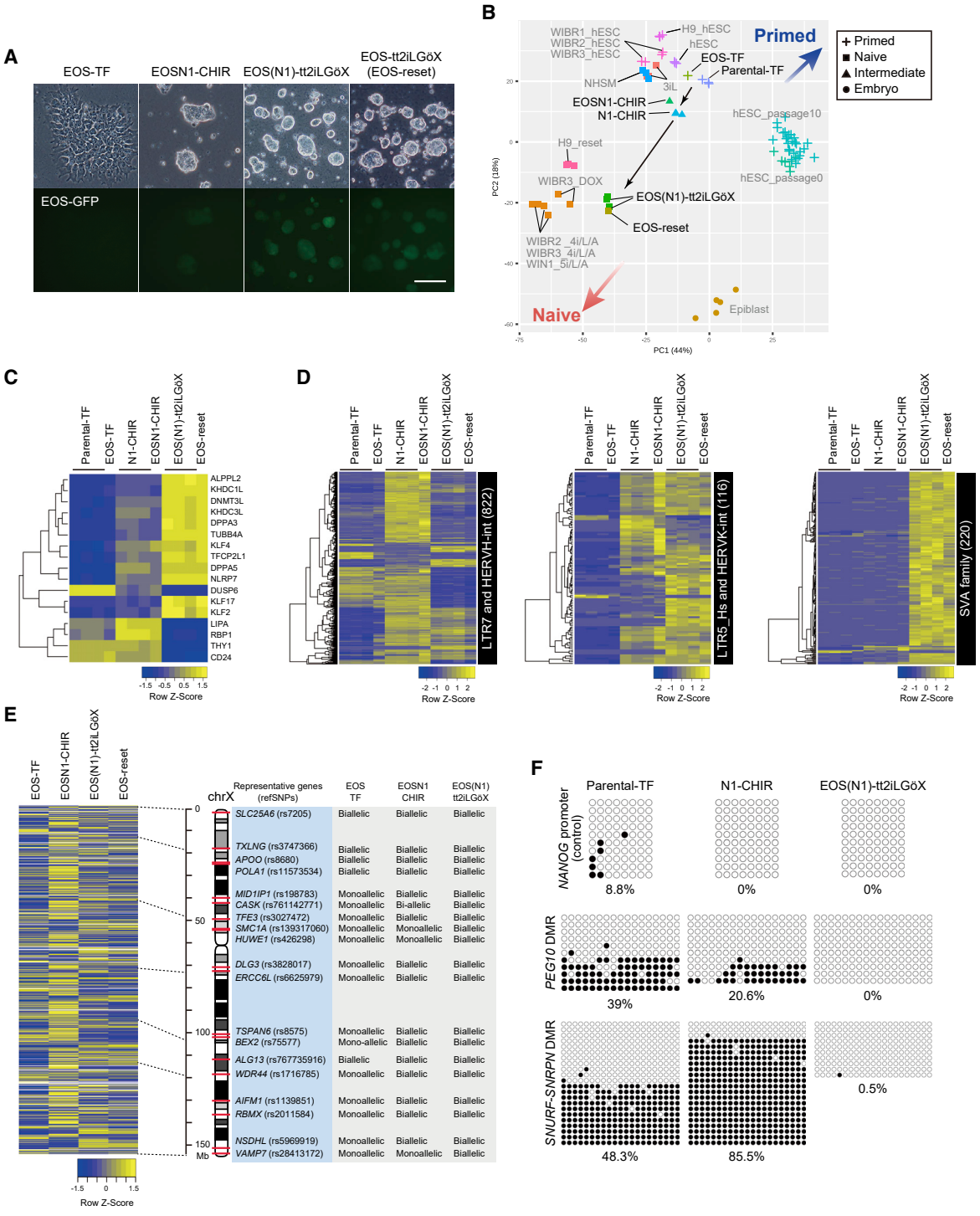


Figure 4. Induction of the Full Naive Pluripotent State in NR5A1-Induced Cells

(A) Representative images of cells cultivated under TF, CHIR, and tt2iLGöX conditions. (N1) represents the absence of NR5A1 expression by DOX withdrawal. EOS-reset cells were established as a control for naive cells. Scale bar, 100 μm.

(B) PCA of RNA-seq data from this study and previously published human naive-like cell lines (3iL, NHSM, WIBR2_4i/L/A, WIBR3_4i/L/A, WIN1_5i/L/A, and WIBR3_DOX, H9_reset), human primed cell lines (hESC_passage0 and 10), and human embryos.

(C and D) Heatmaps of naive-associated genes (C) and TEs (D) FDR < 0.05 among parental, N1, and EOS(N1) cells under each condition.

(legend continued on next page)

NR5A1-Induced Cells Can Progress to the Full Naive Pluripotent State without NR5A1 Overexpression

To examine whether NR5A1-induced cells exist in a state where they can be converted to naive pluripotency, we next investigated whether N1 cells can progress to the full naive state. Analyses of N1 cells revealed that 2i and 2i/LIF conditions had little effect on *DPPA3* expression and proliferation (Figures 2D and 2F), leading us to test a tt2iLGöX condition containing titrated CHIR, PD03, LIF, the protein kinase inhibitor Gö6983, and the tankyrase inhibitor XAV939, which was recently reported to be a more efficient protocol for naive induction (Guo et al., 2017). Because naive induction under the tt2iLGöX condition is accompanied by differentiation and cell death (Guo et al., 2017), the EOS-C(3+)-GFP-IRES-Puro (EOS-GFP) reporter system, which monitors the activity of a multimerized CR4 element derived from the mouse *Oct3/4* distal enhancer driven by an early transposon promoter (Guo et al., 2017; Hotta et al., 2009; Takashima et al., 2014), was used to detect and purify naive cells. GFP expression in DOX-treated parental cells carrying EOS-GFP (EOS-TF cells) was lower than in cells with NR5A1 expression and CHIR treatment (referred to as EOSN1-CHIR cells; Figures 4A and S4D). After switching from CHIR to tt2iLGöX medium, colonies cultured with DOX gradually collapsed and displayed slower proliferation (Figure S4E), whereas DOX-untreated cells (referred to as EOS(N1)-tt2iLGöX cells) formed dome-shaped colonies with robust GFP expression (Figures 4A and S4D). These results suggest that NR5A1 overexpression is dispensable upon conversion, as previously reported in mice (Guo and Smith, 2010). We next performed RNA-seq analysis of EOS(N1)-tt2iLGöX cells for direct comparison with the reset cells (Guo et al., 2017) as well as cells in formative pluripotency, which is an intermediate state of the conversion from naive to primed pluripotency (Rostovskaya et al., 2019). The results demonstrated that EOS(N1)-tt2iLGöX cells shared features with reset cells in terms of both gene and TE expression (Figures 4B–4D; Tables S4 and S6). Notably, N1 cells were clustered closely together with cells at days 7 and 10 of formative transition when expression of formative pluripotency-associated genes was compared (Rostovskaya et al., 2019) (Figure S4F). These findings suggest that NR5A1-induced cells are in an intermediate state between naive and primed states in terms of their gene expression profile and were converted to the full naive state. Moreover, the findings suggest that naive pluripotency can be induced from the intermediate cell state, which is stably maintained by NR5A1 overexpression and CHIR treatment.

X chromosome status is an important hallmark used to define the state of hPSCs (Sahakyan et al., 2017; Theunissen et al., 2016). We therefore investigated this in NR5A1-induced intermediate cells by RNA-fluorescence *in situ* hybridization (FISH) and SNP-based allelic expression analysis of X-linked genes using RNA-seq data. RNA-FISH of *XACT* and *HUWE1*, which are subjected to X chromosome reactivation (Sahakyan et al., 2017), and SNP analyses revealed eroded X inactivation (XaXe) of EOS-TF cells (Figures 4E, S5A, and S5B), as reported previously (Mekhoubad et al., 2012). In EOSN1-CHIR cells, no *XIST* coating was observed (Figure S5B). Almost all examined X-linked genes were biallelically expressed in the cells, although monoallelic expression of *HUWE1* was strongly enforced (Figures S5A and S5B). This result is consistent with the increased transcriptional states of X-linked genes in EOSN1-CHIR cells compared with EOS-TF cells (Figure 4E). In tt2iLGöX medium, the proportion of cells expressing *HUWE1* biallelically was markedly increased with the appearance of monoallelic *XIST* coating, but the majority of cells bore no *XIST* clouds (Figures S5B and S5C), and expression of X-linked genes was decreased (Figure 4E). Similar phenotypes were observed in our reset cells (Figures 4E and S5A–S5C) and in previously established reset cells (Guo et al., 2017; Sahakyan et al., 2017). These results suggest that both X chromosomes are highly activated in the NR5A1-induced intermediate cell state.

Because naive cells are hypomethylated at differentially methylated regions (DMRs) in their imprinted genes (Theunissen et al., 2016), we performed bisulfite sequencing analysis to examine the methylation levels of the DMRs of *PEG10* and *SNURF-SNRPN* (Figure 4F). The DMRs of parental-TF cells were partially methylated, as previously demonstrated in hPSCs (Kim et al., 2007). DNA methylation was maintained on the DMRs of EOSN1-CHIR cells but absent in EOS(N1)-tt2iLGöX cells. These results strongly suggest that hypomethylation of imprinted DMRs is not established at the intermediate cell state, probably because of the sustained high expression of DNMT3A and DNMT3B in N1 cells (Figure S4C).

(E) Transcriptional status of the two X chromosomes under each condition. A heatmap of X-linked genes and SNP-based allelic expression patterns in the RNA-seq data were analyzed. Allelic expression patterns of representative genes (blue box) are shown with their reference SNP IDs (rs), locations on the X chromosome (red bars), and type of expression ("monoallelic" or "biallelic"; gray box). RNA-seq data were obtained from one (EOS-TF, EOSN1-CHIR, and EOS-reset) and three (EOS(N1)-tt2iLGöX) biological samples.

(F) Bisulfite sequencing analysis of DMRs in imprinted genes. Open and closed circles represent unmethylated and methylated CpG sites, respectively. See also Figure S5, Tables S4 and S6.

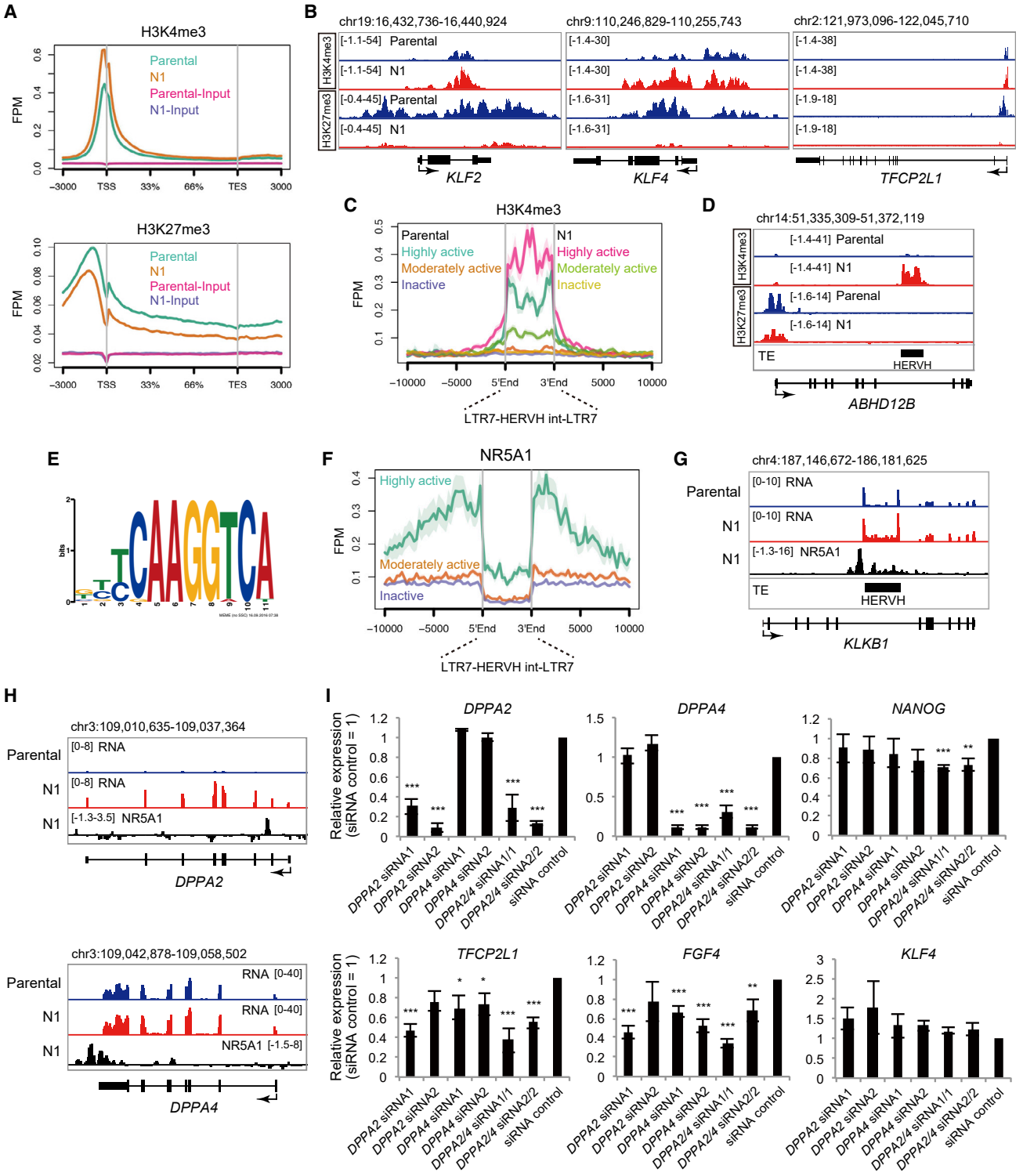


Figure 5. ChIP-Seq Analysis of Histone Modifications and NR5A1 in NR5A1-Induced Cells

(A) Average ChIP-seq signals for H3K4me3 (top) and H3K27me3 (bottom) concentrated around transcription start sites (TSSs) in parental and N1 cells.

(B) H3K4me3 and H3K27me3 signal distributions on individual genes.

(legend continued on next page)

NR5A1 Binds Regions of *DPPA2* and *DPPA4* that Regulate the Expression of *TFCP2L1* and *FGF4* but Not *KLF4*

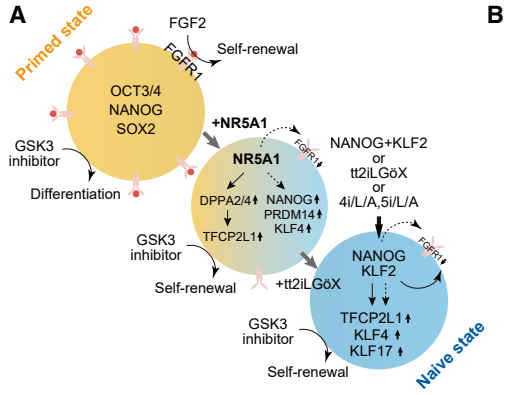
To investigate the epigenetic status of NR5A1-induced intermediate cells, the active histone mark histone 3 lysine 4 trimethylation (H3K4me3) and the repressive mark histone 3 lysine 27 trimethylation (H3K27me3) were analyzed by chromatin immunoprecipitation sequencing (ChIP-seq). A sharp H3K4me3 peak observed in parental cells at the transcription start site (TSS) was increased in N1-CHIR cells (Figure 5A). In contrast, H3K27me3 peaks upstream of the TSS to downstream of the transcriptional end site were decreased in these cells. Naive pluripotency genes, including *KLF2*, *KLF4*, and *TFCP2L1*, which harbored bivalent H3K4me3 and H3K27me3 marks in parental cells, exhibited a loss of H3K27me3 in the intermediate state (Figure 5B). However, a clear reduction in H3K27me3 was detected at the TSSs of only a few development-associated genes, such as *NKX2-5*, in N1-CHIR cells (Figures S6A and S6B), whose phenotypes are different from those of previously established naive cells (Theunissen et al., 2014). This analysis also revealed that H3K4me3 signals were increased on transcriptionally activated HERVHs in NR5A1-induced intermediate cells (Figures 5C and 5D).

