nature communications

Article

https://doi.org/10.1038/s41467-024-52431-1

# ZIC2 and ZIC3 promote SWI/SNF recruitment to safeguard progression towards human primed pluripotency

Received: 19 October 2023

Accepted: 3 September 2024

Published online: 02 October 2024

Check for updates

Ishtiaque Hossain[1], Pierre Priam[2,3], Sofia C. Reynoso[1], Sahil Sahni[1], Xiao X. Zhang[1], Laurence Côté[2,3], Joelle Doumat[1], Candus Chik[1], Tianxin Fu[1], Julie A. Lessard[2,3] & William A. Pastor[1,4]

The primed epiblast acts as a transitional stage between the relatively homogeneous naïve epiblast and the gastrulating embryo. Its formation entails coordinated changes in regulatory circuits driven by transcription factors and epigenetic modifications. Using a multi-omic approach in human embryonic stem cell models across the spectrum of peri-implantation development, we demonstrate that the transcription factors ZIC2 and ZIC3 have overlapping but essential roles in opening primed-specific enhancers. Together, they are essential to facilitate progression to and maintain primed pluripotency. ZIC2/3 accomplish this by recruiting SWI/SNF to chromatin and loss of ZIC2/3 or degradation of SWI/SNF both prevent enhancer activation. Loss of ZIC2/3 also results in transcriptome changes consistent with perturbed Polycomb activity and a shift towards the expression of genes linked to differentiation towards the mesendoderm. Additionally, we find an intriguing dependency on the transcriptional machinery for sustained recruitment of ZIC2/3 over a subset of primed-hESC specific enhancers. Taken together, ZIC2 and ZIC3 regulate highly dynamic lineage-specific enhancers and collectively act as key regulators of human primed pluripotency.

Lineage specification during human embryogenesis is governed by the spatiotemporal interplay between transcription factors (TFs), repressive complexes and epigenetic modifications[1–4]. In the 2nd week post-fertilisation, concurrent with embryonic implantation, the epiblast undergoes a series of dynamic changes—increase in DNA methylation[5] and Polycomb foci[6–8], X-chromosome inactivation[9], silencing of SVA and LTR5Hs elements[10,11] and widespread alteration of the gene expression profile[4–6,12,13]. This series of events primes the epiblast to undergo multi-lineage differentiation at gastrulation and provides instructive cues to extra-embryonic tissues[14].

The pre-implantation epiblast, which is refractory to differentiation cues[15], exists in a "naïve" pluripotent state while the post-implantation epiblast exists in a "primed" pluripotent state[16]. To model these stages of development, human embryonic stem cells (hESCs) can be cultured in distinct media conditions that recapitulate the naïve and primed states, and can transition between these states when the culture media is changed[12,17,18].

Three *ZIC*-family TFs, *ZIC2, ZIC3* and *ZIC5*, are some of the most highly upregulated genes during the peri-implantation window—a phenomenon conserved in mice, primates and humans[18–21]. Haploinsufficiency for *ZIC2* results in holoprosencephaly, and murine mutants of *Zic2* feature severe heart and neurological defects and undergo lethality in mid-[22] or late-gestation[23]. *Zic3*-deficient mice show a variable phenotype; some embryos fail to gastrulate while others give rise to viable offspring[24], but a common feature of *Zic3* deficiency in mice and humans is failure of left–right embryonic patterning and

[1]Department of Biochemistry, McGill University, Montreal, QC, Canada. [2]Institute for Research in Immunology and Cancer (IRIC), University of Montreal, Montreal, QC, Canada. [3]Department of Pathology and Cellular Biology, Faculty of Medicine, University of Montreal, Montreal, QC, Canada. [4]The Rosalind & Morris Goodman Cancer Institute, McGill University, Montreal, QC, Canada. e-mail: william.pastor@mcgill.ca

concomitant heterotaxy[25]. *Zic5*$^{-/-}$ mice show craniofacial and neural tube closure defects and lethality as juveniles[26]. While these phenotypes clearly indicate a role in early development, these single-gene mutants are nonetheless able to form post-implantation epiblast. Likewise, the phenotype observed of *ZIC2*- or *ZIC3*-null ESCs cultured in vitro is surprisingly mild[27–30]. While defects in the differentiation of murine ESCs (mESCs) to neural ectoderm[27,30] or of hESCs in specification to cardiac lineage progenitors[29] have been reported, these are mild and delayed phenotypes considering that *ZIC* TFs are highly expressed in the stem cells themselves. The mechanistic roles of *ZIC* TFs in stem cells are not entirely clear, with ZIC2 implicated in maintaining facultative heterochromatin and ZIC3 reported to cause chromatin opening[27,28].

