ARTICLES
https://doi.org/10.1038/s41592-018-0104-1
nature methods

Reduced MEK inhibition preserves genomic stability in naive human embryonic stem cells

Bruno Di Stefano[1,2,3,4,5], Mai Ueda[6], Shan Sabri[7], Justin Brumbaugh[1,2,3,4,5], Aaron J. Huebner[1,2,3,4,5], Anna Sahakyan[7], Kendell Clement[4,5,8], Katie J. Clowers[9], Alison R. Erickson[9], Keiko Shioda[3], Steven P. Gygi[9], Hongcang Gu[4,5,8], Toshi Shioda[3], Alexander Meissner[4,5,8], Yasuhiro Takashima[6], Kathrin Plath[7] and Konrad Hochedlinger[1,2,3,4,5]*

Human embryonic stem cells (hESCs) can be captured in a primed state in which they resemble the postimplantation epiblast, or in a naive state where they resemble the preimplantation epiblast. Naive-cell-specific culture conditions allow the study of preimplantation development ex vivo but reportedly lead to chromosomal abnormalities, which compromises their utility in research and potential therapeutic applications. Although MEK inhibition is essential for the naive state, here we show that reduced MEK inhibition facilitated the establishment and maintenance of naive hESCs that retained naive-cell-specific features, including global DNA hypomethylation, HERVK expression, and two active X chromosomes. We further show that hESCs cultured under these modified conditions proliferated more rapidly; accrued fewer chromosomal abnormalities; and displayed changes in the phosphorylation levels of MAPK components, regulators of DNA damage/repair, and cell cycle. We thus provide a simple modification to current methods that can enable robust growth and reduced genomic instability in naive hESCs.

hESCs self-renew indefinitely while retaining the capacity for multilineage differentiation, and therefore are a valuable tool for research and potential therapeutic applications. Conventional hESC culture conditions include the use of activin A and basic fibroblast growth factor (FGF) and capture pluripotent cells in a 'primed' pluripotent state in which they resemble the postimplantation epiblast[1,2]. Several laboratories have recently developed protocols to capture pluripotent cells in a more primitive or 'naive' state in which they resemble the preimplantation epiblast[3–5]. Naive stem cells offer a useful system for studies of preimplantation development[6,7] and are more efficient than primed stem cells at producing certain specialized cell types, such as primordial germ cells[8].

Culture conditions used to convert primed hESCs to a naive state typically include a combination of growth factors and small molecules that suppress specific protein kinases involved in differentiation, cell adhesion, and survival[3–5]. Two culture methods appear to be particularly effective[9]: the 't2iLGöY' protocol, which involves transient overexpression of the transcription factors KLF2 and NANOG in the presence of the MEK inhibitor (MEKi) PD0325901 and titrated amounts of GSK3β inhibitor (CHIR99021), supplemented with the PKC inhibitor Gö6983, ROCK inhibitor, and human leukemia inhibitory factor (LIF)[4,10]; and the '5i/LAF' protocol, which requires treatment of primed hESCs with inhibitors that target the GSK3β, ROCK, BRAF, MEK, and SRC kinases, in addition to human LIF, activin A, and FGF[5,7]. Inhibitors of the mitogen-activated protein kinase (MAPK/ERK) pathway are common to all currently available protocols.

Suppression of the MAPK pathway via the MEK1/2 inhibitor PD0325901 (PD03) has previously been shown to erode genomic imprints, lead to chromosomal abnormalities, and compromise the developmental potential of mouse ESCs[11,12]. However, titration of PD03 from 1 μM to 0.3–0.4 μM or replacement with an SRC inhibitor is reportedly sufficient to improve the epigenetic and genomic stability of mouse ESCs, as well as their in vitro and in vivo differentiation potential[11–13]. Considering the effects of MAPK inhibition on mouse ESCs, we examined the consequences of PD03 titration or replacement with alternative MEK inhibitors for the maintenance of naive hESCs cultured in 5i/LAF or t2iLGöY conditions.

Results

Reduced MEK inhibition maintains naive hESCs. We tested whether reduced MEK inhibition maintains naive colony morphology in the 5i/LAF culture system by titrating PD03 in the presence of constant amounts of inhibitors targeting BRAF, SRC, GSK3β, and ROCK, using WIBR3 hESCs carrying a naive-specific ΔPE OCT4–GFP reporter[5] (Fig. 1a). Specifically, we used 0.3 μM, 0.5 μM, 0.6 μM, and 0.8 μM PD03, as these concentrations are lower than the originally used 1 μM[5,7]. Complete omission of PD03 (4i/LAF condition) resulted in downregulation of GFP expression and a concomitant increase in the number of differentiated colonies after ~8 d, in agreement with previous observations[5] (Fig. 1a and Supplementary Fig. 1a). By contrast, hESCs cultured in 4i/LAF and supplemented with reduced amounts of PD03 showed robust GFP expression and undifferentiated colony morphology (Fig. 1a and Supplementary Fig. 1a). Of note, we were able to maintain undifferentiated colonies after continuous passaging of WIBR3 hESCs in 0.5 μM, 0.6 μM, or 0.8 μM PD03, whereas hESCs in 0.3 μM PD03 eventually lost

[1]Department of Molecular Biology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. [2]Center for Regenerative Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. [3]Cancer Center, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. [4]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. [5]Harvard Stem Cell Institute, Cambridge, MA, USA. [6]Center for iPS Cell Research and Application, Kyoto University, Kyoto, Japan. [7]David Geffen School of Medicine, Department of Biological Chemistry, Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, Jonsson Comprehensive Cancer Center, and Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA, USA. [8]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [9]Department of Cell Biology, Harvard Medical School, Boston, MA, USA. *e-mail: khochedlinger@mgh.harvard.edu

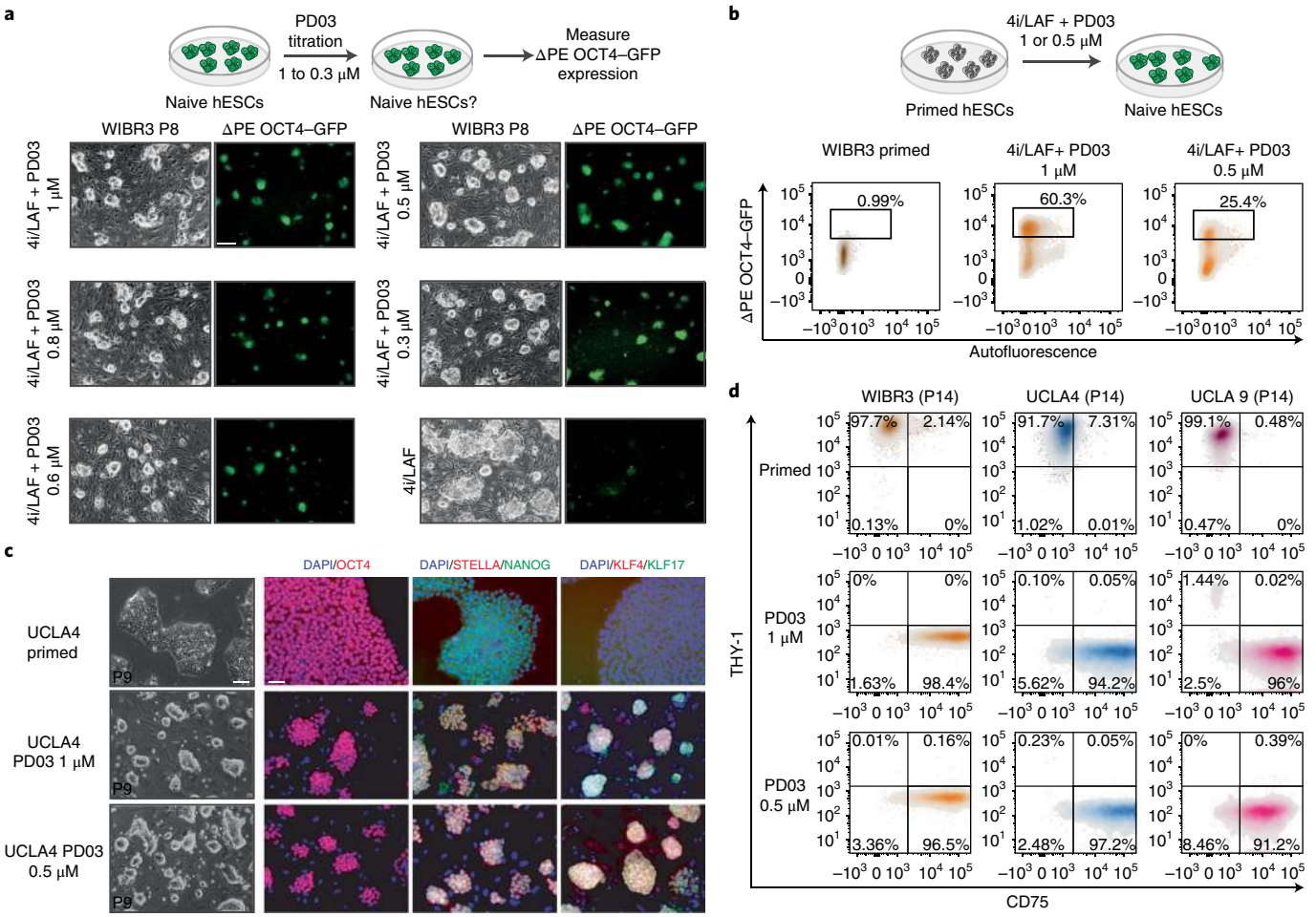


Fig. 1 | Attenuated MEK1/2 inhibition maintains naive pluripotency in hESCs. a, Top, PD03 titration strategy. Bottom, representative phase and fluorescence images of WIBR3 ΔPE OCT4–GFP hESCs at passage 8 (P8) grown in the indicated media. Scale bar, 250 μm. **b**, Flow cytometry analysis of the proportion of ΔPE OCT4–GFP$^+$ cells after reversion of WIBR3-primed hESCs to a naive state. **c**, Representative bright-field (left) and immunofluorescence (right) images for P9 UCLA4 hESCs cultured in the indicated media. Scale bars, 100 μm (bright-field) or 50 μm (immunofluorescence). **d**, Flow cytometry analysis of CD75 and THY-1 protein expression levels in hESC lines cultured as indicated.

their typical dome-shaped morphology (Supplementary Fig. 1b). These results show that reduction of the MEKi concentration from 1 μM to 0.5 μM preserved undifferentiated colony morphology and OCT4–GFP expression of hESCs cultured in 5i/LAF. We refer to this modified culture condition as modified 5i/LAF (m5i/LAF) in the remainder of the text.

