TFAP2C regulates transcription in human naive pluripotency by opening enhancers

William A. Pastor[1,8], **Wanlu Liu**[1,2], **Di Chen**[1], **Jamie Ho**[1], **Rachel Kim**[1], **Timothy J. Hunt**[1], **Anastasia Lukianchikov**[1], **Xiaodong Liu**[3,4,5], **Jose M. Polo**[3,4,5], **Steven E. Jacobsen**[1,6,7], and **Amander T. Clark**[1,6]

[1]Department of Molecular, Cell and Developmental Biology, University of California Los Angeles, Los Angeles, California 90095, USA

[2]Molecular Biology Institute, University of California, Los Angeles, CA 90095

[3]Department of Anatomy and Developmental Biology, Monash University, Wellington Road, Clayton, VIC 3800, Australia

[4]Development and Stem Cells Program, Monash Biomedicine Discovery Institute, Wellington Road, Clayton, VIC 3800, Australia

[5]Australian Regenerative Medicine Institute, Monash University, Wellington Road, Clayton, VIC 3800, Australia

[6]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California Los Angeles, Los Angeles, California 90095, USA

[7]Howard Hughes Medical Institute, University of California Los Angeles, Los Angeles, California 90095, USA

Abstract

Naïve and primed pluripotent hESCs bear transcriptional similarity to pre- and post-implantation epiblast and thus constitute a developmental model for understanding the earliest pluripotent stages in human embryo development. To identify new transcription factors that differentially regulate the unique pluripotent stages, we mapped open chromatin using ATAC-Seq and found enrichment of the AP2 transcription factor binding motif at naïve-specific open chromatin. We determined that the AP2 family member TFAP2C is upregulated during primed to naïve reversion and becomes widespread at naïve-specific enhancers. TFAP2C functions to maintain pluripotency and repress neuroectodermal differentiation during the transition from primed to naïve by facilitating the opening of enhancers proximal to pluripotency factors. Additionally, we identify a previously undiscovered naïve-specific *POU5F1* (OCT4) enhancer enriched for TFAP2C binding.

[*]Correspondence: jacobsen@ucla.edu (S.E.J.), clarka@ucla.edu (A.T.C.).
[8]Present address: Department of Biochemistry, McGill University, Montreal, Quebec H3G 1Y6, Canada

Taken together, TFAP2C establishes and maintains naïve human pluripotency and regulates OCT4 expression by mechanisms that are distinct from mouse.

Introduction

The broad contours of pre-implantation development are conserved between mice and humans[1]. After fertilization to create the zygote, the embryo undergoes cell divisions, compacts to form the morula, then undergoes further cell division and cavitation to form the fluid-filled blastocyst. At this point, the first three cell types, trophoblast, primitive endoderm, and epiblast are specified, with the epiblast destined to give rise to all embryonic tissues. Upon implantation, the epiblast undergoes dramatic changes in gene expression and epigenetic state, "priming" it to differentiate rapidly in response to external cues. As such the epiblast transitions from the *naïve* pluripotent state to the *primed* pluripotent state. Gastrulation then occurs and pluripotency is lost altogether.

Despite this similar overall program, it has become clear that there are dramatic molecular differences between mouse and human embryo development[2–8]. However, given the significant limitations in research using human embryos, it has not been possible to rigorously compare the murine and human naïve epiblast.

In humans, the traditional approach for deriving and culturing human ESCs (hESCs) from pre-implantation embryos results in cells with primed pluripotency similar to EpiSCs. However, new media formulations for transitioning or deriving hESCs in the naïve state have now been developed[9,10]. Critically, naïve hESCs largely recapitulate the transcriptional and epigenetic program of human pre-implantation epiblast cells[6,11,12]. Therefore, naïve and primed hESCs are the only human cell-based models for understanding the critical fate transition between naïve and primed pluripotency in the human embryo and the contrast between murine and human epiblast.

Results

AP2-motifs are strongly enriched in naïve-specific regulatory elements

To identify transcription factors critical for naïve human pluripotency, we mapped open chromatin using assay for transposase-accessible chromatin (ATAC-seq[13]) in naïve and primed hESCs (Supplementary Figure 1A, Supplementary Table 1). Cells were cultured in 5 inhibitors plus LIF, Activin A, and FGF2 (5iLAF) to recapitulate the naïve state and with FGF2 and Knockout serum replacement media (KSR) to recapitulate the primed state[9,12]. As expected, we observed strong enrichment of open chromatin at gene promoters (Supplementary Figure 1B), with enrichment associating with gene expression. We defined sets of ATAC-seq peaks in naïve and primed hESCs, as well as peaks specific to either the naïve or primed states (Supplementary Figure 1C, Supplementary Table 2, and Materials and Methods). While all sets showed enrichment of promoter sequence, this enrichment was much weaker for naïve and primed-specific open sites (Supplementary Figure 1C), consistent with the general trend that enhancer utilization rather than promoter openness is more variable between different cell types[14,15].

Broadly, we observed a strong correlation between the appearance of naive-specific ATAC-seq peaks near a gene, and up-regulation of that gene in the naïve state, and between the appearance of a primed-specific ATAC peak near a gene and down-regulation in the naïve state (Figure 1A,B, Supplementary Figure 1D,E). This was true whether the ATAC-peak was upstream or downstream of the gene TSS (Supplementary Figure 1E,F). For example, naïve-specific ATAC peaks are observed in the vicinity of the naïve-specific *Krupple life factor 5* (*KLF5*) while, primed-specific ATAC-seq peaks are observed in the vicinity of the primed-specific genes *ZIC2* and *ZIC5* (Figure 1C,D). These observations are consistent with a high proportion of ATAC-seq peaks corresponding to enhancers that regulate nearby genes. Comparison to published ChIP-seq data in naïve and primed hESCs[16] revealed enrichment of Mediator over naïve and primed specific ATAC-seq peaks in the corresponding cell type, and we observed strong enrichment of H3K27Ac at the boundaries of these peaks, with a dip in the middle likely explained by nucleosome depletion (Figure 1E). Mediator and H3K27Ac enrichment are predictive features of active enhancers[17,18], further validating the ATAC-seq peaks as regulatory elements.

To identify transcription factors critical for the activity of enhancers in the naïve and primed states, we determined enrichment of known transcription factor binding motifs in the naïve and primed-specific ATAC peaks (Figure 1F, G)[19]. The strongest statistical enrichment in the naïve state corresponded to the KLF motif, consistent with the strong up-regulation of KLF family factors in naïve hESCs and the known role for KLF in naïve-state pluripotency in mouse and human[9,10,20]. Likewise, the motif of the primed-specific[21] ZIC factors was enriched in the primed peaks. Unexpectedly, very strong enrichment for the AP2 transcription factor motif was observed for naïve-specific open chromatin. AP2 transcription factors have been implicated in a number of developmental processes in mice, including placental development[22–24], neural crest development[25] and ectodermal patterning[25–27], but are completely dispensable for murine epiblast formation and mouse pluripotent cell survival[22–24,28]. Hence, there may be a human-specific role for an AP2 factor in the naïve state.

Naïve-specific regulatory elements are present *in vivo*

To determine the *in vivo* relevance of our set of naïve-specific ATAC peaks, we performed ATAC-seq on eight pooled pre-implantation human blastocysts (Supplementary Table 3). We found dramatically increased openness in the human blastocyst over naïve-specific peaks, both relative to surrounding sequence and relative to primed peaks (Figure 2A–C, Supplementary Figure 2A), validating the biological relevance of these peaks. Nonetheless, there were marked differences between the open chromatin patterns in whole blastocyst and naïve hESCs. We reasoned that this was because the day 6 human blastocyst consists primarily of trophoblast, with a much smaller fraction of epiblast and hypoblast[29]. For example, we found that blastocyst showed lower ATAC-seq enrichment in the vicinity of the epiblast-specific gene Nanog but higher enrichment in the vicinity of the trophoblast-specific GATA3 (Figure 2D,E). This trend was apparent when we plotted ATAC-enrichment over epiblast and trophoblast-specific gene bodies as defined from published RNA-seq data (Figure 2F,G).

We found AP2 and KLF motifs strongly enriched in blastocyst and naïve hESC chromatin, consistent with reported activation of AP2 and KLF-family transcription factors in morula and continued expression in human epiblast and trophoblast (Figure 2H,I). GATA TF motif was strongly enriched in blastocyst-specific chromatin while OCT4-SOX2 motif was strongly enriched in naïve and primed hESCs, consistent with preferential expression of GATA2 and GATA3 in the trophoblast and OCT4 in the inner cell mass and epiblast. Our data thus strongly support the idea that naïve hESCs have an open chromatin state similar to pre-implantation epiblast.

Using an alternate approach we further confirmed the *in vivo* relevance of the naïve-specific ATAC peaks by analyzing DNA methylation, given that regulatory elements are typically hypomethylated relative to surrounding sequence[30,31]. Consistent with this trend, we observe strong hypomethylation of naïve-specific ATAC-seq peaks in naïve hESCs cultured in 5iLAF or in t2iLGö[10], a different culture method for generating naive hESCs[6,12] (Supplementary Figure 2B). Likewise, we observe a more pronounced drop in DNA methylation between the oocyte and blastocyst stages of human embryonic development at our defined set of naïve-specific ATAC peaks than over surrounding sequence or primed-specific ATAC peaks (Supplementary Figure 2C,D). Thus, multiple lines of evidence support the proposition that the majority of putative regulatory elements identified in naive hESCs correspond to a hypomethylated regulatory element in human pre-implantation embryos.

TFAP2C supports reversion to the human naïve state

Of the five AP2-family transcription factors present in humans, only TFAP2C is highly expressed in the naïve state (Figure 3A). TFAP2C is upregulated in naïve cells at both the RNA and protein level (Figure 3A, B) and is expressed in human morula and pre-implantation epiblast[3,8,32]. ChIP for TFAP2C showed strong enrichment over naïve-specific ATAC-seq peaks (Figure 3C, D), especially those containing AP2 motifs (Figure 3D). Furthermore, TFAP2C showed stronger enrichment at naïve-specific ATAC-seq peaks than at regions open in both the primed and naïve state (Figure 3E), even though both ATAC peak sets show similar ATAC-seq enrichment in the naïve cells (Figure 3F). Combined with our observation that AP2-motifs are specifically enriched in naïve-specific peaks, this data indicates that TFAP2C may facilitates the opening of regulatory elements during reversion.

We used CRISPR to target TFAP2C in the primed state, and lines containing null mutations of both alleles were confirmed karyotypically normal (Supplementary Figure 3A, B). In the primed state these lines showed normal expression of pluripotency genes and markers (Supplementary Figure 3C–F). The *TFAP2C$^{-/-}$* cells were able to exit pluripotency normally with spontaneous embryoid body (EB) differentiation and showed a skew toward neural lineage, consistent with a known role for TFAP2C in regulating the formation of neural vs. non-neural ectoderm[27] (Supplementary Figure 3G).

Upon reversion in naïve 5iLAF media, we observed a dramatic morphological change in the *TFAP2C$^{-/-}$* hESC lines (Figure 4A, Supplementary Figure 4A, B). Consistent with this rapidly-emerging phenotype we discovered that at day 3 of reversion, TFAP2C protein is strongly induced in the control cells, while being absent in the *TFAP2C$^{-/-}$* lines (Figure 4B, C). TFAP2C ChIP and ATAC-seq at day 5 of reversion shows opening of naïve-specific

enhancers and enrichment of TFAP2C at these sites but no opening of these enhancers in *TFAP2C*[-/-] (Figure 4D,E). Initially, the *TFAP2C*[-/-] cells divided rapidly in 5iLAF, but after one passage, round naïve-like colonies were identified only in controls (Figure 4A). Instead, sparse clusters of small cells were observed in the *TFAP2C*[-/-] lines after the first passage that ceased to divide and disappeared from culture. By day 5 of reversion, the *TFAP2C*[-/-] cells showed dramatic loss of pluripotency factors and upregulation of neural lineage factors (Figure 4F–J), a result confirmed by GO-analysis[33] (Supplementary Figure 4C). Likewise ATAC-seq showed a loss of AP2 and pluripotency transcription factor motifs in open chromatin in the *TFAP2C*[-/-] cells after five days of reversion and instead, a gain of peaks enriched for motifs related to neural development such as SOX and ZIC (Supplementary Figure 4D, E).