In mice, the *Oct3/4*, *Nanog*, and *Klf2* genes are direct targets of NR5A2 during somatic cell reprograming into iPSCs (Heng et al., 2010). To examine whether these genes are targets of NR5A1, we performed ChIP followed by deep sequencing. Motif analysis of NR5A1 revealed a previously reported consensus motif (Figure 5E) (Martin and Tremblay, 2010). ChIP signals were distributed near the TSSs of some genes, such as *NPAS4* (Figure S6C), which were reported previously in the human adrenal cortical carcinoma cell line H295R (Doghman et al., 2013) suggesting that NR5A1 binding to these genes is independent of cell type, and confirming the accuracy of our ChIP-seq analysis. The analysis also revealed strong NR5A1 binding to nearby HERVH regions that are activated in primed hPSCs (Wang et al., 2014) (Figures 5F and 5G), suggesting that NR5A1 is involved in the transcriptional activation of HERVHs, resulting in the stall at the intermediate conversion state. Significant enrichments were not observed on the *OCT3/4*, *NANOG*, or *KLF2* genes, but were observed in regions of *DPPA2* and *DPPA4* (Figure 5H). As both genes are key reprogramming factors in mice (Hernandez et al., 2018), we performed knockdowns using short interfering RNAs (siRNAs) against *DPPA2* and *DPPA4* to confirm that these genes were functional in the NR5A1-mediated pluripotent state (Figures 5I and S6D). Gene expression analysis of N1 cells after knockdown of both genes revealed downregulation of naive-associated genes, such as *NANOG*, *TFCP2L1*, and *FGF4*, although *KLF4* expression was maintained. Furthermore, *PAX6* and *FOXF1* were upregulated. These results suggest that the pluripotent state of N1 cells shifts toward differentiation rather than reverting to the primed state, suggesting that DPPA2 and DPPA4 may be involved in the transcriptional regulation of a subset of naive-associated genes in the NR5A1-induced intermediate state between prime and naive pluripotency (Figure 6A).

DISCUSSION

Accumulating studies on human naive pluripotency have provided ample criteria for evaluating whether established cells are distinct from primed pluripotent cells (Chan et al., 2013; Chen et al., 2015; Duggal et al., 2015; Gafni et al., 2013; Qin et al., 2016; Takashima et al., 2014; Theunissen et al., 2014; Ware et al., 2014; Yang et al., 2015). Based on these criteria, we concluded that NR5A1-induced cells exist in an intermediate state between the primed and naive states (Figure 6B). These cells bear a subset of features of naive pluripotency, such as clonogenicity, GSK3 and MEK-ERK independence, increased mitochondrial respiration capacity, naive-specific gene and LTR5-HERVK expression profiles, and two activated X chromosomes. The results suggest that these phenotypic changes occur early in the process of naive conversion. Features, such as TGF-β/activin/nodal and FGF independence, naive-specific expression profiles of TEs, DNA demethylation of imprinted genes, and *XIST* coating of the active X chromosome were observed under the tt2iLGöX condition, suggesting that these features are acquired during later stages of

(C) Signal enrichment of H3K4me3 marks on moderately and highly active types of HERVHs, characterized based on their expression in primed hPSCs, parental, and N1 cells. FPM, fragments per million mapped fragments.
(D) Distributions of H3K4me3 and H3K27me3 signals on the *ABHD12B* gene in N1 cells.
(E) NR5A1 consensus motif detected by the *de novo* motif discovery algorithm MEME.
(F) NR5A1 signal distributions on the three types of HERVHs in N1 cells.
(G) NR5A1 signal distribution on the *KLKB1* gene in N1 cells.
(H) NR5A1 signal distribution on the *DPPA2* and *DPPA4* genes in N1 cells.
(I) Knockdown of *DPPA2* and *DPPA4* in N1 cells. Data were analyzed by the comparative ΔCt method. Cells treated with non-targeted siRNA (siRNA control) were set to 1. Data are represented as mean ± SD of three biological replicates. *p < 0.05, **p < 0.01, ***p < 0.001 (Student's t test). See also Figure S6.



B

Naive-associated features observed in this study	Primed (TF condition)	Intermediate (CHIR condition)	Naive
Clonogenicity	(-)	(+)	(+)*
Self-renewal in 2i conditions	(-)	(+)	(+)
Mitochondrial respiration capacity	(-)	(+)	(+)*
Expression of *KLF4* and *TFCP2L1*	(-)	(+)	(++)
Expression of LTR5_Hs-HERVKs and SVA family	(-)	Partially (+)	(+)
Expression of EOS-GFP	Low	(+)	(++)
Status of two X chromosomes	XaXe	XaXa	XaXa^{XIST+}
Expression of LTR7_HERVHs	(+)	(++)	Low
Expression of *KLF17*	(-)	(-)	(+)
Dependence of TGF-β familiy signaling	(+)	(+)	(-)
DNA methylation of imprinted DMRs	(+)	(+)	Low
Expression of *DNMT3A* and *DNMT3B*	(+)	(+)	Low
H3K27me3 marks on development-associated genes	(+)	(+)	Low*

Figure 6. Graphical Summary of This Study

(A) Transcription factors induced by NR5A1 overexpression. Arrows and dotted arrows represent direct and indirect potential targets of the indicated genes, respectively.

(B) Representative features of hPSCs in the primed, intermediate, and naive states observed in this study. Asterisks indicate results from previous studies.

conversion. In addition to these results, *KLF17* was expressed at low levels in N1-CHIR cells, but was upregulated after switching to the tt2iLGöX condition. Previous studies have revealed that KLF17 is detectable in cells in the later stages of naive conversion (Kilens et al., 2018), supporting our definition of the pluripotent state of the NR5A1-induced cells. Collectively, our results suggest that naive conversion progresses in a stepwise fashion.

NR5A1 is an orphan nuclear receptor belonging to the nuclear receptor superfamily. In mice, the protein plays central roles in endocrine organ development and steroidogenesis by regulating the transcription of steroidogenic enzyme and glycolytic genes (Lala et al., 1992; Schimmer and White, 2010). Notably, *Nr5a1* is not expressed in the blastocyst (Gu et al., 2005; Ikeda et al., 1994; Pare et al., 2004). Nevertheless, Nr5a1 can replace Oct3/4 as a reprogramming factor in the generation of iPSCs from somatic cells (Heng et al., 2010). Furthermore, it can convert mEpiSCs to a naive pluripotent state (Guo and Smith, 2010). These studies led us to investigate the ability of NR5A1 in inducing human naive pluripotency. Consistent with the previous report on Nr5a1 function in mEpiSCs, continuous NR5A1 expression was dispensable after completing the conversion process in humans. However, NR5A1 overexpression induced only a subset of naive-associated features in hPSCs. The cellular response to suppressed GSK3 and MEK-ERK signaling appears to differ between mice and humans, as extrinsic suppression by 2i is required to establish naive cells from mEpiSCs with Nr5a expression (Guo and Smith, 2010), which is in contrast to our results. Exogenous expression of Nr5a1 enhances the initial expression of *Klf2*, suggesting that *Klf2* could be direct target of Nr5a1 in mEpiSCs (Guo and Smith, 2010), but no significant binding of NR5A1 was detected on *KLF2* in hPSCs. More importantly, whether Nr5a1 is involved is the activation of murine ERVs remains unknown. Our results suggest that the ability of NR5A1 to induce naive pluripotency differs slightly in mice and humans, and careful examination will be necessary to understand its molecular functions in different species.

Overexpression of NANOG and KLF2 induces the naive state in hPSCs (Takashima et al., 2014; Theunissen et al., 2014), and overexpression of KLF4 accelerates naive conversion in the t2iLGö condition (Liu et al., 2017). These reports prompted us to investigate whether *NANOG* and *KLF4* are potential targets of NR5A1, but ChIP-seq analysis showed no significant binding to these genes. However, we found that NR5A1 bound to *DPPA2* and *DPPA4*, whose expression was upregulated in NR5A1-induced cells. Knockdown of both genes indicated that a DPPA2/4-dependent pathway regulated the expression of naive-associated genes, such as *TFCP2L1* and *FGF4*, but not *KLF4*. In addition, the siRNA-mediated depletion of *DPPA2* and *DPPA4* increased the expression of differentiation-associated genes. These observations suggest that these genes play a vital role in maintaining the proper balance between pluripotency and differentiation in the NR5A1-induced state.

It was unexpected that overexpression of NR5A1 elicited a suppressed response to MEK-ERK signaling by extrinsic FGF2 stimuli in a cell-autonomous manner, as such a response has not been reported in humans or mice. In general, FGF-MEK-ERK signaling has diverse functions, including cell-cycle control (Roovers and Assoian, 2000); thus, it is highly possible that in this study the suppressed FGF2 response led to reduced proliferation and G1-phase

cell accumulation in NR5A1-induced cells. Inhibition of ERK phosphorylation by PD03 is required for naive conversion, although *FGFR1* expression is downregulated in reset and 5iLA cells (Takashima et al., 2014; Theunissen et al., 2014). Notably, expression of other *FGFRs*, including *FGFR3*, is upregulated in these cells, but remained low in N1 cells (Figure S2K). MEK-ERK signaling can be activated through FGFR3 as well as FGFR1 (Ornitz et al., 1996), which might explain the necessity for PD03 in the naive state. Direct binding of NR5A1 to *FGFR1* was not observed in our ChIP-seq analysis, suggesting that NR5A1 indirectly regulates *FGFR1* expression.

We expected that human NR5A2 would have a similar ability as NR5A1 because mouse Nr5a2 also induces the naive state of mEpiSCs (Guo and Smith, 2010), and human NR5A2 can function as an iPSC reprogramming factor in humans (Wang et al., 2011). However, despite high sequence similarity between the *NR5A* genes in their DNA-binding domains, NR5A2 only partially mimicked the effects of NR5A1. Primed hPSCs express NR5A2, but show lower expression of PRDM14, KLF4, and TFCP2L1, and unlike NR5A1, NR5A2 overexpression did not promote expression of these genes. These results suggest that NR5A2-specific molecular mechanisms exist in hPSCs.

In summary, NR5A1 has the ability to induce and maintain cells at an intermediate state of naive conversion that has until now been inaccessible for study. Our data suggest that the pluripotent state of N1 cells may be comparable with the state of formative pluripotency, warranting further investigations on the relationship between the NR5A1-induced state and formative transition. Using NR5A1 overexpression to induce the naive state will be advantageous in elucidating the processes that regulate the conversion between naive and primed pluripotency. We cannot exclude the possibility that the pluripotent state induced by NR5A1 does not reflect a particular stage of normal development, as *NR5A1* is not expressed in the human ICM or in primed or naive cells (Yan et al., 2013). However, this model will nonetheless contribute to our understanding of the molecular mechanisms that regulate pluripotency in humans.

EXPERIMENTAL PROCEDURES

To establish a DOX-inducible lentiviral system, the pLV-tTRKRAB and pLVCT-tTR-KRAB vectors were purchased from Addgene (plasmids 12249 and 11643, respectively). To construct the pLV-tTR-KRAB-IRES-Neomycin (Neo) vector, an IRES-Neo cassette was inserted into pLV-tTRKRAB. To construct the pLVCT-NR5A-IRES-Puromycin (Puro) and pLVCT-NR5A-IRES-Zeocin (Zeo) vectors, the backbone of pLVCT carrying the CAG promoter was used after removing the DNA fragment (EGFP-IRES-tTR-KRAB). IRES-Puro and IRES-Zeo cassettes were subcloned into the vectors. FLAG-tagged NR5A1 and NR5A2 were amplified by PCR and each inserted into pLVCT-IRES-Puro. The pLVCT-3×FLAG-tagged NR5A1-IRES-Puro vector was generated for ChIP with an anti-FLAG antibody. Primer sequences are listed in Table S1.

ACCESSION NUMBERS

All microarray and sequencing data have been deposited in the GEO data repository under accession number GSE101089.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.stemcr.2020.01.012.

AUTHOR CONTRIBUTIONS

K.Y. and H.S. conceived the project, designed the experiments, analyzed the data, and wrote the manuscript. K.Y. performed the experiments. T.I. analyzed the microarray, RNA-seq, and ChIP-seq data. M.H. and S.C. developed the RNA-seq and ChIP-seq methods and provided technical assistance. E.K., S.C., and K.H. assisted with data analysis and writing the manuscript. N.N. and H.S. supervised the project.

ACKNOWLEDGMENTS

This study was partially supported by the Research Project for Practical Applications of Regenerative Medicine from the Japan Agency for Medical Research and Development (project number, JP17bk0104043) and a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technologies of Japan (project number, JP19K08539). We thank Takamichi Miyazaki and Kei Takada for helpful discussions, Mari Hamao for karyotype analysis, Fumiyuki Nakagawa for plasmid construction, and Emiko Moribe for sequencing. We are grateful to Masako Tada at Toho University (Japan) and Hiroyuki Kugoh at Tottori University (Japan) for advice and support with RNA-FISH, and the RIKEN BRC DNA Bank at the National BioResource Research Center for providing the NR5A1 plasmid.

Received: July 14, 2019
Revised: January 22, 2020
Accepted: January 23, 2020
Published: February 20, 2020

REFERENCES

Boroviak, T., Loos, R., Bertone, P., Smith, A., and Nichols, J. (2014). The ability of inner cell mass cells to self-renew as embryonic stem cells is acquired upon epiblast specification. Nat. Cell Biol. *16*, 516–528.

Brons, I.G., Smithers, L.E., Trotter, M.W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191–195.

Chan, Y.S., Goke, J., Ng, J.H., Lu, X., Gonzales, K.A., Tan, C.P., Tng, W.Q., Hong, Z.Z., Lim, Y.S., and Ng, H.H. (2013). Induction of a

human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. Cell Stem Cell *13*, 663–675.

Chen, H., Aksoy, I., Gonnot, F., Osteil, P., Aubry, M., Hamela, C., Rognard, C., Hochard, A., Voisin, S., Fontaine, E., et al. (2015). Reinforcement of STAT3 activity reprogrammes human embryonic stem cells to naive-like pluripotency. Nat. Commun. *6*, 7095.

Doghman, M., Figueiredo, B.C., Volante, M., Papotti, M., and Lalli, E. (2013). Integrative analysis of SF-1 transcription factor dosage impact on genome-wide binding and gene expression regulation. Nucleic Acids Res. *41*, 8896–8907.

Duggal, G., Warrier, S., Ghimire, S., Broekaert, D., Van der Jeught, M., Lierman, S., Deroo, T., Peelman, L., Van Soom, A., Cornelissen, R., et al. (2015). Alternative routes to induce naive pluripotency in human embryonic stem cells. Stem Cells *33*, 2686–2698.

Dunn, S.J., Martello, G., Yordanov, B., Emmott, S., and Smith, A.G. (2014). Defining an essential transcription factor program for naive pluripotency. Science *344*, 1156–1160.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154–156.

Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A., Tomlinson, S.R., and Chambers, I. (2012). Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. Cell Stem Cell *11*, 477–490.

Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., et al. (2013). Derivation of novel human ground state naive pluripotent stem cells. Nature *504*, 282–286.

Gillich, A., Bao, S., Grabole, N., Hayashi, K., Trotter, M.W., Pasque, V., Magnusdottir, E., and Surani, M.A. (2012). Epiblast stem cell-based system reveals reprogramming synergy of germline factors. Cell Stem Cell *10*, 425–439.

Gu, P., Goodwin, B., Chung, A.C., Xu, X., Wheeler, D.A., Price, R.R., Galardi, C., Peng, L., Latour, A.M., Koller, B.H., et al. (2005). Orphan nuclear receptor LRH-1 is required to maintain Oct4 expression at the epiblast stage of embryonic development. Mol. Cell. Biol. *25*, 3492–3505.

Guo, G., and Smith, A. (2010). A genome-wide screen in EpiSCs identifies Nr5a nuclear receptors as potent inducers of ground state pluripotency. Development *137*, 3185–3192.

Guo, G., von Meyenn, F., Rostovskaya, M., Clarke, J., Dietmann, S., Baker, D., Sahakyan, A., Myers, S., Bertone, P., Reik, W., et al. (2017). Epigenetic resetting of human pluripotency. Development *144*, 2748–2763.

Guo, G., von Meyenn, F., Santos, F., Chen, Y., Reik, W., Bertone, P., Smith, A., and Nichols, J. (2016). Naive pluripotent stem cells derived directly from isolated cells of the human inner cell mass. Stem Cell Reports *6*, 437–446.

Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development *136*, 1063–1069.

Hall, J., Guo, G., Wray, J., Eyres, I., Nichols, J., Grotewold, L., Morfopoulou, S., Humphreys, P., Mansfield, W., Walker, R., et al. (2009). Oct4 and LIF/Stat3 additively induce Kruppel factors to sustain embryonic stem cell self-renewal. Cell Stem Cell *5*, 597–609.

Hanna, J., Cheng, A.W., Saha, K., Kim, J., Lengner, C.J., Soldner, F., Cassady, J.P., Muffat, J., Carey, B.W., and Jaenisch, R. (2010). Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. Proc. Natl. Acad. Sci. U S A *107*, 9222–9227.

Hanna, J., Markoulaki, S., Mitalipova, M., Cheng, A.W., Cassady, J.P., Staerk, J., Carey, B.W., Lengner, C.J., Foreman, R., Love, J., et al. (2009). Metastable pluripotent states in NOD-mouse-derived ESCs. Cell Stem Cell *4*, 513–524.

Heng, J.C., Feng, B., Han, J., Jiang, J., Kraus, P., Ng, J.H., Orlov, Y.L., Huss, M., Yang, L., Lufkin, T., et al. (2010). The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. Cell Stem Cell *6*, 167–174.

Hernandez, C., Wang, Z., Ramazanov, B., Tang, Y., Mehta, S., Dambrot, C., Lee, Y.W., Tessema, K., Kumar, I., Astudillo, M., et al. (2018). Dppa2/4 facilitate epigenetic remodeling during reprogramming to pluripotency. Cell Stem Cell *23*, 396–411.e8.

Hotta, A., Cheung, A.Y., Farra, N., Vijayaragavan, K., Seguin, C.A., Draper, J.S., Pasceri, P., Maksakova, I.A., Mager, D.L., Rossant, J., et al. (2009). Isolation of human iPS cells using EOS lentiviral vectors to select for pluripotency. Nat. Methods *6*, 370–376.

Huang, K., Maruyama, T., and Fan, G. (2014). The naive state of human pluripotent stem cells: a synthesis of stem cell and preimplantation embryo transcriptome analyses. Cell Stem Cell *15*, 410–415.

Ikeda, Y., Shen, W.H., Ingraham, H.A., and Parker, K.L. (1994). Developmental expression of mouse steroidogenic factor-1, an essential regulator of the steroid hydroxylases. Mol. Endocrinol. *8*, 654–662.

Kilens, S., Meistermann, D., Moreno, D., Chariau, C., Gaignerie, A., Reignier, A., Lelievre, Y., Casanova, M., Vallot, C., Nedellec, S., et al. (2018). Parallel derivation of isogenic human primed and naive induced pluripotent stem cells. Nat. Commun. *9*, 360.

Kim, K.P., Thurston, A., Mummery, C., Ward-van Oostwaard, D., Priddle, H., Allegrucci, C., Denning, C., and Young, L. (2007). Gene-specific vulnerability to imprinting variability in human embryonic stem cell lines. Genome Res. *17*, 1731–1742.

Lala, D.S., Rice, D.A., and Parker, K.L. (1992). Steroidogenic factor I, a key regulator of steroidogenic enzyme expression, is the mouse homolog of fushi tarazu-factor I. Mol. Endocrinol. *6*, 1249–1258.

Liu, X., Nefzger, C.M., Rossello, F.J., Chen, J., Knaupp, A.S., Firas, J., Ford, E., Pflueger, J., Paynter, J.M., Chy, H.S., et al. (2017). Comprehensive characterization of distinct states of human naive pluripotency generated by reprogramming. Nat. Methods *14*, 1055–1062.

Martello, G., Sugimoto, T., Diamanti, E., Joshi, A., Hannah, R., Ohtsuka, S., Gottgens, B., Niwa, H., and Smith, A. (2012). Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. Cell Stem Cell *11*, 491–504.

Martin, L.J., and Tremblay, J.J. (2010). Nuclear receptors in Leydig cell gene expression and function. Biol. Reprod. *83*, 3–14.

Mekhoubad, S., Bock, C., de Boer, A.S., Kiskinis, E., Meissner, A., and Eggan, K. (2012). Erosion of dosage compensation impacts human iPSC disease modeling. Cell Stem Cell *10*, 595–609.

Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. Nature *460*, 118–122.

Ornitz, D.M., Xu, J., Colvin, J.S., McEwen, D.G., MacArthur, C.A., Coulier, F., Gao, G., and Goldfarb, M. (1996). Receptor specificity of the fibroblast growth factor family. J. Biol. Chem. *271*, 15292–15297.

Pare, J.F., Malenfant, D., Courtemanche, C., Jacob-Wagner, M., Roy, S., Allard, D., and Belanger, L. (2004). The fetoprotein transcription factor (FTF) gene is essential to embryogenesis and cholesterol homeostasis and is regulated by a DR4 element. J. Biol. Chem. *279*, 21206–21216.

Pastor, W.A., Chen, D., Liu, W., Kim, R., Sahakyan, A., Lukianchikov, A., Plath, K., Jacobsen, S.E., and Clark, A.T. (2016). Naive human pluripotent cells feature a methylation landscape devoid of blastocyst or germline memory. Cell Stem Cell *18*, 323–329.

Qin, H., Hejna, M., Liu, Y., Percharde, M., Wossidlo, M., Blouin, L., Durruthy-Durruthy, J., Wong, P., Qi, Z., Yu, J., et al. (2016). YAP induces human naive pluripotency. Cell Rep. *14*, 2301–2312.

Roovers, K., and Assoian, R.K. (2000). Integrating the MAP kinase signal into the G1 phase cell cycle machinery. Bioessays *22*, 818–826.

Rostovskaya, M., Stirparo, G.G., and Smith, A. (2019). Capacitation of human naive pluripotent stem cells for multi-lineage differentiation. Development *146*. https://doi.org/10.1242/dev.172916.

Sahakyan, A., Kim, R., Chronis, C., Sabri, S., Bonora, G., Theunissen, T.W., Kuoy, E., Langerman, J., Clark, A.T., Jaenisch, R., et al. (2017). Human naive pluripotent stem cells model X chromosome dampening and X inactivation. Cell Stem Cell *20*, 87–101.

Schimmer, B.P., and White, P.C. (2010). Minireview: steroidogenic factor 1: its roles in differentiation, development, and disease. Mol. Endocrinol. *24*, 1322–1337.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell *138*, 722–737.

Sperber, H., Mathieu, J., Wang, Y., Ferreccio, A., Hesson, J., Xu, Z., Fischer, K.A., Devi, A., Detraux, D., Gu, H., et al. (2015). The metabolome regulates the epigenetic landscape during naive-to-primed human embryonic stem cell transition. Nat. Cell Biol. *17*, 1523–1535.

Sturm, O.E., Orton, R., Grindlay, J., Birtwistle, M., Vyshemirsky, V., Gilbert, D., Calder, M., Pitt, A., Kholodenko, B., and Kolch, W. (2010). The mammalian MAPK/ERK pathway exhibits properties of a negative feedback amplifier. Sci. Signal. *3*, ra90.

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., et al. (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell *158*, 1254–1269.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196–199.

Theunissen, T.W., Friedli, M., He, Y., Planet, E., O'Neil, R.C., Markoulaki, S., Pontis, J., Wang, H., Iouranova, A., Imbeault, M., et al. (2016). Molecular criteria for defining the naive human pluripotent state. Cell Stem Cell *19*, 502–515.

Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., et al. (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell *15*, 471–487.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145–1147.

Vallier, L., Alexander, M., and Pedersen, R.A. (2005). Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. J. Cell Sci. *118*, 4495–4509.

Wang, J., Xie, G., Singh, M., Ghanbarian, A.T., Rasko, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N.V., et al. (2014). Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature *516*, 405–409.

Wang, W., Yang, J., Liu, H., Lu, D., Chen, X., Zenonos, Z., Campos, L.S., Rad, R., Guo, G., Zhang, S., et al. (2011). Rapid and efficient reprogramming of somatic cells to induced pluripotent stem cells by retinoic acid receptor gamma and liver receptor homolog 1. Proc. Natl. Acad. Sci. U S A *108*, 18283–18288.

Ware, C.B., Nelson, A.M., Mecham, B., Hesson, J., Zhou, W., Jonlin, E.C., Jimenez-Caliani, A.J., Deng, X., Cavanaugh, C., Cook, S., et al. (2014). Derivation of naive human embryonic stem cells. Proc. Natl. Acad. Sci. U S A *111*, 4484–4489.

Weinberger, L., Ayyash, M., Novershtern, N., and Hanna, J.H. (2016). Dynamic stem cell states: naive to primed pluripotency in rodents and humans. Nat. Rev. Mol. Cell Biol. *17*, 155–169.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1131–1139.

Yang, Y., Adachi, K., Sheridan, M.A., Alexenko, A.P., Schust, D.J., Schulz, L.C., Ezashi, T., and Roberts, R.M. (2015). Heightened potency of human pluripotent stem cell lines created by transient BMP4 exposure. Proc. Natl. Acad. Sci. U S A *112*, E2337–E2346.

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.

Zhou, W., Choi, M., Margineantu, D., Margaretha, L., Hesson, J., Cavanaugh, C., Blau, C.A., Horwitz, M.S., Hockenbery, D., Ware, C., et al. (2012). HIF1alpha induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. EMBO J. *31*, 2103–2116.

Stem Cell Reports, Volume *14*

Supplemental Information

Overexpression of Nuclear Receptor 5A1 Induces and Maintains an Intermediate State of Conversion between Primed and Naive Pluripotency

Kaori Yamauchi, Tatsuhiko Ikeda, Mihoko Hosokawa, Norio Nakatsuji, Eihachiro Kawase, Shinichiro Chuma, Kouichi Hasegawa, and Hirofumi Suemori

Figure S1

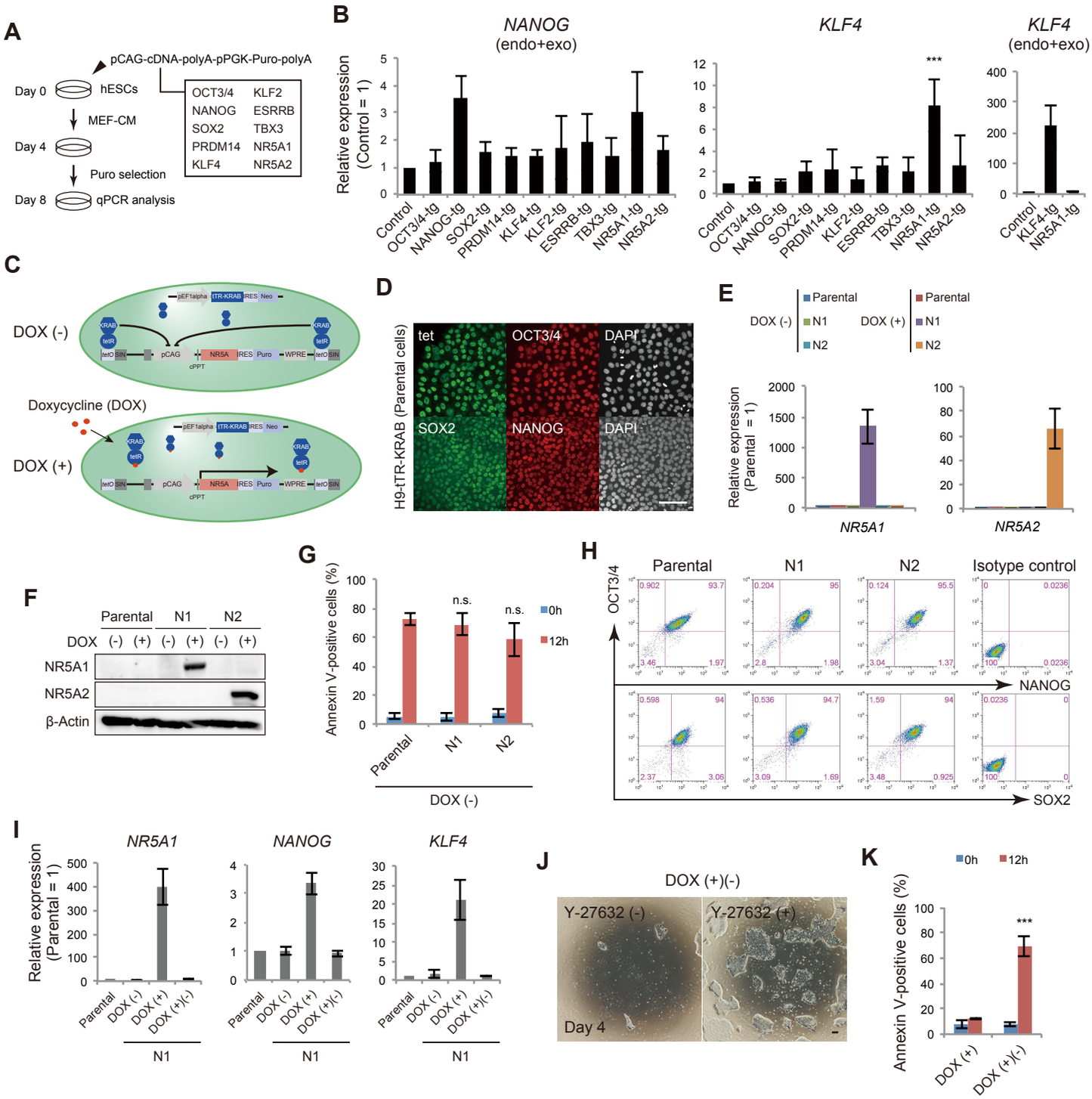

A
pCAG-cDNA-polyA-pPGK-Puro-polyA
Day 0
hESCs
MEF-CM
Day 4
Puro selection
Day 8
qPCR analysis
OCT3/4 KLF2
NANOG ESRRB
SOX2 TBX3
PRDM14 NR5A1
KLF4 NR5A2
B
NANOG (endo+exo)
KLF4
KLF4 (endo+exo)
Relative expression (Control = 1)
***
Control
OCT3/4-tg
NANOG-tg
SOX2-tg
PRDM14-tg
KLF4-tg
KLF2-tg
ESRRB-tg
TBX3-tg
NR5A1-tg
NR5A2-tg
C
DOX (-)
Doxycycline (DOX)
DOX (+)
D
H9-tTR-KRAB (Parental cells)
tet
OCT3/4
DAPI
SOX2
NANOG
DAPI
E
Parental
N1
N2
DOX (-)
DOX (+)
Relative expression (Parental = 1)
NR5A1
NR5A2
F
Parental
N1
N2
DOX (-) (+) (-) (+) (-) (+)
NR5A1
NR5A2
β-Actin
G
Annexin V-positive cells (%)
n.s.
n.s.
0h
12h
Parental
N1
N2
DOX (-)
H
Parental
N1
N2
Isotype control
OCT3/4
NANOG
SOX2
I
NR5A1
NANOG
KLF4
Relative expression (Parental = 1)
Parental
DOX (-)
DOX (+)
DOX (+)(-)
N1
J
DOX (+)(-)
Y-27632 (-)
Y-27632 (+)
Day 4
K
0h
12h
Annexin V-positive cells (%)
***
DOX (+)
DOX (+)(-)