We next exposed primed WIBR3 ΔPE OCT4–GFP hESCs to m5i/LAF to determine whether this culture condition also facilitates conversion to a naive state. Treatment of primed hESCs with either 5i/LAF or m5i/LAF for 10 d led to activation of the ΔPE OCT4–GFP reporter in 60.3% and 25.4% of cells, respectively (Fig. 1b). Despite the lower efficiency of conversion, we were able to establish stable, naive hESC lines from the m5i/LAF culture condition by serially passaging bulk cultures, which expressed the ΔPE OCT4–GFP reporter homogeneously (Supplementary Fig. 1c). We exposed six additional primed hESC lines—UCLA1, -3, -4, -5, -9, and -17[14] (Supplementary Fig. 1d)—to either 5i/LAF or m5i/LAF before assessing naive colony morphology and marker expression. We were again able to derive naive cell lines from UCLA4 and UCLA5 hESCs by serial bulk passaging in m5i/LAF. However, the establishment of naive hESC lines from UCLA1, -9, and -17 cells in m5i/LAF was less efficient and could be achieved only after picking and expansion of individual colonies from heterogeneous cultures (Supplementary Fig. 1d,e). Of note, UCLA3 hESCs were resistant to conversion to the naive state in m5i/LAF medium, which warrants future investigation. For all other examined cell lines, the generation of stable, naive hESC cultures exhibiting homogeneous colony morphology took approximately five to six passages in either 5i/LAF or m5i/LAF media, as confirmed by uniform expression of the pluripotency markers NANOG and OCT4 and the naive-cell markers KLF4, KLF17, and DPPA3 (STELLA) in two representative hESC lines (Fig. 1c and Supplementary Fig. 1f). We found that naive hESC lines downregulated the primed-cell-specific surface marker THY-1[15] and induced the naive-cell-specific surface marker CD75[15] to a similar degree when cultured in either 5i/LAF or m5i/LAF conditions (Fig. 1d). Finally, we confirmed the ability of reprimed m5i/LAF hESCs to access all three germ layers by inducing well-differentiated teratomas in mice (Supplementary Fig. 1g). We conclude that m5i/LAF treatment facilitates not only the maintenance of naive hESCs but also the conversion of primed hESCs to a naive state.

To assess whether titrated PD03 also maintains naive hESCs cultured in t2iLGöY, we initially attempted to transfer naive hESCs from 5i/LAF to the t2iLGöY medium[4,16], as to our knowledge this had not been reported before. Starting with WIBR3 hESCs stably grown in 5i/LAF (for more than eight passages), we switched the medium to t2iLGöY, which indeed allowed us to maintain naive hESCs for at least four passages without extinguishing the ΔPE OCT4–GFP reporter, reducing expression of naive-cell markers

(*KLF2*, *ZFP42* (*REX1*), *DPPA5*, *KLF4*, and *TFCP2L1*), or causing a loss of self-renewal potential (Supplementary Fig. 2a,b). However, reduction of the PD03 concentration from 1 μM to 0.5 μM in t2iLGöY media was insufficient to maintain the pluripotency of transferred hESCs (Supplementary Fig. 2c). Consistent with this, H9 hESCs carrying a naive-cell-specific reporter (EOS-C(3+)-GFP/puro^{R})[4] could not be expanded after naive conversion in t2iLGöY medium supplemented with 0.5 μM PD03 even when *NANOG* and *KLF2* were exogenously expressed (Supplementary Fig. 2d). We also note that only those primed UCLA lines that efficiently converted to a naive state in m5i/LAF (Supplementary Fig. 1d and data not shown) could be adapted to t2iLGöY medium, which points to cell-line-dependent factors that influence the response to MEK inhibition. We conclude that t2iLGöY-cultured hESCs are more sensitive to PD03 concentration than 5i/LAF-cultured hESCs, as they could not be maintained long-term in titrated MEKi despite the expression of exogenous transcription factors.

Alternative MEK inhibitors maintain hESCs in a naive state. We next investigated whether alternative MEK inhibitors are effective in sustaining naive pluripotency in hESCs. Specifically, we tested a panel of nine MEK inhibitors that target MEK1 alone (TAK-733 and cobimetinib), MEK1 and MEK2 together (selumetinib, trametinib, pimasertib, refametinib, binimetinib, and AZD8330), or MEK1 along with BRAF and CRAF (RO5126766). We replaced PD03 in the 5i/LAF condition with one of two different concentrations of these inhibitors (1 μM and 0.5 μM) and then monitored expression of the ΔPE OCT4–GFP transgene (Fig. 2a). All nine inhibitors maintained GFP expression at levels observed in PD03-treated cells (Fig. 2b). However, we detected a loss of dome-shaped morphology in cells cultured with trametinib, AZD8330, or selumetinib; we therefore excluded these small molecules from subsequent experiments. To assess whether alternative MEK inhibitors also facilitate conversion to a naive state, we exposed the primed hESC line UCLA4 to 5i/LAF, m5i/LAF, or 4i/LAF supplemented with 1 or 0.5 μM of TAK-733, refametinib, cobimetinib, pimasertib, or RO526766 (Supplementary Fig. 2e). We obtained homogeneous cell cultures with naive-cell-like morphology that could be expanded for multiple passages with all examined MEK inhibitors at a concentration of 1 μM, although only TAK-733, refametinib, and cobimetinib maintained naive colony morphology at 0.5 μM (Supplementary Fig. 2f). UCLA4 hESCs cultured under these conditions expressed the naive-cell-associated transcription factors KLF4, KLF17, NANOG, and STELLA (Fig. 2c). We quantified conversion efficiency by using flow cytometry to monitor the reactivation of ΔPE OCT4–GFP. Replacement of PD03 with TAK-733, refametinib, or cobimetinib did not affect the percentage of GFP^{+} cells obtained by day 8 of conversion (Fig. 2d). Moreover, treatment of hESCs with TAK-733 and cobimetinib at a low concentration (0.5 μM) activated the ΔPE OCT4–GFP reporter to a similar degree as treatment with a high concentration (1 μM) of PD03 (~40%) (Fig. 2d). Finally, we determined the potential of alternative MEK inhibitors as replacements for PD03 during the maintenance of naive hESCs in t2iLGöY. TAK-733, refametinib, and cobimetinib sustained cell proliferation, clonal growth, and pluripotency-gene expression to a similar extent as PD03 in the context of t2iLGöY culture (Supplementary Fig. 3a–d). Overall, these results suggest that (i) PD03 can be substituted by other MEK inhibitors in two different naive culture systems, (ii) targeting of MEK1 alone is sufficient to sustain naive pluripotency, and (iii) low concentrations of TAK-733 and cobimetinib endow primed cells with a naive-cell-specific phenotype.

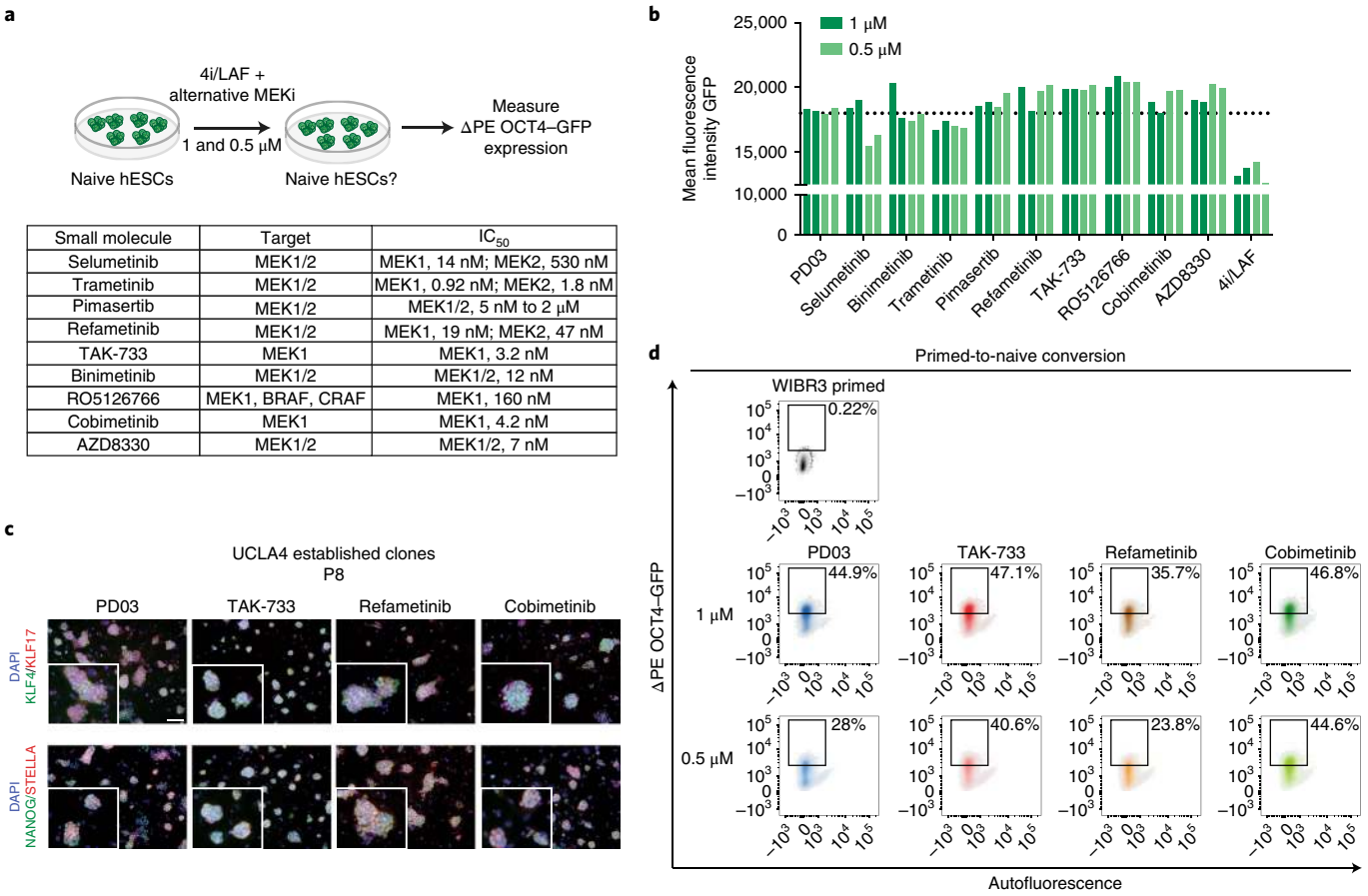
A naive-cell-like transcriptome and proteome in modified 5i/LAF cultures. We carried out RNA-seq of hESC lines maintained in primed conditions and in six naive conditions: (i) 5i/LAF, (ii) m5i/LAF, (iii) 4i/LAF plus 0.5 μM TAK-733, (iv) 4i/LAF plus 0.5 μM cobimetinib, (v) 4i/LAF plus 0.5 μM refametinib, and (vi) hESCs transitioned from 5i/LAF to t2iLGöY for three passages. Principal component analysis (PCA) of these samples showed that the major transcriptional changes along principal component 1 (86% total variability) separated all primed cell lines from all naive cell lines, regardless of culture condition (Fig. 3a and Supplementary Fig. 4a). Principal component 2 seemed to detect differences in the genetic background of cell lines regardless of MEKi or cell state (i.e., primed versus naive) (Fig. 3a). This is consistent with the previous finding that genetic background drives transcriptional variation among different human pluripotent stem cells[17,18].