To confirm that this finding was human specific, we performed ATAC-seq on mESCs cultured in the naïve state (2i+LIF) as well as primed Epiblast Like cells (EpiLCs)[34] (Supplementary Table 4). We discovered that AP2 sites were enriched in naïve-specific open chromatin in 2i + LIF mESCs. However, the degree of enrichment was far lower than human naïve cells (Figure 4K, L). We generated *Tfap2c*[-/-] and *Tfap2a*[-/-]*Tfap2c*[-/-] mESCs (Supplmentary Figure 4F,G) and found normal expression of pluripotency markers in 2i+LIF (Figure 4M). Furthermore, comparing ATAC-seq in control and *Tfap2a*[-/-]*Tfap2c*[-/-] double knockout mESCs, we found only 373 control-specific ATAC-seq peaks, and this set was only moderately enriched for AP2 sites (Figure 4N). Thus, AP2 transcription factors play a more modest role in murine than human naïve states.

Withdrawal of TFAP2C in naïve state causes shift toward primed state

Next, we generated a *TFAP2C*[-/-] mutant line capable of expressing TFAP2C in a doxycycline-dose dependent manner (Figure 5A). Overexpression of TFAP2C in primed media did not result in a pronounced shift toward naïve-gene expression, and TFAP2C primarily honed to regions of chromatin that were already open in primed hESCs (Supplementary Figure 5A,B), arguing that the combinatorial activity of multiple factors is necessary for primed to naïve reversion.

We then reverted the *TFAP2C*[-/-] Dox-inducible line, using 5iLAF media supplemented with various quantities of doxycycline. Induction of TFAP2C rescued the morphological abnormality observed in the mutant, preserved OCT4 expression, repressed SOX1 induction, and allowed formation of colonies with naïve-morphology (Figure 5B–E, Supplementary Figure 5C).

To determine the effect of acute loss of TFAP2C, we cultured cells in 5iLAF + doxycycline until naïve morphology colonies were apparent, then switched to media without doxycycline. No acute phenotype was observed, instead a gradual loss of cells from culture occurred (Supplementary Figure 5D,E). Cells remaining 12 days after doxycycline withdrawal showed increased staining for the primed surface marker SSEA4 and closing of naïve-specific ATAC peaks, especially the subset containing AP2 sites but no KLF sites (Figure 5F, Supplementary Figure 5F). These findings indicate that TFAP2C is essential for maintenance as well as establishment of the naïve state.

TFAP2C$^{-/-}$ in low O$_2$

Because low oxygen conditions can stabilize the pluripotent state and promote human embryogenesis[35], we conducted two independent reversions in 5% oxygen. Similar to the results obtained with reversions under ambient (~20%), morphological differences, loss of OCT4, and gain of SOX1 and PAX6 were all apparent upon culture in 5iLAF in 5% O$_2$ (Figure 6A,B, Supplementary Figure 6A). However, approximately two weeks after onset of culture in 5iLAF under 5% O$_2$ conditions, round colonies appeared in the *TFAP2C$^{-/-}$* cultures and these putative colonies were capable of self-renewal (Figure 6B, Supplementary Figure 6A). However, almost all *TFAP2C$^{-/-}$* cells had high SSEA4 surface expression (Figure 6A), consistent with primed identity [12]. The second reversion featured a substantial population of cells with SSEA4 negative identity, but these cells showed gain of neural and loss of pluripotency markers, indicating that they were differentiated rather than naive (see RPKM in Supplementary Table 5). ATAC-seq of *TFAP2C$^{-/-}$* cells persisting in 5iLAF under 5% O$_2$ showed reduced openness over naïve-specific peaks, and increased openness over primed-specific ATAC-seq peaks compared with controls (Figure 6C). Moreover, principle component analysis of the ATAC-Seq data sets showed a closer similarity to primed cells (Figure 6D), with the transcriptome of the persisting *TFAP2C$^{-/-}$* cells present in 5% O$_2$ shifted toward expression of primed-specific genes (Figure 6E, Supplementary Figures 6B,C). In further support of the finding that persistent TFAP2C$^{-/-}$ colonies in 5% O$_2$ are more primed-like, we compared the RNA-Seq to published primate RNA-seq[6] and found a global reduction in genes specific to pre-implantation epiblast and an increase in genes specific to post-implantation epiblast (Supplementary Figures 6D,E). Finally, we reverted the *TFAP2C$^{-/-}$* in t2iLGöY naïve media[36] in 5% O$_2$, and similar to the results in 5iLAF, the *TFAP2C$^{-/-}$* cells lacked nuclear KLF17, a marker of naïve cells (Supplementary Figure 6F, G). In total, these data support an essential role for TFAP2C in reversion of primed hESCs to the human naïve state.

TFAP2C promotes expression of pluripotency genes

The simple presence of a transcription factor at a locus does not prove a role in regulating nearby genes, and we observe 14,367 distinct TFAP2C peaks throughout the genome (Figure 7A,B), making it difficult to discern which binding events are important for gene regulation. Compared with the striking correlation observed between presence of a naïve-specific enhancer and upregulation of a nearby gene (Figure 1A), we observed only a modest correlation between the presence of a TFAP2C ChIP-peak near a gene and the up-regulation of that gene in the naïve state or downregulation in *TFAP2C$^{-/-}$* (Figure 7A). To the extent an effect was discernable, the presence of a TFAP2C peak at an enhancer adjacent to the gene was predictive of upregulation in the naïve state, but the presence of a TFAP2C peak at a gene TSS had a very little effect on the expression of that gene, which was surprising given that the promoter is a key site of gene regulation.

We therefore sought to identify direct targets of TFAP2C by combining RNA-seq, ATAC-seq and ChIP-seq data. First, we looked at the set of genes specific to the naïve or primed state and focused on the subset that showed >4-fold changes in expression in *TFAP2C$^{-/-}$* (Figure 7C). Second, we defined a set of TFAP2C-dependent regulatory elements: TFAP2C ChIP-seq peaks that overlapped with naïve-specific ATAC-peaks and showed reduced openness in

TFAP2C⁻/⁻ (Figure 7D,E). We found an extremely strong relationship between downregulation of a gene in *TFAP2C⁻/⁻* and the presence of a TFAP2C-dependent regulatory element nearby (Figure 7F, Supplementary Table 6). The vast majority of TFAP2C-dependent regulatory elements did not overlap with a gene TSS and were thus likely to be enhancers rather than promoters (Figure 7G). By contrast, TFAP2C ChIP-seq peaks in regions of openness conserved between naïve and primed state, had virtually no predictive effect on gene expression in *TFAP2C⁻/⁻* (Supplementary Figure 7A,B). In other words, TFAP2C's primary effect in naïve hESCs is most likely to open a discreet set of regulatory elements, mainly enhancers (Figure 7H).

GREAT analysis[37] showed that genes within 50kb of a TFAP2C-dependent regulatory element were upregulated in Theiller stage 3 and 4 embryos (morula and early blastocyst) and that mutations of these genes were associated with abnormal embryogenesis (Supplementary Tables 6,7). Adjacent genes included *CBFA2T2, TFCP2L1, KLF5, SOX2, FGF4, NANOG, DPPA3, DPPA5* and *TFAP2C* itself (Figure 7I, Supplementary Figure 7C–E, Supplementary Table 7), supporting a role for TFAP2C in directly promoting the naïve pluripotent program.

An intronic enhancer for OCT4 is active in naïve hESCs

One of the characteristic properties that distinguishes naive and primed states is different enhancer utilization at *POU5F1* (OCT4). In mouse, the proximal enhancer upstream of *POU5F1* is critical for expression in the post-implantation epiblast while the distal enhancer farther upstream drives expression in primordial germ cells and ICM[38]. In human pre-implantation blastocyst however, neither enhancer appears open, whereas two putative enhancers appear downstream of the *POU5F1* TSS (Figure 8A). Each of these peaks contains a cluster of AP2 sites and a KLF site, indicating that they could be opened by the combinatorial activity of these transcription factors during pre-implantation development (Figure 8B). "Intron element 1" shows evolutionary conservation across placental mammals (Supplementary Figure 8A) and is open and enriched for TFAP2C in naïve hESCs (Figure 8A,B) but is not open in naïve mESC (Figure 8C). We do not observe any reads emanating from this element spliced into the *OCT4* transcript, ruling out the possibility that it is actually an alternative promoter (Supplementary Figure 8B). Furthermore, we observe enhancer activity for this region in a luciferase assay, which is largely eliminated by the loss of either the AP2 or KLF site (Figure 8D).

To examine the role of this enhancer in naïve pluripotency, we ablated this sequence using CRISPR/Cas9 and confirmed normal karyotype (Figure 8B, Supplementary Figure 8C). We found normal expression of OCT4 (Supplementary Figure 8D) and self-renewal in the primed state, but a dramatic loss of OCT4 expression accompanied by differentiation upon reversion to the naïve state (Figure 8E,F). This indicates a potential direct role for TFAP2C in regulating the pluripotency master-regulator OCT4 by binding to a previously unknown enhancer, which in turn is likely to be important for pre-implantation OCT4 expression.

Discussion

We present strong evidence that TFAP2C is critical for the opening of a set of enhancers in naïve hESCs. Furthermore, we show that most of these enhancers are present in human embryo and therefore biologically relevant, and likely to directly regulate genes critical for human naïve pluripotency.

TFAP2C has been implicated in both activation and repression of target loci[39–41], which may explain the limited effect of TFAP2C at promoters where it is already present. However, the enrichment of AP2 motifs in naïve-specific ATAC-peaks, the failure of many of these enhancers to open in the absence of TFAP2C, and the strong association between TFAP2C-dependent enhancers and expression of nearby genes is indicative of a critical role for TFAP2C in regulating gene expression by opening enhancers. TFAP2C is known to interact with members of the CITED family of proteins, which in turn recruit the histone acetyltransferase p300[42–44], suggesting a model in which TFAP2C facilities enhancer opening by promoting histone acetylation. Because TFAP2C is expressed in the morula before blastocyst formation, it could have a role in resetting the chromatin landscape prior to the establishment of naïve pluripotency, analogous to what happens in the artificial system of *in vitro* reversion.

The observation that TFAP2C is critical in naïve hESCs *in vitro* would lead us to predict that TFAP2C is critical for gene regulation in pre-implantation epiblast *in vivo*. This is surprising in light of results in mouse, where TFAP2C is clearly dispensable for ICM and epiblast specification. *Tfap2c* homozygous null mice develop to the blastocyst stage[22–24], as do mutants generated using *Tfap2c*[fl/fl] Zp3-Cre in which the maternal *Tfap2c* transcript is absent[23]. *Tfap2c* deficient mESCs have been successfully derived from embryos[22,28] and generate viable mice in tetraploid complementation[22], indicating that the gene is non-essential in ICM. Redundancy with other AP2 factors is unlikely to explain this non-essential role, as *Tfap2a*[−/−] *Tfap2c*[−/−] double mutant embryos also develop an epiblast, and the other AP2 factors are expressed at very low levels in morula and blastocyst[23]. The major role for Tfap2c in mouse pre-implantation embryo development is specification and differentiation of trophoblast, with Tfap2c null mutant mice dying from placental defects[45]. Notably, while Tfap2c is strongly enriched in the trophoblast relative to ICM in mouse blastocysts, human ICM and pre-implantation epiblast retain high levels of TFAP2C[3,8,32]. Tfap2c has also been reported in porcine ICM[46], indicating that loss of Tfap2c from the ICM may be specific to mice. TFAP2C direct targets in human embryos include both genes general to pre-implantation embryo as well as genes specific to epiblast such as *CBFA2T2, FGF4*, and *MEG3*.

