nature cell biology

ARTICLES

https://doi.org/10.1038/s41556-018-0089-0

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*]

Naive and primed pluripotent human embryonic stem cells bear transcriptional similarity to pre- and post-implantation epiblast and thus constitute a developmental model for understanding the 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 activator protein-2 (AP2) transcription factor binding motif at naive-specific open chromatin. We determined that the AP2 family member TFAP2C is upregulated during primed to naive reversion and becomes widespread at naive-specific enhancers. TFAP2C functions to maintain pluripotency and repress neuroectodermal differentiation during the transition from primed to naive by facilitating the opening of enhancers proximal to pluripotency factors. Additionally, we identify a previously undiscovered naive-specific *POU5F1 (OCT4)* enhancer enriched for TFAP2C binding. Taken together, TFAP2C establishes and maintains naive human pluripotency and regulates OCT4 expression by mechanisms that are distinct from mouse.

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 naive 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 naive epiblast.

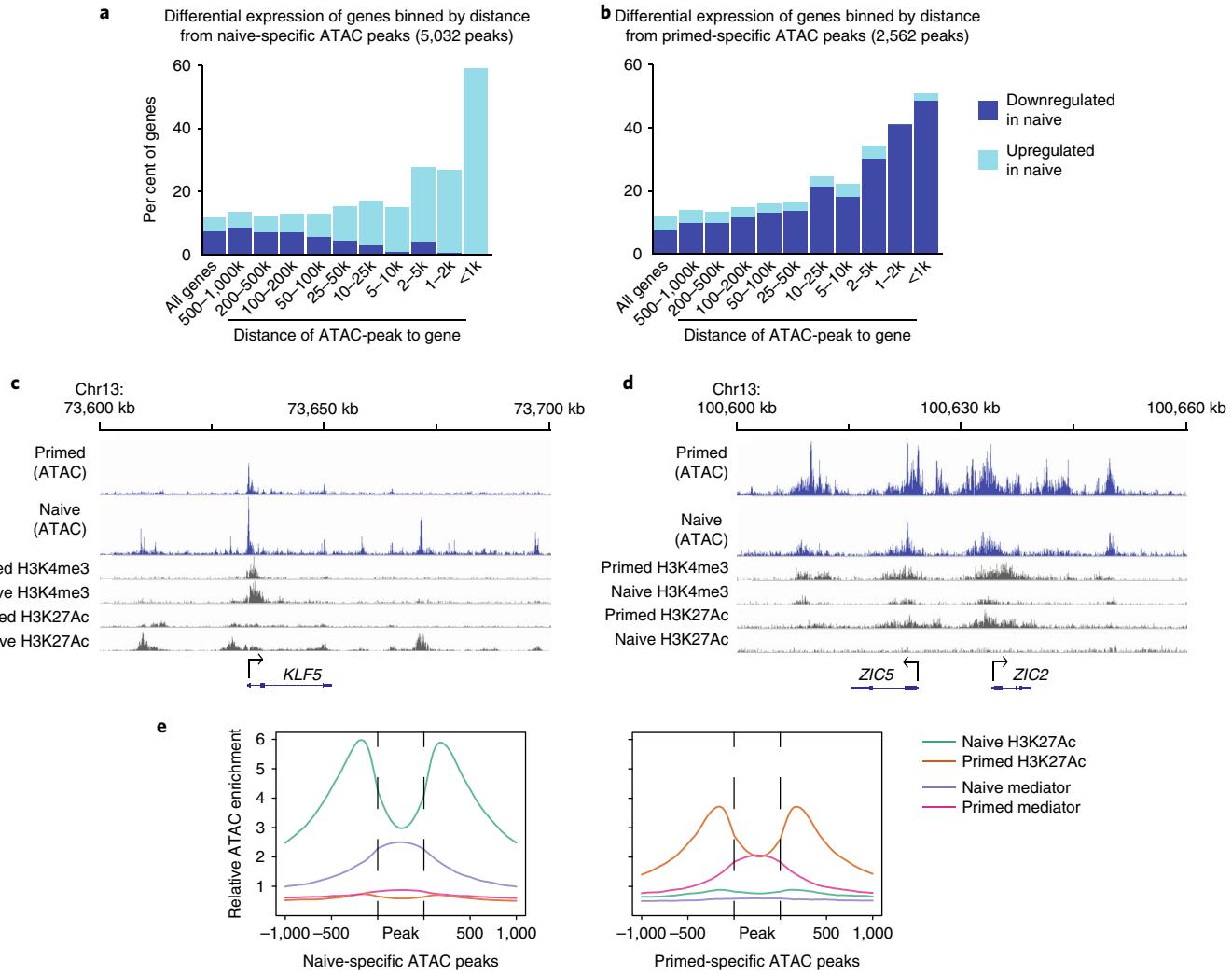
The traditional approach for deriving and culturing human embryonic stem cells (hESCs) from pre-implantation embryos results in cells with primed pluripotency similar to murine post-implantation epiblast stem cells (EpiSCs). However, new medium formulations for transitioning or deriving hESCs in the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

Results

Activator protein-2 motifs are strongly enriched in naive-specific regulatory elements. To identify transcription factors critical for naive human pluripotency, we mapped open chromatin using the assay for transposase-accessible chromatin (ATAC-seq[13]) in naive and primed hESCs (Supplementary Fig. 1a and Supplementary Table 1). Cells were cultured in five inhibitors plus LIF, Activin A and FGF2 (5iLAF) to recapitulate the naive 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 Fig. 1b), with enrichment associating with gene expression. We defined sets of ATAC-seq peaks in naive and primed hESCs, as well as peaks specific to either the naive or primed states (Supplementary Fig. 1c, Supplementary Table 2 and Methods). Although all sets showed enrichment of the promoter sequence, this enrichment was much weaker for naive and primed-specific open sites (Supplementary Fig. 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 transcription start site (TSS), and upregulation of that gene in the naive state, and between the appearance of a primed-specific ATAC peak near a gene TSS and downregulation in the naive state (Fig. 1a,b and Supplementary Fig. 1d,e,f). This was true whether the ATAC peak was upstream or downstream of the gene transcription start site (Supplementary Fig. 1e,f). For example, naive-specific ATAC peaks are observed in the vicinity of the naive-specific Kruppel-like

[1]Department of Molecular, Cell and Developmental Biology, University of California Los Angeles, Los Angeles, CA, USA. [2]Molecular Biology Institute, University of California, Los Angeles, CA, USA. [3]Department of Anatomy and Developmental Biology, Monash University, Clayton, Victoria, Australia. [4]Development and Stem Cells Program, Monash Biomedicine Discovery Institute, Clayton, Victoria, Australia. [5]Australian Regenerative Medicine Institute, Monash University, Clayton, Victoria, Australia. [6]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California Los Angeles, Los Angeles, CA, USA. [7]Howard Hughes Medical Institute, University of California Los Angeles, Los Angeles, CA, USA. Present address: [8]Department of Biochemistry, McGill University, Montreal, Quebec, Canada. *e-mail: jacobsen@ucla.edu; clarka@ucla.edu



f Top motifs enriched over naive-specific ATAC peaks

Motif	% Targets	% Background	Fold enriched	*P* value
KLF	51.49	16.47	3.12	$P < 10^{-708}$
AP2	42.03	12.88	3.26	$P < 10^{-572}$
OCT4	11.96	3.11	3.85	$P < 10^{-169}$
SOX	27.33	14.23	1.92	$P < 10^{-128}$

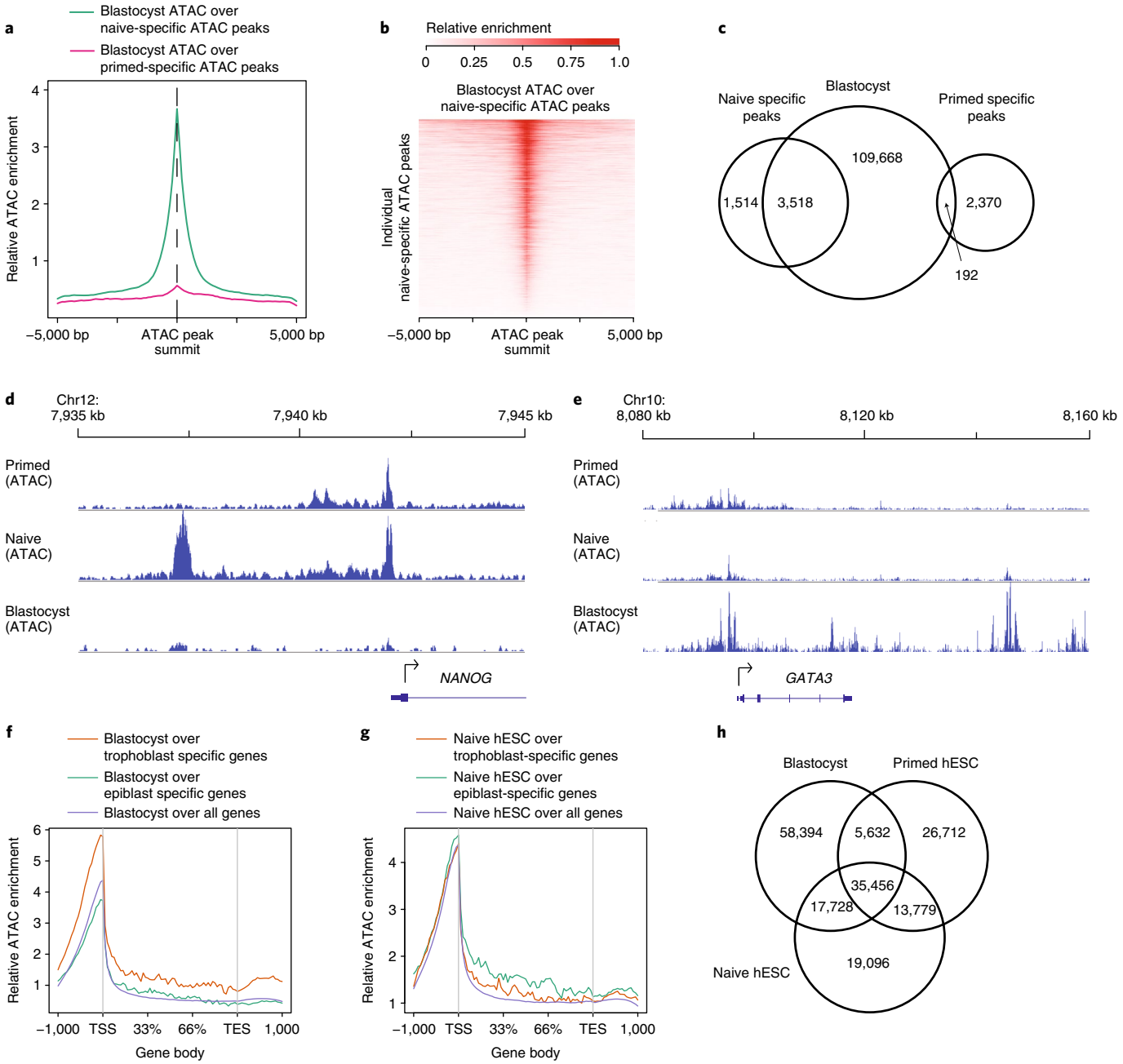
g Top motifs enriched over primed-specific ATAC peaks