Figure S2

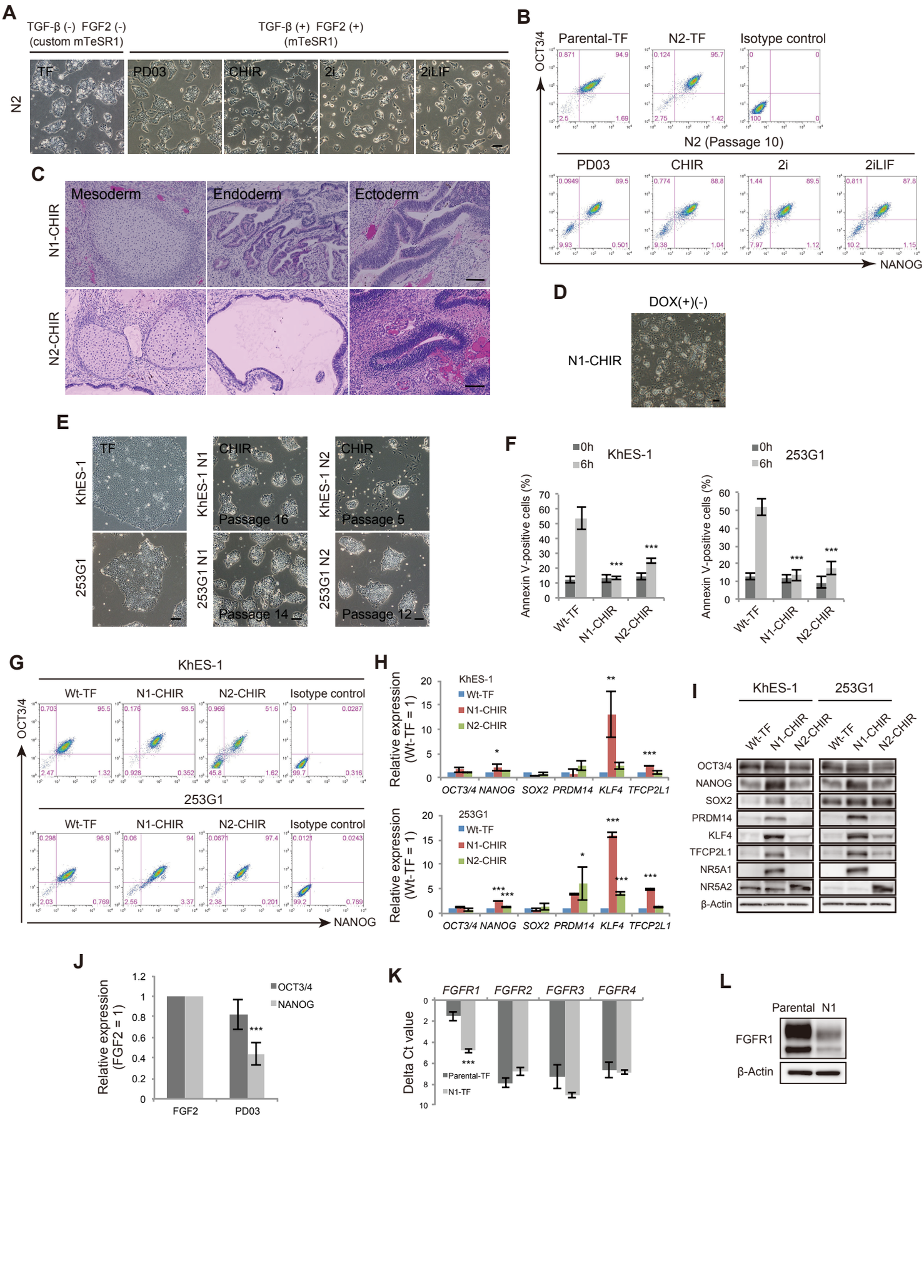

A
TGF-β (-) FGF2 (-)
(custom mTeSR1)
TGF-β (+) FGF2 (+)
(mTeSR1)
N2
TF
PD03
CHIR
2i
2iLIF
B
Parental-TF
N2-TF
Isotype control
OCT3/4
N2 (Passage 10)
PD03
CHIR
2i
2iLIF
NANOG
C
N1-CHIR
N2-CHIR
Mesoderm
Endoderm
Ectoderm
D
DOX(+)(-)
N1-CHIR
E
KhES-1
253G1
TF
KhES-1 N1
253G1 N1
CHIR
Passage 16
Passage 14
KhES-1 N2
253G1 N2
CHIR
Passage 5
Passage 12
F
0h
6h
KhES-1
253G1
Annexin V-positive cells (%)
Wt-TF
N1-CHIR
N2-CHIR
G
KhES-1
Wt-TF
N1-CHIR
N2-CHIR
Isotype control
OCT3/4
253G1
NANOG
H
KhES-1
Wt-TF
N1-CHIR
N2-CHIR
Relative expression
(Wt-TF = 1)
OCT3/4 NANOG SOX2 PRDM14 KLF4 TFCP2L1
253G1
I
KhES-1
253G1
Wt-TF
N1-CHIR
N2-CHIR
OCT3/4
NANOG
SOX2
PRDM14
KLF4
TFCP2L1
NR5A1
NR5A2
β-Actin
J
Relative expression
(FGF2 = 1)
OCT3/4
NANOG
FGF2
PD03
K
FGFR1 FGFR2 FGFR3 FGFR4
Delta Ct value
Parental-TF
N1-TF
L
Parental N1
FGFR1
β-Actin

Figure S3

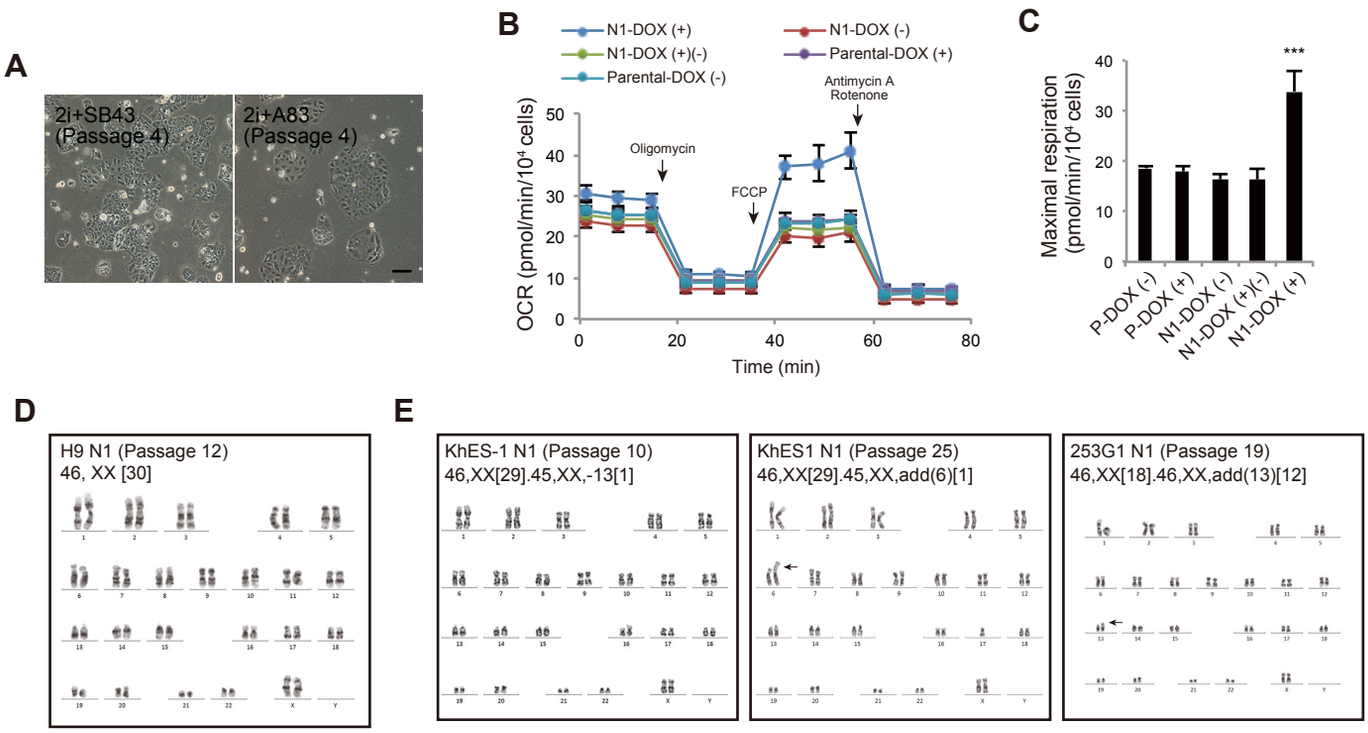

A
2i+SB43
(Passage 4)
2i+A83
(Passage 4)
B
N1-DOX (+)
N1-DOX (-)
N1-DOX (+)(-)
Parental-DOX (+)
Parental-DOX (-)
Antimycin A
Rotenone
Oligomycin
FCCP
OCR (pmol/min/10^4 cells)
Time (min)
C
***
Maximal respiration
(pmol/min/10^4 cells)
P-DOX (-)
P-DOX (+)
N1-DOX (-)
N1-DOX (+)(-)
N1-DOX (+)
D
H9 N1 (Passage 12)
46, XX [30]
E
KhES-1 N1 (Passage 10)
46,XX[29].45,XX,-13[1]
KhES1 N1 (Passage 25)
46,XX[29].45,XX,add(6)[1]
253G1 N1 (Passage 19)
46,XX[18].46,XX,add(13)[12]

Figure S4

A

Upregulated genes in parental-CHIR

Category	Term	Count	PValue
GOTERM_BP_FAT	GO:0030182~neuron differentiation	90	1.70E-29
	GO:0000904~cell morphogenesis involved in differentiation	56	9.79E-21
	GO:0007423~sensory organ development	54	1.41E-20
GOTERM_CC_FAT	GO:0031012~extracellular matrix	60	2.04E-17
	GO:0044421~extracellular region part	109	2.12E-16
	GO:0005578~proteinaceous extracellular matrix	56	3.19E-16
GOTERM_MF_FAT	GO:0043565~sequence-specific DNA binding	86	2.56E-17
	GO:0003700~transcription factor activity	107	2.01E-13
	GO:0030528~transcription regulator activity	142	2.25E-12
KEGG_PATHWAY	hsa05217:Basal cell carcinoma	21	2.18E-13
	hsa04340:Hedgehog signaling pathway	19	4.05E-11
	hsa04310:Wnt signaling pathway	25	1.44E-07

Upregulated genes in N1-CHIR

Category	Term	Count	PValue
GOTERM_BP_FAT	GO:0051240~positive regulation of multicellular organismal process	35	5.38E-08
	GO:0009611~response to wounding	57	8.02E-08
	GO:0006952~defense response	62	2.03E-07
GOTERM_CC_FAT	GO:0005576~extracellular region	175	1.94E-15
	GO:0044421~extracellular region part	100	4.44E-13
	GO:0005615~extracellular space	71	2.67E-09
GOTERM_MF_FAT	GO:0005125~cytokine activity	27	5.72E-06
	GO:0008083~growth factor activity	24	6.19E-06
	GO:0030246~carbohydrate binding	35	2.49E-04
KEGG_PATHWAY	hsa04060:Cytokine-cytokine receptor interaction	33	1.20.E-05
	hsa00980:Metabolism of xenobiotics by cytochrome P450	11	1.38.E-03
	hsa05200:Pathways in cancer	30	6.54.E-03

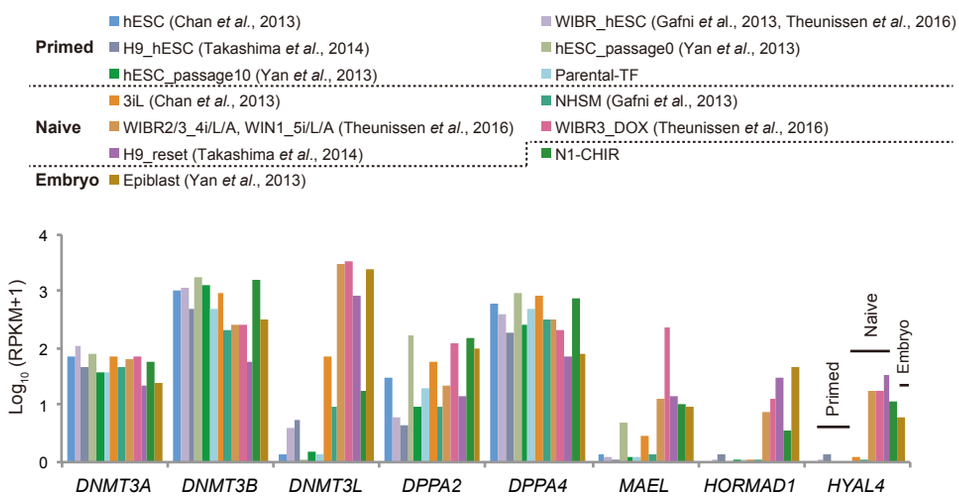
Downregulated genes in parental-CHIR

Category	Term	Count	PValue
GOTERM_BP_FAT	GO:0009611~response to wounding	66	4.03E-07
	GO:0042127~regulation of cell proliferation	85	3.07E-06
	GO:0050865~regulation of cell activation	29	7.45E-06
GOTERM_CC_FAT	GO:0044459~plasma membrane part	241	2.36E-16
	GO:0005887~integral to plasma membrane	140	1.69E-11
	GO:0031226~intrinsic to plasma membrane	140	8.47E-11
GOTERM_MF_FAT	GO:0046870~cadmium ion binding	8	5.11E-07
	GO:0005507~copper ion binding	15	1.27E-04
	GO:0008289~lipid binding	51	1.62E-04
KEGG_PATHWAY	hsa04060:Cytokine-cytokine receptor interaction	35	3.95.E-04
	hsa04514:Cell adhesion molecules (CAMs)	21	9.64.E-04
	hsa04080:Neuroactive ligand-receptor interaction	33	1.12.E-03