Genes associated with naive pluripotency (*KLF4*/*5*/*17*, *TFCP2L1*, and *DPPA5*) were expressed at high levels in all naive cell lines, whereas genes associated with primed pluripotency (*ZIC3*, *THY1*, and *PODXL*) were expressed only in primed hESCs (Fig. 3b and Supplementary Fig. 4b). We confirmed the induction of naive-state-specific genes independently by qRT-PCR in UCLA17 hESCs cultured in either 5i/LAF or m5i/LAF conditions (Supplementary Fig. 4c). Of note, UCLA4 naive cells cultured in 4i/LAF and supplemented with TAK-733 expressed slightly higher levels of naive markers compared with levels in UCLA4 cells cultured in 5i/LAF (Supplementary Fig. 4d). Reactome pathway enrichment analysis revealed an overrepresentation of MAPK targets among the upregulated genes in the m5i/LAF condition compared with levels in the 5i/LAF condition (Supplementary Fig. 4e), consistent with attenuated MAPK suppression in m5i/LAF-cultured cells. Together, these results demonstrate that hESCs cultured in m5i/LAF or with alternative MEK inhibitors had preserved and in some cases boosted transcription patterns characteristic of naive pluripotent cells.

We next carried out a large-scale quantitative proteomics analysis of UCLA4, UCLA9, and WIBR3 hESC lines derived and maintained in primed conditions, 5i/LAF, or m5i/LAF. In addition, we analyzed the proteome of UCLA4 hESCs cultured in t2iLGöY or in 4i/LAF supplemented with 0.5 μM TAK-733. We quantified >8,649 proteins with high expression in primed and naive cells (Supplementary Fig. 5a and Supplementary Table 1). We further observed good correlation (Pearson correlation: ~0.56) between RNA and protein expression in our samples, in line with previous studies that used other cell lines[19] (Supplementary Fig. 5b). Unsupervised hierarchical clustering of proteome data accurately separated naive from primed samples regardless of culture condition (Fig. 3c). In further agreement with the RNA-seq data, naive hESCs upregulated naive-specific proteins, including FGF4 and KLF17, and downregulated proteins associated with the primed state, such as ZIC3 and THY-1 (Fig. 3d and Supplementary Fig. 5c). We again observed higher expression of some naive-cell markers (e.g., DNMT3L and DPPA5) in hESCs cultured in the presence of TAK-733 (Fig. 3d). Notably, when we compared proteomics data with RNA-seq data, we found that the expression of ~40% of genes that were upregulated at the protein level in either naive or primed conditions did not change at the RNA level, which suggests that a large fraction of genes are regulated post-transcriptionally (Supplementary Fig. 5d). These proteins included chromatin regulators, transcription factors, and RNA-binding proteins, including known pluripotency regulators such as SOX2 and MBNL1 (Supplementary Fig. 5d,e).

Collectively, these data show that our modification of the naive culture condition had minimal effects on the proteome of naive hESCs.

Modified 5i/LAF maintains HERVK expression, genomic hypomethylation, and X chromosome inactivation state. Our preparation of RNA-seq libraries from ribodepleted RNA allowed us to measure the expression of transposable elements (TEs). Prior research has suggested that TEs permit sensitive discrimination between primed and naive hESCs[7,20]. TE transcriptional patterns indeed separated primed samples from all of the naive samples



Small molecule	Target	IC_{50}
Selumetinib	MEK1/2	MEK1, 14 nM; MEK2, 530 nM
Trametinib	MEK1/2	MEK1, 0.92 nM; MEK2, 1.8 nM
Pimasertib	MEK1/2	MEK1/2, 5 nM to 2 μM
Refametinib	MEK1/2	MEK1, 19 nM; MEK2, 47 nM
TAK-733	MEK1	MEK1, 3.2 nM
Binimetinib	MEK1/2	MEK1/2, 12 nM
RO5126766	MEK1, BRAF, CRAF	MEK1, 160 nM
Cobimetinib	MEK1	MEK1, 4.2 nM
AZD8330	MEK1/2	MEK1/2, 7 nM

Fig. 2 | MEK1 inhibition is sufficient to sustain naive pluripotency in 5i/LAF. a, Alternative MEK inhibitors used in this study. IC_{50}, half-maximal inhibitory concentration. **b**, Flow cytometry analysis of the proportion of WIBR3 ΔPE OCT4–GFP$^+$ hESCs under the indicated culture conditions. The black dotted line indicates the GFP intensity threshold in 5i/LAF conditions. **c**, Representative immunofluorescence images of UCLA4 hESCs grown with the indicated inhibitors (P8). Scale bar, 100 μm. The insets are at 2× magnification relative to the primary images. **d**, Flow cytometry analysis of the proportion of ΔPE OCT4–GFP$^+$ cells after primed-to-naive conversion of WIBR3 hESCs in 5i/LAF supplemented with the indicated MEK inhibitors.

(Supplementary Fig. 6a). Inspection of TE subclasses in our cell lines further revealed that members of the HERVK family were exclusively induced in the naive state, in agreement with a recent report[7] and in support of the interpretation that our conditions endow hESCs with a bona fide naive state (Fig. 4a). hESCs cultured in medium containing 0.5 μM TAK-733 showed the highest HERVK expression among cells from all tested naive conditions, including 5i/LAF (Fig. 4a). This finding correlates with higher levels of naive transcription factors in TAK-733-cultured hESCs and suggests that suppression of MEK1 with this small molecule is most effective at inducing 'naive-like' molecular characteristics.

Global DNA hypomethylation is a key epigenetic hallmark of naive pluripotent cells[21,22]. We used reduced-representation bisulfite sequencing to assess the effect of naive culture conditions on global DNA methylation relative to that in primed hESCs. All examined naive samples exhibited global hypomethylation compared with that of their primed counterparts (Fig. 4b and Supplementary Fig. 6b). In agreement, PCA analysis separated primed from naive samples on the basis of 95,239 1-kb tiles (Supplementary Fig. 6c) or promoter regions (Supplementary Fig. 6d). We also observed that imprint control regions became demethylated in naive conditions (Supplementary Fig. 6e), in agreement with recent reports[7,15,23]. We found that naive cells cultured in m5i/LAF had slightly higher global methylation levels compared with those of 5i/LAF cultures, consistent with our previous finding that PD03 drives genomic hypomethylation in naive mouse ESCs[11]. Closer inspection of these regions revealed 984 sites that were specifically hypermethylated in m5i/LAF relative to 5i/LAF cultures (Supplementary Fig. 6f). Most of these regions were associated with repetitive DNA elements of the SINE and LINE families (Supplementary Fig. 6g). UCLA4 hESCs exposed to 4i/LAF plus 0.5 μM TAK-733 showed the lowest overall methylation levels among the tested cell cultures, including 5i/LAF cultures (Fig. 4b and Supplementary Fig. 6b), which concurs with the relative upregulation of naive-cell-specific markers in TAK-733-treated hESCs (Figs. 3d and 4a and Supplementary Fig. 4d).

X-chromosome inactivation (XCI) is another tightly regulated epigenetic process that distinguishes primed from naive female hESCs[6,24]. We used RNA fluorescence in situ hybridization (RNA-FISH) for X-linked transcripts to assess how the m5i/LAF condition influences XCI status. Although all tested female primed hESCs lacked *XIST* expression because of XCI erosion, we confirmed the presence of two X chromosomes by RNA-FISH for the X-linked gene *UTX* (*KDM6A*), which generally escapes XCI (Fig. 4c and Supplementary Fig. 7a,b). We detected monoallelic transcription of *HUWE1*, an X-linked gene commonly subject to XCI, in four of five primed hESC lines (UCLA1, -4, and -5 and WIBR3) (Fig. 4c and Supplementary Fig. 7a–c). UCLA9 cells showed *HUWE1* signal from both X chromosomes, as previously observed for this line (Supplementary Fig. 7c and ref. [14]). Adaptation of all primed hESC lines to either 5i/LAF or m5i/LAF conditions resulted in biallelic transcription of *HUWE1*,

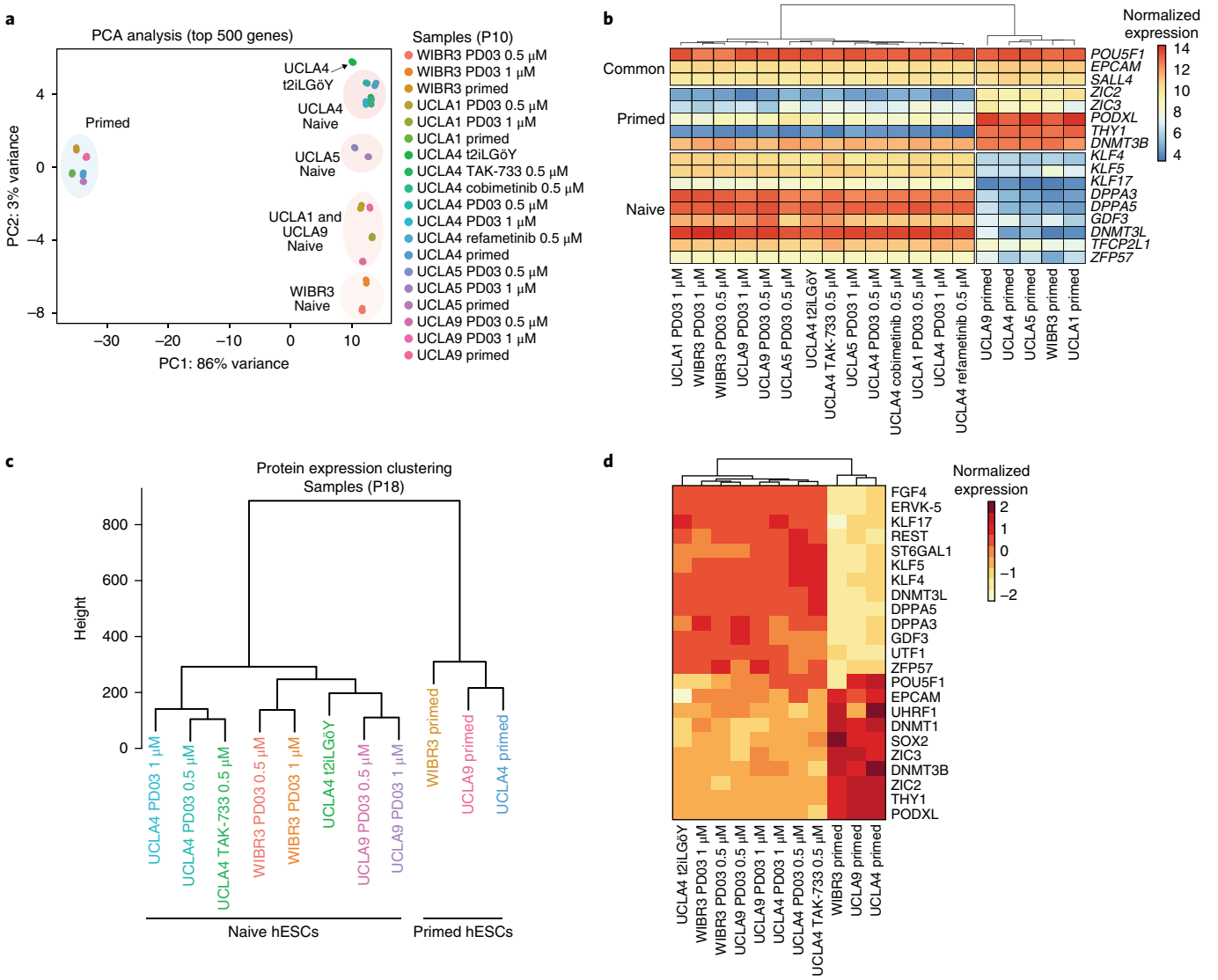


Fig. 3 | Transcriptome and proteome analysis of hESCs cultured in modified conditions. a, PCA analysis of RNA-seq data for the indicated hESC samples (P10) using the 500 most variable genes. **b**, Heat map of selected primed and naive pluripotency genes. **c**, Unsupervised hierarchical clustering of proteomic samples (P18). **d**, Heat map of selected primed and naive pluripotency proteins.

indicative of X-chromosome reactivation in the naive state (Fig. 4c and Supplementary Fig. 7a–c). We next sought to quantify *XIST* signal in primed, 5i/LAF-cultured, and m5i/LAF-cultured hESCs. Colonies in which *XIST* signal was present or absent in more than 80% of cells were considered XIST$^+$ and XIST$^-$, respectively. Colonies with heterogeneous patterns of *XIST* expression were designated as XIST$^{+/-}$. The vast majority (~95%) of naive colonies that we derived and expanded in regular 5i/LAF media expressed *XIST* from one of the active X chromosomes (XIST$^+$), in agreement with a previous report[6]. Naive cells cultured in m5i/LAF showed a lower fraction (20–40%) of XIST$^+$ colonies (Fig. 4d). Collectively, these results indicate that the m5i/LAF culture condition captures a bona fide naive state that extends from the gene expression and proteomic level to the level of TE transcription and epigenetic modification.