Motif	% Targets	% Background	Fold enriched	*P* value
SOX	36.49	14.02	2.60	$P < 10^{-176}$
OCT4	14.05	2.75	5.11	$P < 10^{-138}$
ZIC	22.48	8.99	2.50	$P < 10^{-92}$

Fig. 1 | Determination of regulatory elements specific to the naive and primed states in humans. a,**b**, Percentage of the time a gene whose TSS is a given distance from a naive-specific ATAC peak (**a**) or primed-specific ATAC peak (**b**) is upregulated or downregulated in the naive state. **c**,**d**, Many naive-specific ATAC-seq peaks appear proximal to *KLF5* (**c**), a gene highly upregulated in the naive state, while primed-specific ATAC peaks are present near the primed-specific *ZIC* genes (**d**). 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 naive-specific (left) and primed specific (right) ATAC peaks. Distance on *x* axis is delineated in base pairs. **f**,**g**, Most statistically significant transcription factor binding motifs enriched in naive-specific (**f**) or primed-specific (**g**) ATAC peaks were calculated using a cumulative binomial distribution[19]. Pooled data from four naive and four primed biological replicates were used.

factor 5 (*KLF5*), and primed-specific ATAC-seq peaks are observed in the vicinity of the primed-specific genes *ZIC2* and *ZIC5* (Fig. 1c,d). These observations are consistent with a high proportion of ATAC-seq peaks corresponding to enhancers that regulate nearby genes. A comparison with data from a published chromatin immunoprecipitation assay with sequencing (ChIP-seq) in naive and primed hESCs[16] revealed enrichment of Mediator over naive 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 probably explained by nucleosome depletion (Fig. 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 naive and primed states, we determined enrichment of known transcription factor binding motifs in the naive and primed-specific ATAC peaks (Fig. 1f,g)[19]. The strongest statistical enrichment in the naive state corresponded to the KLF



i

		Blastocyst	Naive hESC	Primed hESC	Naive+blast intersect (pre-implant)	Primed blast intersect	Naive primed intersect (hESC)	Triple intersect
GATA3	AGATAASA	2.64	1.12	No enrich	1.96	1.42	No enrich	1.09
AP2	SCCCSAGGSSAT	2.35	1.99	No enrich	2.50	1.11	No enrich	1.08
KLF	AGGGTGTGGC	1.65	2.47	1.07	2.32	1.14	1.23	1.44
OCT-SOX	ATTTGCATAACAATG	1.24	7.80	8.47	4.27	3.52	8.03	4.43

Fig. 2 | Most naive-specific ATAC peaks are present in other naive human cells and the human embryo. a, Normalized ATAC-seq reads from the human blastocyst plotted relative to naive-specific and primed-specific peaks. Note the far greater enrichment over naive-specific peaks. **b**, Blastocyst ATAC-seq plotted relative to all naive-specific ATAC peaks. Note the enrichment over almost all naive-specific peaks, indicating that they are open in the blastocyst. **c**, Most naive-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, naive hESCs and blastocyst in the viscinity of *NANOG* (**d**) and *GATA3* (**e**). Peak height is normalized to the 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. TES, transcription end site. **h**, Venn diagram showing overlap of all ATAC-seq peaks in blastocyst, naive hESCs and primed hESCs. **i**, Enrichment of GATA, AP2, KLF and OCT-SOX motifs in each set identified in **h**. Note the enrichment of AP2 and KLF motifs in both blastocyst and naive hESCs, stronger enrichment of GATA in blastocyst, and stronger enrichment of OCT-SOX in ESCs.

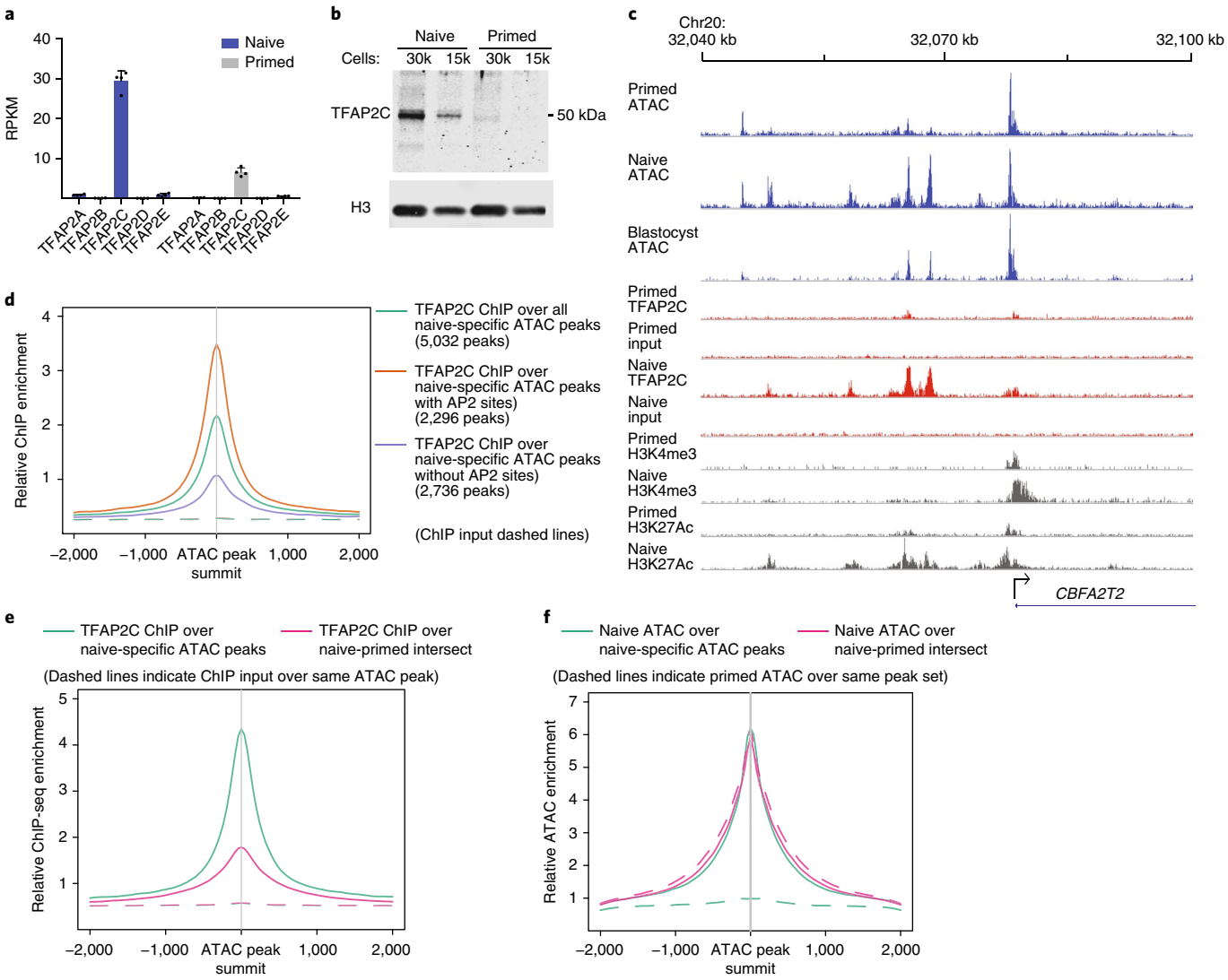


Fig. 3 | TFAP2C is highly enriched over naive-specific open chromatin in humans. a, *TFAP2C* is highly expressed in naive cells, both relative to other AP2 transcription factors and relative to primed cells. Mean and standard deviation are shown, with dots representing each replicate ($n = 4$ independent experiments). **b**, TFAP2C protein is highly upregulated in the naive pluripotent state. Data represent one out of five independent experiments with similar results. H3, histone 3. **c**, Strong co-enrichment of TFAP2C with naive-specific ATAC peaks at the *CBFA2T2* locus. **d**, Global enrichment of TFAP2C relative to the summits of different categories of naive-specific ATAC peaks. TFAP2C is enriched over naive-specific ATAC peaks, especially those with AP2 motifs. **e**,**f**, TFAP2C is strongly enriched over naive-specific ATAC peak summits compared with enrichment over regions that show ATAC enrichment in both naive and primed cells (naive-primed intersect) (**e**), even though both peak sets show similar ATAC enrichment (**f**). Uncropped western blot images are provided in Supplementary Fig. 9. Source data for **a** are provided in Supplementary Table 8.

motif, consistent with the strong upregulation of KLF family factors in naive hESCs and the known role for KLF in naive-state pluripotency in mouse and human[9,10,20]. Similarly, the motif of the primed-specific[21] ZIC factors was enriched in the primed peaks. Unexpectedly, very strong enrichment for the activator protein-2 (AP2) transcription factor motif was observed for naive-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 naive state.

Naive-specific regulatory elements are present in vivo. To determine the in vivo relevance of our set of naive-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 naive-specific peaks, both relative to the surrounding sequence and relative to primed peaks (Fig. 2a-c and Supplementary Fig. 2a), validating the biological relevance of these peaks. Nonetheless, there were marked differences between the open chromatin patterns in whole blastocyst and naive 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* (Fig. 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 (Fig. 2f,g).