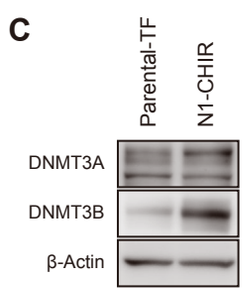
Downregulated genes in N1-CHIR

Category	Term	Count	PValue
GOTERM_BP_FAT	GO:0022610~biological adhesion	56	2.65E-08
	GO:0007155~cell adhesion	55	6.45E-08
	GO:0007507~heart development	24	2.57E-06
GOTERM_CC_FAT	GO:0044459~plasma membrane part	131	2.60E-07
	GO:0005578~proteinaceous extracellular matrix	33	9.72E-07
	GO:0031012~extracellular matrix	34	1.80E-06
GOTERM_MF_FAT	GO:0004714~transmembrane receptor protein tyrosine kinase activity	12	3.22E-05
	GO:0005509~calcium ion binding	55	4.94E-04
	GO:0019838~growth factor binding	13	5.08E-04
KEGG_PATHWAY	hsa04510:Focal adhesion	17	1.37.E-03
	hsa04514:Cell adhesion molecules (CAMs)	12	5.44.E-03
	hsa04512:ECM-receptor interaction	8	2.48.E-02

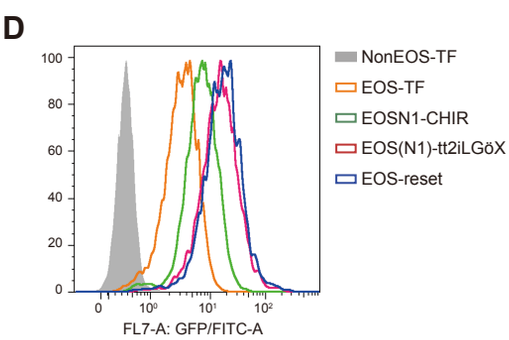
B



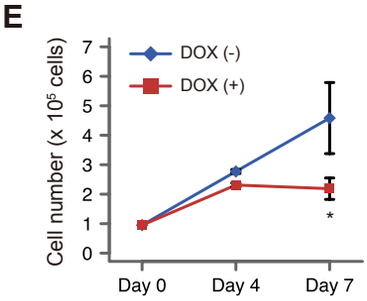
C



D



E



F

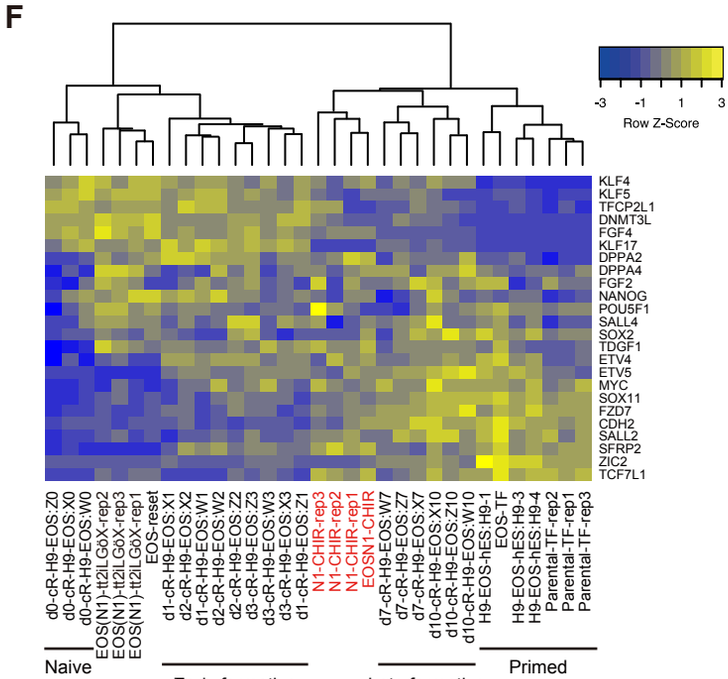


Figure S5

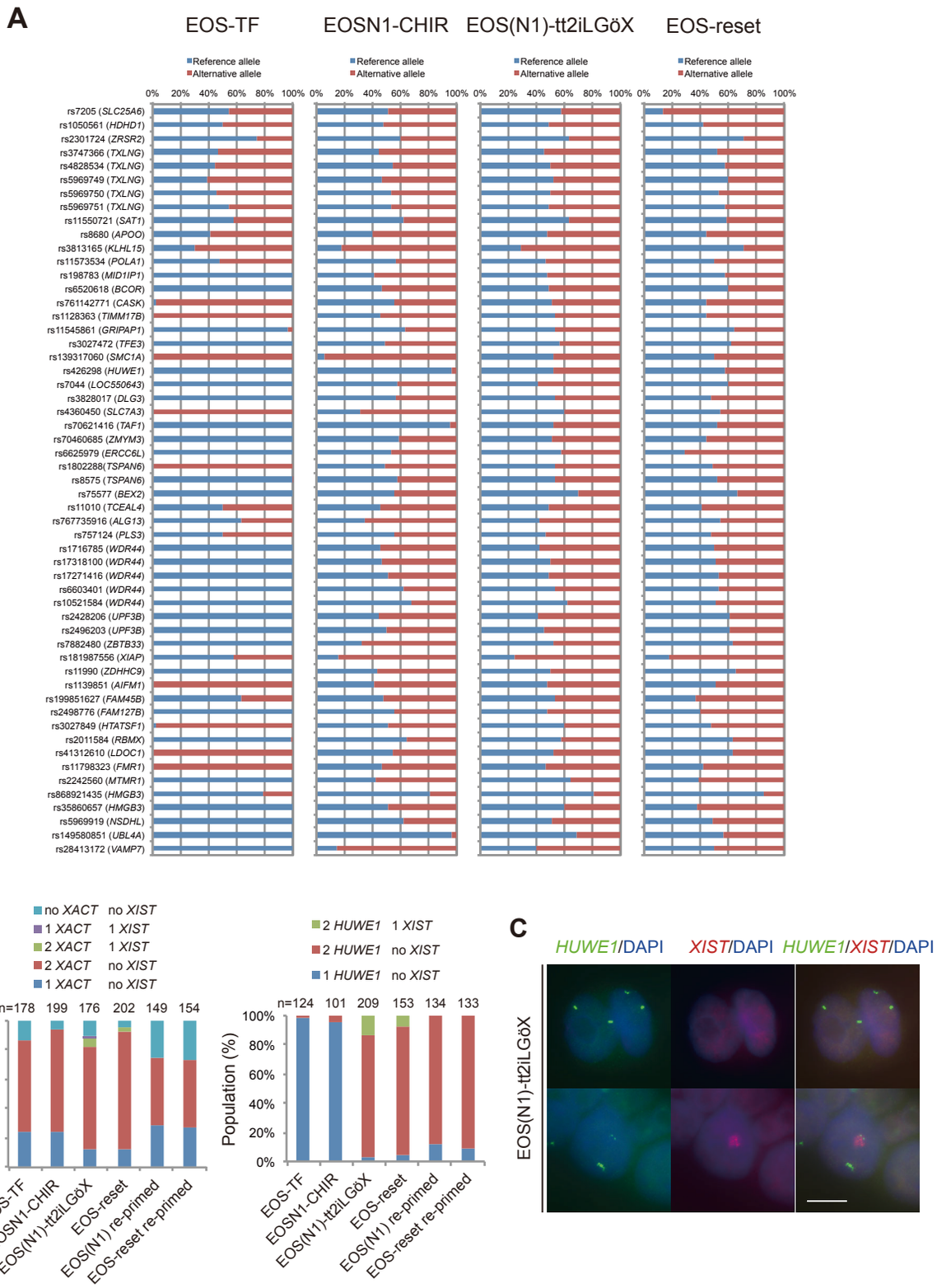
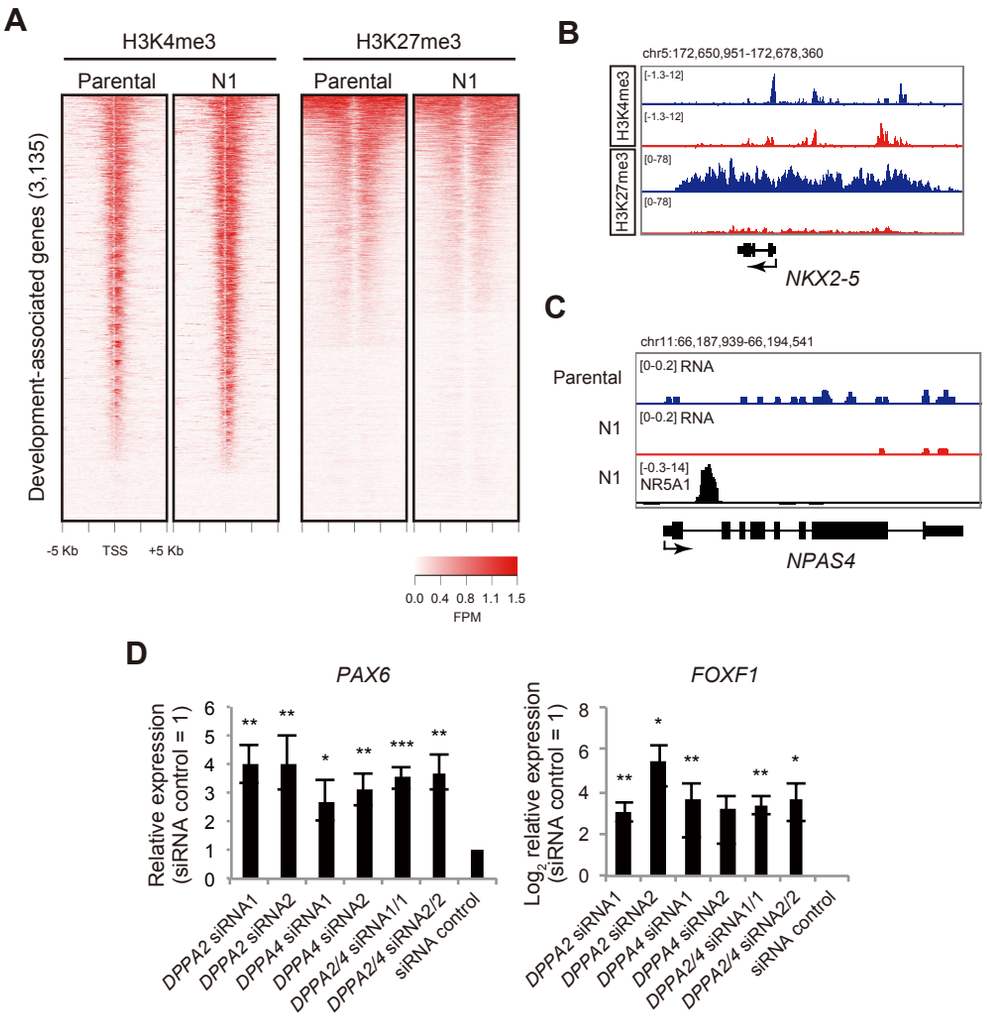

A
EOS-TF
EOSN1-CHIR
EOS(N1)-tt2iLGöX
EOS-reset
Reference allele
Alternative allele
0% 20% 40% 60% 80% 100%
rs7205 (SLC25A6)
rs1050561 (HDHD1)
rs2301724 (ZRSR2)
rs3747366 (TXLNG)
rs4828534 (TXLNG)
rs5969749 (TXLNG)
rs5969750 (TXLNG)
rs5969751 (TXLNG)
rs11550721 (SAT1)
rs8680 (APOO)
rs3813165 (KLHL15)
rs11573534 (POLA1)
rs198783 (MID1IP1)
rs6520618 (BCOR)
rs761142771 (CASK)
rs1128363 (TIMM17B)
rs11545861 (GRIPAP1)
rs3027472 (TFE3)
rs139317060 (SMC1A)
rs426298 (HUWE1)
rs7044 (LOC550643)
rs3828017 (DLG3)
rs4360450 (SLC7A3)
rs70621416 (TAF1)
rs70460685 (ZMYM3)
rs6625979 (ERCC6L)
rs1802288(TSPAN6)
rs8575 (TSPAN6)
rs75577 (BEX2)
rs11010 (TCEAL4)
rs767735916 (ALG13)
rs757124 (PLS3)
rs1716785 (WDR44)
rs17318100 (WDR44)
rs17271416 (WDR44)
rs6603401 (WDR44)
rs10521584 (WDR44)
rs2428206 (UPF3B)
rs2496203 (UPF3B)
rs7882480 (ZBTB33)
rs181987556 (XIAP)
rs11990 (ZDHHC9)
rs1139851 (AIFM1)
rs199851627 (FAM45B)
rs2498776 (FAM127B)
rs3027849 (HTATSF1)
rs2011584 (RBMX)
rs41312610 (LDOC1)
rs11798323 (FMR1)
rs2242560 (MTMR1)
rs868921435 (HMGB3)
rs35860657 (HMGB3)
rs5969919 (NSDHL)
rs149580851 (UBL4A)
rs28413172 (VAMP7)
B
no XACT no XIST
1 XACT 1 XIST
2 XACT 1 XIST
2 XACT no XIST
1 XACT no XIST
n=178 199 176 202 149 154
Population (%)
EOS-TF
EOSN1-CHIR
EOS(N1)-tt2iLGöX
EOS-reset
EOS(N1) re-primed
EOS-reset re-primed
2 HUWE1 1 XIST
2 HUWE1 no XIST
1 HUWE1 no XIST
n=124 101 209 153 134 133
C
HUWE1/DAPI
XIST/DAPI
HUWE1/XIST/DAPI
EOS(N1)-tt2iLGöX

Figure S6


A
H3K4me3
H3K27me3
Parental
N1
Parental
N1
Development-associated genes (3,135)
-5 Kb
TSS
+5 Kb
0.0 0.4 0.8 1.1 1.5
FPM
B
chr5:172,650,951-172,678,360
H3K4me3
[-1.3-12]
[-1.3-12]
H3K27me3
[0-78]
[0-78]
NKX2-5
C
chr11:66,187,939-66,194,541
Parental
[0-0.2] RNA
N1
[0-0.2] RNA
N1
[-0.3-14]
NR5A1
NPAS4
D
PAX6
Relative expression (siRNA control = 1)
FOXF1
Log2 relative expression (siRNA control = 1)
DPPA2 siRNA1
DPPA2 siRNA2
DPPA4 siRNA1
DPPA4 siRNA2
DPPA2/4 siRNA1/1
DPPA2/4 siRNA2/2
siRNA control

Supplemental Figure Legends

Figure S1. Analysis of NR5A-induced cells under the TF condition (related to Figure 1).

(A) Scheme of the gene expression system used to screen the indicated transcription factors.

(B) Expression of *NANOG* and *KLF4* in control cells and transfectants. Data were analyzed by the comparative ΔCt method. Expression in control cells transfected with empty vector was set to 1. Endo and exo represent endogenous and exogenous expression, respectively. ***$p < 0.001$ (Dunnett's test).

(C) Scheme of the doxycycline (DOX)-inducible gene expression system used in this study. In the absence of DOX [DOX (–)], a fusion protein of tetracycline (tet) repressor (tTR) and a Krüppel-associated box (KRAB) domain binds to tet operator (*tetO*) sequences and suppresses expression of NR5A. Addition of DOX [DOX (+)] blocks the binding of tTR-KRAB to *tetO*, resulting in transgene expression.

(D) Immunocytochemical staining of tet and pluripotency markers, such as OCT3/4, NANOG, and SOX2, in a parental clone expressing tTR-KRAB.

(E) NR5A expression analyzed by qPCR. Data were analyzed by the comparative ΔCt method. Parental cell expression was set to 1. NR5A1- and NR5A2-expressing cells are referred to as N1 and N2 cells, respectively.

(F) NR5A expression analyzed by western blotting. β-Actin was used as a loading control.

(G) Percentages of annexin V-positive cells before (0 h) and 12 h after single cell dissociation. Parental, N1, and N2 cells were cultured in the absence of DOX. The numbers of positive cells were analyzed by flow cytometry. Statistical significance vs. parental cells at 0 or 12 h (Dunnett's test). n.s., not significant.

(H) Flow cytometric analysis of pluripotency markers in parental, N1, and N2 cells.