TFAP2C-dependent regulation of OCT4 may also be different in mouse and human, as is role of OCT4 itself. In mice, OCT4 is essential for pluripotency and for repression of trophoblast genes in the ICM[47]. CRISPR ablation of OCT4 in human embryos by contrast results in outright failure to form blastocyst or express genes associated with trophoblast or epiblast lineage[48]. Thus, OCT4 plays an essential role in humans as early as morula. Our data are consistent with a model in which OCT4 expression is initially regulated by TFAP2C and KLF family TFs via the intronic enhancers and only later is regulated from the naïve-

specific distal enhancer. However, alternative possibilities cannot be ruled out, such as the distal enhancer being active in morula and decommissioned in trophoblast, which makes up the bulk of early blastocyst.

Morphological and molecular evidence supports the phenomenon of the "developmental hourglass", the idea that the developmental program is actually most evolutionarily conserved in mid-embryogenesis, and both early and late stages of development feature high levels of variation across different species[49,50]. The discovery of a human-specific naïve pluripotency factor fits into this paradigm, and therefore model organisms may only reveal some of the story of how human embryos develop.

Methods

Human cell culture

Culture of primed and naïve hESCs, and reversion from primed to naïve state, were conducted as previously reported[12], the only modification being the inclusion of 1x Primocin (Invivogen) in all medias. All cells were cultured in 5% CO_2 and ambient oxygen unless otherwise indicated. Where indicated, doxycycline was added.

The reversion of primed hESCs to naïve t2iLGöY state was performed as previously described[36], generating the t2iLGöY media as described in a previous report[51]. All cell lines were cultured in a 37°C, 5% O_2 and 5% CO_2 for the t2iLGöY reversion experiments.

Murine Cell culture

During routine passage and CRISPR editing, murine ESCs were cultured in Serum+LIF media: 15% Hyclone FBS (ThermoFisher), 1x Penicillin/Streptomycin/Glutamine (ThermoFisher), 1x Non-essential amino acids (ThermoFisher), 55µM β–mercaptoethanol (ThermoFisher), 1xPrimocin (Invivogen) and 1000U/mL ESGRO LIF (Millipore) in KnockOut DMEM (ThermoFisher). Cells were passaged with 0.25% Trypsin every three days and cultured on MEFs. Prior to all RNA-seq or ATAC-seq experiments, cells were cultured for at least five passages in 2i+LIF media: 1x N2 supplement, 1xB27 supplement, 1x Penicillin/Streptomycin/Glutamine (ThermoFisher), 1x Non-essential amino acids (ThermoFisher), 55µM β–mercaptoethanol (ThermoFisher), 1xPrimocin (Invivogen), 3µM CHIR99021 (Stemgent), 0.5µM PD0325901 (Stemgent) and 1000U/mL ESGRO LIF (Millipore) in a 50%/50% mixture of DMEM/F12 without HEPES (ThermoFisher) and Neurobasal media (ThermoFisher). Cells passaged in 2i+LIF were passaged every three days with 0.25% trypsin and plated at 50k cells/well onto wells pretreated with poly-L-Ornithine (Sigma) and Laminin.

Murine EpiSCs were a gift from P. Tesar and were cultured in Primed hESC media[12]. EpiSCs were passaged with 1xCollagenase Type IV (Life Technologies) every three days.

Collecting cell populations for sequencing experiments

For sorting primed and naïve hESCs in steady state for RNA-seq and ATAC-seq TRA-1-85+ SSEA4+ and TRA-1-85+ SSEA4− cells were sorted as previously described[12]. For the first replicate of RNA-seq from day 5 reversion cells, MEFs were removed by twice plating the

cells for five minutes on a gelatinized plate to allow MEFs to attach, but for the second replicate of RNA-seq and for ATAC-seq, MEFs were removed by sorting for all TRA-1-85[+] cells. The isolated human cells were then processed for sequencing as discussed below

To separate mESCs or EpiLCs from MEFs, cells were detached with 1xTrypsin, quenched and then washed with 1xFACS buffer, stained with 1:150 anti-SSEA1 in 1xFACS buffer, then washed and stained with DAPI immediately prior to sort. SSEA1[+] SSC[lo] DAPI[−] cells were sorted and then used for RNA-seq or ATAC-seq.

For human or murine ChIP experiments, cells were harvested using Accutase, quenched, and washed with 1xPBS prior to fixation.

Embryoid body differentiation

Primed human embryonic stem cells 7 days after plating were washed with 1xPBS and treated with 1xCollagenase Type IV at 37°C for 1hr, then removed from the plate with short strokes by a cells scraper. 4mL of MEF media (see above) was added to the well and the colonies were allowed to settle in a 50mL conical tube. Media was then aspirated by pipet, and the cells were resuspended in 3mL mTESR media with ROCKi and plated in a low attachment 6-well plate. At the 24 hour time point, the EBs were transferred into a 50mL conical tube and allowed to settle. Media was removed and replaced with primed hESC media[12] lacking FGF. Media was changed again at hour 72 and the embryoid bodies were harvested at the 144 hour timepoint.

Embryo isolation and generation

Day 6 vitrified blastocysts were thawed using Vit Kit-Thaw (Irvine Scientific) according to manufacturer protocol. Embryos were cultured in drops of Continuous Single Culture media (Irvine Scientific) supplemented with 20% Serum Substitute Supplement (Irvine Scientific) under mineral oil for 2–3 hours at 37°C, 6% CO_2 and 5% O_2. Embryos with good morphology were used for ATAC-seq.

Targeting of loci with guide RNA

Guide RNA were designed using crispr.mit.edu. The highest scoring appropriately situated gDNA sequences were used, with bases removed from the 5'end as necessary so that the guide RNA sequence started with the base "G". Human TFAP2C was targeted with the guide sequence GCTTAAATGCCTCGTTAC. The Human OCT4 intronic enhancer was targeted with the guides GGCACCCCTTGTAGAAAGC and GTAATGAGTGACCAGACCCT. Murine TFAP2C was targeted with the guide sequence GTTACTTTGTACTTCGACG. Murine TFAP2A was targeted with GGGACTATCGGCGGCACG.

CRISPR Editing of hESCs

UCLA1 hESCs[52] were cultured at least two passages in mTESR1 media (StemCell Technologies) on Matrigel (Corning). Cells in exponential growth phase were harvested with Accutase, and 800k hESCs were electroporated with 4μg plasmid DNA using the CA-137, Primary Cell 3 program in the Amaxa 4-D Nucleofector X-subunit (Lonza). After

transfection, cells were transferred to a single well of a 24-well plate containing primed hESC media (see above) supplemented with 1x Y-27632 (Stemgent). Prior to transferring the electroporated cells, the 24-well plate was coated with gelatin and with MEFs. The hESCs were cultured in MEFs in all later steps.

For generation of the TFAP2C deficient hESC lines, cells were passaged with Accutase and plated on 10cm plates for colony picking the day after transfection. This resulted in heterogenous colonies, probably because CRISPR-mediated cleavage continued after single cells were plated for colony-picking, requiring later subcloning. Pure *TFAP2C*$^{-/-}$ lines were only generated later by subcloning. The OCT4 Intronic Enhancer line was plated with on 10cm plates three days after transfection, and did not require later subcloning.

To obtain clonal and physically separate colonies, cells were harvested with Accutase and 10k cells were plated on 10cm plates to allow physically separated colonies to grow. Cells were fed with primed hESC media starting two days after plating and media was subsequently changed every day. Nine to 11 days after plating, colonies were scored with a syringe and the pieces were transferred to a 24-well plate, where they were allowed to grow an additional 6–7 days in primed hESC media. Cells were then split with Accutase. Two thirds of the material was used for DNA extraction and screening (see *Screening for mutations* below), the remaining third was plated in primed hESC media with ROCKi in a well of a 12-well plate. After two days, media was changed to primed hESC media without ROCKi and the cells were passaged using normal primed conditions described above in the *cell culture* subsection above.

For the *TFAP2C*$^{-/-}$ lines, a further round of colony picking, expansion and genotyping was conducted to generate pure knockout populations. Both *TFAP2C*$^{-/-}$Lines 1 and 2 were generated from the same round of transfection of UCLA1 hESCs. Control Line 1 was generated by transfection of UCLA1 hESCs with pMaxGFP plasmid and no CRISPR construct, with cloning and subcloning performed in parallel.

CRISPR Editing of mESCs

Murine ESCs were plated the day before transfection at a density of 150k cells per well in a 6-well plate for each transfection sample. On day of transfection, cells were harvested with trypsin, precipitate, and then resuspended in 2.5mL of serum+LIF media (see above).

In a separate tube, 5μg of DNA (1.43μg GFP plasmid + 3.57μg CAS9/gDNA construct, 5μg GFP plasmid for controls) was diluted to 375μL with Opti-MEM media (ThermoFisher). In another separate tube 12.5μL of Lipofectamine 2000 (ThermoFisher) was combined with 375μL Opti-MEM. The Lipofectamine/Opti-MEM solution was incubated 5 minutes, combined with the DNA solution, and incubated a further 20 minutes at room temperature. The DNA/Lipofectamine/Opti-MEM mix was added to the suspended cells and the cells were rotated for 4 hours at 37°C. The transfected cells were then spun down, resuspended in fresh serum+LIF media, and plated on MEFs.

After 48 hours, cells were harvested with Trypsin and GFP$^+$ cells were sorted and cultured in 96-well plates on MEFs. Three days after sorting, media was changed. Six days after

sorting, wells with colonies were split with trypsin and split onto 24-well plates with MEFs. After another three days, cells were split again, with 12.5% of the cells split onto a 24-well plate with MEFs to propagate the line, 25% split onto a gelatin-treated plate without MEFs to grow cells for DNA extraction, and the rest frozen to create stocks. After another three days, the cells on gelatin were harvested for DNA extraction.

To obtain pure clonal population, the targeted mESCs were later subcloned by sorting for individual SSEA1[+] cells and plated.

Screening for mutations

DNA was extracted using the Quick gDNA Miniprep kit and the region containing the targeted allele was amplified by PCR. To screen human $TFAP2C^{-/-}$ and murine $TFAP2C^{-/-}$ and $TFAP2A^{-/-}$ $TFAP2C^{-/-}$ mutant lines, the Surveyor Mutant Detection Kit (IDT) was used to identify point mutants, though some point mutations in the murine lines were large enough to be apparent by agarose electrophoresis even without Surveyor cutting. For targeting of the OCT4 Naïve Enhancer, mutant alleles were identified based on the reduced size of the targeted region.

To determine the identity of the mutations and confirm clonality of the targeted lines, several strategies were undertaken. First, bulk PCR product was subjected to Sanger sequencing, to determine if there was any visible trace from WT product. Second, PCR product was cloned into the TopoTA vector and at least eight clones sequenced to identify the mutations in both alleles and confirm no WT allele. Third, for human $TFAP2C^{-/-}$ and murine $Tfap2c^{-/-}$ lack of protein was confirmed by Western blot. For the OCT4 Intron element targeting, clonal deletion was also confirmed both by the lack of a WT-sized band in the initial screening PCR and by the failure to amplify with primers internal to the deleted region.

Generation of Doxycycline inducible line

TFAP2C was cloned into a construct facilitating expression under a doxycycline inducible (tetON) promoter, followed by an auto-cleaving "2A" linker and Red Fluorescent Protein (RFP) to allow detection. This tetON-TFAP2C-2A-RFP construct was made by cloning TFAP2C-2A-RFP to replace the hNANOG in FUW-tetO-lox-hNANOG (Addgene 60849). VSVG coated lentiviruses including tetON-TFAP2C-2A-RFP and FUW-lox-M2rtTA were generated in HEK293T cells. TFAP2C mutant line 1 hESCs were treated with Accutase to make single cell suspension in 100uL hESC media with ROCKi with 100,000 single cells. Cells were transduced with 1:1 ratio of tetON-TFAP2C-2A-RFP and FUW-lox-M2rtTA and plated in 10cM dishes at different concentrations. Individual colonies were picked and genotyped for tetON-TFAP2C-2A-RFP and FUW-lox-M2rtTA.