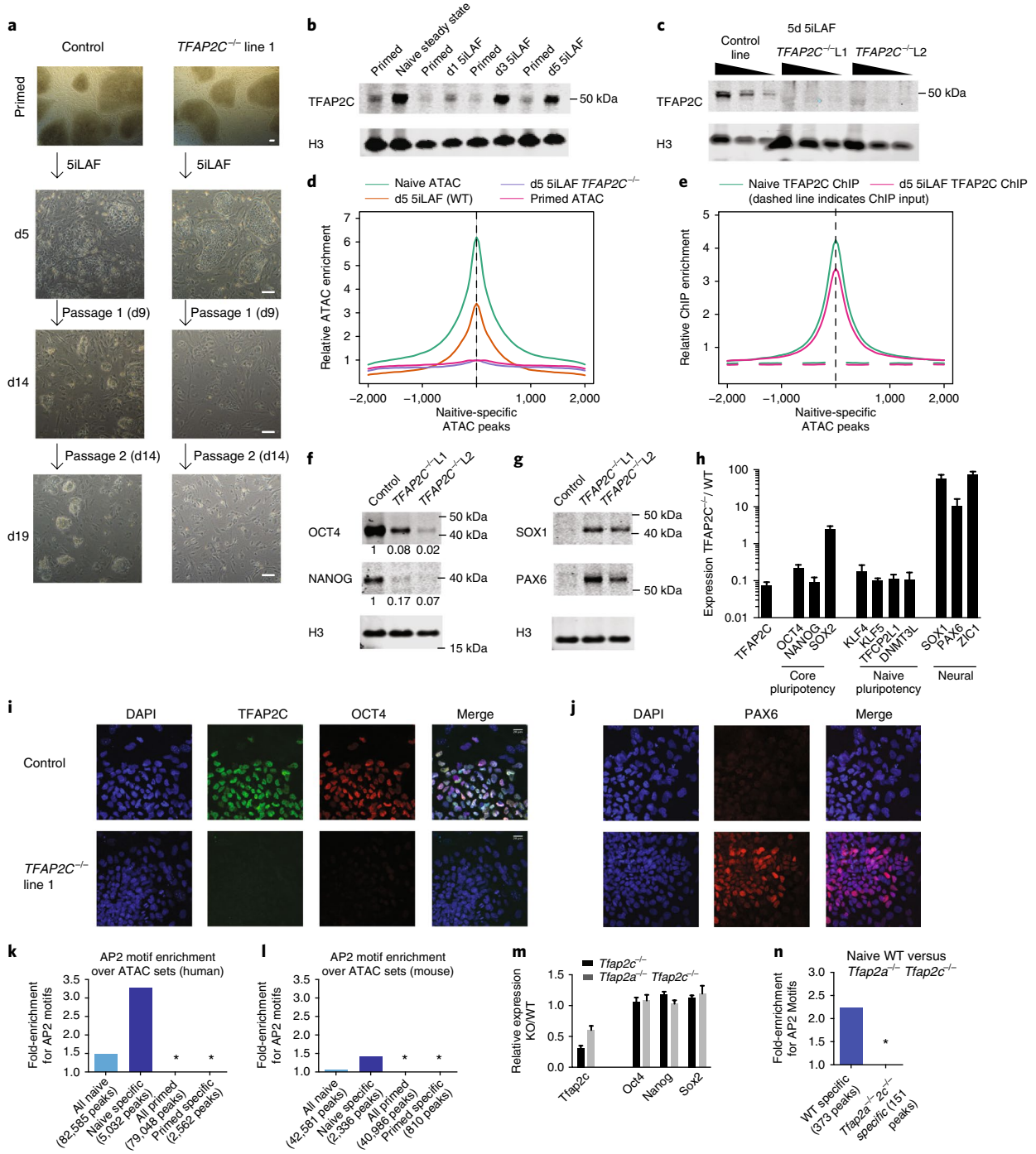


Fig. 4 | *TFAP2C*$^{-/-}$ cells differentiate in naive media. a, *TFAP2C*$^{-/-}$ hESCs self-renew in primed conditions but differentiate and fail to self-renew on treatment in naive (5iLAF) media. Scale bars, 100 µm. Data represent one of four independent experiments with similar results. **b**, Western blot of TFAP2C upon culture in primed or 5iLAF conditions. TFAP2C is strongly induced within 3 days of treatment with 5iLAF. Data represent one of two independent experiments with similar results. **c**, Western blot for TFAP2C after 5 days of 5iLAF culture. TFAP2C is absent from *TFAP2C*$^{-/-}$-deficient lines. **d**, ATAC-seq openness of naive, d5 5iLAF wild type (WT) and *TFAP2C*$^{-/-}$, and primed cells over naive-specific ATAC peaks. Note that, after 5 days of reversion, substantial opening of the naive-specific ATAC peaks has already occurred, but not in the *TFAP2C*$^{-/-}$ cells. **e**, TFAP2C ChIP enrichment shown over naive and d5 5iLAF samples. In **e**, the ChIP input for each set is shown as a dashed line. **f**, Western blot for OCT4 and NANOG in control and *TFAP2C*$^{-/-}$ cells after 5 days of culture in 5iLAF. Quantification is normalized to histone below. **g**, Western blot for SOX1 and PAX6 in control and *TFAP2C*$^{-/-}$ cells. **h**, Relative RPKM (reads per kilobase per million mapped reads) of pluripotency and neural markers in RNA-seq. Data are 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 bars, 20 µm. **k**,**l**, Fold enrichment for AP2 motifs in the specified peak sets in humans (**k**) and mouse (**l**). *No enrichment. Although AP2 motifs are enriched in naive-specific peaks in both species, the enrichment is much stronger in the human naive-specific set. **m**, Expression of key pluripotency markers in WT, *Tfap2c*$^{-/-}$ and *Tfap2a*$^{-/-}$*Tfap2c*$^{-/-}$ cells in 2i + LIF conditions. $n=4$ biological replicates 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 was determined. *No enrichment. Uncropped western blots are provided in Supplementary Fig. 9. Source data for **h** and **m** are provided in Supplementary Table 8.

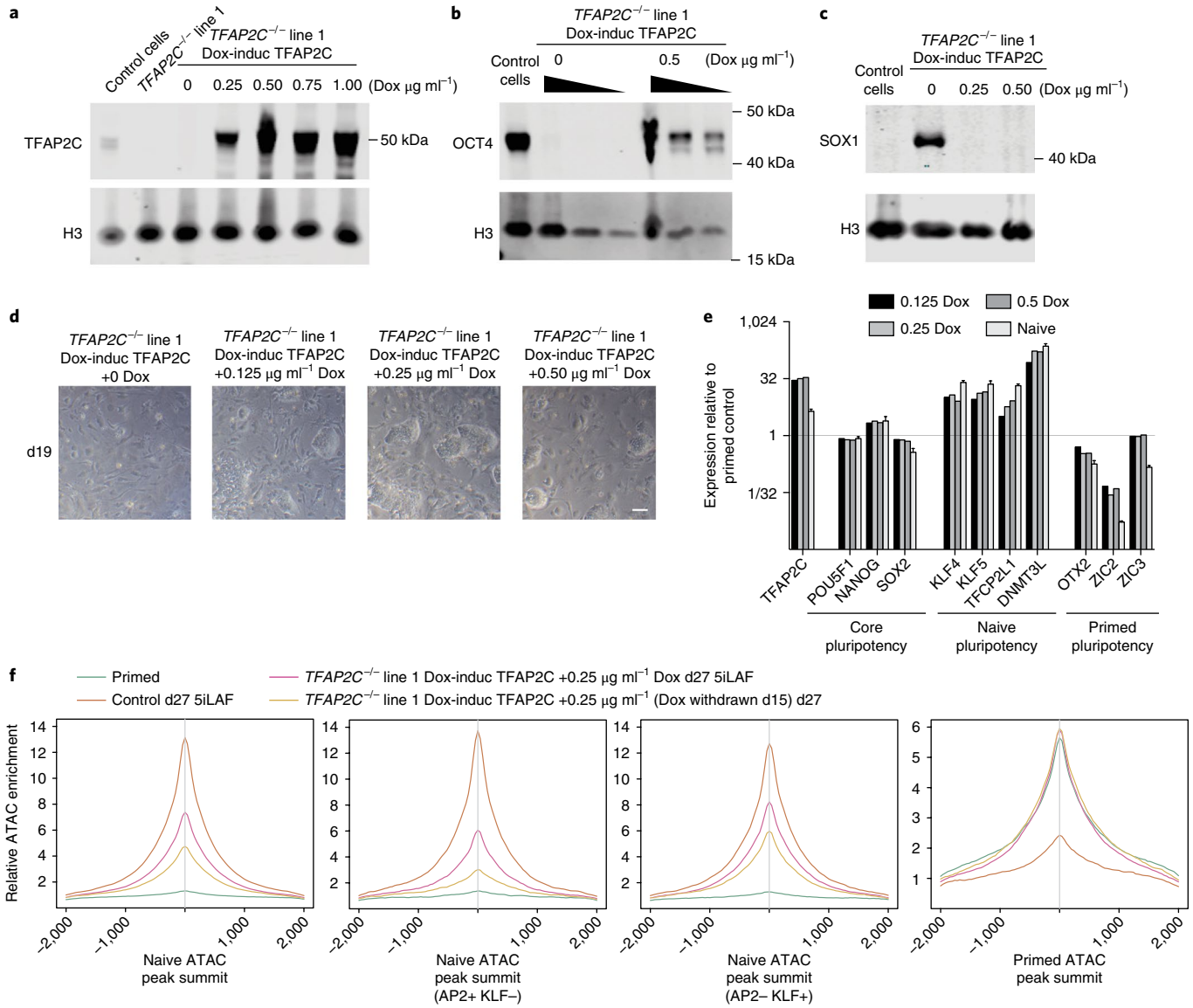


Fig. 5 | Ectopic expression of TFAP2C partially rescues the *TFAP2C*$^{-/-}$ phenotype. a, Quantitative western blot showing tunable TFAP2C induction in *TFAP2C*$^{-/-}$ background in primed conditions. **b**,**c**, Western blots show rescue of OCT4 expression and SOX1 repression upon doxycycline-inducible TFAP2C expression. Lysates were collected after 5 days of treatment with 5iLAF and the indicated concentration of doxycycline. **d**, Appearance of round naive-like colonies in line with ectopic TFAP2C expression. Scale bar, 100 μm. Results represent one of four independent experiments with similar results. **e**, Partial rescue of upregulation of naive pluripotency factors, downregulation of primed-factors with ectopic TFAP2C expression; one replicate for dox induction samples and primed control, four for naive 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 naive-specific peaks (5,032 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 (1,551 peaks), and primed-specific peaks (2,562 peaks). Reduced ATAC-seq density over naive specific peaks and increased density over primed-specific peaks, in the sample in which doxycycline had been withdrawn. Closing of naive specific peaks is especially pronounced over the subset of peaks that contain AP2 sites but no KLF sites ($AP2^{+}$ KLF^{-}). Peak subsets are listed in Supplementary Table 2. Uncropped western blot images are provided in Supplementary Fig. 9. Source data for **e** are provided in Supplementary Table 8.