Modified 5i/LAF condition preserves hESC genomic stability. We examined the functional consequences of continual treatment of naive hESCs with titrated PD03 by measuring growth potential, apoptosis, and genomic stability. hESC lines cultured in m5i/LAF proliferated more rapidly than hESCs cultured in 5i/LAF ($n = 3$ representative hESC lines), indicating that strong MAPK inhibition limits the growth potential of naive hESCs (Fig. 5a). These differences in growth behavior were more pronounced in some hESC lines (e.g., UCLA5; approximately fourfold) than in others (e.g., UCLA4; approximately twofold), which pointed once more to cell-line-intrinsic responses to MAPK modulation. We found that m5i/LAF-cultured WIBR3 hESCs gave rise to twice as many colonies as 5i/LAF-cultured hESCs in a clonal colony formation assay, suggesting increased viability of dissociated hESCs that were maintained with low concentrations of MEKi (Supplementary Fig. 8a). Consistent with these observations, we detected a twofold decrease in the fraction of annexin V$^+$ cells among m5i/LAF-cultured WIBR3 hESCs compared with that in 5iLAF cultures (Supplementary Fig. 8b). hESCs cultured in t2iLGöY and subsequently plated in reduced concentrations of PD03 (0.6 µM) for 7 d also exhibited higher clonogenic potential (Supplementary Fig. 8c). Together, these results show that the use of reduced concentrations of MEKi in two naive culture conditions led to increased proliferation and decreased apoptosis.

Recent studies have shown that suppression of MAPK/ERK signaling compromises genomic stability in mouse ESCs[25] and

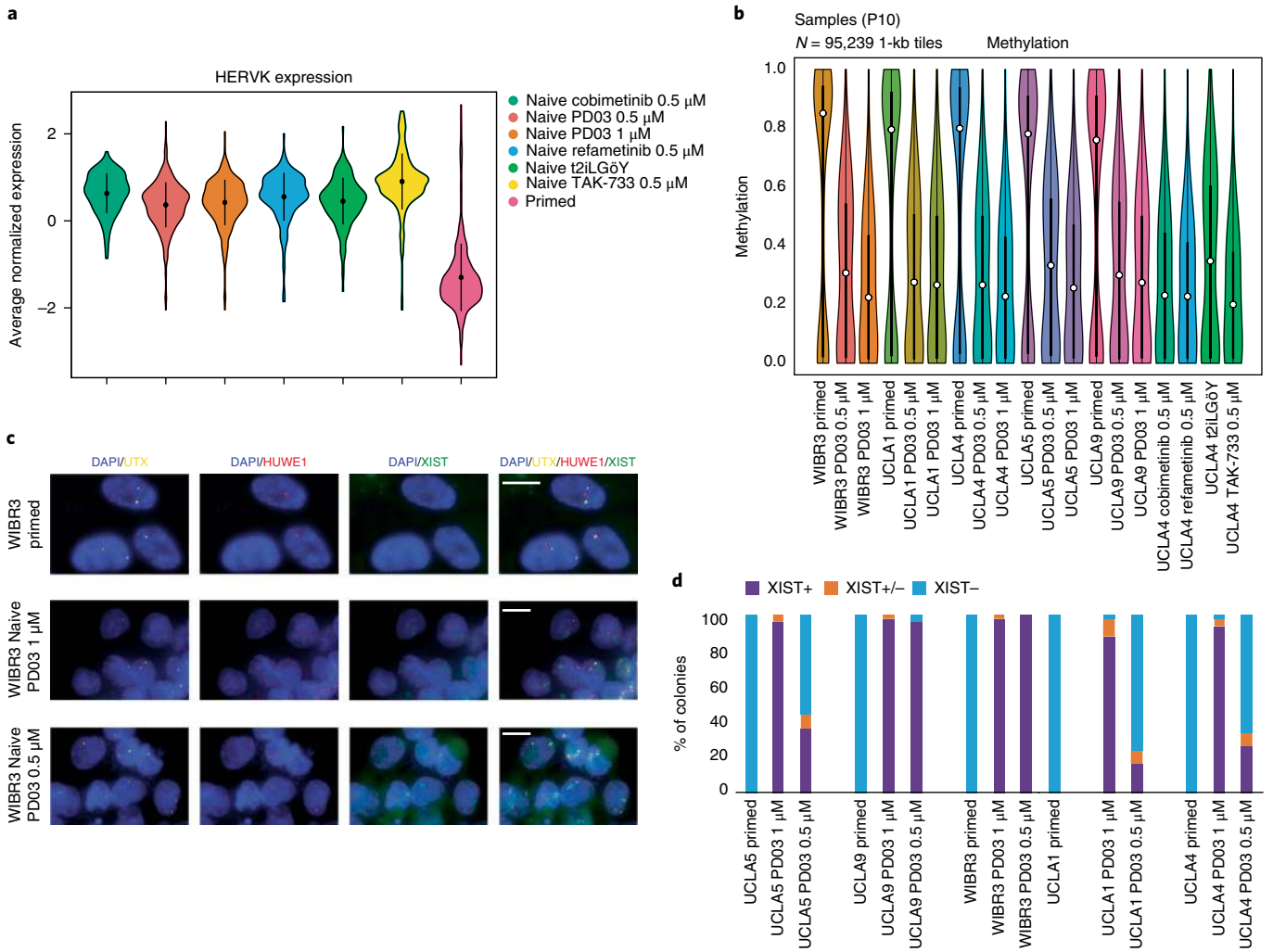


Fig. 4 | hESCs cultured in m5i/LAF are hypomethylated and have two active X chromosomes. a, Violin plot of HERVK expression ($n = 64$ elements) in hESCs expanded in the indicated culture conditions (P10). For primed, naive PD03 (1 μM), and naive PD03 (0.5 μM) conditions, $n = 5$ independent hESC lines (UCLA1, -4, -5, and -9 and WIBR3); for naive cobimetinib (0.5 μM), refametinib, TAK-733, and t2iLGöY, $n = 1$ hESC line (UCLA4). In each violin, the black dot represents the mean normalized expression value, and the black line extending from this point represents 1 s.d. from the mean. **b**, Violin plots showing global methylation of primed and naive hESCs, based on the results of reduced-representation bisulfite sequencing analysis. In each violin, the circle represents the median value, the heavy black rectangle extends from the 25th to the 75th percentile of the data (the interquartile range), and whiskers extend to the most extreme data point (no more than 1.5× the interquartile range). **c**, Representative RNA-FISH images of primed and naive hESCs at P16, showing the indicated transcripts. Scale bars, 10 μm. **d**, Quantification of XIST RNA-FISH patterns.

that this is accompanied by an increased number of γ-H2AX foci, which correlate with DNA double-strand breaks[26]. We used flow cytometry to examine the frequency of γ-H2AX foci in primed hESCs and naive hESCs cultured in either 5i/LAF or m5i/LAF. hESCs expanded in 5i/LAF showed a fourfold increase in the number of γ-H2AX foci (~20%) compared with that in hESCs maintained in primed conditions (~5%) (Fig. 5b). Decreasing the amount of PD03 from 1 μM to 0.5 μM reduced the percentage of γ-H2AX foci (~10%; $P = 0.0005$), thus supporting the interpretation that strong MAPK inhibition via PD03 triggers a DNA damage response.

On the basis of our observations, we hypothesized that naive hESCs in m5i/LAF might retain a relatively normal karyotype owing to reduced selective pressure. Indeed, we found that naive WIBR3 hESCs cultured in m5i/LAF for ten passages retained a largely normal karyotype that was similar to that of the parental hESCs cultured in primed medium (Fig. 5c and Supplementary Fig. 8d,e). By contrast, hESCs cultured in conventional 5i/LAF media for the same period of time acquired multiple additional chromosomal abnormalities, and none of the karyotyped cells ($n = 20$) were normal (Supplementary Fig. 8f).

To test the generality of our findings, we carried out shallow whole-genome sequencing of UCLA1, -4, -5, and -17 hESCs after continual propagation of these cell lines for ten passages in primed media, 5i/LAF, m5i/LAF, and 4i/LAF supplemented with TAK-733. We found that UCLA5 hESCs cultured in 5i/LAF conditions accrued a greater number of chromosomal aberrations (trisomy 5; partial amplification of chromosome 7; loss of one X chromosome; and partial loss of chromosomes 8, 9, and 15) compared with cells cultured in the m5i/LAF conditions (trisomy 5 and partial amplification of chromosome 7) (Fig. 5d). Similarly, UCLA1 and UCLA17 hESCs cultured in 5i/LAF showed trisomy 5 and 7, whereas they remained mostly normal in m5i/LAF conditions (Fig. 5d and Supplementary Fig. 8g). Naive hESCs cultured in 4i/LAF and TAK-733, which showed higher HERVK transcription, increased RNA and protein expression of naive-cell markers, and