Reporter assay

The OCT4 intronic enhancer (hg19 chr6 31,137,269-31,137,697) was amplified and cloned into pGL3 Promoter vector (Promega). Versions with the three AP2 sites (hg19 31,137,370-31,137,378, 31,137,529-31,137,537 and 31,137,547-31,137,555) or KLF site (hg19 31,137,477-31,137,485) deleted were synthesized by Genewiz and cloned into pGL3. 200,000 naïve UCLA19n (first replicate) or UCLA20n (second replicate) hESCs were then

transfected with 800ng of either empty pGL3 promoter vector or one of the three constructs described above, along with 200ng of pRL-TK Renilla vector as a transfection efficiency control. We used Amaxa nucleofection with P3 buffer and the program CA-137 and plated the cells onto a 12-well plate well of MEFs. The cells were then detached from the well with Accutase and lysed, and luminescence was detected using the Dual-Glo Luciferase system (Promega).

Immunostaining

For Figures 4I and J, immunofluorescence was conducted as published[12], with using anti-TFAP2C (SantaCruz 8977, 1:100), anti-OCT4 (sc8628-X, 1:100), anti-PAX6 (R&D Systems AF3369, 1:100). For Supplementary Figure 6F, immunostaining was performed as previously described[51]. Primary antibodies used: rabbit anti-KLF17 polyclonal (1:500, Sigma HPA024629), mouse anti-TRA-1-60 IgM (1:300, BD), and secondary antibodies used: goat anti-rabbit IgG AF555 secondary (1:400, ThermoFisher), goat anti-mouse IgM AF488 secondary (1:400, ThermoFisher).

Western blotting

Western blots and quantitation with the Odyssey Infrared Imager (Licor) were conducted as described previously[12]. Antibodies used include anti-OCT4 (SantaCruz sc8628)**,** NANOG (R&D Systems AF1997), TFAP2C (SantaCruz sc8977 and Abcam ab76007), SOX1 (R&D Systems AF3369), SOX2 (R&D Systems AF2018), PAX6 (R&D Systems AF2085). Western signals were normalized to signal from anti-H3 antibody (Abcam ab10799 or Abcam ab1791). All antibodies were used at concentrations of 1:1000 except Santa Cruz anti-TFAP2C (1:700) and H3 (1:3000).

RNA isolation and library generation

RNA was isolated using the RNeasy Mini Kit (Qiagen). 5ng to 50ng total RNA input was used to generate sequencing libraries using the Ovation Ultralow Library System V2 (Nugen) and then Ovation Rapid Library System (Nugen) protocols.

ATAC-seq library preparation

In all experiments using cultured cells, between 25k and 50k sorted cells were subjected to ATAC-seq as previously reported[53]. To perform ATAC-seq on embryos, the embryos were incubated in the reported ATAC-seq lysis buffer for ten minutes, during which they were vortexed for ten seconds every three to four minutes, after which the protocol was conducted identically to the previous report.

ChIP protocol and library generation

Cells were fixed with 1% paraformaldehyde (Sigma) and incubated with rotation for 10 minutes at room temperature. The paraformaldehyde was quenched by adding glycine to a final concentration of 0.14M and rotated another 10 minutes at room temperature. The cells were then centrifuged 735g for 5 minutes and then flash frozen with liquid nitrogen and stored at −80°C until ChIP was conducted.

To lyse the cells for ChIP, cells were thawed and resuspended with 1mL lysis buffer (10mM TrisHCl pH 8.0, 0.25% Triton X-100, 10mM EDTA, 0.5mM EGTA, 1x Protease inhibitors (Roche) and 1mM PMSF), then rotated 15 minutes. Nuclei were pelleted by centrifugation at 1500g for 5 minutes at 4°C. Nuclei were then resuspended with 1mL 10mM TrisHCl pH 8.0, 200mM NaCl, 10mM EDTA, 0.5mM EGTA, 1x Protease inhibitor, 1mM PMSF and rotated 10 minutes. Nuclei were then pelleted and resuspended in 650μL 10mM TrisHCl pH 8.0, 10mM EDTA, 0.5mM EGTA,1x Protease inhibitor, 1mM PMSF and sonicated in a 12mm × 12mm Sonication Tube (Covaris) in a Covaris S2 (Intensity = 5; Cycles/burst = 200; Duty Cycle = 5%; 8× (30″ on/30″ off) for four minutes effective sonication). The sonicated lysate is then centrifuged for 10 minutes at 14200g, and the supernatant retained. 10% of the supernatant is saved as "Input", the rest is used for ChIP.

30μL of Protein A Dynabeads (ThermoFisher) are washed three times with ChIP buffer (16.7mM TrisHCl pH 8.0, 0.01% SDS, 1.1 Triton X-100, 1.2mM EDTA, 167mM NaCl), each wash consisting of addition of 1mL of buffer and collection of the beads on a magnetic rack (Diagenode). The 30uL of beads were then resuspended in 650μL of ChIP buffer and combined with the ChIP sample to pre-Clear the sample. Beads and chromatin were rotated 2hrs at 4°C, and the beads were collected and the supernatant retained. 3μL of anti-TFAP2C antibody (sc-8977) was added to the ChIP sample. The samples were then rotated overnight at 4°C.

60μL of Protein A Dynabeads were then added to the ChIP samples and rotated 2 hours at 4°C. The beads were then washed 2×4 minutes with 500μL Wash Buffer A (50mM HEPES pH 7.9, 1% TritonX-100, 0.1% Deoxycholate, 1mM EDTA, 140mM NaCl), 500μL Wash Buffer B (50m HEPES pH 7.9, 0.1% SDS, 1% Triton X100, 0.1% Deoxycholate, 1mM EDTA, 500mM NaCl), and 500μL TE buffer (10mM TrisHCl pH 8.0, 1mM EDTA). Each wash consists of resuspension in 500μL buffer and rotation at 4 minutes, followed by collection of beads and removal of supernatant. DNA was eluted in 100μL elution buffer (50mM TrisHCl pH 8.0, 1mM EDTA, 1% SDS) at 65°C for 10 minutes in a ThermoMixer (Eppendorf) shaking at 1400RPM. The eluent was collected and the beads were subjected to a second round of elution with 150μL elution buffer

The ChIP eluants were pooled, and the input sample was diluted to 250μL with elution buffer. These samples was incubated 65°C overnight to promote decrosslinking. The samples were then allowed to cool to room temperature, 15μg of RNAse A (Purelink, ThermoFisher) was added, and the samples were incubated 30 minutes at 37°C to degrade RNA. 100 μg Proteinase K was then added and the samples were incubated 56°C for two hours. DNA was purified using a MinElute PCR Purification kit (Qiagen).

DNA was sonicated again to 150bp average fragment size with a Covaris S2, concentrated with Agencourt AMPure XP beads (Beckman Coulter) and libraries were generated using the Ovation Ultralow Library System V2 (Nugen).

Replicates and data pooling

All replicates are listed in Supplementary Table 1 and are biological replicates except where otherwise noted. For determination of ChIP or ATAC-seq peaks or display of ChIP or ATAC

data in figures, all reads from a given condition (e.g. d5 human ATAC-seq control samples) were merged to increase coverage. RNA-seq reads for a given condition were merged when comparing RPKM across conditions or analyzing splicing but were considered separately when calculating differentially expressed genes (see RNA-seq data analysis below).

RNA-seq data analysis

RNAseq data was mapped to hg19 using Tophat[54] and read counts per gene were determined using HTSeq[55] as previously described[12]. Differentially expressed genes were calculated using DESeq[56], and RPKM values were calculated with a custom script. Once differential expressed genes were determined, they were analyzed for GO terms were called using GOrilla, which calculates p-values and q-values using a hypergeometric test[33].

Correlation between changes in gene expression and proximity of ATAC and ChIP peaks was also calculated by a custom script.

ATAC-seq data analysis

ATAC-seq data was mapped using Bowtie as previously described[53]. Peaks were defined in each condition using MACS2 callpeaks tool[57] with appropriate genome size. To find peaks specific to one condition (e.g. Naïve Specific), we use the predictd module of MACS2 to determine predicted extension size of each data set being compared, callpeaks for each dataset the –B and –– no model options and with the extension size specified as the average of the two samples, and the bgddiff module using the generated pileup and lambda files with the options –g 60 –l 120. An eight-fold relative enrichment cutoff was used to define peaks specific to each state, except when comparing murine WT and $Tfap2a^{-/-}$ $Tfap2c^{-/-}$, in which a six-fold cutoff was used due to the relatively small number of peaks different in the two conditions.

To identify peaks in common between the primed and naïve states, and overlap between different peak sets, we used the Bedtools intersect tool[58].

ChIP-seq data analysis

ChIP-seq data was mapped using Bowtie2 with default settings and clonal reads were removed using samtools rmdup. Reads from all replicates for a given condition were merged and peaks were using MACS2 callpeaks[57], comparing ChIP against input reads and using appropriate genome size and default settings. To determine overlap with ATAC-seq peaks, we used the bedtools intersect tool[58].

Motif Analysis

Enriched motifs in peak sets were identified using the HOMER findMotifsGenome tool with appropriate genome and default settings[19].

Principal component analysis

For principal component analysis (PCA) for RNA-seq data, RPKM for each sample were used as input. Variance of each gene's RPKM in different samples were then calculated (rowVars function in R). PCA analysis (prcomp function in R) was performed on genes with

the top 1000 variance across samples. PCA plot were then plotted with ggplot2 package in R (http://ggplot2.org).

For principal component analysis (PCA) for ATAC-seq data, peaks of in each ATAC-seq samples were first defined with MACS2(v.2.1.1) with default parameters. Then, ATAC-seq peaks cross all samples were merged into one union ATAC-seq peak set and ATAC-seq reads in each sample were calculated over the union ATAC-seq peak set. ATAC-seq reads in different samples were then normalized over its sequencing depth and this matrix is used as input for PCA analysis. Variance of normalized ATAC-seq reads over each peak were then calculated (rowVars function in R). PCA analysis (prcomp function in R) was performed on peaks with the top 1000 variance across samples. PCA were then plotted with ggplot2 package in R (http://ggplot2.org).

Analysis of peak location

Annotation of peak location (Promoter, intragenic, intergenic etc.) and calculation of distance to nearest promoter was calculated using HOMER annotatepeaks tool with the appropriate genome[19].

GREAT Analysis

We used GREAT analysis[37] calling genes near the midpoints of the TFAPC dependent regulatory elements. We instructed the program to use the two nearest genes to the peaks, provided they are within 50kb of the peak. GREAT calculates p and q values using a binomial test, as well as a hypergeometric test for comparison[37].

Reverse Transcription and real-time PCR

For Figure 8E, real time PCR was perfomed as described[59]. For Supplementary Figure 6G, RNA extraction and cDNA synthesis were performed as described previously[51], RT-PCR was carried out on the 7500 Real-Time PCR system (ThermoFisher). TaqMan probes used: ACTB: Hs01060665_g1 (ThermoFisher) and KLF17: Hs00703004_s1 (ThermoFisher).

Alterations to images

Overall brightness of some brightfield microscopy images was increased to improve visibility, and Supplementary Figure 5C was converted to grayscale to improve visibility.

Animal work

MEFs were derived following UCLA institutional animal care and use committee (IACUC) approval.

Embryo donation

Use of human embryos in this research project followed California State law, which required review by two committees, the Institutional Review Board (IRB) and the human embryonic stem cell research oversight committee (ESCRO), which approved the process of informed consent, and experiments using human embryos for research purposes. Following approval and outreach to fertility clinics with a flyer advertising the study, individuals and couples

with stored frozen IVF embryos contacted us to donate surplus embryos either through referral or through initiating contact with the UCLA Broad Stem Cell Research Center. Patients were not paid for participation, and all donors were informed that the embryos would be destroyed as part of the research study. Participants were also informed that they could withdraw consent at any time and if the embryos had been shipped to UCLA, they would be destroyed. Participants were also informed that donated embryos would not be used to make a baby. All research with human embryos in this study complied with the principles laid out in the International Society for Stem Cell Research guidelines.

Statistics and reproducibility

Statistical analysis was performed using existing programs (MACS2[57], Homer[19], DESeq[56], GOrilla[33]). Mean and standard error were calculated using standard statistical formulas. Where the ratio of expression for two sets is shown, as in Figures 4H, 4M, and 5E, standard error is calculated by the formula SE(quotient) =quotient*SQRT((SE(Set1)/Mean(Set1))^2+ (SE(Set2)/Mean(Set2))^2).