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

Using an alternative approach we further confirmed the in vivo relevance of the naive-specific ATAC peaks by analysing DNA methylation, given that regulatory elements are typically hypomethylated relative to the surrounding sequence[30,31]. Consistent with this trend, we observe strong hypomethylation of naive-specific ATAC-seq peaks in naive hESCs cultured in 5iLAF or in t2iLGö[10], a different culture method for generating naive hESCs[6,12] (Supplementary Fig. 2b).

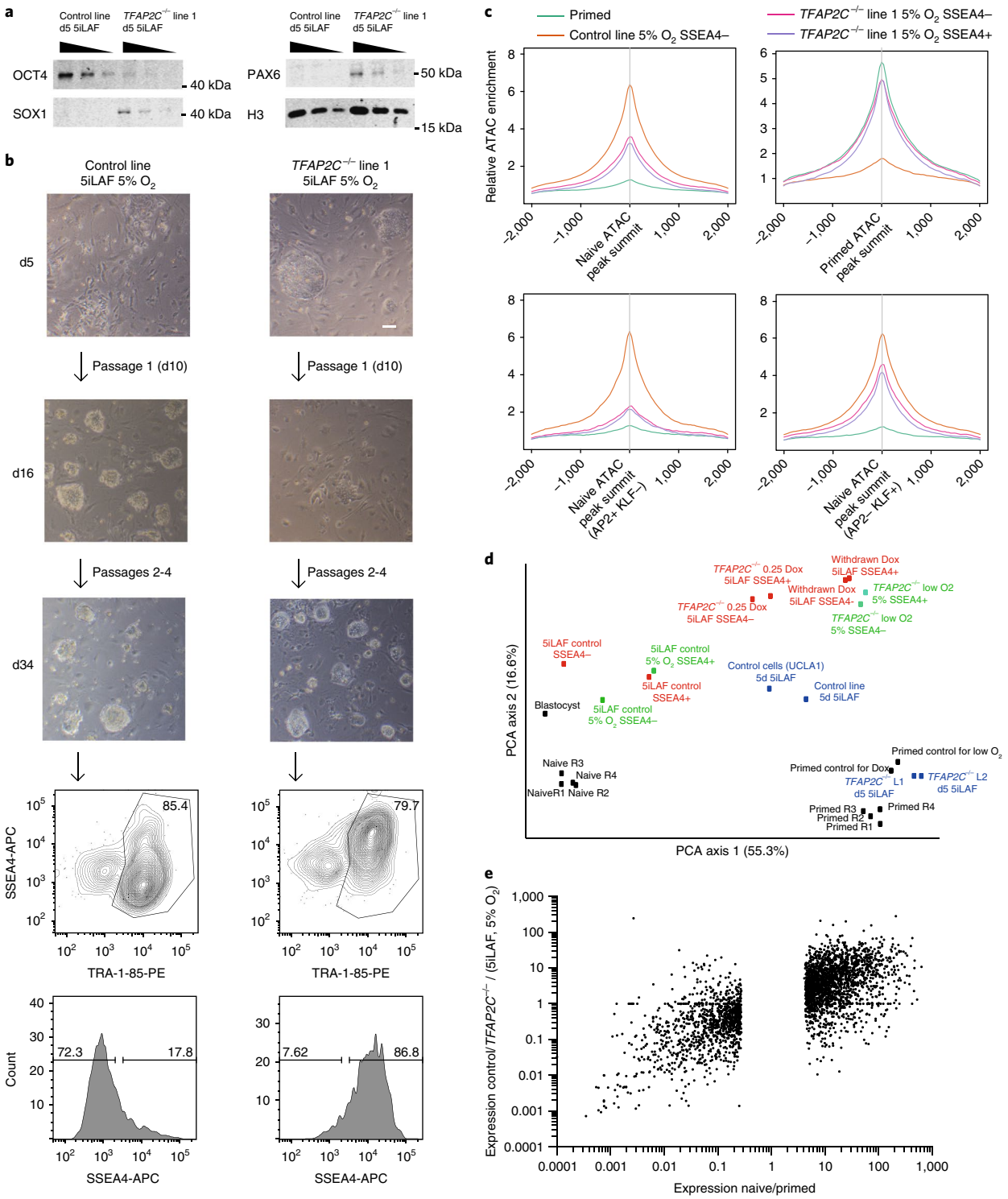


Fig. 6 | *TFAP2C*$^{-/-}$ cells survive in 5iLAF in 5% O_2 conditions but do not transition to naive state. a, Western blots for the pluripotency marker OCT4 and the neural markers SOX1 and PAX6 in WT and *TFAP2C*$^{-/-}$ cells after 5 days in 5iLAF at 5% O_2. **b**,**c**, Bright-field images of control and *TFAP2C*$^{-/-}$ cells in 5iLAF culture. Initially the *TFAP2C*$^{-/-}$ cells show morphology similar to what is observed under ambient oxygen concentration conditions (compare to Fig. 4a). However, some colonies are observable after passaging. These colonies show a shift toward the SSEA4^{+} (primed) state. Scale bar, 100 µm. **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 data sets generated in this work. Blue dots: after 5 days in 5iLAF, WT control cells show an ATAC-seq landscape part way between primed and naive, whereas *TFAP2C*$^{-/-}$ cells show no change toward naive. Green dots: although *TFAP2C*$^{-/-}$ cells survive in low oxygen conditions, they have an ATAC-seq landscape much more similar to primed than naive cells. Red dots: ectopic doxycycline-dependent expression of TFAP2C in *TFAP2C*$^{-/-}$ partially rescues the naive-landscape, and withdrawal of doxycycline induces a shift towards primed identity. Shown for comparison are control cells reverted at the same time. **e**, Genes differentially regulated in naive versus primed hESCs are plotted. Note that genes more highly expressed in naive cells are expressed lower in *TFAP2C*$^{-/-}$. The RPKM values correspond to SSEA4^{-} cells in control (average of two biological replicates) and SSEA4^{+} in *TFAP2C*$^{-/-}$ (average of three biological replicates). Uncropped western blot images are provided in Supplementary Fig. 9.

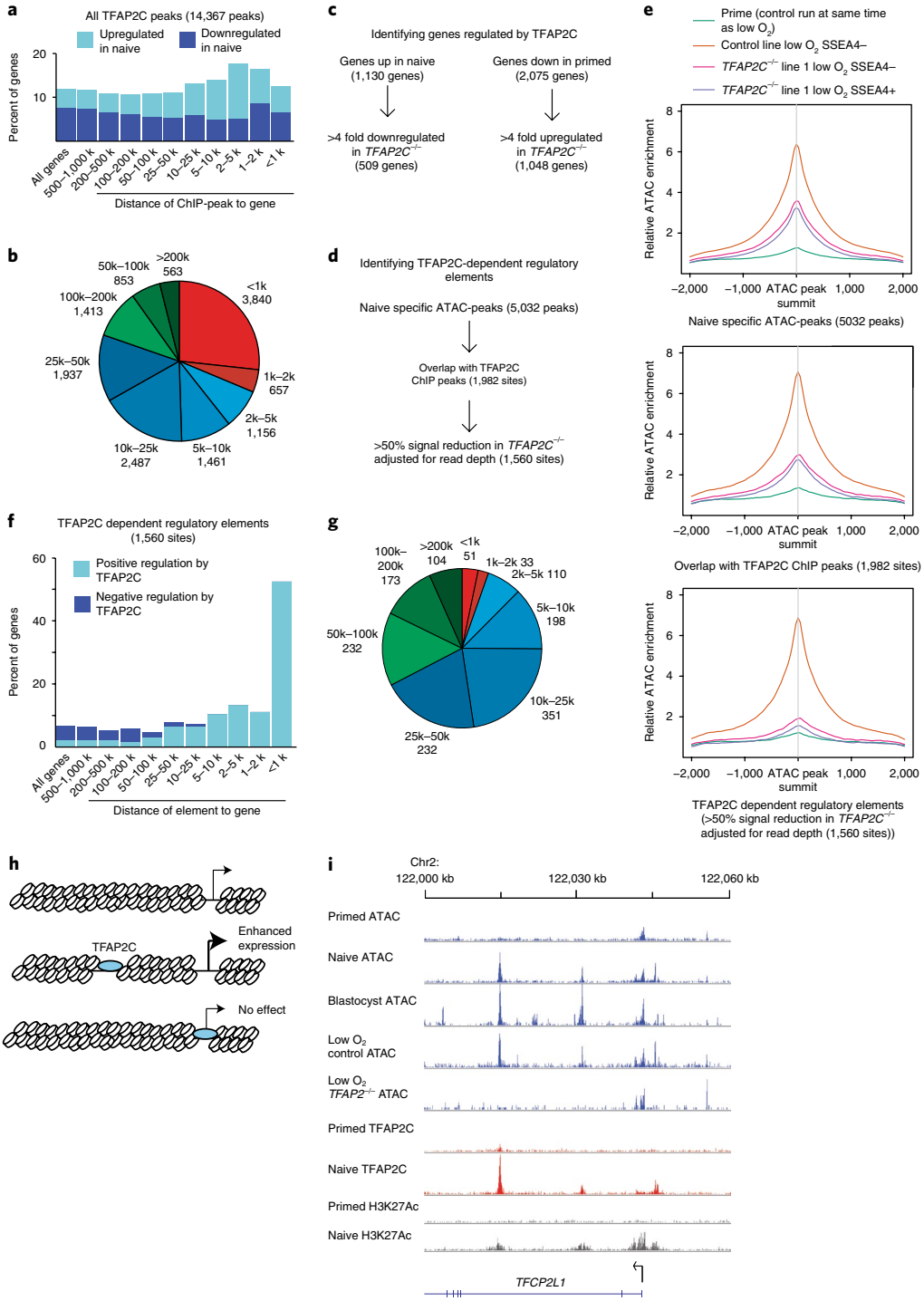


Fig. 7 | Identifying direct regulatory targets of TFAP2C. a, Percentage of the time a gene whose TSS is a given distance from a TFAP2C ChIP-seq peak is upregulated or downregulated in naive hESCs. Notice the much weaker correspondence compared with Fig. 1a, and the lack of any effect at the promoter. **b**, Distance of TFAP2C ChIP-seq peak to nearest TSS. **c**, To identify pluripotency-state genes positively or negatively regulated by TFAP2C, we identified the subset of naive-specific genes downregulated less than fourfold in *TFAP2C*$^{-/-}$ (positively related by TFAP2C) and primed-specific genes upregulated more than fourfold 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 naive-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 naive-specific ATAC peaks, naive-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 the nearest gene TSS. Note that the vast majority of such elements are enhancers. **h**, Schematic demonstrating the typical regulatory role of TFAP2C in naive 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 naive-pluripotency factor *TFCP2L1*.