(I) Expression of the indicated genes analyzed by qPCR. Data were analyzed by the comparative ΔCt method. Parental cell expression was set to 1. N1-DOX (+)(–) represents cells after withdrawal of DOX.

(J) Representative images of the NR5A1 transfectant after DOX withdrawal. Cells were seeded in the absence or presence of Y-27632 (+).

(K) Percentages of annexin V-positive cells before (0 h) and 12 h after single cell dissociation. ***$p < 0.001$ (Student's *t*-test).

Data are represented as mean ± SD (B, E, G, and I) or SEM (G and K) of three biological replicates. Scale bars = 100 μm.

Figure S2. NR5A-induced cells maintain pluripotency in the presence of a GSK3 inhibitor (related to Figure 2).

(A) Representative images of N2 cells cultured under TF or 2i conditions.

(B) Flow cytometric analysis of pluripotency markers in parental and N2 cells under each condition.

(C) Teratoma formation of N1 and N2 cells after passages 10.

(D) Representative image of N1 cells after withdrawal of DOX under the CHIR condition.

(E) Representative images of control cells and NR5A transfectants derived from KhES-1 and 253G1 cells. The cells were cultured under TF or CHIR conditions.

(F) Percentages of annexin V-positive cells among control cells and NR5A transfectants derived from KhES-1 and 253G1 cells 6 h after single cell dissociation. ***$p < 0.001$ vs. wild type cells (Dunnett's test).

(G) Flow cytometric analysis of pluripotency markers in control cells and NR5A transfectants derived from KhES-1 and 253G1 under each condition.

(H and I) Expression of transcripts and proteins associated with the naive pluripotent state in control cells and NR5A transfectants derived from KhES-1 and 253G1. Data were analyzed by the comparative ΔCt method. Expression in unmodified hPSCs cultured in the TF condition (Wt) was set to 1. β-Actin was used as a loading control.

(J) Expression of *POU5F1* and *NANOG* in parental cells 24 h after PD03 treatment. Data were analyzed by the comparative ΔCt method. Expression in the FGF condition was set to 1.

(K) Expression of *FGFRs* in parental and NR5A1-induced cells cultured under the TF condition. Data analyzed by qPCR are shown as ΔCt values.

(L) Western blotting of FGFR1 in parental and NR5A1-induced cells cultured under the TF condition. β-Actin was used as a loading control.

Scale bars = 100 μm. Data are represented as mean ± SD (G–K) or SEM (F) of three biological replicates. ***$p < 0.001$ (Student's *t*-test; J, K).

Figure S3. Characterizations of NR5A-induced cells cultured in the presence of a GSK3 inhibitor (related to Figure 2)

(A) Representative images of NR5A1-induced cells cultured with inhibitors of TGF-β/activin/nodal signaling.

(B and C) Oxygen consumption rate (OCR) assay under mitochondrial stress. N1-DOX (+) and N1-DOX (–) cells were cultured under CHIR or TF conditions. N1-DOX (+)(–) represents cells after withdrawal of DOX. Data are represented as mean ± SD of five biological replicates ***$p < 0.001$ vs. P-DOX (–) cells (Dunnett's test).

(D and E) Karyotype analysis of H9 N1 (M), KhES-1 N1 (N, left and middle panels), and 253G1 N1 (N, right panel) cells cultured with CHIR. Thirty metaphases from cells maintained by single cell passaging for the indicated passage numbers were analyzed.

Scale bars = 100 μm.

Figure S4. Global analysis of gene expression under NR5A1 regulation (related to Figure 3).

(A) Gene Ontology (GO) analysis of up- and downregulated genes in CHIR-treated parental and NR5A1-induced cells.

(B) RNA-seq analysis of the indicated genes. Data are shown as the $\log_{10}$ (RPKM+1). RPKM, reads per kilobase of exon per million mapped reads.

(C) DNMT3A and DNMT3B expression analyzed by western blotting. β-Actin was used as a loading control.

(D) Flow cytometric analysis of GFP expression under each condition.

(E) Proliferation analysis of EOS-N1 cells in the absence or presence of DOX under the tt2iLGöX condition. Data are represented as mean ± SD of three biological replicates. *$p < 0.005$ (Student's *t*-test).

(F) Heat map of genes selected from RNA-seq data of this study and previously published study relating to formative transition (Rostovskaya et al., 2019).

Figure S5. Conversion of NR5A1-induced cells to the full naive pluripotent state under the tt2iLGöX condition (related to Figure 4).

(A) Single nucleotide polymorphism (SNP)-based allelic expression patterns of the two X chromosomes. The allelic expression patterns of representative genes analyzed in RNA-seq data are shown with their reference SNP IDs (rs).

(B) Percentages of cells bearing different X chromosome statuses. Data were analyzed by RNA- fluorescence in situ hybridization (FISH) of *XACT*, *HUWE1*, and *XIST* in each condition.

(C) Representative RNA-FISH images of *HUWE1* and *XIST* in EOS(N1)-tt2iLGöX cells. Scale bars = 10 μm.

Figure S6. Chromatin immunoprecipitation sequencing (ChIP-seq) analysis of NR5A1 in CHIR-treated NR5A1-induced cells (related to Figure 5).

(A) Heat maps based on H3K4me3 and H3K27me3 signal enrichment around the transcriptional start sites (TSSs) of development-associated genes (Theunissen et al., 2014).

(B) NR5A1 signal distribution on the *NKX2-5* gene in NR5A1-induced cells. Data analyzed by RNA-seq are also shown.

(C) NR5A1 signal distribution on the *NPAS4* gene in NR5A1-induced cells. Data analyzed by RNA-seq are also shown.

(D) Depletion of *DPPA2* and *DPPA4* in NR5A1-induced cells by siRNA. Data were analyzed by the comparative ΔCt method. Expression in cells treated with control siRNA were set to 1. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (Student's *t*-test).

Supplemental Tables

Table S1. Primer sequences used in this study (related to the Experimental Procedures)

Gene	Forward	Reverse	Use
NR5A1	atttaaatatggactacaaggacgatgac	atttaaattcaagtctgcttggcttgca	Cloning
NR5A1	gaagcttgactattcgtacgacga	ggatcctcaagtctgcttggctt	3×FLAG
NR5A2	atttaaatctaagaatgtcttctaattcag	atttaaatttaaaccttatcgtcgtcatc	Cloning
NR5A2	caaggttaccaaacatatggccactttccta	taggaaagtggccatatgtttggtaaccttg	Mutagenesis
GAPDH	gaaggtgaaggtcggagtc	gaagatggtgatgggatttc	RT-qPCR
ACTB	catgtacgttgctatccaggc	ctccttaatgtcacgcacga	RT-qPCR
HPRT	tggtcaggcagtataatccaaaga	tcaaatccaacaaagtctggctta	RT-qPCR
RPL37A	ccaaacgtaccaagaaagtcgg	gcgtgctggctgatttcaa	RT-qPCR
OCT3/4	gaaggagaagctggagcaaa	catcggcctgtgtatatccc	RT-qPCR
NANOG	ctgctgagatgcctcacacg	tgcctttgggactggtgga	RT-qPCR
SOX2	ggcagctacagcatgatgcaggacc	tggtcatggagttgtactgcagg	RT-qPCR
PRDM14	tgagccttcaggtcacagag	atttcctatcgcccttgtcc	RT-qPCR
KLF4	gggcccaattacccatcctt	ctttggcttgggctcctctg	RT-qPCR
TFCP2L1	gctcttcaacgccatcaaa	caggggcactcgattctg	RT-qPCR
DPPA2	accctgaacaacggcaag	ttgcgtttcctcgaacatc	RT-qPCR
DPPA4	cctcctgggcgagaattt	gaccacacaccacctgacac	RT-qPCR
DPPA3	gaccaacaaacaaggagcctaag	agaaggatccatccattagaca	RT-qPCR
ZFP42	cagaacagaagaggccttcac	tctgagtaagctgtcttcagcaa	RT-qPCR
NR5A1	gcaggtgcatggtcttcaa	agttctgcagcagcgtcat	RT-qPCR
NR5A2	ccgacaagtggtacatggaa	tccggcttgtgatgctatta	RT-qPCR
FGFR1	actccggcctctatgcttg	aggaggggagagcatctga	RT-qPCR
FGFR2	cctgccaaaacagcaagc	aagacccctatgcagtaaatgg	RT-qPCR
FGFR3	tcctcgggagatgacgaa	cagcagcttcttgtccatcc	RT-qPCR
FGFR4	gaggggccgcctagagatt	caggacgatcatggagcct	RT-qPCR
FGF4	gcaagggcaagctctatgg	tgtaggactcgtaggcgttgt	RT-qPCR
PAX6	gcttcaccatggcaaataacc	ggcagcatgcaggagtatga	RT-qPCR
FOXF1	acagcggcgcctcttatatc	ctcctttcggtcacacatgc	RT-qPCR

Genes	Forward	Reverse		References
NANOG	tggttaggttggttttaaatttttg	aacccacccttataaattctcaatta	Bisulfite-PCR	(Takahashi et al., 2007)
PEG10	ggtgtaatttatataaggtttatagtttg	aacaaaaaaaataaaatcccacac	Bisulfite-PCR	(Kim et al., 2007)
SNURF/ SNRPN	taggttgtttttgagagaagttat	aaaaaaactaaaacccctacactac	Bisulfite-PCR	(Kim et al., 2007)

RT-qPCR, quantitative reverse transcription polymerase chain reaction.

Table S2. Antibodies used in this study

Antibody	Company	Catalog Number	Dilution
OCT3/4 (C-10)	Santa Cruz Biotechnology	sc-5279	1:300 for immunocytochemistry 1:1000 for western blotting
NANOG	Cell Signaling Technology	4903	1:300 for immunocytochemistry 1:1000 for western blotting
KLF4 (H-180)	Santa Cruz Biotechnology	sc-20691	1:300 for immunocytochemistry 1:1000 for western blotting
SOX2	R&D Systems	245610	1:500 for immunocytochemistry 1:1000 for western blotting
PRDM14	Abgent	AP1214a	1:500 for western blotting
Tet Repressor	Molecular Biotechnology	TET01	1:200 for immunocytochemistry
DDDDK-tag (FLA-1)	MBL	M185-3L	1:5000 for immunocytochemistry 1:10000 for western blotting ChIP
NR5A1	Perseus Proteomics	PP-N1665-00	1:1000 for western blotting
NR5A2	Perseus Proteomics	PP-H2325-00	1:1000 for western blotting
TFCP2L1	R&D Systems	AF5726	1:1000 for western blotting
β-Actin	Sigma	AC-15	1:5000 for western blotting
Phosho-MEK1/2 (Ser217/221)	Cell Signaling Technology	9121	1:1000 for western blotting
MEK1/2 (L38C12)	Cell Signaling Technology	4694	1:1000 for western blotting
Phospho-p44/42 MAPK (Erk1/2)	Cell Signaling Technology	9106	1:1000 for western blotting

(Thr202/Tyr204) (E10)			
p44/42 MAPK (Erk1/2)	Cell Signaling Technology	9102	1:1000 for western blotting
FGFR1	Abcam	Ab76464	1:500 for western blotting
DNMT3A	Santa Cruz Biotechnology	sc-365769	1:300 for western blotting
DNMT3B	Santa Cruz Biotechnology	sc-376043	1:300 for western blotting
H3K4me3	Abcam	ab8580	ChIP
H3K27me3	Merck Millipore	07-449	ChIP
Alexa Fluor-488 conjugate secondary antibody	Thermo Fisher Scientific		1:500 for immunocytochemistry
Alexa Fluor-546 conjugate secondary antibody	Thermo Fisher Scientific		1:500 for immunocytochemistry
Alexa Fluor-555 conjugate secondary antibody	Thermo Fisher Scientific		1:500 for immunocytochemistry
Goat anti-mouse IgG-HRP	Santa Cruz Biotechnology	sc-2055	1:10000 for western blotting
Goat anti-rabbit IgG-HRP	Agilent	P044801-2	1:10000 for western blotting
Donkey anti-goat IgG-HRP	Santa Cruz Biotechnology	sc-2056	1:10000 for western blotting

HRP, horseradish peroxidase.

Table S3. Microarray analysis of parental and NR5A1-induced cells cultured under TF and CHIR conditions (related to Figure 3)

Comparisons between two samples are as follows,

Comparison 1: Parental-DOX (+) and NR5A1-DOX (+)

Comparison 2: Parental-DOX (+)_CHIR and NR5A1-DOX (+)_CHIR

Comparison 3: NR5A1-DOX (+) and NR5A1-DOX (+)_CHIR

Comparison 4: Parental-DOX (+) and NR5A1-DOX (+)_CHIR

Comparison 5: Parental-DOX (+) and Parental-DOX (+)_CHIR

Table S4. RNA-seq analysis in this study, in human naive-like cell lines (Gafni et al., 2013; Chan et al., 2013; Takashima et al., 2014; Theunissen et al., 2014), and in human embryos (Yan et al., 2013) (related to Figures 3 and 4)

Table S5. RNA-seq analysis of transposable elements (TEs) in parental and NR5A1-induced cells cultured under TF and CHIR conditions (related to Figure 3)

Raw count data obtained from featureCounts and the results of DESeq2 analysis [adjusted p-value < 0.05, $\log_2$ fold change (FC) > 1.5] are shown.

Table S6. RNA-seq analysis of TEs in all samples of this study (related to Figure 4)

Raw count data obtained from featureCounts are shown. The DESeq2 analysis was performed in parental-TF, N1-CHIR, and EOS(N1)-tt2iLGöX cells; false discovery rate (FDR) < 0.05.

Supplemental Experimental Procedures

Plasmid construction

For transcription factor screening, the pLVSIN-EF1α Pur vector was purchased from TaKaRa Bio. After removing the EF1α promoter from the vector, the CAG promoter was inserted to construct the pLVSIN-CAG Pur vector. POU5F1, SOX2, NANOG, and KLF4 cDNAs were obtained from the pDON5 OKSLN vector (TaKaRa Bio) by PCR. The PRDM14 cDNA (Tsuneyoshi et al., 2008) was cloned by PCR, and the cDNAs of TBX3 (IRAL042P10), KLF2 (SC127849), ESRRB (SC327896), NR5A1 (RDB06299), and NR5A2 (RC213887) were purchased from the RIKEN BRC DNA Bank (TBX3 and NR5A1) and OriGENE Technologies (KLF2, ESRRB, and NR5A2). Site-directed mutagenesis of KLF2 and NR5A2 was performed by PCR using primers matching the reference sequences (NP_057354 and NP_003813, respectively).

Cell culture

All hESC lines were used according to the Guidelines on the Distribution and Utilization of Human Embryonic Stem Cells of the Ministry of Education, Culture, Sports, Science, and Technology of Japan. The hESC lines KhES-1 (Kyoto University, Kyoto, Japan) (Suemori et al., 2006) and H9 (WA09; WiCell Research Institute, Madison, WI) (Thomson et al., 1998) and the hiPSC line 253G1 (Kyoto University) (Nakagawa et al., 2008; Takahashi et al., 2007) were maintained on mitomycin-C-treated MEFs in conventional hPSC medium consisting of Dulbecco's modified Eagle's medium (DMEM)/F12 (D6421; Sigma-Aldrich, St. Louis, MO), 20% KSR (Thermo Fisher Scientific, Waltham, MA), 2 mM L-glutamine (G7513; Sigma-Aldrich), 1X non-essential amino acids (M7145; Sigma-Aldrich), 0.1 mM 2-mercaptoethanol, and 5 ng/mL recombinant human FGF2 (Wako, Osaka, Japan).