The observation that TFAP2C is upregulated at the protein level in naïve over primed cells (as shown in Figures 3B and 4B) was observed for Western blots generated from five independent lysate preparations. Induction of TFAP2C within three days of culture in 5iLAF (as shown in Figure 4B) was observed in two independent reversions.

The finding that $TFAP2C^{-/-}$ hESCs express OCT4 normally in the primed state (Supplementary Figure 3C) was demonstrated in Western blots from three independent lysate preparations, similar findings for NANOG and SOX2 (Supplementary Figures 3D, E) were performed once. The flow cytometry plots in Supplementary Figure 3F showing pluripotency surface markers on $TFAP2C^{-/-}$ hESCs are representative of three experiments.

The finding that $TFAP2C^{-/-}$ hESCs colonies showed morphological abnormality in 5iLAF (as shown in Figures 4A and Supplementary Figure 4 A,B) by day 5 was observed in nine independent reversions. Four of these reversions were continued long enough to confirm loss of $TFAP2C^{-/-}$ upon prolonged culture, as shown in Figures 4A and 5D. Loss of TFAP2C in $TFAP2C^{-/-}$ mutant (as shown in Figure 4C) was shown by Western blot from five independent reversions and by immunofluorescence (as shown in Figure 4I) in two independent reversions. Loss of pluripotency factors. and gain of neural factors was shown in by Western blot (as shown in Figure 4F and J) in four independent reversions and by immunofluorescence (as shown in Figure 4I and 4J) in two independent reversions. Loss of Tfap2c in $Tfap2C^{-/-}$ murine ESCs (as shown in Supplementary Figure 4G) was confirmed via Western blots from two independent lysate preparations.

Rescue of the $TFAP2C^{-/-}$ differentiation phenotype in 5iLAF by doxycycline inducible TFAP2C expression (as shown in Figure 5A–D and Supplementary Figure 5C) was observed in four experiments. Reduced cell proliferation after withdrawal of doxycycline (as shown in Supplementary Figure 5D and 5E) was observed in two experiments.

Bulk differentiation of $TFAP2C^{-/-}$ in 5iLAF at 5% O_2, with a small population maintaining self-renewal (as shown in Figure 6, Supplementary Figure 6) was observed in two

independent reversions. Loss of KLF17 in *TFAP2C*$^{-/-}$ t2iLGöY in Supplementary Figure 6F was demonstrated from two reversions.

The finding that OCT4 protein is expressed normally in the OCT4 Intronic enhancer mutant, as shown in Supplementary Figure 8D, was based on one Western blot. Loss of cells in the OCT4 Intronic enhancer mutant upon treatment with 5iLAF (as shown in Figure 8F) was observed in three independent reversions, with qRT-PCR data (as shown in Figure 8E) derived from one such reversion.

Code Availability

Custom scripts used for demultiplexing NGS reads, calculating RPKM, generating DNA methylation metaplots and comparing distribution of peaks to expression of nearby genes will be made available upon request.

Data Availability

All high throughput sequencing datasets described in Supplementary Table 1 have been deposited in the Gene Expression Omnibus (GEO) under the accession number GSE101074. RNA-seq data from naive and primed hESCs was gleaned from our previously published data[12] (GSE76970). ChIP-seq data on H3K4me3, H3K27Ac3 and Mediator from naïve and primed hESCs was obtained from published sources[9,16] (GSE69647). Peaks sets used in this paper are included in Supplmentary Tables 2 and 4, and RPKMs from RNA-seq data are included in Supplementary Table 5. Source data for Figures 3A, 4H, 4M, 5E and Supplementary Figures 3G, 5A and 6B are available in Supplementary Table 8. Additional data is available upon reasonable request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Niakan KK, Han J, Pedersen RA, Simon C, Pera RA. Human pre-implantation embryo development. Development. 2012; 139:829–841. DOI: 10.1242/dev.060426 [PubMed: 22318624]

2. Okamoto I, et al. Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. Nature. 2011; 472:370–374. DOI: 10.1038/nature09872 [PubMed: 21471966]

3. Petropoulos S, et al. Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. Cell. 2016; 165:1012–1026. DOI: 10.1016/j.cell.2016.03.023 [PubMed: 27062923]

4. Sahakyan A, et al. Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation. Cell Stem Cell. 2017; 20:87–101. DOI: 10.1016/j.stem.2016.10.006 [PubMed: 27989770]

5. Kunath T, et al. Developmental differences in the expression of FGF receptors between human and mouse embryos. Placenta. 2014; 35:1079–1088. DOI: 10.1016/j.placenta.2014.09.008 [PubMed: 25443433]

6. Nakamura T, et al. A developmental coordinate of pluripotency among mice, monkeys and humans. Nature. 2016; 537:57–62. DOI: 10.1038/nature19096 [PubMed: 27556940]

7. Deglincerti A, et al. Self-organization of the in vitro attached human embryo. Nature. 2016; 533:251–254. DOI: 10.1038/nature17948 [PubMed: 27144363]

8. Blakeley P, et al. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. Development. 2015; 142:3151–3165. DOI: 10.1242/dev.123547 [PubMed: 26293300]

9. Theunissen TW, et al. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell. 2014; 15:471–487. DOI: 10.1016/j.stem.2014.07.002 [PubMed: 25090446]

10. Takashima Y, et al. Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell. 2014; 158:1254–1269. DOI: 10.1016/j.cell.2014.08.029 [PubMed: 25215486]

11. 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. 2014; 15:410–415. DOI: 10.1016/j.stem.2014.09.014 [PubMed: 25280217]

12. Pastor WA, et al. Naive Human Pluripotent Cells Feature a Methylation Landscape Devoid of Blastocyst or Germline Memory. Cell Stem Cell. 2016; 18:323–329. DOI: 10.1016/j.stem.2016.01.019 [PubMed: 26853856]

13. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013; 10:1213–1218. DOI: 10.1038/nmeth.2688 [PubMed: 24097267]

14. Song L, et al. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome Res. 2011; 21:1757–1767. DOI: 10.1101/gr.121541.111 [PubMed: 21750106]

15. Thurman RE, et al. The accessible chromatin landscape of the human genome. Nature. 2012; 489:75–82. DOI: 10.1038/nature11232 [PubMed: 22955617]

16. Ji X, et al. 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. Cell Stem Cell. 2016; 18:262–275. DOI: 10.1016/j.stem.2015.11.007 [PubMed: 26686465]

17. Rada-Iglesias A, et al. A unique chromatin signature uncovers early developmental enhancers in humans. Nature. 2011; 470:279–283. DOI: 10.1038/nature09692 [PubMed: 21160473]

18. Whyte WA, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013; 153:307–319. DOI: 10.1016/j.cell.2013.03.035 [PubMed: 23582322]

19. Heinz S, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010; 38:576–589. DOI: 10.1016/j.molcel.2010.05.004 [PubMed: 20513432]

20. Guo G, et al. Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development. 2009; 136:1063–1069. DOI: 10.1242/dev.030957 [PubMed: 19224983]

21. Buecker C, et al. Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell. 2014; 14:838–853. DOI: 10.1016/j.stem.2014.04.003 [PubMed: 24905168]

22. Auman HJ, et al. Transcription factor AP-2gamma is essential in the extra-embryonic lineages for early postimplantation development. Development. 2002; 129:2733–2747. [PubMed: 12015300]

23. Winger Q, Huang J, Auman HJ, Lewandoski M, Williams T. Analysis of transcription factor AP-2 expression and function during mouse preimplantation development. Biol Reprod. 2006; 75:324–333. DOI: 10.1095/biolreprod.106.052407 [PubMed: 16672719]

24. Werling U, Schorle H. Transcription factor gene AP-2 gamma essential for early murine development. Molecular and cellular biology. 2002; 22:3149–3156. [PubMed: 11940672]

25. Hoffman TL, Javier AL, Campeau SA, Knight RD, Schilling TF. Tfap2 transcription factors in zebrafish neural crest development and ectodermal evolution. J Exp Zool B Mol Dev Evol. 2007; 308:679–691. DOI: 10.1002/jez.b.21189 [PubMed: 17724731]

26. Guttormsen J, et al. Disruption of epidermal specific gene expression and delayed skin development in AP-2 gamma mutant mice. Dev Biol. 2008; 317:187–195. DOI: 10.1016/j.ydbio. 2008.02.017 [PubMed: 18353300]

27. Qiao Y, et al. AP2gamma regulates neural and epidermal development downstream of the BMP pathway at early stages of ectodermal patterning. Cell Res. 2012; 22:1546–1561. DOI: 10.1038/cr. 2012.122 [PubMed: 22945355]

28. Schemmer J, et al. Transcription factor TFAP2C regulates major programs required for murine fetal germ cell maintenance and haploinsufficiency predisposes to teratomas in male mice. PLoS One. 2013; 8:e71113. [PubMed: 23967156]

29. Niakan KK, Eggan K. Analysis of human embryos from zygote to blastocyst reveals distinct gene expression patterns relative to the mouse. Dev Biol. 2013; 375:54–64. DOI: 10.1016/j.ydbio. 2012.12.008 [PubMed: 23261930]

30. Xie W, et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell. 2013; 153:1134–1148. DOI: 10.1016/j.cell.2013.04.022 [PubMed: 23664764]

31. Tsankov AM, et al. Transcription factor binding dynamics during human ES cell differentiation. Nature. 2015; 518:344–349. DOI: 10.1038/nature14233 [PubMed: 25693565]

32. Yan L, et al. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat Struct Mol Biol. 2013; 20:1131–1139. DOI: 10.1038/nsmb.2660 [PubMed: 23934149]

33. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC bioinformatics. 2009; 10:48. [PubMed: 19192299]

34. Tesar PJ, et al. New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature. 2007; 448:196–199. DOI: 10.1038/nature05972 [PubMed: 17597760]

35. Xie H, Simon MC. Oxygen availability and metabolic reprogramming in cancer. J Biol Chem. 2017; 292:16825–16832. DOI: 10.1074/jbc.R117.799973 [PubMed: 28842498]

36. Guo G, et al. Epigenetic resetting of human pluripotency. Development. 2017; 144:2748–2763. DOI: 10.1242/dev.146811 [PubMed: 28765214]

37. McLean CY, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010; 28:495–501. DOI: 10.1038/nbt.1630 [PubMed: 20436461]

38. Yeom YI, et al. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development. 1996; 122:881–894. [PubMed: 8631266]

39. Eloranta JJ, Hurst HC. Transcription factor AP-2 interacts with the SUMO-conjugating enzyme UBC9 and is sumolated in vivo. J Biol Chem. 2002; 277:30798–30804. DOI: 10.1074/jbc.M202780200 [PubMed: 12072434]

40. Wong PP, et al. Histone demethylase KDM5B collaborates with TFAP2C and Myc to repress the cell cycle inhibitor p21(cip) (CDKN1A). Molecular and cellular biology. 2012; 32:1633–1644. DOI: 10.1128/MCB.06373-11 [PubMed: 22371483]

41. Magnusdottir E, et al. A tripartite transcription factor network regulates primordial germ cell specification in mice. Nat Cell Biol. 2013; 15:905–915. DOI: 10.1038/ncb2798 [PubMed: 23851488]

42. Braganca J, et al. Human CREB-binding protein/p300-interacting transactivator with ED-rich tail (CITED) 4, a new member of the CITED family, functions as a co-activator for transcription factor AP-2. J Biol Chem. 2002; 277:8559–8565. DOI: 10.1074/jbc.M110850200 [PubMed: 11744733]

43. Braganca J, et al. Physical and functional interactions among AP-2 transcription factors, p300/CREB-binding protein, and CITED2. J Biol Chem. 2003; 278:16021–16029. DOI: 10.1074/jbc.M208144200 [PubMed: 12586840]

44. Bamforth SD, et al. Cardiac malformations, adrenal agenesis, neural crest defects and exencephaly in mice lacking Cited2, a new Tfap2 co-activator. Nat Genet. 2001; 29:469–474. DOI: 10.1038/ng768 [PubMed: 11694877]