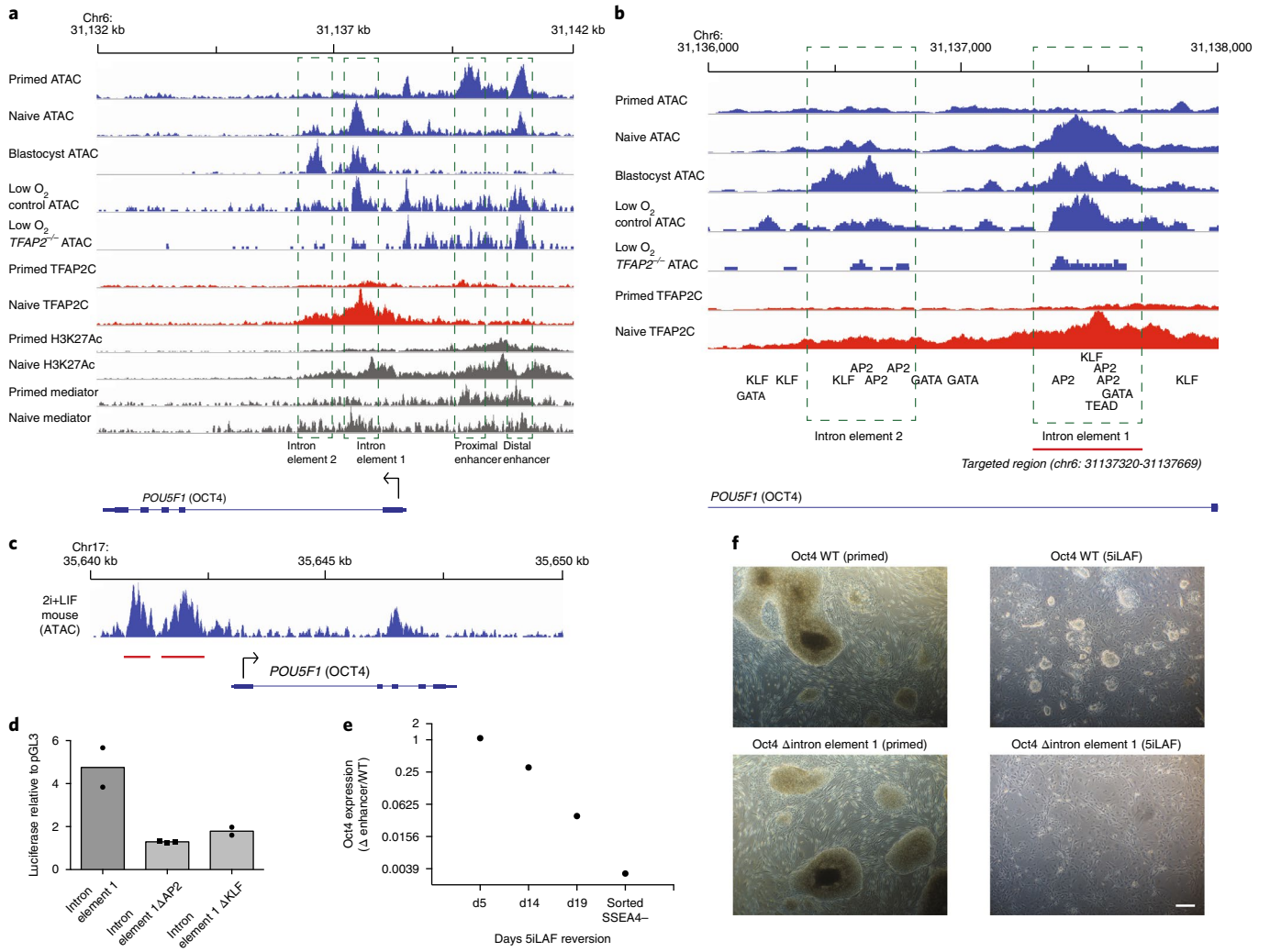


Fig. 8 | A TFAP2C$^+$ intronic enhancer of OCT4. a, Chromatin landscape of OCT4. Two putative enhancers, intron element 1 and intron element 2, are present in blastocyst. Intron element 1 is also strongly enriched in naive cells and lost in *TFAP2C*$^{-/-}$. **b**, The location of consensus motifs for 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 naive (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 experiments, except for the ΔAP2 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 naive 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 naive colonies on reversion. Scale bar, 200 µm. Images are representative of three independent reversions. Source data for **d** are available in Supplementary Table 8.

Similarly, we observe a more pronounced drop in DNA methylation between the oocyte and blastocyst stages of human embryonic development at our defined set of naive-specific ATAC peaks than over the surrounding sequence or primed-specific ATAC peaks (Supplementary Fig. 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 naive state. Of the five AP2-family transcription factors present in humans, only TFAP2C is highly expressed in the naive state (Fig. 3a). TFAP2C is upregulated in naive cells at both the RNA and protein level (Fig. 3a,b) and is expressed in human morula and pre-implantation epiblast[3,8,32]. ChIP for TFAP2C showed strong enrichment over naive-specific ATAC-seq peaks (Fig. 3c,d), especially those containing AP2 motifs (Fig. 3d). Furthermore, TFAP2C showed stronger enrichment at naive-specific ATAC-seq peaks than at regions open in both the primed and naive state (Fig. 3e), even though both ATAC peak sets show similar ATAC-seq enrichment in the naive cells (Fig. 3f). Combined with our observation that AP2 motifs are specifically enriched in naive-specific peaks, these data indicate that TFAP2C may facilitate the opening of naive-specific regulatory elements.

We used CRISPR to target *TFAP2C* in the primed state, and lines containing null mutations of both alleles were confirmed karyotypically normal (Supplementary Fig. 3a,b). In the primed state these lines showed normal expression of pluripotency genes and markers (Supplementary Fig. 3c–f). The *TFAP2C*$^{-/-}$ cells were able to exit pluripotency normally with spontaneous embryoid body (EB) differentiation, and showed a skew towards neural lineage, consistent

with a known role for TFAP2C in regulating the formation of neural versus non-neural ectoderm[27] (Supplementary Fig. 3g).

Upon reversion in naive 5iLAF medium, we observed a dramatic morphological change in the *TFAP2C*$^{-/-}$ hESC lines (Fig. 4a and Supplementary Fig. 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, but is absent in the *TFAP2C*$^{-/-}$ lines (Fig. 4b,c). TFAP2C ChIP and ATAC-seq at day 5 of reversion show opening of naive-specific enhancers and enrichment of TFAP2C at these sites, but no opening of these enhancers in *TFAP2C*$^{-/-}$ (Fig. 4d,e). Initially, the *TFAP2C*$^{-/-}$ cells divided rapidly in 5iLAF, but after one passage, round naive-like colonies were identified only in controls (Fig. 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 (Fig. 4f–j), a result confirmed by gene ontology (GO) analysis[33] (Supplementary Fig. 4c). Similarly, ATAC-seq showed a loss of AP2 and pluripotency transcription factor motifs in open chromatin in the *TFAP2C*$^{-/-}$ cells after 5 days of reversion, and instead, a gain of peaks enriched for motifs related to neural development such as SOX and ZIC (Supplementary Fig. 4d,e).

To confirm that this finding was human specific, we performed ATAC-seq on murine ESCs (mESCs) cultured in the naive state (2i + LIF) as well as primed EpiSCs[34] (Supplementary Table 4). We discovered that AP2 sites were enriched in naive-specific open chromatin in 2i + LIF mESCs. However, the degree of enrichment was far lower than for human naive cells (Fig. 4k,l). We generated *Tfap2c*$^{-/-}$ and *Tfap2a*$^{-/-}$*Tfap2c*$^{-/-}$ mESCs (Supplementary Fig. 4f,g) and found normal expression of pluripotency markers in 2i + LIF (Fig. 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 (Fig. 4n). Thus, AP2 transcription factors play a more modest role in murine than human naive states.

Withdrawal of TFAP2C in naive state causes shift towards the primed state. Next, we generated a *TFAP2C*$^{-/-}$ mutant line capable of expressing TFAP2C in a doxycycline-dose-dependent manner (Fig. 5a). Overexpression of TFAP2C in primed media did not result in a pronounced shift toward naive-gene expression, and TFAP2C primarily honed to regions of chromatin that were already open in primed hESCs (Supplementary Fig. 5a,b), arguing that the combinatorial activity of multiple factors is necessary for primed to naive reversion.

We then reverted the *TFAP2C*$^{-/-}$ Dox-inducible line, using 5iLAF medium 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 the formation of colonies with naive morphology (Fig. 5b–e and Supplementary Fig. 5c).