For feeder-free culture, small clumps of cells were transferred into Matrigel Matrix (growth factor reduced; Corning Inc.)-coated dishes and cultured in mTeSR1 medium (ST-05850; Stemcell Technologies). The dishes were coated with 25 μg/cm^2 Matrigel for at least 1 h at room temperature (RT) before cell seeding. Parental cells were passaged every 3–4 days using 2 mg/mL Dispase (Thermo Fisher Scientific). To culture NR5A transfectants, 0.1 μg/mL DOX (MP Biomedicals) was added, and the medium was changed daily. To select for transgene-expressing cells, puromycin (Sigma-Aldrich) and Zeocin (Thermo Fisher Scientific) were added at 1 and 5 μg/mL, respectively, after at least 5 days of DOX treatment. After cultivation with DOX and antibiotics for 10 days, single cell dissociation was initiated. The cells were incubated with 4 mM ethylenediaminetetraacetic acid (EDTA)/phosphate buffered saline (PBS) for 3–5 min and then with TrypLE Select for 5–10 min at 37°C (the EDTA treatment can be omitted if desired). The dissociated cells were seeded in Matrigel-coated dishes at 1.25–2.5 × 10^4 cells/cm^2 and passaged every 4–5 days.

For cell culture with small molecules, custom mTeSR1 without FGF2 and TGF-β was purchased from Stem Cell Technologies (ST-05896). Note that the previous version of the medium contained three components, LiCl, gamma-aminobutyric acid, and pipecolic acid, but the current version does not. When this study was initiated, only the previous version was available. We have confirmed that the current version is also suitable for this experiment. Small molecules and growth factors were used at the following concentrations, 3 μM CHIR99021 (CHIR; Axon), 1 μM

PD0325901 (PD03; Axon), 2 μM SB431542 (Tocris), 0.25 μM A83-01 (Tocris), and 10 ng/mL recombinant human leukemia inhibitory factor (LIF; Merck Millipore). Induction of naive cells from EOS cells was implemented according to a previously reported chemical resetting protocol (Guo et al., 2017). Briefly, cells cultured on mitomycin-C-treated MEFs were treated with 1 mM valproic acid sodium salt (VPA; Sigma-Aldrich), 1 μM PD03, and 10 ng/mL LIF for the first 3 days in N2B27 medium containing DMEM/F12, Neurobasal (Thermo Fisher Scientific), 0.5× B27 (Thermo Fisher Scientific), 0.5× N2 (Thermo Fisher Scientific), 2 mM L-glutamine, and 0.1 mM 2-mercaptoethanol. The cells were then cultured with 1 μM PD03, 2 μM Gö6983 (Tocris), 2 μM XAV939 (Sigma-Aldrich), and 10 ng/mL LIF for 6 days. For feeder-free culture, single dissociated cells were transferred to tt2iLGö plus XAV939 (tt2iLGöX) medium (0.3 μM CHIR). Geltrex (1 μl/cm^2; Thermo Fisher Scientific) and 10 μM Y-27632 were added to the dishes at each routine passaging. To induce naive cells from EOS-N1 cells, cells cultured under the CHIR condition were transferred to tt2iLGöX medium and cultivated as described above. Experiments were performed in a 5% CO_2 atmosphere under TF and CHIR conditions or in 7% CO_2 and 5% O_2 under the naive condition.

Lentiviral infection

The Lenti-X HTX Packaging System (TaKaRa Bio) was used to transfect pLVSIN-CAG Pur vectors carrying each cDNA, according to manufacturer's instructions. The pMD2.G and psPAX2 vectors were obtained from Addgene (plasmids 12259 and 12260, respectively) to generate vesicular stomatitis virus G-pseudotyped viral particles. Either pLV-tTR-KRAB-IRES-Neo or pLVCT-NR5A-IRES-Puro was co-transfected with pMD2.G and psPAX2 into Lenti-X 293T cells (TaKaRa Bio) using Lipofectamine 2000 Reagent (Thermo Fisher Scientific), according to manufacturers' instructions. The collected virus-containing supernatants were centrifuged and then filtered through a 0.45-μm polyvinylidene fluoride filter (Merck Millipore) to remove cellular debris. The viral stocks were concentrated using a Lenti-X Concentrator (TaKaRa Bio). After centrifugation, the pellets were resuspended in mouse embryonic fibroblast (MEF)-conditioned medium (CM). One day before viral infection, human embryonic stem cells (hESCs) cultured in MEF-CM were dissociated into single cells with TrypLE Select (Thermo Fisher Scientific) and transferred into Matrigel-coated 24-well plates at 5 × 10^4 cells/cm^2 with 10 μM Y-27632 (Wako). The next day, the culture medium was exchanged with the concentrated viral suspension in the presence of 10 μg/mL polybrene. After 24 h of culture, the medium was replaced with fresh MEF-CM, and the cells were maintained on MEFs in CM for human pluripotent stem cell (hPSC) culture. To clone parental cells, tTR-KRAB-Neo-expressing cells were selected on Neo-resistant MEFs with 50 μg/mL G418 (Thermo Fisher Scientific) and seeded at a density of 5,000 cells/60-mm Neo-MEF-containing plate with Y-27632. Formed colonies were isolated and used to examine tet expression by immunocytochemistry.

EOS-GFP transfection

The pPB-EOS-GFP-IP (Plasmid 60439; (Guo et al., 2017; Takashima et al., 2014) and pCMV-hyPBase (Yusa et al., 2011) vectors were obtained from Addgene and the Wellcome Trust Sanger Institute, respectively. Plasmids were transfected into H9 parental cells by electroporation. Puromycin was used to select for GFP-expressing cells at 0.5

μg/mL for EOS and EOS-N1 cells and 1 μg/mL for EOS-reset and EOS(N1) cells.

siRNA transfection

Parental and NR5A1-induced cells were seeded in Matrigel-coated dishes at 5×10^4 and 2.5×10^4 cells/cm^2, respectively. Parental cells were cultured in mTeSR1 in the presence of 10 μM Y-27632 and 0.1 μg/mL DOX. NR5A1-induced cells were cultivated in custom mTeSR1 containing 3 μM CHIR99021 and DOX. Either 10 or 20 nM siRNA was transfected using Lipofectamine RNAiMAX Transfection Reagent (Thermo Fisher Scientific) 24 and 48 h after cell seeding. Cells were treated with lysis buffer ussing a RNeasy Micro Kit (Qiagen) for RNA extraction 24 h after the second transfection. Control (cat. no. 1027286 and 1022076), *DPPA2* (cat. no. SI05171740 and SI04349961), and *DPPA4*, (cat. no. SI00373282 and SI03223024) siRNAs were purchased from Qiagen.

Cell growth rate determination and cell cycle analysis

To analyze the cell growth rate, dissociated cells were seeded on Matrigel-coated 24-well plates at 2.5×10^4 cells/cm^2. Parental cells were cultured in mTeSR1 with 10 μM Y-27632. NR5A1-induced cells were cultured in mTeSR1 or custom mTeSR1 with 2i and LIF in the presence of DOX. Cell viabilities were analyzed every 24 h with acridine orange and propidium iodide staining using a NucleoCounter NC-3000 (ChemoMetec). The doubling time (DT) was calculated according to the American Type Culture Collection Animal Cell Culture Guide (https://www.atcc.org/) using the equation DT = T In2/In (Xe/Xb), where T is the incubation time, Xb is the cell number at the beginning of the incubation time, and Xb is the number at the end of the incubation time. For cell cycle analysis, dissociated cells were seeded onto Matrigel-coated 12-well plates at 2.5×10^4 cells/cm^2 for 24 h. The medium was exchanged with prewarmed fresh medium, and the cells were treated with 10 μM Click-iT EdU (Thermo Fisher Scientific) for 1 h. After labeling, the collected cells were reacted with Alexa Fluor 647 azide and FxCycle Violet Stain (Thermo Fisher Scientific) for flow cytometric analysis. These procedures were performed as described in the instructions for the Click-iT Plus EdU Alexa Fluor 647 Flow Cytometry Kit (Thermo Fisher Scientific).

Teratoma formation assay

Animal experiments were conducted in compliance with the regulations of animal experimentation at Kyoto University with approval by the animal ethics committee. Cells were dissociated by treatment with TrypLE Select and then washed in PBS (-). The collected cell pellets were injected into the testes of 6–8-week-old SCID mice (n = 2 for each cell line). Teratomas were collected 2–3 months after injection. For hematoxylin and eosin staining, the tissues were fixed in Bouin's fixative, embedded in paraffin, and sectioned into 5-μm thick sections. The stained sections were observed using a BZ9000 all-in-one microscope (Keyence). Images were merged using Keyence Analysis Software (Keyence).

Karyotype analysis

Cells were treated with 0.5 μg/mL KaryoMAX Colcemid Solution in PBS (Thermo Fisher Scientific) for 2 h and then

collected by trypsinization. Dissociated cells were incubated in 0.075 M KCl for 12 min at 37°C and then fixed in Carnoy's fixative. After Giemsa staining, at least 30 metaphase spreads were counted.

Oxygen consumption test

Dissociated cells were seeded on a Matrigel-coated 96-well assay plate (Seahorse) at 4×10^4 cells/well 1 day before the assay. Parental cells were cultured in mTeSR1 with 0.1 μg/mL DOX and 10 μM Y-27632. NR5A1-induced cells were cultivated in custom mTeSR1 with DOX and 3 μM CHIR99021. In DOX (–), both parental and NR5A1-induced cells were cultured in mTeSR1 with 10 μM Y-27632. NR5A1-induced cells after withdrawal of DOX were cultured with mTeSR1 with 10 μM Y-27632. An XF Cell Mito Stress Test Kit and XFe96 Analyzer (Seahorse) were used for this experiment. Small molecules were prepared at the following concentrations, 2 μM oligomycin (Abcam), 0.25 μM FCCP (Sigma-Aldrich), 1 μM rotenone (Sigma-Aldrich), and 1 μM antimycin (Sigma-Aldrich) in assay medium consisting of DMEM (D5030; Sigma-Aldrich) with 13.7 mM D-glucose, 0.392 mM sodium pyruvate (Sigma-Aldrich), and 2.94 mM L-glutamine (Sigma-Aldrich). The medium was exchanged with assay medium 1 h before the assay, and the cells were cultured at 37°C at an atmospheric concentration of CO_2. The oxygen consumption rate (OCR) was analyzed using XFe Wave analysis software (Seahorse).

Bisulfite PCR and sequencing

Genomic DNA was extracted using a Gentra Puregene Cell Kit (Qiagen), according to manufacturer's instructions, and 1 μg of DNA was used for bisulfite conversion, followed by PCR amplification. The PCR products were cloned into pGEM®-T easy (Promega). At least 10 clones of each sample were sequenced with T7 universal primer. PCR primer sequences are listed in Table S1. Methylated and unmethylated CpGs were analyzed using the Quantification Tool for Methylation Analysis (Center for Developmental Biology, RIKEN, Japan; http://quma.cdb.riken.jp/index_j.html).

RNA extraction, cDNA synthesis, and real-time PCR analysis

Total RNA was extracted using RNeasy Mini and Micro Kits (Qiagen) according to manufacturer's instructions. Synthesis of cDNA was performed using an Omniscript Reverse Transcription kit (Qiagen) with random primers (TaKaRa Bio). Fast SYBR Green Master Mix (Thermo Fisher Scientific) was used for real-time PCR analysis using a StepOnePlus Real-Time PCR System (Thermo Fisher Scientific). *GAPDH*, *HPRT, ACTB*, and *RPL37A* expression was analyzed using the comparative ΔCt method (Livak and Schmittgen, 2001). Primer sequences used in this study are listed in Table S1.

RNA-fluorescence in situ hybridization (FISH)

RNA-FISH was performed as described previously (Vallot and Rougeulle, 2016). Briefly, cells seeded on Matrigel-coated glass coverslips were incubated in ice-cold CSK buffer (100 mM NaCl, 300 mM sucrose, 3 mM $MgCl_2$, and 10 mM PIPES; pH 6.8) containing 0.5% Triton X-100 and 2 mM Ribonucleoside Vanadyl Complex (VRC;

BioLabs) for 5 min on ice, fixed with 3% paraformaldehyde containing 2 mM VRC for 10 min at RT, and washed three times for 4 min in ice-cold 70% EtOH. After dehydration in 90% and 100% EtOH for 4 min each, the fixed cells were incubated with green 496- (Enzo Life Sciences) or Cy3-labeled (GE Healthcare) probes at 37°C overnight. The next day, the cells were washed twice for 10 min in 50% formamide/2X SSC at 42°C and then washed twice for 5 min in 2X SSC and mounted with Vectashield containing 4,6-diamidino-2-phenylindole (DAPI; Vector Laboratories). Fluorescence was observed under a Zeiss fluorescence microscope (Axio Imager. Z2; Carl Zeiss). The probes were prepared using a Nick Translation Kit (Abbott Molecular) with bacterial artificial chromosome DNAs for *XIST* (CH17-218B21), *XACT* (RP11-35D3), and *HUWE1* (RP11-579N19) obtained from the BACPAC Resources Center. The proportion of cells bearing mono- or bi-allelic expression was determined manually.

Immunocytochemistry

Cells were cultured on dishes for 2–3 days and then fixed in 4% paraformaldehyde for 20 min. After washing with PBS (-), the fixed cells were permeabilized with 0.2% Triton X-100 for 15 min and then blocked with 10% bovine serum albumin and 5% serum in PBS (-) for 30 min. Then, the cells were incubated with primary antibodies for 1 h at RT, following by washing with PBS (-). After incubation with secondary antibodies for 1 h, the washed samples were mounted on glass slides with Vectashield containing DAPI. Fluorescence was observed under the Zeiss fluorescence microscope. Antibodies used in this study are listed in Table S2.

Western blotting

Cells were lysed in radioimmunoprecipitation assay (RIPA) buffer [50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1% NP40, 0.5% sodium deoxycholate, 0.1% SDS, 2 mM EDTA, and proteinase inhibitor cocktail (PIC; Roche)]. Protein concentrations were determined using a BCA protein assay kit (Thermo Fisher Scientific). Proteins (5–20 μg) were separated by sodium dodecyl sulfate (SDS)-polyacrylamide electrophoresis, then transferred onto polyvinylidene fluoride membranes and incubated in Blocking One (Nacalai Tesque Inc., Kyoto, Japan) at RT for 1 h. Then, the membranes were incubated in Can Get Signal Solution 1 (Toyobo) with primary antibodies at 4°C overnight, following by washing with Tris-buffered saline containing 0.1% Tween 20. Treatment with the secondary antibodies was performed in Can Get Signal Solution 2 (Toyobo) for 1 h at RT. The signals were detected by luminol-based enhanced chemiluminescence (SuperSignal West Dura Extended Duration Substrate; Thermo Fisher Scientific) using a CCD imaging system (LAS3000; FUJIFILM). Antibodies used in this study are listed in Table S2.

Flow cytometric analysis

Expression of pluripotency markers, such as OCT3/4, NANOG, and SOX2, was analyzed using a Human Pluripotent Stem Cell Transcription Factor Analysis Kit (BD Biosciences) and a BD FACS Aria II flow cytometer (BD Biosciences). Percentages of apoptotic populations induced by single cell dissociation were analyzed using an Annexin V Apoptosis Detection Kit I (BD Biosciences). Cell cycle analysis was performed using a Click-iT Plus EdU Alexa

Fluor 647 Flow Cytometry Kit (Thermo Fisher Scientific) and FxCycle Violet Stain. A MACS Quant Analyzer VYB (Miltenyi Biotec) was used for these experiments. All procedures followed the manufacturers' instructions. Flow cytometry data were analyzed using FlowJo software (version 9.9.5; FlowJo LLC).