45. Kuckenberg P, et al. The transcription factor TCFAP2C/AP-2gamma cooperates with CDX2 to maintain trophectoderm formation. Molecular and cellular biology. 2010; 30:3310–3320. DOI: 10.1128/MCB.01215-09 [PubMed: 20404091]

46. Lee SH, Kwon JW, Choi I, Kim NH. Expression and function of transcription factor AP-2? in early embryonic development of porcine parthenotes. Reprod Fertil Dev. 2015

47. Nichols J, et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell. 1998; 95:379–391. [PubMed: 9814708]

48. Fogarty NME, et al. Genome editing reveals a role for OCT4 in human embryogenesis. Nature. 2017; 550:67–73. DOI: 10.1038/nature24033 [PubMed: 28953884]

49. Kalinka AT, Tomancak P. The evolution of early animal embryos: conservation or divergence? Trends Ecol Evol. 2012; 27:385–393. DOI: 10.1016/j.tree.2012.03.007 [PubMed: 22520868]

50. Irie N, Kuratani S. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. Nat Commun. 2011; 2:248. [PubMed: 21427719]

51. Liu X, et al. Comprehensive characterization of distinct states of human naive pluripotency generated by reprogramming. Nat Methods. 2017; 14:1055–1062. DOI: 10.1038/nmeth.4436 [PubMed: 28945704]

52. Diaz Perez SV, et al. Derivation of new human embryonic stem cell lines reveals rapid epigenetic progression in vitro that can be prevented by chemical modification of chromatin. Hum Mol Genet. 2012; 21:751–764. DOI: 10.1093/hmg/ddr506 [PubMed: 22058289]

53. Pastor WA, et al. MORC1 represses transposable elements in the mouse male germline. Nat Commun. 2014; 5:5795. [PubMed: 25503965]

54. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009; 25:1105–1111. DOI: 10.1093/bioinformatics/btp120 [PubMed: 19289445]

55. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015; 31:166–169. DOI: 10.1093/bioinformatics/btu638 [PubMed: 25260700]

56. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11:R106. [PubMed: 20979621]

57. Zhang Y, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008; 9:R137. [PubMed: 18798982]

58. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. DOI: 10.1093/bioinformatics/btq033 [PubMed: 20110278]

59. Gkountela S, et al. The ontogeny of cKIT+ human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and in vitro differentiation. Nature cell biology. 2013; 15:113–122. DOI: 10.1038/ncb2638 [PubMed: 23242216]
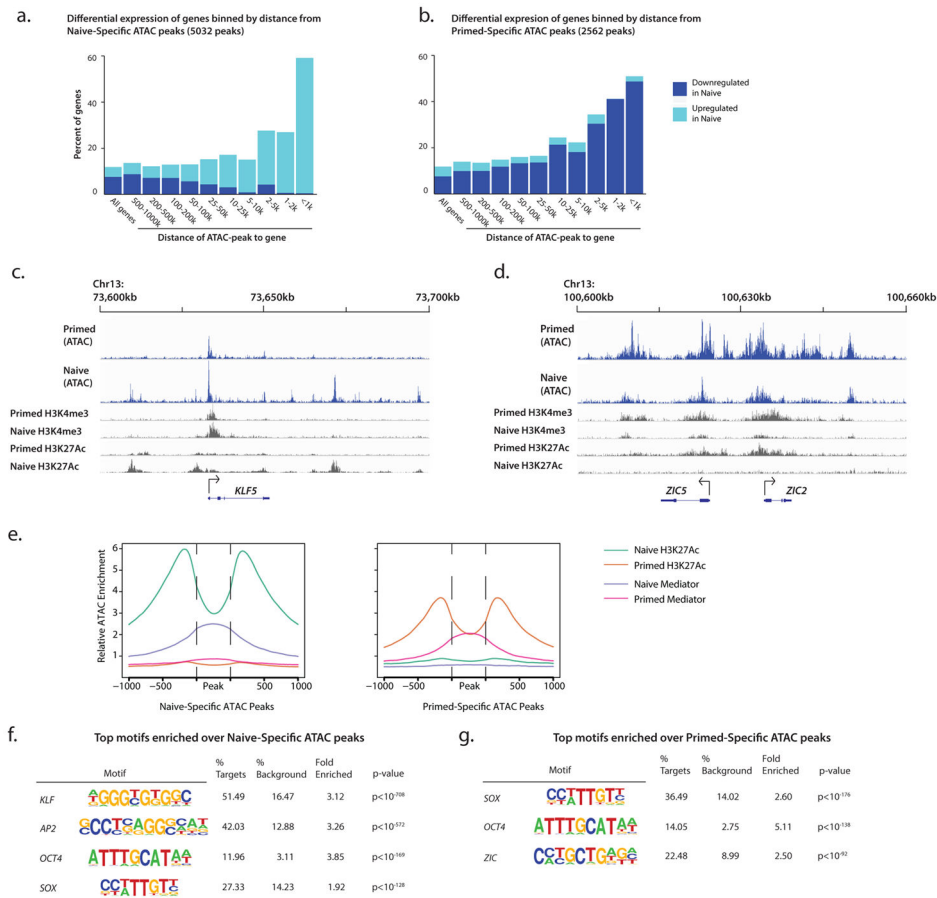
Figure 1. Determination of regulatory elements specific to the naïve and primed states in humans
a,b Percentage of the time a gene whose transcriptional start site is a given distance from a naïve-specific ATAC-peak (**a**) or primed-specific ATAC-peak (**b**) is upregulated or downregulated in the naïve state. **c,d** Many naïve-specific ATAC-seq peaks appear proximal to *KLF5* (**c**), a gene highly upregulated in the naïve state, while primed-specific ATAC-peaks are present near the primed-specific *ZIC* genes. Note the strong enrichment of H3K27Ac near ATAC-seq peaks. H3K4me3 and H3K27Ac ChIP-seq data come from published sources[9,16]. **e,** Metaplot of H3K27Ac and Mediator over naïve-specific (left) and primed specific (right) ATAC-peaks. **f,g** Most statistically significant transcription factor binding motifs enriched in naïve-specific (**f**) or primed-specific (**g**) ATAC peaks were calculated using a cumulative binomial distribution[19]. Pooled data from 4 naïve and 4 primed biological replicates were used.
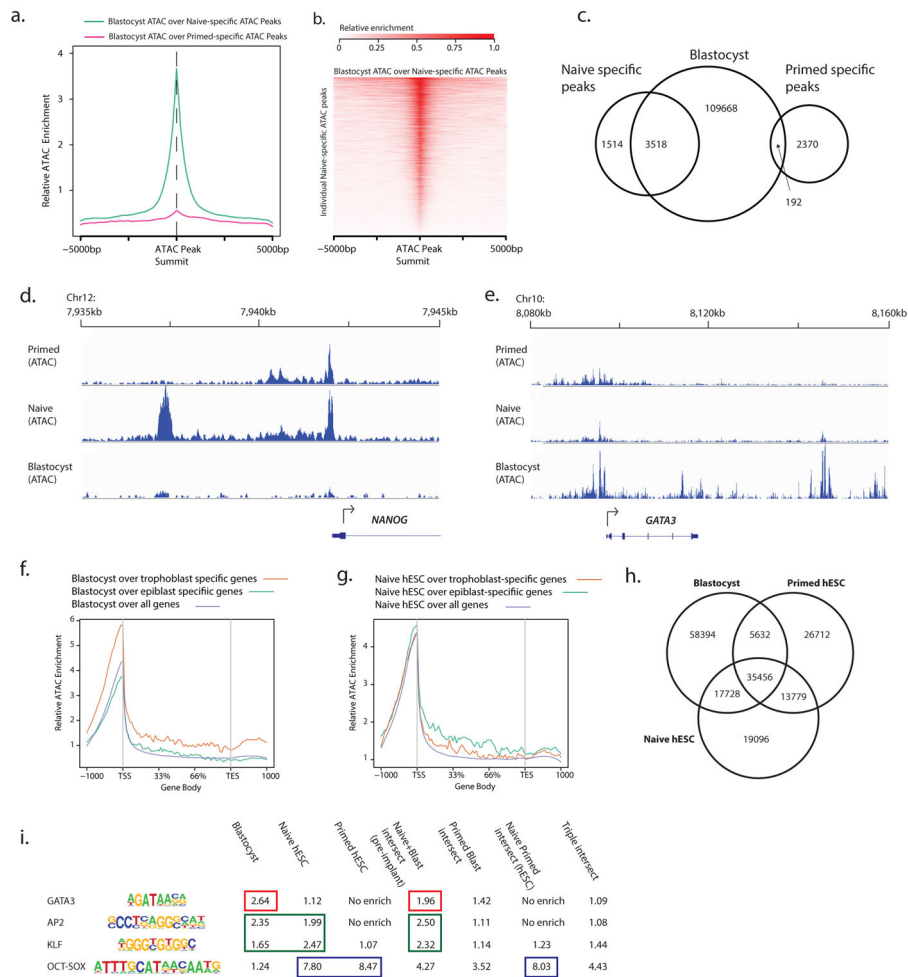
Figure 2. Most naïve-specific ATAC-peaks are present in other naïve human cells and the human embryo