To determine the effect of acute loss of TFAP2C, we cultured cells in 5iLAF + doxycycline until naive morphology colonies were apparent (day 15), then switched to media without doxycycline. No acute phenotype was observed; instead, a gradual loss of cells from culture occurred (Supplementary Fig. 5d,e). Cells remaining 12 days after doxycycline withdrawal showed increased staining for the primed surface marker SSEA4 and closing of naive-specific ATAC peaks, especially the subset containing AP2 sites but no KLF sites (Fig. 5f and Supplementary Fig. 5f). These findings indicate that TFAP2C is essential for maintenance as well as establishment of the naive 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% O_2), morphological differences, loss of OCT4, and gain of SOX1 and PAX6 were all apparent upon culture in 5iLAF in 5% O_2 (Fig. 6a,b and Supplementary Fig. 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 (Fig. 6b and Supplementary Fig. 6a). However, almost all *TFAP2C*$^{-/-}$ cells had high SSEA4 surface expression (Fig. 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 naive-specific peaks, and increased openness over primed-specific ATAC-seq peaks compared with controls (Fig. 6c). Moreover, principle component analysis of the ATAC-Seq data sets showed a closer similarity to primed cells (Fig. 6d), with the transcriptome of the persisting *TFAP2C*$^{-/-}$ cells present in 5% O_2 shifted towards expression of primed-specific genes (Fig. 6e and Supplementary Fig. 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 Fig. 6d,e). Finally, we reverted the *TFAP2C*$^{-/-}$ in t2iLGöY naive media[36] in 5% O_2, and similar to the results in 5iLAF, the *TFAP2C*$^{-/-}$ cells lacked nuclear KLF17, a marker of naive cells (Supplementary Fig. 6f,g). In total, these data support an essential role for TFAP2C in the reversion of primed hESCs to the naive 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 (Fig. 7a,b), making it difficult to discern which binding events are important for gene regulation. Compared with the striking correlation observed between the presence of a naive-specific enhancer and upregulation of a nearby gene (Fig. 1a), we observed only a modest correlation between the presence of a TFAP2C ChIP peak near a gene and the upregulation of that gene in the naive state or downregulation in *TFAP2C*$^{-/-}$ (Fig. 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 naive 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 naive or primed state and focused on the subset that showed more than fourfold changes in expression in *TFAP2C*$^{-/-}$ (Fig. 7c). Second, we defined a set of TFAP2C-dependent regulatory elements: TFAP2C ChIP-seq peaks that overlapped with naive-specific ATAC peaks and showed reduced openness in *TFAP2C*$^{-/-}$ (Fig. 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 (Fig. 7f and 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 (Fig. 7g). By contrast, TFAP2C ChIP-seq peaks in regions of openness conserved between naive and primed state had virtually no predictive effect on gene expression in *TFAP2C*$^{-/-}$ (Supplementary Fig. 7a,b). In other words, the primary effect of TFAP2C in naive hESCs is most likely to open a discrete set of regulatory elements, mainly enhancers (Fig. 7h).

Genomic Regions Enrichment of Annotations Tool (GREAT) analysis[37] showed that genes within 50 kb 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 and 7). Adjacent genes included *CBFA2T2*, *TFCP2L1*, *KLF5*, *SOX2*, *FGF4*, *NANOG*, *DPPA3*, *DPPA5* and *TFAP2C* itself (Fig. 7i, Supplementary Fig. 7c–e and Supplementary Table 7), supporting a role for TFAP2C in directly promoting the naive pluripotent program.

An intronic enhancer for OCT4 is active in naive 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 further 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 (Fig. 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 (Fig. 8b). Intron element 1 shows evolutionary conservation across placental mammals (Supplementary Fig. 8a) and is open and enriched for TFAP2C in naive hESCs (Fig. 8a,b), but is not open in naive mESC (Fig. 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 Fig. 8b). Furthermore, we observe enhancer activity for this region in a luciferase assay, which is largely eliminated by the loss of either the AP2 sites or KLF site (Fig. 8d).

To examine the role of this enhancer in naive pluripotency, we ablated this sequence using CRISPR–Cas9 and confirmed normal karyotype (Fig. 8b and Supplementary Fig. 8c). We found normal expression of OCT4 (Supplementary Fig. 8d) and self-renewal in the primed state, but a dramatic loss of OCT4 expression accompanied by differentiation upon reversion to the naive state (Fig. 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 naive hESCs. Furthermore, we show that most of these enhancers are present in human embryo and therefore biologically relevant, and are likely to directly regulate genes critical for human naive 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 naive-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 facilitates 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 naive pluripotency, analogous to what happens in the artificial system of in vitro reversion.

The observation that TFAP2C is critical in naive 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 the 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 the 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 naive hESCs include both genes general to the 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 the 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 transcription factors via the intronic enhancers, and only later is regulated from the naive-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 naive pluripotency factor fits into this paradigm, and therefore model organisms may only reveal some of the story of how human embryos develop.

Received: 22 August 2017; Accepted: 20 March 2018;
Published online: 25 April 2018

References

1. Niakan, K. K., Han, J., Pedersen, R. A., Simon, C. & Pera, R. A. Human pre-implantation embryo development. *Development* **139**, 829–841 (2012).
2. Okamoto, I. et al. Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature* **472**, 370–374 (2011).
3. Petropoulos, S. et al. Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* **165**, 1012–1026 (2016).
4. Sahakyan, A. et al. Human naive pluripotent stem cells model X chromosome dampening and X inactivation. *Cell Stem Cell* **20**, 87–101 (2017).
5. Kunath, T. et al. Developmental differences in the expression of FGF receptors between human and mouse embryos. *Placenta* **35**, 1079–1088 (2014).
6. Nakamura, T. et al. A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature* **537**, 57–62 (2016).
7. Deglincerti, A. et al. Self-organization of the in vitro attached human embryo. *Nature* **533**, 251–254 (2016).
8. Blakeley, P. et al. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* **142**, 3151–3165 (2015).
9. 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).
10. Takashima, Y. et al. Resetting transcription factor control circuitry toward ground-state pluripotency in human.*Cell* **158**, 1254–1269 (2014).

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* **15**, 410–415 (2014).
12. 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).
13. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
14. Song, L. et al. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res* **21**, 1757–1767 (2011).
15. Thurman, R. E. et al. The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
16. Ji, X. et al. 3D chromosome regulatory landscape of human pluripotent cells. *Cell Stem Cell* **18**, 262–275 (2016).
17. Rada-Iglesias, A. et al. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
18. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
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* **38**, 576–589 (2010).
20. Guo, G. et al. Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* **136**, 1063–1069 (2009).
21. Buecker, C. et al. Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 838–853 (2014).
22. Auman, H. J. et al. Transcription factor AP-2gamma is essential in the extra-embryonic lineages for early postimplantation development. *Development* **129**, 2733–2747 (2002).
23. Winger, Q., Huang, J., Auman, H. J., Lewandoski, M. & Williams, T. Analysis of transcription factor AP-2 expression and function during mouse preimplantation development. *Biol. Reprod.* **75**, 324–333 (2006).
24. Werling, U. & Schorle, H. Transcription factor gene AP-2 gamma essential for early murine development. *Mol. Cell. Biol.* **22**, 3149–3156 (2002).
25. Hoffman, T. L., Javier, A. L., Campeau, S. A., Knight, R. D. & Schilling, T. F. Tfap2 transcription factors in zebrafish neural crest development and ectodermal evolution. *J. Exp. Zool. B Mol. Dev. Evol.* **308**, 679–691 (2007).
26. Guttormsen, J. et al. Disruption of epidermal specific gene expression and delayed skin development in AP-2 gamma mutant mice. *Dev. Biol.* **317**, 187–195 (2008).
27. Qiao, Y. et al. AP2gamma regulates neural and epidermal development downstream of the BMP pathway at early stages of ectodermal patterning. *Cell Res.* **22**, 1546–1561 (2012).
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* **8**, e71113 (2013).
29. Niakan, K. K. & Eggan, K. Analysis of human embryos from zygote to blastocyst reveals distinct gene expression patterns relative to the mouse. *Dev. Biol.* **375**, 54–64 (2013).
30. Xie, W. et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).
31. Tsankov, A. M. et al. Transcription factor binding dynamics during human ES cell differentiation. *Nature* **518**, 344–349 (2015).
32. Yan, L. et al. Single-cell RNA-seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1131–1139 (2013).
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* **10**, 48 (2009).
34. Tesar, P. J. et al. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196–199 (2007).
35. Xie, H. & Simon, M. C. Oxygen availability and metabolic reprogramming in cancer. *J. Biol. Chem.* **292**, 16825–16832 (2017).
36. Guo, G. et al. Epigenetic resetting of human pluripotency. *Development* **144**, 2748–2763 (2017).
37. McLean, C. Y. et al. GREAT improves functional interpretation of *cis*-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
38. Yeom, Y. I. et al. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881–894 (1996).
39. Eloranta, J. J. & Hurst, H. C. Transcription factor AP-2 interacts with the SUMO-conjugating enzyme UBC9 and is sumolated in vivo. *J. Biol. Chem.* **277**, 30798–30804 (2002).
40. Wong, P. P. et al. Histone demethylase KDM5B collaborates with TFAP2C and Myc to repress the cell cycle inhibitorp21(cip) (CDKN1A). *Mol. Cell. Biol.* **32**, 1633–1644 (2012).
41. Magnusdottir, E. et al. A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nat. Cell Biol.* **15**, 905–915 (2013).
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.* **277**, 8559–8565 (2002).
43. Braganca, J. et al. Physical and functional interactions among AP-2 transcription factors, p300/CREB-binding protein, and CITED2. *J. Biol. Chem.* **278**, 16021–16029 (2003).
44. Bamforth, S. D. et al. Cardiac malformations, adrenal agenesis, neural crest defects and exencephaly in mice lacking Cited2, a new Tfap2 co-activator. *Nat. Genet.* **29**, 469–474 (2001).
45. Kuckenberg, P. et al. The transcription factor TCFAP2C/AP-2gamma cooperates with CDX2 to maintain trophectoderm formation. *Mol. Cell. Biol.* **30**, 3310–3320 (2010).
46. Lee, S. H., Kwon, J. W., Choi, I. & Kim, N. H. Expression and function of transcriptional factor AP-2γ in early embryonic development of porcine parthenotes. *Reprod. Fertil. Dev.* **28**, 1197–1205 (2015).
47. Nichols, J. et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379–391 (1998).
48. Fogarty, N. M. E. et al. Genome editing reveals a role for OCT4 in human embryogenesis. *Nature* **550**, 67–73 (2017).
49. Kalinka, A. T. & Tomancak, P. The evolution of early animal embryos: conservation or divergence?. *Trends Ecol. Evol.* **27**, 385–393 (2012).
50. Irie, N. & Kuratani, S. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat. Commun.* **2**, 248 (2011).

Acknowledgements

The authors thank the UCLA Broad Stem Cell Research Center (BSCRC) Flow Cytometry core and the UCLA BSCRC High Throughput Sequencing Core for technical assistance. W.A.P. was supported by the Jane Coffin Childs Memorial Fund for Medical Research and a UCLA BSCRC Postdoctoral Training Fellowship. D.C. is supported by a UCLA BSCRC Postdoctoral Training Fellowship. W.L. is supported by the Philip J. Whitcome Fellowship from the UCLA Molecular Biology Institute and a scholarship from the Chinese Scholarship Council. Work was funded by R01 HD079546 (ATC) and a NHMRC project grant APP1104560 (to J.M.P.) and a Sylvia and Charles Viertel Senior Medical Research Fellowships (to J.M.P.). All work with human pre-implantation embryos was funded by UCLA BSCRC and not the National Institute of Health. S.E.J. is a fellow of the Howard Hughes Medical Institute.