Microarray data analysis and filter criteria

Total RNA was isolated from tissues using TRIzol Reagent (Invitrogen) and purified using SV Total RNA Isolation System (Promega). cRNA was amplified and labelled using the Low Input Quick Amp Labelling Kit (Agilent Technologies). cRNA was hybridized to a 60K 60-mer oligomicroarray (SurePrint G3 Human Gene Expression Microarray 8x60K v2; Agilent Technologies) according to manufacturer's instructions. Hybridized microarray slides were scanned using an Agilent scanner and the relative hybridization intensities and background hybridization values were calculated using Feature Extraction Software version 9.5.1.1 (Agilent Technologies). Scanned images were analyzed with Feature Extraction Software 9.5.1.1 (Agilent) using default parameters to obtain background subtracted and spatially detrended Processed Signal intensities. Raw signal intensities and flags for each probe were calculated from hybridization intensities (gProcessedSignal) and spot information (gIsSaturated) according to the procedures recommended by Agilent [Flag criteria on GeneSpring Software: Absent (A): "Feature is not positive and significant" and "Feature is not above background". Marginal (M): "Feature is not Uniform", "Feature is Saturated", and "Feature is a population outlier". Present (P): others]. The raw signal intensities of 12 samples were $\log_2$-transformed and normalized by the quantile algorithm using the "preprocessCore" library package (Bolstad et al., 2003) in Bioconductor software (Gentleman et al., 2004). Probes that called the "P" flag in at least one sample were selected, excluding lincRNA probes. We next applied the Linear Models for Microarray Analysis (limma) package (Smyth, 2017) of Bioconductor and obtained 24,169 genes. The criteria were a limma p-value < 0.05 and absolute log-fold-change ($|\log FC| > 1$ (non-log-transformed intensities and ratios are shown in Table S3). Heat maps were generated using MeV software (Saeed et al., 2003) and a principal component analysis (PCA) plot was generated using the ggplot2 package (https://github.com/tidyverse/ggplot2). The hierarchical clustering method was used to sort genes. The color indicates the distance from the median of each row (the distance metric was "Euclidean distance" and the linkage method was "average linkage clustering"). For GO analysis, genes up- or downregulated by > 4-fold were analyzed in the Database of Annotation, Visualization, and Integrated Discovery (DAVID; https://david.ncifcrf.gov/).

RNA extraction, library preparation, and RNA-seq

Total RNA was extracted using TRIzol-LS Reagent (Thermo Fisher Scientific) according to manufacturer's instructions. RNA integrity was evaluated using an Agilent 2100 Bioanalyzer with an Agilent 6000 RNA Pico Kit (Agilent). Libraries for sequencing were prepared using a TruSeq® Stranded mRNA Library Preparation Kit (Illumina) according to manufacturer's instructions. After adapter ligation, purified cDNA was amplified by 15 cycles of PCR. The libraries were evaluated using the Agilent 2100 Bioanalyzer with an Agilent High-Sensitivity DNA Kit (Agilent). Validated libraries were loaded into a MiSeq Reagent Kit v3 (150 cycles) at a final concentration of 10 pM. Sequencing was

performed using a MiSeq sequencer (Illumina) in paired-end mode.

RNA-seq data analysis

RNA-seq data from human naive-like cell lines and embryos were downloaded from the Gene Expression Omnibus (GEO; GSE83765 for 5i/L/A cells, GSE85689 for NHSM cells, GSE123055 for cells in formative transition, and GSE36552 for embryos) and ArrayExpress (E-MTAB-2857 for reset cells and E-MTAB-2031 for 3iL cells). To reduce technical variability in the data, downsampling was performed for datasets with more than 30 million reads (except GSE36552) using the seqtk tool (https://github.com/lh3/seqtk). All RNA-seq datasets included in this study were aligned using STAR (version 2.5.3) to the human genome (UCSC hg19) with 50 bp single-end mode and command options for 2-pass and unique mapping (Dobin et al., 2013). The count data were obtained using htseq-count in HTSeq (version 0.9.1; (Anders and Huber, 2010) with annotation from UCSC hg19. PCA plots were generated by plotPCA (ntop = 500) in the DESeq2 package with the function to calculate variance-stabilizing transformation (Love et al., 2014). For expression analysis, RPKM values were obtained using R (version 3.3.1). Genes were excluded when the sum of the values from all samples was zero (Table S4). Graphical representations were generated by the Integrative Genomics Viewer (Broad Institute) with coverage data normalized to total read counts.

To analyze transposable elements (TEs), RNA-seq datasets were aligned using TopHat (version 2.1.1; (Trapnell et al., 2009) to the human genome (UCSC hg19) with unique mapping and a 76 bp paired-end mode. We used featureCounts in Subread (version 1.5.2(Liao et al., 2014) to calculate read counts on TEs. RepeatMasker annotations were obtained from the University of California Santa Cruz Table Browser. Differentially expressed TEs covered by at least 10 reads between the parental and N1 cell (Table S5) or among parental, N1, and EOS(N1) cells (Table S6) were sorted by the DESeq2 package. The TEs of LTR7-HERVH, LTR5_Hs-HERVK, and SVA elements (adjusted $p < 0.05$, $\log_2$ FC > 1.5 in Fig. 3F, H; FDR < 0.05 in Fig. 4D) are shown in heat maps. All heat maps were generated using the heatmap.2 function of R. Allelic expression of X-linked genes was analyzed in terms of informative SNPs covered by at least 10 reads among the samples used for analysis. Expression of a gene was defined as biallelic when at least 25% of reads from the minor allele were observed, as previously described (Vallot et al., 2017).

ChIP

Parental and NR5A1-induced cells were treated with 4 mM EDTA and TrypLE Select, and then suspended in PBS (-) at 2.5×10^6 cells/mL. For H3K4me3 and H3K27me3 ChIP-seq, the collected cells were crosslinked with 0.5% formaldehyde for 10 min at RT and then quenched with 0.2 M glycine for 5 min. The crosslinked cells were washed with ice-cold PBS (-). To isolate the nuclear fraction, the cell pellet was suspended in ice-cold buffer 1 [0.3 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 15 mM Tris-HCl (pH 7.4), 0.5 mM DTT, and 1X PIC (Promega)], and then ice-cold buffer 2 (0.3 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 15 mM Tris-HCl pH 7.4, 0.5 mM DTT, and 1X PIC) was added to the cell suspension. After incubation for 10 min on ice, ice-cold buffer 3 (1.2 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 15 mM Tris-HCl pH 7.4,

0.5 mM DTT, and 1X PIC) was added, followed by centrifugation at 10,000 × *g* for 30 min at 4°C (himac CP80WX; Hitachi Koki). The collected nuclear extract was suspended in 1X micrococcal nuclease (MNase) digestion buffer (0.32 M sucrose, 50 mM Tris-HCl pH 7.4, 4 mM $MgCl_2$, 1 mM $CaCl_2$, and 1X PIC) and incubated with 75 U (per 1.5×10^7 cells) MNase for 15–20 min at 37°C. EDTA (5 mM) was added to the suspension to stop the reaction. After sonication and centrifugation, the supernatant (Sup 1) was frozen with liquid N_2. The cell pellet was dissolved in dialysis buffer (1 mM Tri-HCl pH. 7.4, 0.2 mM EDTA, and 1X PIC), and the suspension was transferred into a Slide-A-Lyzer Dialysis Cassette (3.5 K MWCO; Thermo Fisher Scientific) and dialyzed against the dialysis buffer overnight. After centrifugation, the supernatant (Sup 2) was frozen in liquid N_2. The pellet was sheared in an S220 Focused-ultrasonicator (Covaris) using the following settings, duty factor: 10%; peak incident power: 175 W; cycles per burst: 200; and time: 180 s. The sheared sample was centrifuged, and the supernatant (Sup 3) was frozen in liquid N_2 until use. For preclearing, a mixture of Sup 1–3 was suspended in ChIP dilution buffer (0.01% SDS, 1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.1) and 167 mM NaCl) and incubated with Protein A/G PLUS-Agarose Immunoprecipitation Reagent (Santa Cruz Biotechnology) at 4°C for 1 h. After centrifugation, the supernatant was transferred to a new tube. One percent of the input from the supernatant was stored at 4°C. The remainder was incubated with 10 μg antibodies (Table S2) and normal rabbit IgG (Cell Signaling Technology) per 1×10^7 cells at 4°C overnight. Then, Protein A/G PLUS-Agarose was added to the lysate, followed by incubation at 4°C for 4 h. After centrifugation, the pelleted beads were washed once with low salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, and 150 mM NaCl), once with high salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, and 500 mM NaCl), once with LiCl wash buffer (0.25 M LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA, and 10 mM Tris-HCl pH 8.1), and twice with TE buffer (10 mM Tris-HCl and 1 mM EDTA pH 8.0). The samples and inputs were then eluted with elution buffer (1% SDS and 100 mM $NaHCO_3$). DNA-protein crosslinks were reversed in 250 mM NaCl at 65°C overnight. The samples were then treated with RNase A and proteinase K. DNA was purified by phenol chloroform extraction and ethanol precipitation. The dsDNA was evaluated using an Agilent 2100 Bioanalyzer with the Agilent High-Sensitivity DNA Kit.

For NR5A1 ChIP-seq, the collected cells were crosslinked with 1% formaldehyde for 10 min at RT and quenched with 0.2 M glycine for 5 min. To isolate the nuclear fraction, the crosslinked sample was incubated in lysis buffer 1 (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP40, 0.25% Triton X-100, and 1X PIC) at 4°C for 10 min and then incubated in lysis buffer 2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, and 1X PIC) at 4°C for 10 min. The collected cell pellet was dissolved in lysis buffer 3 (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate, 0.5% n-lauroylsarcosine, and 1X PIC) and sheared in the Covaris S220 Focused-ultrasonicator using the following settings, duty factor: 10%; peak incident power: 140 W; cycles per burst: 200; and time: 80 s. For preclearing, the sample was incubated with Protein A/G PLUS-Agarose at 4°C for 1 h. After centrifugation, the supernatant was transferred to a new tube; one percent of the input supernatant was stored at 4°C, while the rest was incubated with 10 μg mouse anti-DDDDK-tag (Table S2) and normal mouse IgG (Santa Cruz Biotechnology) per 1×10^7 cells at 4°C overnight. Subsequent steps

were performed as described above.

Library preparation and ChIP-seq

Libraries for sequencing were prepared using a TruSeq® DNA Sample Preparation Kit (Illumina) according to manufacturer's instructions. After adapter ligation, DNA fragments of 300–400 bp were excised from 2% agarose gels. The purified DNA was amplified by 15–18 cycles of PCR. Libraries were evaluated using an Agilent 2100 Bioanalyzer with the Agilent High-Sensitivity DNA Kit. Validated libraries were loaded into the MiSeq Reagent Kit v3 (150 cycles) at a final concentration of 10 pM. Sequencing was performed in a MiSeq sequencer using the paired-end mode.

ChIP-seq data analysis

All ChIP-seq datasets were aligned to the human genome (UCSC hg19) with up to a single mismatch using Bowtie2 (version 2.2.8; (Langmead et al., 2009). We used the MACS2 (version 2.1.1; (Zhang et al., 2008) peak-finding algorithm to identify regions of ChIP-seq enrichment above the background. Graphical representations were generated using the $\log_{10}$ likelihood ratio. We used ngs.plot (Shen et al., 2014) to construct heat maps and the average profile (avgprof) of the ChIP-seq read density. Heat maps of H3K4me3 and H3K27me3 were prepared spanning the regions ± 5 kb up- and downstream of the TSSs of annotated polycomb-associated genes (3,135 genes; (Theunissen et al., 2014). We also calculated the avgprof of approximately ± 3 kb up- and downstream of polycomb-associated genes for H3K4me3 and H3K27me3. To confirm the bivalent association between H3K4me3 and H3K27me3, we calculated the avgprof of approximately ± 3 kb up- and downstream of peak centers detected by MACS2. To detect H3K4me3 and NR5A1 ChIP signal enrichments on HERVHs, we calculated the avgprof of approximately ± 10 kb up- and downstream of 1,225 HERVH regions (Wang et al., 2014). MEME-ChIP was used to search for the motif (Bailey et al., 2009).

Statistical analysis

Error bars in figures represent the standard deviation (s.d.) or standard error of the mean (s.e.m) of three biological replicates, unless otherwise indicated. Statistical significance was determined using the two-tailed unpaired Student's *t*-test for two group comparisons. Multiple group comparisons were performed using one-way analysis of variance with Dunnett's multiple comparison test. P values < 0.05 were considered statistically significant. No samples were excluded from the analyses, and no blinding or randomization was performed in this study.

Supplemental References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202–208.

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185–193.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J.*, et al.* (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. *5*, R80.

Guo, G., von Meyenn, F., Rostovskaya, M., Clarke, J., Dietmann, S., Baker, D., Sahakyan, A., Myers, S., Bertone, P., Reik, W.*, et al.* (2017). Epigenetic resetting of human pluripotency. Development *144*, 2748–2763.

Kim, K.P., Thurston, A., Mummery, C., Ward-van Oostwaard, D., Priddle, H., Allegrucci, C., Denning, C., and Young, L. (2007). Gene-specific vulnerability to imprinting variability in human embryonic stem cell lines. Genome Res. *17*, 1731–1742.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

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.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat. Biotechnol. *26*, 101–106.

Rostovskaya, M., Stirparo, G.G., and Smith, A. (2019). Capacitation of human naive pluripotent stem cells for multi-lineage differentiation. Development *146*.

Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M.*, et al.* (2003). TM4: a free, open-source system for microarray data management and analysis. Biotechniques *34*, 374–378.

Shen, L., Shao, N., Liu, X., and Nestler, E. (2014). ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics *15*, 284.

Smyth, G.K. (2017). Bioinformatics and Computational Biology Solutions Using R | Robert Gentleman | Springer.

Suemori, H., Yasuchika, K., Hasegawa, K., Fujioka, T., Tsuneyoshi, N., and Nakatsuji, N. (2006). Efficient establishment of human embryonic stem cell lines and long-term maintenance with stable karyotype by enzymatic bulk passage. Biochem. Biophys. Res. Commun. *345*, 926–932.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W.*, et*

al. (2014). Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. Cell *158*, 1254–1269.

Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L.*, et al.* (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell *15*, 471–487.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145–1147.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Tsuneyoshi, N., Sumi, T., Onda, H., Nojima, H., Nakatsuji, N., and Suemori, H. (2008). PRDM14 suppresses expression of differentiation marker genes in human embryonic stem cells. Biochem. Biophys. Res. Commun. *367*, 899–905.

Vallot, C., Patrat, C., Collier, A.J., Huret, C., Casanova, M., Liyakat Ali, T.M., Tosolini, M., Frydman, N., Heard, E., Rugg-Gunn, P.J.*, et al.* (2017). XACT Noncoding RNA Competes with XIST in the Control of X Chromosome Activity during Human Early Development. Cell Stem Cell *20*, 102–111.

Vallot, C., and Rougeulle, C. (2016). Single-cell Visualization of Chromosome Transcriptional Territories by RNA-paint. Bio. Protoc. *6*.

Wang, J., Xie, G., Singh, M., Ghanbarian, A.T., Rasko, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N.V.*, et al.* (2014). Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature *516*, 405–409.

Yusa, K., Zhou, L., Li, M.A., Bradley, A., and Craig, N.L. (2011). A hyperactive piggyBac transposase for mammalian applications. Proc. Natl. Acad. Sci. U S A *108*, 1531–1536.

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.