a, Normalized ATAC-seq reads from the human blastocyst plotted relative to Naïve-specific and primed-specific peaks. Note far greater enrichment over naïve-specific peaks. **b,** Blastocyst ATAC-seq plotted relative to all Naïve-specific ATAC peaks. Note enrichment over almost all naïve-specific peaks, indicating that they are open in the blastocyst. **c,** Most Naïve-specific ATAC-seq peaks overlap with a blastocyst-ATAC peak, but most primed-specific peaks do not. **d,e** ATAC-seq signal for primed hESCs, naïve hESCs and blastocyst in the viscinity of NANOG (**d**) and GATA3 (**e**). Peak height is normalized to total number of reads in each sample. **f,g** Metaplot of ATAC-seq read density over the gene bodies of 100 genes most highly specific to trophoblast or epiblast, as defined from single-cell RNA-seq data in human[3], as well as all genes. **h,** Venn diagram showing overlap of all ATAC-seq peaks in blastocyst, naïve hESCs and primed hESCs. **i,** Enrichment of GATA, AP2, KLF and OCT-SOX motifs in each set identified in part e. Note enrichment of AP2 and KLF motifs in both blastocyst and naïve hESCs, stronger enrichment of GATA in blastocyst and stronger enrichment of OCT-SOX in ESCs.
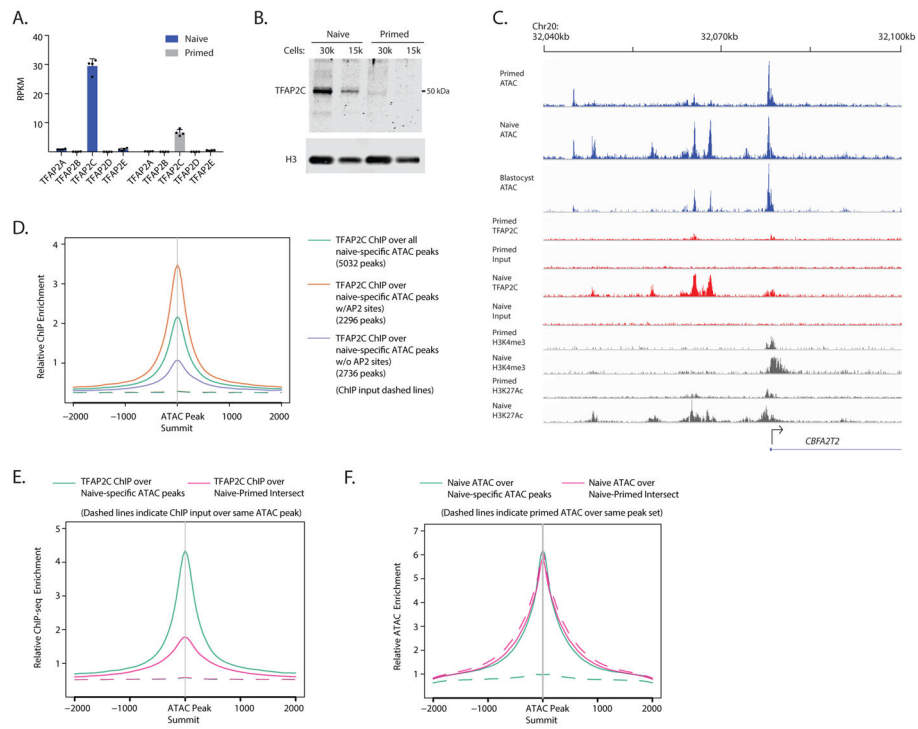
Figure 3. TFAP2C is highly enriched over naïve-specific open chromatin in humans
a, *TFAP2C* is highly expressed in naïve cells, both relative to other AP2 transcription factors and relative to primed cells. Mean and standard deviation are shown, with dots representing each replicate (n=4 independent experiments). **b,** TFAP2C protein is highly upregulated in the naïve-pluripotent state. Data represent 1 out of 5 independent experiments with similar results. **c,** Strong co-enrichment of TFAP2C with naïve-specific ATAC peaks at the *CBFA2T2* locus. **d,** Global enrichment of TFAP2C relative to the summits of different categories of naïve-specific ATAC peaks. TFAP2C is enriched over naïve-specific ATAC peaks, especially those with AP2 motifs. **e,f** TFAP2C is strongly enriched over Naïve-specific ATAC peak summits compared with enrichment over regions that show ATAC enrichment in both naïve and primed cells (Naïve-primed Intersect) (**e**), even though both peak sets show similar ATAC enrichment (**f**). Uncropped Western blot images are available in Supplementary Figure 9. Source data for **a** is available in Supplementary Table 8.
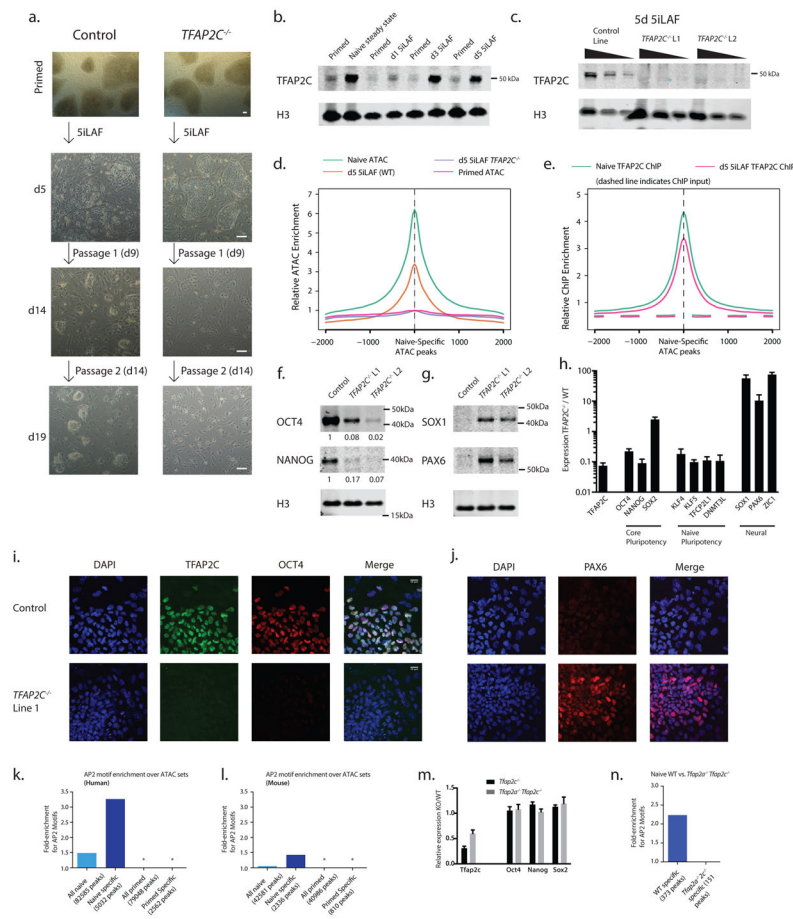
Figure 4. *TFAP2C*^{−/−} cells differentiate in naïve media

a, *TFAP2C*^{−/−} hESCs, self-renew in primed conditions but differentiate and fail to self-renew upon treatment in naïve (5iLAF) media. Scale bars indicate 100 μm. Data represent 1 out of 4 independent experiments with similar results. **b,** Western blot of TFAP2C upon culture in primed or 5iLAF conditions. TFAP2C is strongly induced within three days of treatment with 5iLAF. Data represent 1 out of 2 independent experiments with similar results. **c,** Western blot for TFAP2C after five days of 5iLAF culture. TFAP2C is absent from *TFAP2C*^{−/−} deficient lines. **d,** ATAC-seq openness of Naïve, d5 5iLAF WT and *TFAP2C*^{−/−}, and primed cells over naïve-specific ATAC peaks. Note that after five days of reversion, substantial opening of the naïve-specific ATAC peaks has already occurred, but not in the *TFAP2C*^{−/−}cells. **e,** TFAP2C ChIP enrichment shown over naïve and d5 5iLAF samples. ChIP input for each set is shown as a dashed line. **f,** Western blot for OCT4 and NANOG in control and *TFAP2C*^{−/−} cells after five days of culture in 5iLAF. Quantitation normalized to histone below. **g,** Western Blot for SOX1 and PAX6 in control and *TFAP2C*^{−/−} cells. **h,** Relative RPKM of pluripotency and neural markers in RNA-seq. Data is from n=3 WT and n=4 *TFAP2C*^{−/−} independent biological replicates (mean +/− s.e). **i,j,** Immunofluorescent staining for TFAP2C, OCT4 (**i**) and PAX6 (**j**) in control and *TFAP2C*^{−/−} cells. Scale bar indicates 20 μm. **k,l** Fold enrichment for AP2 motifs in the specified peak sets in humans (**k**) and mouse (**l**). Asterix indicates no enrichment. Although AP2 motifs are enriched in naïve-specific peaks in both species, the enrichment is much stronger in the

human naïve-specific set. **m,** Expression of key pluripotency markers in WT, *Tfap2c*$^{-/-}$, and *Tfap2a*$^{-/-}$ *Tfap2c*$^{-/-}$ cells in 2i+LIF conditions. n=4 biological replicates in for *Tfap2c*$^{-/-}$ and controls and n=6 biological replicates for *Tfap2a*$^{-/-}$ *Tfap2c*$^{-/-}$ and controls (mean +/− s.e.) **n,** ATAC-seq peaks specific to WT and *Tfap2a*$^{-/-}$ *Tfap2c*$^{-/-}$ were calculated and enrichment for AP2 motifs determined. Asterisk indicates no enrichment. Uncropped Western blots in Supplementary Figure 9. Source data for **h** and **m** in Supplementary Table 8.
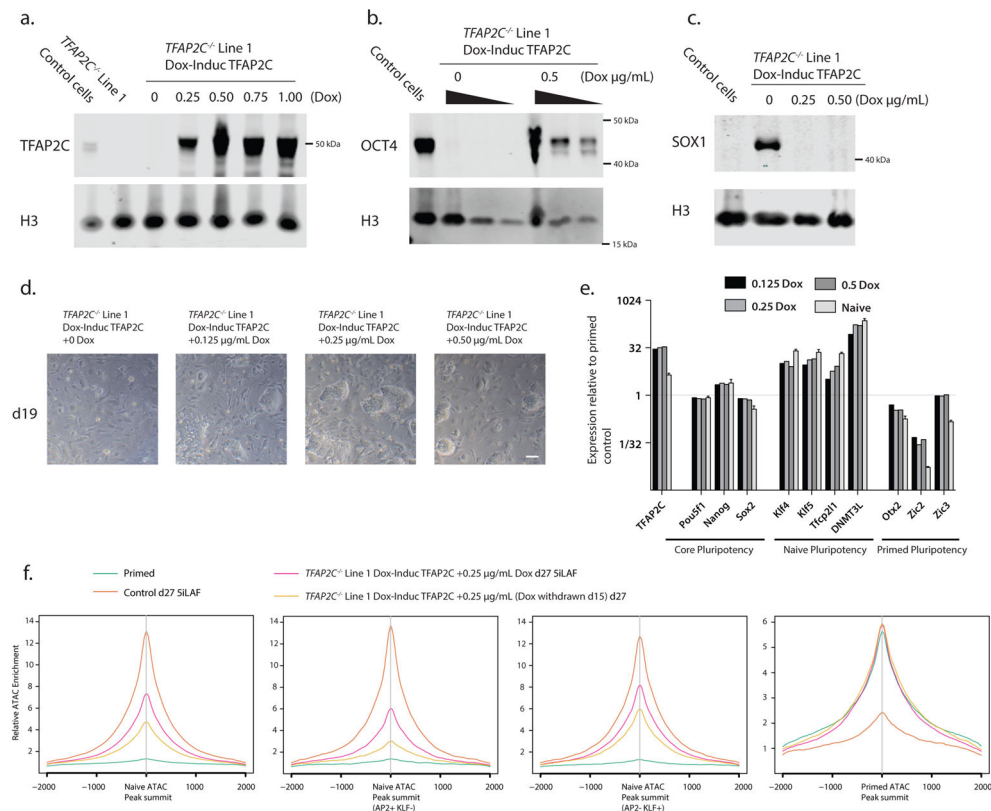
Figure 5. Ectopic expression of TFAP2C partially rescues *TFAP2C*^{−/−} phenotype

a, Quantitative western blot showing tunable TFAP2C induction in *TFAP2C*⁻/⁻ background in primed conditions. **b,c** Western blots show rescue of OCT4 expression and SOX1 repression upon doxycycline inducible TFAP2C expression. Lysates were collected after five days of treatment with 5iLAF and the indicated concentration of doxycycline. **d,** Appearance of round naïve-like colonies in lines with ectopic TFAP2C expression. Scale bar indicates 100μm. Results represent 1 out of 4 independent experiments with similar results. **e,** Partial rescue of upregulation of naïve pluripotency factors, downregulation of primed-factors with ectopic TFAP2C expression. 1 replicate for dox induction samples and primed control, 4 for naïve samples and primed control. **f,** TFAP2C was ectopically expressed for the first 15 days of reversion, then removed in some cells to induce acute loss of TFAP2C. ATAC-seq plotted from these cells is plotted over naïve-specific peaks (5032 peaks), a subset that contained an AP2 motif but no KLF motif (1054 peaks), a subset that contained a KLF motif but no AP2 motif (1551 peaks) and primed-specific peaks (2562 peaks). Reduced ATAC-seq density over naïve specific peaks and increased density over primed-specific peaks, in the sample in which doxycycline had been withdrawn. Closing of naïve specific peaks is especially pronounced over the subset of peaks that contain AP2 sites but no KLF sites (AP2+ KLF-). Peaks subsets are listed in Supplementary Table 2. Uncropped Western blot images are available in Supplementary Figure 9. Source data for **e** is available in Supplementary Table 8
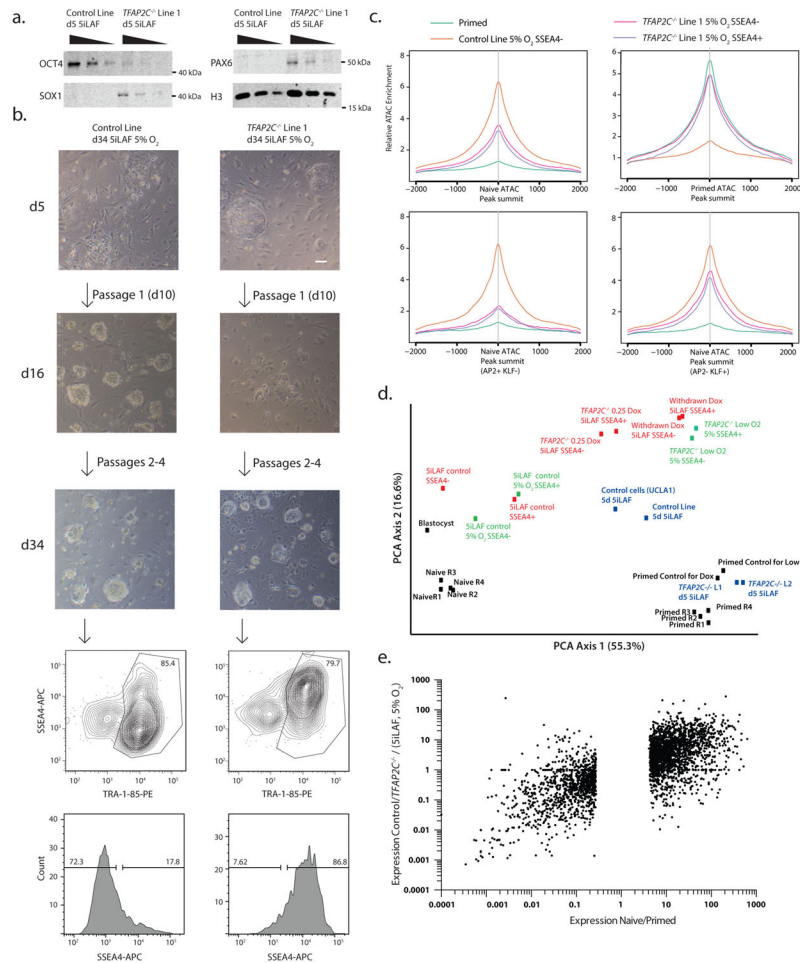
Figure 6. *TFAP2C$^{-/-}$* cells survive in 5iLAF in 5% O$_2$ conditions but do not transition to naïve state