Author contributions

W.A.P., D.C., J.H. R.K., T.J.H., A.L. and X.L. conducted experiments. W.A.P. and W.L. conducted bioinformatics analysis. W.A.P. and A.T.C. wrote the manuscript. J.M.P., S.E.J. and A.T.C. supervised the research.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41556-018-0089-0.

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

Correspondence and requests for materials should be addressed to S.E.J. or A.T.C.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Human cell culture. Culturing of primed and naive hESCs, and reversion from primed to naive state, were conducted as previously reported[12], the only modification being the inclusion of 1× Primocin (Invivogen) in all media. 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 naive t2iLGöY state was performed as previously described[36], generating the t2iLGöY medium 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), 1× penicillin/streptomycin/glutamine (ThermoFisher), 1× non-essential amino acids (ThermoFisher), 55 µM β-mercaptoethanol (ThermoFisher), 1× Primocin (Invivogen) and 1,000 U ml^{-1} ESGRO LIF (Millipore) in knockout DMEM (ThermoFisher). Cells were passaged with 0.25% trypsin every 3 days and cultured on mouse embryonic fibroblast (MEFs). Before all RNA-seq or ATAC-seq experiments, cells were cultured for at least five passages in 2i + LIF media: 1× N2 supplement, 1× B27 supplement, 1× penicillin/streptomycin/glutamine (ThermoFisher), 1× non-essential amino acids (ThermoFisher), 55 µM β-mercaptoethanol (ThermoFisher), 1× Primocin (Invivogen), 3 µM CHIR99021 (Stemgent), 0.5 µM PD0325901 (Stemgent) and 1,000 U ml^{-1} 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 3 days with 0.25% trypsin and plated at 50,000 cells per 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 medium[12]. EpiSCs were passaged with 1× collagenase type IV (Life Technologies) every 3 days.

Collecting cell populations for sequencing experiments. To sort primed and naive hESCs in the 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 5 min 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 in the following.

To separate mESCs or EpiLCs from MEFs, cells were detached with 1× trypsin, quenched, and then washed with 1× FACS buffer, stained with 1:150 anti-SSEA1 in 1× FACS buffer, then washed and stained with DAPI immediately before sorting. SSEA1$^+$ SSC^{lo} DAPI$^-$ cells were sorted and then used for RNA-seq or ATAC-seq.

For human or murine ChIP experiments, cells were collected using Accutase, quenched, and washed with 1× PBS before fixation.

EB differentiation. Primed hESCs 7 days after plating were washed with 1× PBS and treated with 1× collagenase type IV at 37 °C for 1 h, then removed from the plate with short strokes by a cell scraper. A 4 ml volume of MEF medium (10% qualified fetal bovine serum (ThermoFisher), Penicillin/Streptomycin/Glutamine (ThermoFisher), 1x Primocin (Invivogen) in KnockOut DMEM) was added to the well and the colonies were allowed to settle in a 50 ml conical tube. Medium was then aspirated by pipette, and the cells were resuspended in 3 ml mTESR medium with ROCKi and plated in a low-attachment six-well plate. At the 24 h time point, the EBs were transferred into a 50 ml conical tube and allowed to settle. Medium was removed and replaced with primed hESC medium[12] lacking FGF. The medium was changed again at 72 h and the EBs were collected at the 144 h timepoint.

Embryo isolation and generation. Day 6 vitrified blastocysts were thawed using Vit Kit-Thaw (Irvine Scientific) according to the manufacturer's protocol. Embryos were cultured in drops of continuous single culture medium (Irvine Scientific) supplemented with 20% serum substitute supplement (Irvine Scientific) under mineral oil for 2–3 h 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 base G. Human *TFAP2C* was targeted with the guide sequence GCTTAAATGCCTCGTTAC. The human *OCT4* intronic enhancer was targeted with 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 for at least two passages in mTESR1 medium (StemCell Technologies) on Matrigel (Corning). Cells in exponential growth phase were collected with Accutase, and 800,000 hESCs were electroporated with 4 µg plasmid DNA using the CA-137, Primary Cell 3 program in an Amaxa 4-D Nucleofector X-subunit (Lonza). After transfection, cells were transferred to a single well of a 24-well plate containing primed hESC medium[12] supplemented with 1× Y-27632 (Stemgent). Before 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 10 cm 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 10 cm plates 3 days after transfection, and did not require later subcloning.

To obtain clonal and physically separate colonies, cells were collected with Accutase and 10,000 cells were plated on 10 cm plates to allow physically separated colonies to grow. Cells were fed with primed hESC medium starting 2 days after plating and the medium 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 for an additional six to seven days in primed hESC medium. Cells were then split with Accutase. Two-thirds of the material was used for DNA extraction and screening (see section 'Screening for mutations'), the remaining third was plated in primed hESC medium with ROCKi in a well of a 12-well plate. After two days, the medium was changed to primed hESC media without ROCKi and the cells were passaged using normal primed conditions described above in the 'Cell culture' sections.

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. mESCs were plated the day before transfection at a density of 150,000 cells per well in a six-well plate for each transfection sample. On the day of transfection, cells were collected with trypsin, precipitated, and then resuspended in 2.5 ml of serum + LIF medium (see Murine cell culture above).

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

After 48 h, cells were collected with trypsin and GFP$^+$ cells were sorted and cultured in 96-well plates on MEFs. Three days after sorting, the medium was changed. Six days after sorting, wells with colonies were split with trypsin and split onto 24-well plates with MEFs. After another 3 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 3 days, the cells on gelatin were collected 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, although 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 naive 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 autocleaving '2 A' 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). Vesicular stomatitis Indiana virus glycoprotein (VSVG)-coated lentiviruses including

tetON-TFAP2C-2A-RFP and FUW-lox- M2rtTA were generated in HEK 293T cells. TFAP2C mutant line 1 hESCs were treated with Accutase to make a single cell suspension in 100 µl hESC medium with ROCKi with 100,000 single cells. Cells were transduced with a1:1 ratio of tetON-TFAP2C-2A-RFP and FUW-lox-M2rtTA and plated in 10 cm 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 naive UCLA19n (first replicate) or UCLA20n (second replicate) hESCs were then transfected with 800 ng of either empty pGL3 promoter vector or one of the three constructs described above, along with 200 ng 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 Fig. 4i,j, immunofluorescence was conducted as published in ref. [12], using anti-TFAP2C (SantaCruz 8977, 1:100), anti-OCT4 (sc8628-X, 1:100) and anti-PAX6 (R&D Systems AF8150, 1:100). For Supplementary Fig. 6f, immunostaining was performed as previously described[51]. The following primary antibodies were used: rabbit anti-KLF17 polyclonal (1:500, Sigma HPA024629) and mouse anti-TRA-1-60 IgM (1:300, BD). The following secondary antibodies were used: goat anti-rabbit IgG AF555 secondary (1:400, ThermoFisher) and 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 MAB2018) and PAX6 (R&D Systems AF8150). Western signals were normalized to the signal from anti-H3 antibody (Abcam ab10799 or Abcam ab1791). All antibodies were used at concentrations of 1:1,000 except Santa Cruz anti-TFAP2C (1:700) and H3 (1:3,000).

RNA isolation and library generation. RNA was isolated using the RNeasy Mini Kit (Qiagen). A 5–50 ng 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 25,000 and 50,000 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 10 min, during which they were vortexed for 10 s every 3–4 min, 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 min at room temperature. The paraformaldehyde was quenched by adding glycine to a final concentration of 0.14 M and rotated another 10 min at room temperature. The cells were then centrifuged at 735*g* for 5 min 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 1 ml lysis buffer (10 mM TrisHCl pH 8.0, 0.25% Triton X-100, 10 mM EDTA, 0.5 mM EGTA, 1× protease inhibitors (Roche) and 1 mM PMSF), then rotated for 15 min. Nuclei were pelleted by centrifugation at 1,500*g* for 5 min at 4 °C. Nuclei were then resuspended with 1 ml 10 mM TrisHCl pH 8.0, 200 mM NaCl, 10 mM EDTA, 0.5 mM EGTA, 1× protease inhibitor, 1 mM PMSF and rotated for 10 min. Nuclei were then pelleted and resuspended in 650 µl 10 mM TrisHCl pH 8.0, 10 mM EDTA, 0.5 mM EGTA,1× protease inhibitor, 1 mM PMSF and sonicated in a 12 mm × 12 mm sonication tube (Covaris) in a Covaris S2 (intensity = 5; cycles per burst = 200; duty cycle = 5%; 8 × (30 s on/30 s off) for 4 min effective sonication). The sonicated lysate was then centrifuged for 10 min at 14,200*g*, and the supernatant retained. 10% of the supernatant was saved as 'Input' and the rest was used for ChIP.

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

Protein A Dynabeads (60 µl) were added to the ChIP samples and rotated for 2 h at 4 °C. The beads were then washed 2 × 4 min with 500 µl wash buffer A (50 mM HEPES pH 7.9, 1% TritonX-100, 0.1% deoxycholate, 1 mM EDTA, 140 mM NaCl), 500 µl wash buffer B (50 m HEPES pH 7.9, 0.1% SDS, 1% Triton X100, 0.1% deoxycholate, 1 mM EDTA, 500 mM NaCl) and 500 µl TE buffer (10 mM TrisHCl pH 8.0, 1 mM EDTA). Each wash consisted of resuspension in 500 µl buffer and rotation at 4 min, followed by collection of beads and removal of supernatant. DNA was eluted in 100 µl elution buffer (50 mM TrisHCl pH 8.0, 1 mM EDTA, 1% SDS) at 65 °C for 10 min in a ThermoMixer (Eppendorf) shaking at 1,400 r.p.m. The eluant was collected and the beads were subjected to a second round of elution with 150 µl elution buffer.

The ChIP eluants were pooled, and the input sample was diluted to 250 µl with elution buffer. The samples were incubated 65 °C overnight to promote decrosslinking. The samples were then allowed to cool to room temperature, 15 µg of RNAse A (Purelink, ThermoFisher) was added, and the samples were incubated for 30 min at 37 °C to degrade RNA. Proteinase K (100 µg) was then added and the samples were incubated at 56 °C for 2 h. DNA was purified using a MinElute PCR purification kit (Qiagen).