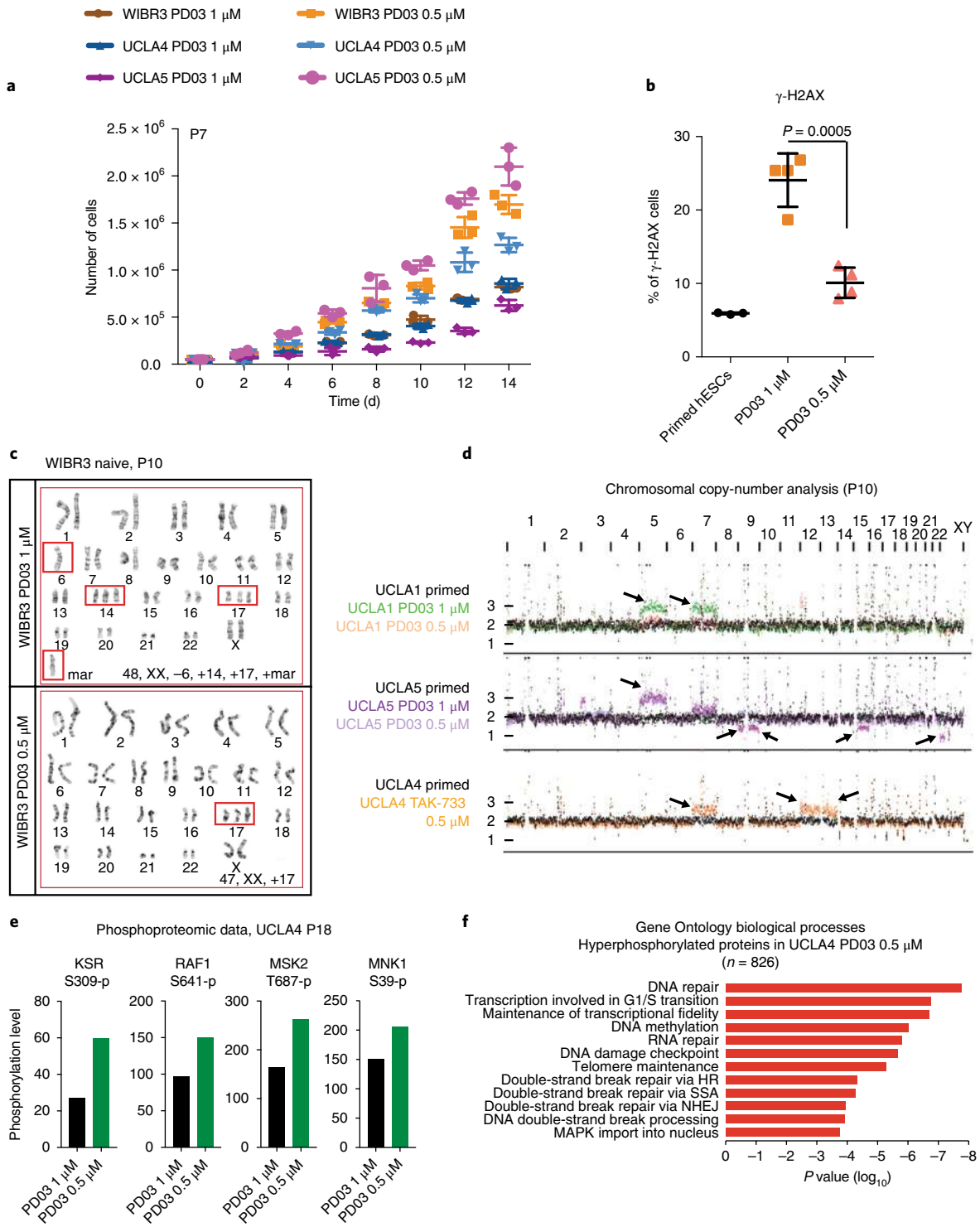


Fig. 5 | Naive hESCs cultured in m5i/LAF retain a more stable karyotype. a, Cell numbers during the expansion of naive hESCs under the indicated conditions. $n = 3$ biologically independent samples per condition (cell cultures). **b**, Flow cytometry analysis for γ-H2AX$^+$ cells using primed UCLA1 hESCs. $n = 3$ biologically independent samples for primed hESCs; $n = 4$ independent clones for naive UCLA1 hESCs cultured in 5i/LAF or m5i/LAF. Statistical significance was determined by two-tailed unpaired Student's *t*-test. Data in **a**,**b** are shown as mean ± s.d. **c**, Representative karyotypes at P10 of hESCs converted and cultured as indicated. Red boxes highlight chromosomal abnormalities. Mar, marker chromosome. **d**, Chromosomal copy-number analysis by whole-genome sequencing. Arrows indicate abnormalities. **e**, Phosphorylation levels at the indicated residues, assessed by phosphoproteomics in UCLA4 cells cultured as indicated. **f**, Gene Ontology analysis of hyperphosphorylated proteins (>1.5-fold) in UCLA4 hESCs cultured as indicated. HR, homologous recombination; SSA, single-strand annealing; NHEJ, nonhomologous end-joining.

reduced global methylation compared with hESCs in the other MEKi conditions, became highly abnormal after ten passages (trisomy 7, 12, and 13) (Fig. 5d). These results imply that reduced MAPK inhibition relieves selection pressure for the acquisition of chromosomal abnormalities that promote the growth and/or survival of naive hESCs.

Molecular effects of reduced MEK inhibition. In an effort to understand the mechanisms by which increased MEK activity affects naive hESC properties, we generated phosphoproteomic data for UCLA4 hESCs cultured in 5i/LAF, m5iLAF, and t2iLGöY (Supplementary Table 2). We confirmed that reduction of the PD03 concentration to 0.5 μM resulted in increased ERK phosphorylation in UCLA4 hESCs, which we verified independently by western blot analysis (Supplementary Figs. 9a,b and 10). We also observed increased phosphorylation of other components of the MAPK pathway (Fig. 5e).

We next focused on proteins that gained phosphorylation (>1.5-fold) in m5i/LAF relative to 5i/LAF culture conditions but were not directly related to MAPK signaling. Most of these proteins were associated with functional categories related to "DNA repair," "DNA methylation," "DNA damage checkpoints," and "G1/S transition of cell cycle," which implies that MAPK inhibition by PD03 affects molecular regulators of genomic stability and DNA repair (Fig. 5f). We confirmed overrepresentation of these functional categories by using an independent hESC line (UCLA9) (Supplementary Fig. 9c,d). Finally, we selected proteins with differential phosphorylation between hESCs cultured in 5i/LAF and hESCs cultured in t2iLGöY and carried out Gene Ontology enrichment analysis. Surprisingly, we also observed a clear enrichment of categories related to the MAPK pathway. Considering that equal concentrations of PD03 are present in 5i/LAF and t2iLGöY, this observation points to MEKi-independent modulation of the MAPK pathway in one of the two naive culture conditions (Supplementary Fig. 9e).

Discussion

Through systematic characterization of primed and naive hESCs maintained under different culture conditions, we uncovered an unexpected sensitivity of naive pluripotent stem cells to MAPK signal inhibition, which affects their proliferation rate, survival, and genome integrity. We propose that minimal MAPK activity is beneficial for the preservation of both robust growth potential and genomic stability in naive human pluripotent stem cells maintained in 5i/LAF (model in Supplementary Fig. 9f). Our data also suggest that different hESC lines respond differently to MEK inhibition. For instance, we found that the concentration of PD03 could be reduced to 0.5 μM in WIBR3 hESCs and to 0.25 μM in UCLA4 hESCs without adverse effects (data not shown). We suggest that MEK inhibitors should ideally be titrated in a cell-line-dependent fashion. The observation that attenuated MEK inhibition affects the phosphorylation levels of proteins that regulate DNA damage and repair and leads to a reduced number of γ-H2AX foci provides a possible molecular explanation for the observed phenotypes.

Our finding that naive hESCs cultured in 5i/LAF and t2iLGöY exhibited differences in MAPK activity despite equivalent concentrations of MEKi suggests that the BRAF and SRC inhibitors in 5i/LAF media further dampen the MAPK pathway compared with t2iLGöY media. The notion that MAPK signaling is less repressed in t2iLGöY-cultured hESCs than in 5i/LAF-cultured hESCs parallels our observations in m5i/LAF-cultured hESCs and is consistent with the following. First, t2iLGöY-cultured cells, like m5i/LAF-cultured cells, divide more rapidly and are karyotypically more stable than 5i/LAF-cultured hESCs[10,27]. Second, we found that we could not reduce the concentration of PD03 to 0.5 μM in hESCs continually cultured in t2iLGöY without inducing differentiation (Supplementary Fig. 2c,d). Third, UCLA lines that are refractory to naive conversion with bulk passaging in 0.5 μM PD03 could not be transferred from 5i/LAF to t2iLGöY, which suggests a requirement for stronger suppression of the MAPK pathway in certain hESC lines. Fourth, the majority of cells cultured in t2iLGöY media have two active X chromosomes but lack *XIST* transcription[6,10], similar to cells expanded in m5i/LAF medium. Finally, hESCs maintained in t2iLGöY had levels of global methylation that were higher than those in naive cells expanded in 5i/LAF medium but similar to those in m5i/LAF-cultured hESCs (Fig. 4b, Supplementary Fig. 6b, and ref. [27]).

The transcriptional, proteomic, phosphoproteomic, and genomic methylation data we have generated here will provide a useful resource for future studies aimed at dissecting the mechanisms underlying naive and primed pluripotency. Our results may also have a bearing on potential therapeutic applications of naive pluripotent stem cells, as chromosomal stability is of the utmost importance in cell and gene therapy applications.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41592-018-0104-1.

Received: 1 June 2017; Accepted: 6 July 2018;
Published online: 20 August 2018

References

1. Weinberger, L., Ayyash, M., Novershtern, N. & Hanna, J. H. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat. Rev. Mol. Cell Biol.* **17**, 155–169 (2016).
2. Nichols, J. & Smith, A. Pluripotency in the embryo and in culture. *Cold Spring Harb. Perspect. Biol.* **4**, a008128 (2012).
3. Gafni, O. et al. Derivation of novel human ground state naive pluripotent stem cells. *Nature* **504**, 282–286 (2013).
4. Takashima, Y. et al. Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* **158**, 1254–1269 (2014).
5. Theunissen, T. W. et al. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* **15**, 471–487 (2014).
6. Sahakyan, A. et al. Human naive pluripotent stem cells model X chromosome dampening and X inactivation. *Cell Stem Cell* **20**, 87–101 (2017).
7. Theunissen, T. W. et al. Molecular criteria for defining the naive human pluripotent state. *Cell Stem Cell* **19**, 502–515 (2016).
8. Irie, N. et al. SOX17 is a critical specifier of human primordial germ cell fate. *Cell* **160**, 253–268 (2015).
9. Huang, K., Maruyama, T. & Fan, G. The naive state of human pluripotent stem cells: a synthesis of stem cell and preimplantation embryo transcriptome analyses. *Cell Stem Cell* **15**, 410–415 (2014).
10. Guo, G. et al. Epigenetic resetting of human pluripotency. *Development* **144**, 2748–2763 (2017).
11. Choi, J. et al. Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. *Nature* **548**, 219–223 (2017).
12. Yagi, M. et al. Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation. *Nature* **548**, 224–227 (2017).
13. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532 (2011).
14. Patel, S. et al. Human embryonic stem cells do not change their X inactivation status during differentiation. *Cell Rep.* **18**, 54–67 (2017).
15. Collier, A. J. et al. Comprehensive cell surface protein profiling identifies specific markers of human naive and primed pluripotent states. *Cell Stem Cell* **20**, 874–890 (2017).
16. Guo, G. et al. Naive pluripotent stem cells derived directly from isolated cells of the human inner cell mass. *Stem Cell Rep.* **6**, 437–446 (2016).
17. Choi, J. et al. A comparison of genetically matched cell lines reveals the equivalence of human iPSCs and ESCs. *Nat. Biotechnol.* **33**, 1173–1181 (2015).
18. Rouhani, F. et al. Genetic background drives transcriptional variation in human induced pluripotent stem cells. *PLoS Genet.* **10**, e1004432 (2014).
19. Liu, Y., Beyer, A. & Aebersold, R. On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**, 535–550 (2016).
20. Friedli, M. & Trono, D. The developmental control of transposable elements and the evolution of higher species. *Annu. Rev. Cell Dev. Biol.* **31**, 429–451 (2015).
21. Lee, H. J., Hore, T. A. & Reik, W. Reprogramming the methylome: erasing memory and creating diversity. *Cell Stem Cell* **14**, 710–719 (2014).
22. Smith, Z. D. et al. DNA methylation dynamics of the human preimplantation embryo. *Nature* **511**, 611–615 (2014).
23. Pastor, W. A. et al. Naive human pluripotent cells feature a methylation landscape devoid of blastocyst or germline memory. *Cell Stem Cell* **18**, 323–329 (2016).
24. Petropoulos, S. et al. Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* **165**, 1012–1026 (2016).