a, Western blots for the pluripotency marker OCT4 and the neural markers SOX1 and PAX6 in WT and *TFAP2C$^{-/-}$* cells after 5 days in 5iLAF at 5% O$_2$. **b,c** Brightfield images of control and *TFAP2C$^{-/-}$* cells in 5iLAF culture. Initially the *TFAP2C$^{-/-}$* cells show morphology similar to what is observed in ambient oxygen concentration conditions (compare to Figure 4A). However, some colonies are observable after passaging. These colonies show a shift toward SSEA4$^+$ (primed) state. Scale bars indicates 100μm. **c,** ATAC-seq data from control and *TFAP2C$^{-/-}$* cells in 5% O$_2$ plotted over ATAC-seq peak sets. **d,** Principle component analysis comparing ATAC-seq datasets generated in this work. Blue dots: after five days in 5iLAF, WT control cells show an ATAC-seq landscape part-way between primed and naïve, whereas *TFAP2C$^{-/-}$* cells show no change toward naïve. Green dots: although *TFAP2C$^{-/-}$* cells survive in low oxygen conditions, they have an ATAC-seq landscape much more similar to primed than naïve cells. Red dots: ectopic doxycycline dependent-expression of TFAP2C in *TFAP2C$^{-/-}$* partially rescues the naïve-landscape, and withdrawal of doxycycline induces a shift toward primed identity. Shown for comparison are control cells reverted at the same time. **e,** Genes differentially regulated in naïve vs. primed hESCs are plotted. Note that genes higher expressed in naïve cells are lower expressed in *TFAP2C$^{-/-}$*. The RPKM values correspond to SSEA4$^-$ cells in control (average of 2

biological replicates) and SSEA4$^+$ in *TFAP2C*$^{-/-}$ (average of 3 biological replicates). Uncropped Western blot images are available in Supplementary Figure 9.
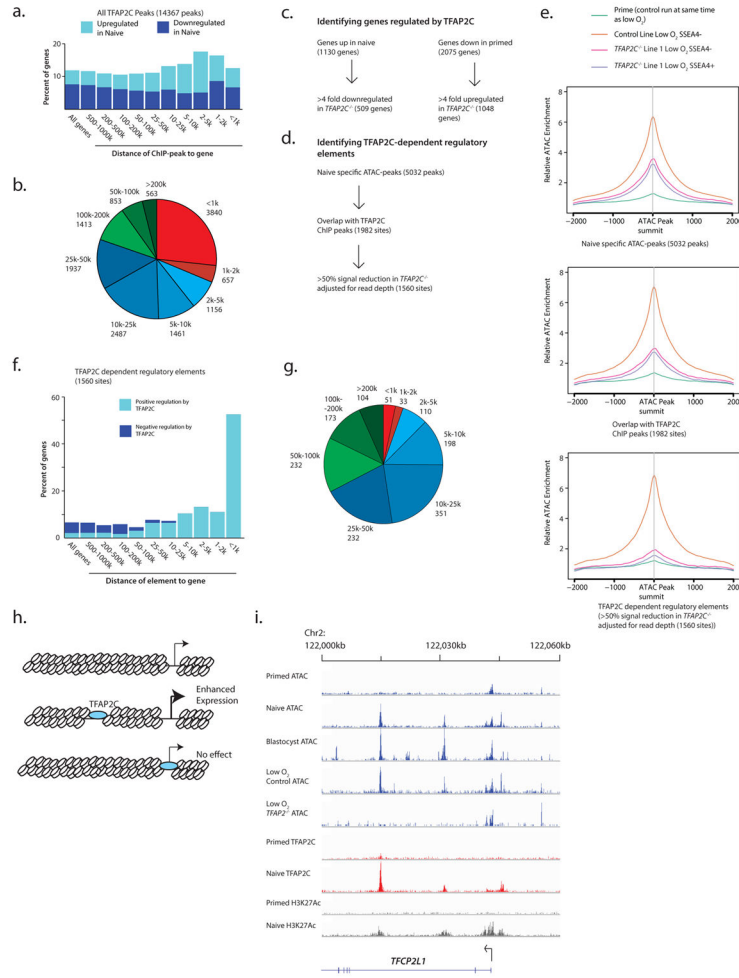
Figure 7. Identifying direct regulatory targets of TFAP2C

a, Percentage of the time a gene whose transcriptional start site is a given distance from a TFAP2C ChIP-seq peak is upregulated or downregulated in naïve hESCs. Notice the much weaker correspondence compared with Figure 1A, and the lack of any effect at the promoter. **b,** Distance of TFAP2C ChIP-seq peak to nearest promoter. **c,** To identify pluripotency-state genes positively or negatively regulated by TFAP2C, we identified the subset of naïve-specific genes downregulated <4 fold in *TFAP2C*$^{-/-}$ (positively related by TFAP2C) and primed-specific genes upregulated >4 fold in *TFAP2C*$^{-/-}$ (negatively regulated by TFAP2C). Because they were the predominant pluripotent populations, we compared expression of SSEA4$^-$ control cells (n=2 biological replicates) to SSEA4$^+$ *TFAP2C*$^{-/-}$ cells (n=3 biological replicates) **d,** To identify TFAP2C-dependent enhancers, we identified the overlap of the naïve-specific and TFAP2C ChIP-seq peaks, then took the subset of peaks that showed >50% density reduction in *TFAP2C*$^{-/-}$ SSEA4$^+$ as compared with control SSEA4$^-$ cells, normalized for total read depth. These were classified as TFAP2C-dependent regulatory elements. **e,** ATAC-seq read density over all naïve-specific ATAC-peaks, naïve-specific ATAC peaks overlapping with TFAP2C ChIP-seq peaks, and the TFAP2C-dependent regulatory element set identified in (**d**). Note dramatic loss of signal in *TFAP2C*$^{-/-}$ over the TFAP2C-dependent set. **f,** Frequency with which a gene a given distance from a

TFAP2C-dependent ATAC-seq peak is positively or negatively regulated by TFAP2C. **g,** Distance of TFAP2C-dependent ATAC-seq peak to nearest gene. Note that the vast majority of such elements are enhancers. **h,** Schematic demonstrating the typical regulatory role of TFAP2C in naïve hESCs. Where TFAP2C facilitates the opening of a new enhancer, it has a positive regulatory role. Where it hones to chromatin that is already open, it has no tangible effect on transcription. **i,** ATAC-seq and ChIP-seq data are shown in the vicinity of naïve-pluripotency factor *TFCP2L1*.
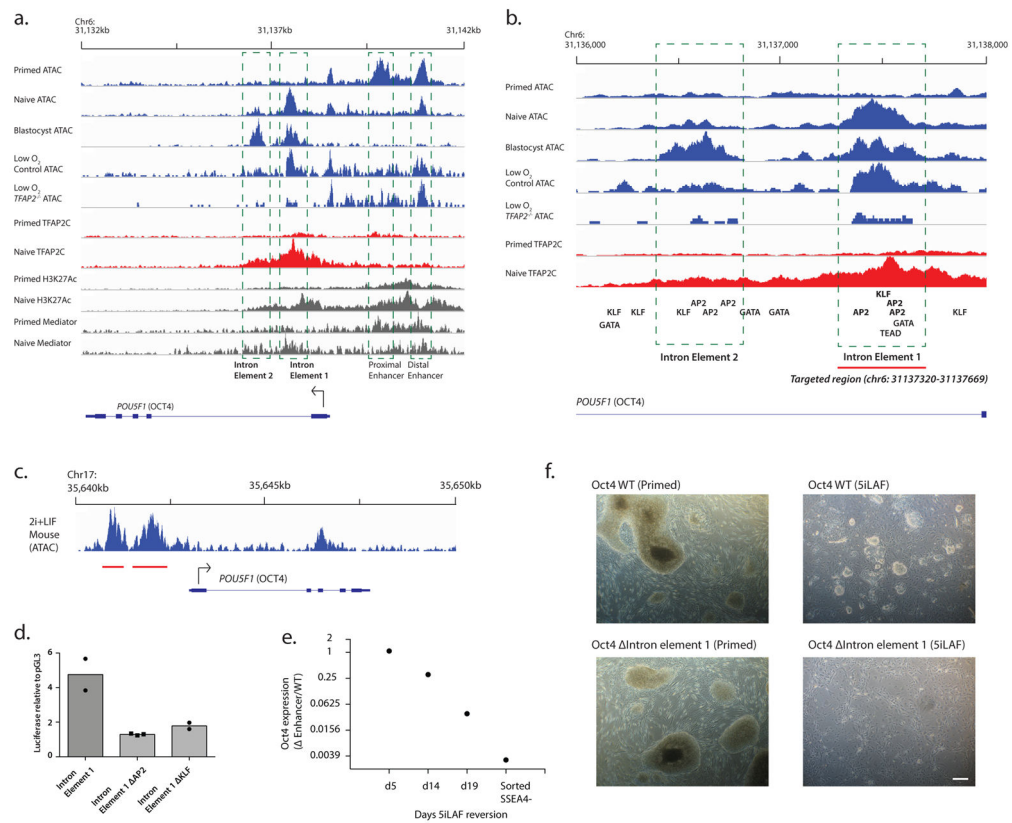
Figure 8. A TFAP2C+ intronic enhancer of OCT4

a, Chromatin landscape of OCT4. Two putative enhancers "intron element 1" and "intron element 2" are present in blastocyst. Intron element 1 is also strongly enriched in naïve cells and lost in $TFAP2C^{-/-}$. **b,** The location of consensus motifs for key pre-implantation transcription factors is shown in the vicinity of Intron Elements 1 and 2. Note the clustering of AP2 sites at each element. The control low O_2 track is the SSEA4− population, the $TFAP2C^{-/-}$ low O_2 is the SSEA4+ population. The region targeted for CRISPR deletion is shown. **c,** ATAC-seq reads over the murine POU5F1 locus in naïve (2i+LIF) conditions. Note the absence of either intronic enhancer. **d,** Luciferase activity from a pGL3 construct in which WT or mutant Intron Element 1 had been cloned, normalized to signal from a pGL3 construct with no enhancer. Results are shown from two independent experiment, except for the AP2 sample, for which there are n=3 replicates from two experiments. All signals were first normalized for Renilla signal. **e,** OCT4 expression is lost over time upon reversion of the intron element 1-deleted mutant, indicating differentiation. Sorting for SSEA4− cells in 5iLAF culture typically produces a pure population of naïve hESCs, but this population has lost OCT4 expression in the intron element 1-deleted mutant. Mean of n=2 technical replicates is shown. **f,** A line in which the intron element 1 is deleted appears normal in primed conditions but fails to yield naïve colonies upon reversion. Scale bar indicates 200μm. Images are representative of 3 independent reversions. Source data for **d** is available in Supplementary Table 8.