DNA was sonicated again to 150 bp 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 (for example, 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 analysing splicing but were considered separately when calculating differentially expressed genes (see next section).

RNA-seq data analysis. RNA-seq data were 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 differentially expressed genes were determined, they were analysed for GO terms 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 were mapped using Bowtie as previously described[53]. Peaks were defined in each condition using the MACS2 callpeaks tool[57] with appropriate genome size. To find peaks specific to one condition (for example, naive specific), we uses the predictd module of MACS2 to determine the predicted extension size of each data set being compared, callpeaks for each data set with 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 eightfold relative enrichment cutoff was used to define peaks specific to each state, except when comparing murine WT and *Tfap2a*$^{-/-}$*Tfap2c*$^{-/-}$, in which a sixfold 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 naive states, and overlap between different peak sets, we used the Bedtools intersect tool[58].

ChIP-seq data analysis. ChIP-seq data were 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 determined 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 values for each sample were used as input. The variance of each gene's RPKM in different samples was then calculated (rowVars function in R). PCA analysis (prcomp function in R) was performed on genes with the top 1,000 variances across samples. PCA plots were then plotted with the ggplot2 package in R (http://ggplot2.org).

For PCA for ATAC-seq data, peaks for each ATAC-seq sample 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 the sequencing depth and this matrix was used as input for PCA analysis. The variance of normalized ATAC-seq reads over each peak was then calculated (rowVars function in R). PCA analysis (prcomp function in R) was performed on peaks with the top 1,000 variances across samples. PCA plots were then plotted with ggplot2 package in R (http://ggplot2.org).

Analysis of peak location. Annotation of peak location (promoter, intragenic, intergenic, and so on) and calculation of distance to the nearest promoter were carried out using the 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 were within 50 kb 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 Fig. 8e, real-time PCR was performed as described in ref. [59]. For Supplementary Fig. 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. The overall brightness of some bright-field microscopy images was increased to improve visibility, and Supplementary Fig. 5c was converted to greyscale 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 by 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 formulae. Where the ratio of expression for two sets is shown, as in Fig. 4h,m 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 naive over primed cells (as shown in Figs. 3b and 4b) was observed for western blots generated from five independent lysate preparations. Induction of TFAP2C within 3 days of culture in 5iLAF (as shown in Fig. 4b) was observed in two independent reversions.

The finding that *TFAP2C*$^{-/-}$ hESCs express OCT4 normally in the primed state (Supplementary Fig. 3c) was demonstrated in western blots from three independent lysate preparations, and similar findings for NANOG and SOX2 (Supplementary Fig. 3d,e) were made once. The flow cytometry plots in Supplementary Fig. 3f showing pluripotency surface markers on *TFAP2C*$^{-/-}$ hESCs are representative of three experiments.

The finding that *TFAP2C*$^{-/-}$ hESC colonies showed morphological abnormality in 5iLAF (as shown in Fig. 4a and Supplementary Fig. 4a,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 Figs. 4a and 5d. Loss of TFAP2C in *TFAP2C*$^{-/-}$ mutant (Fig. 4c) was shown by western blot from five independent reversions and by immunofluorescence (Fig. 4i) in two independent reversions. Loss of pluripotency factors and gain of neural factors was shown by western blot (Fig. 4f,j) in four independent reversions and by immunofluorescence (Fig. 4i,j) in two independent reversions. Loss of Tfap2c in *Tfap2C*$^{-/-}$ mESCs (Supplementary Fig. 4g) was confirmed via western blots from two independent lysate preparations.

Rescue of the *TFAP2C*$^{-/-}$ differentiation phenotype in 5iLAF by doxycycline-inducible TFAP2C expression (Fig. 5a–d and Supplementary Fig. 5c) was observed in four experiments. Reduced cell proliferation after withdrawal of doxycycline (Supplementary Fig. 5d,e) was observed in two experiments.

Bulk differentiation of *TFAP2C*$^{-/-}$ in 5iLAF at 5% O_2, with a small population maintaining self-renewal (Fig. 6 and Supplementary Fig. 6), was observed in two independent reversions. Loss of KLF17 in *TFAP2C*$^{-/-}$ t2iLGöY in Supplementary Fig. 6f was demonstrated from two reversions.

The finding that OCT4 protein is expressed normally in the OCT4 ΔIntronic enhancer mutant (Supplementary Fig. 8d) was based on one western blot. Loss of cells in the OCT4 ΔIntronic enhancer mutant upon treatment with 5iLAF (Fig. 8f) was observed in three independent reversions, with qRT–PCR data (Fig. 8e) derived from one such reversion.

Code availability. Custom scripts used for demultiplexing NGS reads, calculating RPKM, generating DNA methylation metaplots and comparing the distribution of peaks to expression of nearby genes is available upon request.

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

Data availability. All high-throughput sequencing data sets described in Supplementary Table 1 have been deposited in the Gene Expression Omnibus (GEO) under accession no. GSE101074. RNA-seq data from naive and primed hESCs were gleaned from our previously published data[12] (GSE76970). ChIP-seq data on H3K4me3, H3K27Ac3 and Mediator from naive and primed hESCs were obtained from published sources[9,16] (GSE69647). Peak sets used are included in Supplementary Tables 2 and 4, and RPKMs from RNA-seq data are included in Supplementary Table 5. Source data for Figs. 3a, 4h,m, 5e and 8d Supplementary Figs. 3g, 5a and 6b are available in Supplementary Table 8. Additional data are available upon reasonable request.

References

51. Liu, X. et al. Comprehensive characterization of distinct states of human naive pluripotency generated by reprogramming. *Nat. Methods* **14**, 1055–1062 (2017).
52. Diaz Perez, S. V. 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.* **21**, 751–764 (2012).
53. Pastor, W. A. et al. MORC1 represses transposable elements in the mouse male germline. *Nat. Commun.* **5**, 5795 (2014).
54. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
55. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
56. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
57. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS).*Genome Biol.* **9**, R137 (2008).
58. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
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. *Nat. Cell Biol.* **15**, 113–122 (2013).

Corresponding author(s): Amander Clark

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.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Experimental design

1. Sample size

Describe how sample size was determined.

No sample size calculation was performed. For the key experiments in the publication (comparison of ATAC-seq data in naive and primed hESCs, RNA-seq in WT and TFAP2C-/- cells, TFAP2C ChIP-seq in naive hESCs) at least replicates were conducted.

2. Data exclusions

Describe any data exclusions.

No data was excluded from the analysis. For all high throughput sequencing datasets (ChIP, ATAC, RNA-seq) all replicates sequenced for a given condition were included.

3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

Statistical analysis was performed using existing programs (MACS257, Homer19, DESeq56, GOrilla33). 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% O2, 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 ΔIntronic enhancer mutant, as shown in Supplementary Figure 8D, was based on one Western blot. Loss of cells in the OCT4 ΔIntronic 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.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Randomization is not relevant, as at no point was an experiment conducted in which one set of organisms received a treatment different from another set.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Blinding was not performed given the complexity of the experiments. However, key experiments were repeated in a different laboratory as an alternate approach to blinding.

Note: all in vivo studies must report how sample size was determined and 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 *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ Test values indicating whether an effect is present *Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes 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 in all relevant figure captions (with explicit mention of central tendency and variation)

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.

Bowtie v0.12.8, Bowtie2 v2.2.9, samtools v0.1.19, MACS v2.1.1, HOMER v4.8.2, Tophat v2.0.14, HTSeq v0.6.1p1, DESeq v1.28.0, ngsplot v2.47, bedtools v2.26.0, GOrilla, GREAT v3.0.0, ggplot2. Custom scripts for demultiplexing, calculating RPKM, methylation metaplotting, and comparing peak sites to expression changes in nearby genes are available to reviewers or readers upon request.

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 third party.

All materials will be freely available with Material Transfer Agreement. In addition, all human embryonic stem cell lines from this paper will be disbursed following Institutional Embryonic Stem Cell Research Oversight Committee Approval.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

For ChIP: anti-TFAP2C sc-8977 was used. We confirmed 1) It recognized a strong band of TFAP2C's approximate molecular weight. 2) This band was absent in cells deficient for TFAP2C. 3) TFAP2C ChIP peaks showed extremely strong enrichment for TFAP2C's binding motif.
Antibodies used for Western blots: anti-Oct4 Santa Cruz (sc8628, 1:1000), anti-Nanog (R&D Systems AF1997, 1:1000), anti-TFAP2C (sc-8977 1:700 or ab76007 1:1000), anti-Sox1 (R&D Systems AF3369 1:1000), anti-Sox2 (R&D Systems AF2018 1:1000), anti-Pax6 (R&D Systems AF8150 1:1000) and anti-H3 (Abcam 10799 1:3000 or Abcam 1791 1:3000) have all been used extensively in other publications, recognize strong bands of correct molecular weight, and show changes in intensity that correspond to expression changes measured by RNA-seq.
Antibodies used for Immunofluorescence: anti-OCT4 (SantaCruz sc-8628x 1:100), anti-TFAP2C (SantaCruz sc-8977 1:100), anti-Pax6 (R&D Systems AF8150 1:100) anti-KLF17 (Sigma HPA024629 1:500) and TRA-1-60 (BD Biosciences 56007 1:300) have all been validated in other publications.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

The UCLA hESC lines were generated at UCLA from single blastocysts.
The V6.5 mouse ESC line was obtained from the lab of Robert Blelloch at UCSF
The EpiSCs were obtained from the lab of Paul Tesar at Case Western Reserve.

b. Describe the method of cell line authentication used.

Human hESCs were authenticated using SNP/CNV analysis.

To confirm clonality of the CRISPR lines generated, 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-/-, lack of protein was confirmed by Western blot. For the OCT4 Naïve Enhancer mutation, deletion was also confirmed clonal deletion 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.

c. Report whether the cell lines were tested for mycoplasma contamination.

All lines were routinely tested 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 lines used are on the ICLAC database.

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 all relevant details on animals and/or animal-derived materials used in the study.

Time mated pregnant female mice (Charles River CF-1) were purchased from a vendor prior to euthanization. Murine embryonic feeder cells (MEFs) were generated from the embryos to support ESC culture. MEFs were derived following UCLA institutional animal care and use committee (IACUC) approval.

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.

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.