25. Chen, H. et al. Erk signaling is indispensable for genomic stability and self-renewal of mouse embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **112**, E5936–E5943 (2015).
26. Kuo, L. J. & Yang, L. X. Gamma-H2AX—a novel biomarker for DNA double-strand breaks. *In Vivo* **22**, 305–309 (2008).
27. Liu, X. et al. Comprehensive characterization of distinct states of human naive pluripotency generated by reprogramming. *Nat. Methods* **14**, 1055–1062 (2017).

Acknowledgements

We thank T.W. Theunissen and R. Jaenisch for primed WIBR3 hESCs, the MGH next-generation sequencing facility for technical assistance, and members of the Hochedlinger lab for discussions. This work was supported by MGH (K.H.), the NIH (R01 HD058013 and P01 GM099134 to K.H.; P01 GM099134 to K.P.; GM67945 to S.P.G.), the Gerald and Darlene Jordan Chair in Regenerative Medicine (K.H.), EMBO (long-term fellowship #ALTF 1143-2015 to B.D.S.), UCLA (UCLA Broad Stem Cell Research Center–Rose Hills Foundation Training Award to S.S.; UCLA Dissertation Year Fellowship to A.S.; UCLA Broad Stem Cell Research Center (to K.P.); the David Geffen School of Medicine (to K.P.); the Jonsson Comprehensive Cancer Center (to K.P.)), the Howard Hughes Medical Institute (Faculty Scholar grant to K.P.), and JSPS KAKENHI (JP16K15489 and JP16H02465 to Y.T.).

Author contributions

B.D.S. and K.H. conceived the study and wrote the manuscript. B.D.S., M.U., J.B., A.J.H., A.S., K.P., and Y.T. performed the experiments and analyzed the data. S.S. performed the bioinformatics analysis. K.C., H.G., and A.M. performed and analyzed the reduced-representation bisulfite sequencing experiments. K.S. and T.S. performed and analyzed the whole-genome sequencing experiments. K.J.C., A.R.E., and S.P.G. performed and analyzed the proteomics experiments.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-018-0104-1.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to K.H.

Methods

Mice. All procedures involving mice adhered to the guidelines of approved Massachusetts General Hospital Institutional Animal Care and Use Committee (IACUC) protocol no. 2006N000104. NOD/SCID mice were acquired from Jackson Laboratory.

Cell culture. Conventional primed hESCs (WIBR3, ΔPE OCT4–GFP[5], UCLA1, UCLA3, UCLA4, UCLA5, UCLA9, and UCLA17 cell lines) were maintained on Matrigel (Corning)-coated dishes in E8 medium (Stem Cell Technologies) and passaged using Versene solution (Life Technologies). For maintenance, cells were passaged every 4–5 d. Primed H1 and H9 hESC lines were cultured as previously reported[4]. To revert cells to a naive state, we washed hESCs that had been passaged 6 d earlier once with 1× PBS (Life Technologies) and treated them for 3 min with TrypLE express enzyme (1×; Life Technologies). Cells were dissociated into a single-cell suspension and plated at a density of 30,000 cells per 9.5 cm^2 on irradiated CF-1 mouse embryonic fibroblast (MEF) feeder cells (MTI Globalstem) in E8 medium supplemented with 10 μM Y-27632 (Axon Medchem). Two days later, the medium was changed to 5i/LAF, and thereafter it was replaced daily. 5i/LAF medium contained a 50:50 mixture of DMEM/F-12 (Life Technologies) and Neurobasal medium (Life Technologies) containing 1× N2 supplement (Life Technologies), 1× B27 supplement (Life Technologies), 10 ng/mL basic FGF (Peprotech), 1% nonessential amino acids (Life Technologies), 1 mM GlutaMAX (Life Technologies), penicillin–streptomycin (Life Technologies), 0.1 mM β-mercaptoethanol (Life Technologies), 50 μg/mL BSA (Life Technologies), 0.5 μM IM-12 (Axon Medchem), 0.5 μM SB590885 (Axon Medchem), 1 μM WH-4-023 (Axon Medchem), 10 μM Y-27632 (Axon Medchem), 20 ng/mL activin A (Peprotech), 20 ng/mL recombinant human LIF (Peprotech), 0.5% KSR (Life Technologies), and 0.5 or 1 μM PD0325901 (Axon Medchem) as indicated in the main text. After roughly 8–10 d, cells were dissociated with Accutase (Life Technologies) and centrifuged in fibroblast medium (DMEM (Life Technologies) supplemented with 10% FBS (Hyclone), 1 mM GlutaMAX (Life Technologies), 1% nonessential amino acids (Life Technologies), penicillin–streptomycin (Life Technologies), 0.1 mM β-mercaptoethanol (Life Technologies)) and replated after being passed through a 40-μm cell strainer in 5i/LAF medium on irradiated CF-1 MEFs.

Established naive hESC lines were cultured on irradiated CF-1 MEFs (2.5×10^6 cells per 9.5 cm^2) in 4i/LAF medium with different concentrations of PD0325901 and passaged every 6–7 d. Cells were fed daily with fresh medium.

Human naive hESCs were also cultured under t2iLGöY conditions for the experiments indicated in the main text. t2iLGöY medium contained a 50:50 mixture of DMEM/F-12 (Life Technologies) and Neurobasal medium (Life Technologies) containing 1× N2 supplement (Life Technologies), 1× B27 supplement (Life Technologies), 10 ng/mL basic FGF (Peprotech), 1% nonessential amino acids (Life Technologies), 1 mM GlutaMAX (Life Technologies), penicillin–streptomycin (Life Technologies), 0.1 mM β-mercaptoethanol (Life Technologies), 50 μg/mL BSA (Life Technologies), 1 μM or 0.5 μM PD0325901 (Axon Medchem), 1 μM CHIR99021 (Axon Medchem), 2.5 μM Gö6983 (Axon Medchem), 10 μM Y-27632 (Axon Medchem), 20 ng/mL recombinant human LIF (Peprotech), and 50 μg/ml ascorbic acid (Sigma-Aldrich).

H9 hESCs and H9 hESCs carrying the EOS-C(3+)-GFP/puro^R piggyBac reporter (EOS) were reset to the naive state as reported previously[4]. Briefly, cells were reset by overexpression of *NANOG* and *KLF2* under t2iL + Dox conditions. After about five passages, the medium was switched to t2iLGöY and cells were cultured for at least five additional passages. For colony-formation assay of reset hESCs, 3,500 cells were seeded on irradiated MEFs and cultured with various concentrations of PD0325901 (–Dox). Cells were stained at day 7 for alkaline phosphatase with a Leukocyte Alkaline Phosphatase Kit (Sigma-Aldrich; 86R).

All experiments in this study were performed under low-oxygen conditions (5% O_2).

Human naive and primed hESCs were sent for karyotyping to the Cytogenomics Core Laboratory at the Brigham and Women's Hospital in Boston, MA.

A detailed protocol for the primed-to-naive conversion experiments can be found at Protocol Exchange[28].

MEK inhibitors. We used the following MEK inhibitors to culture naive hESCs at the concentrations indicated in the main text: PD0325901 (Axon Medchem), selumetinib (Selleck Chemicals), binimetinib (Selleck Chemicals), trametinib (Selleck Chemicals), pimasertib (Selleck Chemicals), refametinib (Selleck Chemicals), TAK-733 (Selleck Chemicals), RO5126766 (Selleck Chemicals), cobimetinib (Selleck Chemicals), and AZD8330 (Selleck Chemicals).

RNA-FISH and microscopy. RNA-FISH was performed as previously described[6]. Briefly, cells grown on gelatinized glass coverslips were fixed with 4% formaldehyde, permeabilized with 0.5% Triton X-100 (10 min each), and serially dehydrated with 70–100% ethanol. Coverslips were hybridized with labeled DNA probes generated from bacterial artificial chromosomes. We used an Imager M1 microscope (Zeiss) for acquisition and ImageJ software (NIH) for processing of *z*-stack images.

Western blot analysis. In preparation for western blot analysis, cells were lysed in modified RIPA buffer (50 mM Tris-HCl, pH 8 (Sigma-Aldrich), 150 mM NaCl (Sigma-Aldrich), 0.1% SDS (Sigma-Aldrich), 0.5% sodium deoxycholate (Sigma-Aldrich), 1% Triton X-100 (Sigma-Aldrich), 1 mM EDTA (Sigma-Aldrich), 1× protease inhibitor cocktail (Roche Diagnostics), 1× PhosSTOP phosphatase inhibitors (Roche Diagnostics), and 0.01 U/μl benzonase (Novagen)). Lysates were sonicated for 5 min with a Bioruptor sonication device (Diginode Diagnostics). The following antibodies were used for western blotting: β-actin (Cell Signaling Technology; clone 13E5; 5125), ERK1/2 (Cell Signaling Technology; 137F5; 4695), and p-ERK1/2 (Cell Signaling Technology; D13.14.4E; 4370).

Alkaline phosphatase staining. Alkaline phosphatase staining was performed with the Vector Red substrate kit (Vector Labs). Colony quantifications were performed using ImageJ.

RNA preparation. RNA was isolated with the miRNeasy mini kit (Qiagen). RNA was eluted from the columns with RNase-free water and quantified on a Nanodrop ND-1000. cDNA was produced with the High Capacity RNA-to-cDNA kit (Applied Biosystems). To generate material for sequencing, we sorted cells for cell viability (DAPI staining).

qRT-PCR analyses. qRT-PCR reactions were set up in triplicate with Brilliant III SYBR Master Mix (Agilent Genomics) with the primers listed in Supplementary Table 3. Reactions were run on a LightCycler 480 (Roche) PCR machine with 40 cycles of 30 s at 95 °C, 30 s at 60 °C, and 30 s at 72 °C.

DNA preparation. DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen) and quantified using the Qubit dsDNA High Sensitivity Kit (Life Technologies).

Reduced-representation bisulfite sequencing. Reduced-representation bisulfite sequencing (RRBS) libraries were prepared with a commercial kit (Ovation RRBS Methyl-Seq System; NuGen, San Carlos, CA) according to the manufacturer's protocol, except that we pooled 12 individually barcoded reactions after the final repair step and carried out bisulfite conversion and library amplification using that pool. Libraries were sequenced on an Illumina HiSeq 2500 high-output flow cell without a PhiX spike-in using Nugen's custom sequencing primer for read 1 (50 bases) and standard Illumina sequencing primers to read the 8-base sample barcodes. Sequencing reads were trimmed with Trim Galore! (default parameters; http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and a NuGEN-provided script for diversity trimming and filtering. Trimmed reads were aligned to the hg19 human genome using BSmap[29] with flags -v 0.05 -s 16 -w 100 -S 1 -p 8 -u. The methylation status of CpGs was called by observation of bisulfite conversion in reads at locations of cytosines in the reference sequence. Region methylation averages were called using CpGs that were covered with at least three reads in at least 80% of samples. Imprinting control regions were defined as in ref. [30].

FACS analyses. We analyzed levels of GFP, annexin V (allophycocyanin; BD Bioscience; 550474), γ-H2AX (Alexa Fluor 647; BD Bioscience; 560447), THY-1 (phycoerythrin; anti-human CD90 (Biolegend; 328110)), and CD75 (purified anti-human CD75 (LN-1; 326901; Biolegend)) with a BD LSR II FACS (BD Biosciences) using Diva v6.1.2 (BD Biosciences) and FlowJo software v10 (TreeStar). Cell permeabilization was carried out with the Fix and Perm Cell fixation and cell permeabilization kit (Thermo Fisher Scientific; GAS003) according to the manufacturer's instructions. We used goat anti-mouse IgM heavy chain secondary antibody, Alexa Fluor 647 conjugate (1:500) (A-21238; Thermo Fisher), as a secondary antibody.

Immunofluorescence. For immunostaining, the cells were fixed with 4% paraformaldehyde, blocked, and incubated with primary antibodies overnight at 4 °C. They were then stained with secondary antibodies goat anti-rabbit IgG (H + L) (1:500) (Alexa Fluor 488; A11008; Thermo Fisher) and goat anti-mouse IgG (H + L) (1:500) (Alexa Fluor 546; A11003; Thermo Fisher) at room temperature for 1 h. Nuclei were stained with DAPI (1:1,000; BD Bioscience; 564907). The following primary antibodies were used in this study: anti-KLF17 produced in rabbit (1:100; HPA024629; Sigma), GKLF/EKLF/LKLF antibody (F-8; 1:100; sc-166238; Santa Cruz), mouse anti-Stella (1:100; MAB4388; EMD Millipore), Nanog (D73G4) XP rabbit (mAb 4903; 1:300; 4903 S; Cell Signaling), and Oct-3/4 antibody (C-10; 1:300; sc-5279; Santa Cruz).

Teratoma assay. 2×10^6 cells were resuspended in 300 μl of Matrigel (Corning) and injected subcutaneously into the flanks of an immunocompromised NOD/SCID mouse. Teratomas were monitored and removed once tumor size reached 1 cm or tumors became ulcerated. Tumors were then weighed and processed for hematoxylin and eosin staining analysis. The tumor sections were blindly scored for the presence of each germ layer. Human naive cells were re-primed in hES primed media and cultured as primed cells before injection.

Chromosomal integrity analysis by shallow-coverage whole-genome resequencing. Chromosomal aneuploidy and partial amplification/deletion were examined by shallow-coverage whole-genome resequencing as described

in our previous study, with modifications[11]. Briefly, genomic DNA (1.1 µg in 55 µL) was sonicated with a Covaris S2 sonicator (Covaris, Woburn, MA) to ~350-bp peak fragments and subjected to PCR-free synthesis of Illumina NextSeq deep sequencing libraries using the TruSeq DNA PCR-free LT sample preparation kit, set A (Illumina; FC-121-3001) according to the protocol provided by the manufacturer (Illumina FC-121-9006DOC, part 15036187 Rev. D). During the library synthesis, we used the adaptor without index sequences from NEBNext Multiplex Oligos for Illumina Dual Index Primer Set 1 (7600S; New England Biolabs, Ipswich, MA), and the loops of the adaptor ends were cut with the USER enzyme provided in the kit. The index-free libraries were amplified with adaptors including the i7 and i5 dual-index sequences for 12 cycles and quantified with a KAPA Illumina library quantification qPCR kit (KAPA Biosystems, Wilmington, MA), and equal concentrations (4 nM) of libraries were pooled. Size distributions of gDNA fragments and deep-sequencing libraries were examined with Agilent Tapestation with Genomic DNA screening tapes (Agilent Technologies, Santa Clara, CA) and passed our standard quality-control criteria. Single-read deep sequencing (76 nt) was carried out on an Illumina NextSeq 500 sequencer with the high output v2 kit, and the machine-generated base call (bcl) files were converted to raw read sequence files (fastq) using the Illumina bcl2fastq tool (2.17.1.14). Fastq files were subjected to a quality-control process using Trim Galore! and fastQC (Babraham Institute, Cambridge, UK), which excluded raw quality reads (Phred score < 20) and deleted the Illumina adaptor sequences and the last nucleotide at the 3′ end to avoid nucleotide base biases. The trimmed/quality-controlled fastq reads were aligned to the mouse GRCm38/mm10 reference genome sequence using the STAR aligner (ver. 2.4.2.a) with the genomic DNA alignment configuration, thus generating the BAM format aligned read files. The BAM files generated from the four lanes of the sequencer were sorted, merged, and subjected to uniquely mapped read extraction using the sambamba toolkit[31]. The deduplication criteria used by sambamba markdup are identical to those used by the Broad Institute toolkit Picard. Numbers of uniquely mapped reads were 9.7–13.3 million per sample, which corresponded to about 0.25–0.35× average depth of genome coverage. Quality-control measurements of the uniquely mapped reads (BAM files) obtained with fastQC did not detect any bias with the cell types, either. The aligned unique read files (.merged.sorted.uniq.bam) were converted into normalized bigWig files, which were visualized by the Integrative Genomics Viewer (v. 2.3.25). Reads assigned to each chromosome were counted to generate .cnt files. The chromosome count data were subjected to ploidy analysis as described by Treff et al.[32], who reported successful evaluation of chromosomal aneuploidy in early-stage mouse embryos with only 0.04–0.16× average depth of coverage.

Library preparation and sequencing. RNA-sequencing libraries were prepared with the Ribo-Zero rRNA Removal Kit (H/M/R) (Illumina; MRZH11124) and NEBNext Ultra Directional RNA Library Prep Kit for Illumina (E7420). The total RNA input amount for the RiboZero kit was 1 µm. The rRNA input for library construction was 50 ng in total. Libraries were amplified for 14 cycles. After library construction, the samples were validated with the 2200 Tapestation System and High Sensitivity D1000 ScreenTape kit. Libraries were quantified with the Kapa Biosystems Library Quantification Kit (KK4828) and the Bio-Rad CFX96 instrument. Each lane of sequencing was pooled into a 19-plex (19 samples per lane) with unique barcodes. Pooled libraries were also quantified with the Kapa Biosystems Library Quantification Kit (KK4828) and the Bio-Rad CFX96 instrument. These pools were then denatured to 16 pM with 1% phix and sequenced on the Illumina HiSeq 2000 instrument. The sequencing length read was 50 bp paired-end.

Bioinformatics analysis. RNA-sequencing library reads were mapped to the human hg19 reference genome with TopHat v2.1.0[33]. The annotation of aligned reads with genomic features was done with HTSeq[34]. Repbase database annotations for TE superfamilies ERV1, ERVK, and ERVL were used to intersect the alignment and generate counts using the multiBamCov tool from Bedtools v.2.27[35]. TEs that overlapped exons were discarded. TEs that did not have at least one sample with 20 reads and genes that did not have at least one sample with 5 reads were removed from all downstream analysis. Count normalization was done with size factors to correct for differences in sequencing depth and was followed by variance stabilizing transformation scaling and differential expression analysis by DESeq2[36]. A gene (or TE) was considered to be differentially expressed when abs($\log_2$(fold change)) > 3 and false discovery rate (FDR) < 0.01. Statistically significant ($P < 0.01$) proteins for primed and naive conditions were determined by paired *t*-test, and Pearson correlation was used as a metric to compare differential fold changes between genes and proteins. R studio (https://www.rstudio.com/) was used to run custom R scripts for PCA, hierarchical clustering (stats package) and variance analysis, and to construct heat maps (pheatmap package), scatter plots, violin plots, and dendrograms. Generally, the ggplot2 package was used to visualize data graphs. Gene Ontology analysis was done with EnrichR.

Proteomics and phosphoproteomics. Cells were syringe-lysed in a buffer containing 2% SDS, 150 mM NaCl, 50 mM Tris, pH 8.5, 2 mM sodium vanadate, and protease and phosphatase inhibitors. Protein concentrations in the clarified lysates were estimated via bicinchoninic acid (BCA) protein assay (Thermo Fisher Scientific). Lysates were reduced with 5 mM DTT for 30 min at 60 °C. After cooling to room temperature, cysteine residues were alkylated with 14 mM iodoacetamide in the dark for 45 min at room temperature. Excess iodoacetamide was quenched with 5 mM DTT for 15 min in the dark. For each sample, approximately 1 mg of total protein was precipitated with methanol/chloroform and washed three times with methanol. Protein pellets were dried and solubilized in 4 M urea, 25 mM HEPES, pH 8.5, before digestion with 15 µg of LysC for 1.5 h at 37 °C. Digestion proceeded with another 15 µg of LysC overnight at room temperature. Samples were diluted to 1 M urea with 25 mM HEPES, pH 8.5, and trypsin was added to a final 1/125 enzyme/protein ratio for 12 h at 37 °C. Peptides were desalted with 50-mg Sep-Paks (Waters) and quantified by Pierce Quantitative Colorimetric Peptide Assay. Fifty micrograms of each sample were reserved for protein-level quantitation before vacuum centrifugation.

Phosphopeptide enrichment. Phosphopeptides were enriched with the High-Select Fe-NTA phosphopeptide enrichment kit (Thermo Scientific). Briefly, 750 µg of peptides were solubilized in 200 µL of binding/wash buffer and clarified via centrifugation before binding equilibrated spin columns for 40 min with intermittent gentle mixing. Phosphopeptide-bound columns were washed three times with 200 µL of binding/wash buffer and once with 200 µL of HPLC-grade water. Phosphopeptides were eluted three times with 200 µL of elution buffer into trifluoroacetic acid (0.1% final concentration) and immediately dried via vacuum centrifugation. Phosphopeptides were desalted with SOLA HRP 10-mg Sep-Pak cartridges (Thermo Fisher Scientific) before tandem mass tag (TMT) labeling.

TMT labeling. Peptides and phosphopeptides were resuspended in 200 mM EPPS, pH 8.5, before labeling. TMT 11-plex reagents (0.8 mg) were dissolved in 40 µL of anhydrous acetonitrile. For protein-level and phosphoproteome-level samples, 5 µL of appropriate TMT reagent was added per sample with anhydrous acetonitrile (final concentration, 30% (v/v)). The labeling reaction was carried out for 1 h at room temperature. All labeling reactions were quenched with 0.5% hydroxylamine, combined, acidified with formic acid (final concentration, 1.0%), and desalted with 50-mg and 10-mg Sep-Paks (Waters) for unenriched and enriched phosphopeptides, respectively. Labeled phosphopeptides were mixed according to protein-level normalization factors rather than phosphopeptide abundance.

Offline basic pH reversed-phase fractionation. TMT-labeled peptides were fractionated by basic pH reversed-phase HPLC with an Agilent 1260 Infinity pump equipped with a de-gasser and a single wavelength detector (set at 220 nm). Using an Agilent 300 Extend C18 column (3.5-µm particles, 4.6-mm inner diameter, and 250-mm length), we separated peptides with a 50-min linear gradient from 8% to 40% acetonitrile (proteome) or 5% to 35% acetonitrile (phosphoproteome) in 10 mM ammonium bicarbonate, pH 8, at a flow rate of 0.6 mL/min. For each experiment, a total of 96 fractions were collected and consolidated into 24 fractions and vacuum-dried. Twelve of the 24 proteome fractions and all 24 phosphoproteome fractions were desalted via StageTip, dried via vacuum centrifugation, and reconstituted in 5% acetonitrile, 5% formic acid for LC-MS/MS processing.

Liquid chromatography and tandem mass spectrometry. All mass spectra were collected on an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific) equipped with a Proxeon EASY-nLC 1000 liquid chromatography system (Thermo Fisher Scientific). Peptides and phosphopeptides were separated on a 75-µm-inner-diameter microcapillary column packed with ~35 cm of Sepax GPC18 resin (1.8 µm, 150 Å; Thermo Fisher Scientific). Approximately 2 µg of peptides and 2.5 µg of phosphopeptides were separated via 3-h and 2-h gradients, respectively, of acidic acetonitrile. We used the multinotch MS3-based TMT method[37].

The proteome-wide analysis (12 fractions) and phosphoproteome-wide analysis (24 fractions) scan sequences began with a full MS spectrum (MS1) acquired in the Orbitrap (resolution, 120,000; mass range, 400–1,400 Th) followed by up to 10 MS/MS spectra (MS2). MS2 settings included collision-induced dissociation (collision energy: 35) with a maximum ion injection time of either 50 ms (proteome) or 200 ms (phosphoproteome) and an isolation window of 0.5 Da. For quantitative information, MS3 precursors were fragmented by high-energy collision-induced dissociation (collision energy: 55) and analyzed in the Orbitrap (resolution of 50,000 at 200 Th) with a maximum ion injection time of either 110 ms (proteome) or 300 ms (phosphoproteome) and an isolation window of 1.2[38].

Mass spectrometry data analysis. Mass spectra were processed via a SEQUEST-based software pipeline[39]. A modified version of ReAdW.exe was used to convert spectra to mzXML. All spectra were searched against a database containing the human proteome downloaded from UniProt in both forward and reverse orientations and common contaminant protein sequences. Database searches were performed with a peptide mass tolerance of 20 p.p.m. and a fragment ion tolerance of 0.9 Da. These wide mass-tolerance windows were chosen to maximize sensitivity in conjunction with SEQUEST searches and linear discriminant analysis[39,40]. TMTs on lysine residues and peptide N termini

(+229.163 Da) and carbamidomethylation of cysteine residues (+57.021 Da) were set as static modifications, and oxidation of methionine residues (+15.995 Da) was set as a variable modification. Phosphoproteome fractions also included variable modifications of phosphorylation on serine, threonine, and tyrosine residues (+79.966 Da).

Peptide–spectrum matches were filtered by linear discriminant analysis to an FDR of 1% at the peptide level on the basis of matches to reversed sequences, as described previously[40]. Linear discriminant analysis considered the following parameters: cross-correlation, ΔC_n, missed cleavages, adjusted p.p.m., peptide length, fraction of ions matched, charge state, and precursor mass accuracy. We considered only peptides containing phosphorylation modifications when setting the FDR for phosphopeptides. Filtered peptides were collapsed further to a final protein-level FDR of 1%. We quantified peptides from MS3 scans after filtering out those with poor quality (required total TMT reporter signal-to-noise ratio > 200 and isolation specificity of 0.5). Localization of phosphorylation sites was determined by AScore, and sites with an AScore > 13 were selected for further analysis. For protein quantitation, the signal-to-noise values for all peptides for a given protein were summed, and each TMT channel was summed across all quantified proteins and normalized to enforce equal protein loading. Prior to downstream analysis, relative protein measurements across the 11 samples were normalized to sum to 100. Hierarchical clustering was performed in R on Euclidean distances via the ward.D2 method.

Statistics and reproducibility. Exact *n* values are defined and described in relevant figure legends. Statistical significance was determined by two-tailed unpaired Student's *t*-test using Prism software (GraphPad), and exact *P* values are indicated in figures/figure legends. Violin plots in Fig. 4a show the distribution of HERVK repeat elements ($n = 64$) for each replicate in each condition. In each violin, the black dot represents the mean normalized expression value, and the black line extending from this point represents one standard deviation from the mean. Violin plots in Fig. 4b show the density of methylation values ($n = 95{,}239$ 1-kb tiles) in each individual cell line. In each violin plot, the white dot represents the median value. The box extends from the 25th to the 75th percentile of the data (the interquartile range). Whiskers extend from the box to the most extreme data point, which is no more than 1.5 times the interquartile range. Representative data are shown only if results were similar for two independent experiments. RNA-seq, proteomic, phosphoproteomic, and RRBS data were generated for each independent hESC line.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Code availability. Code used in this study can be obtained from the corresponding author upon reasonable request.

Data availability. The original RNA-seq and RRBS data have been deposited in GEO under accession number GSE111020. Proteomic data from this study are included in Supplementary Tables 1 and 2.

References

28. Di Stefano, B. & Hochedlinger, K. Reduced MEK inhibition preserves genomic stability in naïve human ES cells. *Protoc. Exch.* https://doi.org/10.1038/protex.2018.062 (2018).
29. Xi, Y. & Li, W. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* **10**, 232 (2009).
30. Woodfine, K., Huddleston, J. E. & Murrell, A. Quantitative analysis of DNA methylation at all human imprinted regions reveals preservation of epigenetic stability in adult somatic tissue. *Epigenetics Chromatin* **4**, 1 (2011).
31. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
32. Treff, N. R. et al. Next generation sequencing-based comprehensive chromosome screening in mouse polar bodies, oocytes, and embryos. *Biol. Reprod.* **94**, 76 (2016).
33. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
34. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
35. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
36. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
37. McAlister, G. C. et al. Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Anal. Chem.* **84**, 7469–7478 (2012).
38. Erickson, B. K. et al. Evaluating multiplexed quantitative phosphopeptide analysis on a hybrid quadrupole mass filter/linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* **87**, 1241–1249 (2015).
39. Huttlin, E. L. et al. A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174–1189 (2010).
40. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214 (2007).

nature research

Corresponding author(s): Konrad Hochedlinger

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.	we used minimal sample size for statistical comparisons

2. Data exclusions

Describe any data exclusions.	No data were excluded

3. Replication

Describe whether the experimental findings were reliably reproduced.	All attempt at replication were successful

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.	Randomization is not relevant to this study because no comparisons between experimental groups were made.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.	For teratoma assay the investigator was blinded to the group allocation (with no prior knowledge about the cell lines and treatments) during the experiments.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☒	☐	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Softwares used in this study for data analysis were described in the method section.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	no unique materials were used

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	Primary antibodies: Anti-KLF17 antibody produced in rabbit (1:100) (HPA024629, SIGMA), GKLF/EKLF/LKLF Antibody (F-8) (1:100) (sc-166238, Santa Cruz), Mouse Anti-Stella Antibody (1:100) (MAB4388, EMD Millipore), Nanog (D73G4) XP® Rabbit mAb #4903 (1:300) (4903S, Cell Signaling), Oct-3/4 Antibody (C-10) (1:300) (sc-5279, Santa Cruz), Annexin V (1:100) (APC, BD Bioscience 550474), ?H2AX (1:100) (Alexa Fluor 647, BD Bioscience 560447), THY-1 (1:100) (PE anti-human CD90 (Biolegend 328110)) and CD75 (1:100) (Purified anti-human CD75 (LN-1, 326901 Biolegend). β-ACTIN (1:10000) (Cell Signaling Technology, clone 13E5, cat. #5125); ERK1/2 (1:1000) (Cell Signaling Technology, 137F5, 4695); p-ERK1/2 (1:1000) (Cell Signaling Technology, D13.14.4E, 4370). Secondary antibodies: Goat anti-Rabbit IgG (H+L) (Alexa Fluor 488, A11008, Thermo Fisher), Goat anti-Mouse IgG (H+L) (Alexa Fluor 546, A11003, Thermo Fisher), Goat anti-Mouse IgM Heavy Chain Secondary Antibody, Alexa Fluor® 647 conjugate (A-21238, Thermo Fisher). Antibodies were validated by the company; refer to the company website for detailed validation analysis.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	Human primed WIBR3 cells were obtained from Dr Rudolf Jaenisch (Whitehead Institute, Boston). UCLA1, UCLA3, UCLA4, UCLA5, UCLA9, UCLA17 hESC lines from Dr. Kathrin Plath lab (UCLA, Los Angeles).
b. Describe the method of cell line authentication used.	UCLA1, UCLA3, UCLA4, UCLA5, UCLA9, UCLA17 hESC lines were authenticated by the Human Embryonic and Induced Pluripotent Stem Cell Core at UCLA. The WIBR3 cell line by the Whitehead Institute.
c. Report whether the cell lines were tested for mycoplasma contamination.	Cell lines were tested weekly for mycoplasma contamination using the LONZA Mycoplasma Detection Kit. All cell lines tested negative for Mycoplasma contamination.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No commonly misidentified cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	9 immunodeficient NOD/SCID mice from Jackson laboratory were used for Teratoma assay.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	This study did not involve human research participants

nature research

Corresponding author(s): Konrad Hochedlinger

☐ Initial submission ☐ Revised version ☒ Final submission

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ 3. All plots are contour plots with outliers or pseudocolor plots.

☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.	In all flow experiments cells were detached using Accutase, centrifuged in DMEM medium containing 10% FBS, filtered through a 0.45uM strained and resuspended in FACS buffer (PBS + 5% FBS). Cell permeabilization was performed using the Fix and Perm Cell fixation and cell permeabilization Kit (ThermoFisher Scientific, GAS003) following the manufacturer instructions.
6. Identify the instrument used for data collection.	LSRII FACS BD Bioscience
7. Describe the software used to collect and analyze the flow cytometry data.	Collection: Diva v6.1.2 (BD Biosciences). Analysis: FlowJo software v10 (TreeStar)
8. Describe the abundance of the relevant cell populations within post-sort fractions.	N/A
9. Describe the gating strategy used.	Cell debris was excluded using a FSC vs SSC gate; aggregates were excluded via a FSC-H vs FSC-A approach; dead cells were defined as DAPI high/positve and gated out. Data were represented as Empty-channel/GFP;Empty-channel/APC or CD75-APC/THY-1-PE

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☐