The mammalian *ZIC* family consists of five paralogs—all of which show high sequence similarity and have highly conserved DNA-binding domains[31]. *ZIC2*, *ZIC3* and *ZIC5* are co-expressed in primed hESCs and in the post-implantation epiblast, so we reasoned that *ZIC2/3/5* might have largely overlapping roles. To address this and discern paralog-specific roles from roles shared by multiple *ZIC* paralogs, we comprehensively studied loss-of-function *ZIC2/3/5* single- and compound knockouts in primed hESCs at cellular, epigenomic and transcriptomic levels. We demonstrate that ZIC2 and ZIC3 act redundantly in primed hESCs and are integral for their viability. We also show that ectopic expression of *ZIC2* in naïve hESCs results in exit from naïve pluripotency towards a primed-like state, and that ZIC2/3 are important for specification of the primed lineage. At a molecular level, ZIC2/3 support primed pluripotency directly by positively regulating epiblast-specific genes through recruitment of the ATPase-dependent chromatin remodelling SWI/SNF complex and indirectly by repressing drivers of gastrulation and extra-embryonic specification.

In summary, we identify ZIC2 and ZIC3 as important regulators of the peri-implantation epiblast epigenome, highlighting their critical role and essentiality during this transitory phase.

## Results

### ZIC2/3 are associated with changes in chromatin state during the naïve-to-primed transition

To identify TFs essential for progression towards primed pluripotency, we conducted pairwise analysis of ATAC-seq datasets[4] from naïve and primed hESCs. We found motifs corresponding to *ZIC* TFs as being most highly enriched in regions specifically open in primed hESCs, alongside pan pluripotency TFs like *POU5F1*, *SOX2* and *NANOG* (Fig. 1a, Supplementary Fig. S1a and Supplementary Data 1). The *ZIC*-family TFs *ZIC2, ZIC3* and *ZIC5* show high expression in primed hESCs, with dramatically higher expression relative to the naïve state both in human stem cells and in primates (Fig. 1b, c)[4,7,8,15,18,32–35].

To delineate between specific- and overlapping roles among *ZIC2/3/5* we attempted to iteratively generate single- and combinatorial knockouts of all three relevant *ZIC* TFs in primed hESCs. All deletions were designed to disrupt the protein within or before the DNA-binding domain of each factor. We were able to efficiently generate single knockout *ZIC2*$^{-/-}$, *ZIC3*$^{-/-}$ and *ZIC5*$^{-/-}$ and combinatorial knockouts *ZIC2*$^{-/-}$*ZIC5*$^{-/-}$ and *ZIC3*$^{-/-}$*ZIC5*$^{-/-}$ in primed hESCs (Supplementary Fig. S1b–d). However, after several attempts we were unable to generate homozygous concomitant deletion of both *ZIC2* and *ZIC3* (Fig. 1d), which have a high degree of amino acid sequence homology (Supplementary Fig. S1e–g), indicating shared yet essential role for ZIC2 and ZIC3. We thus conducted ChIP-seq for both ZIC2 and ZIC3. ChIP-seq for ZIC3 was conducted in *ZIC2*$^{-/-}$*ZIC5*$^{-/-}$ hESCs because of evidence that the anti-ZIC3 antibody recognised ZIC2 or ZIC5 on chromatin (Supplementary Fig. S2a–c). We observed that ZIC2 and ZIC3 have similar DNA occupancy patterns, although ZIC2 ChIP showed higher enrichment over background (Fig. 1e and Supplementary Fig. S2d). Ectopically expressed Flag-ZIC2 and Flag-ZIC3 also broadly showed enrichment over the same ZIC2-enriched regions (Fig. 1e, Supplementary Fig. S2d).

We observed high enrichment of ZIC2 and ZIC3 specifically over primed hESC-specific ATAC peaks (Fig. 1f), in line with enrichment of motifs corresponding to these factors over the same loci. Correspondingly, we observed striking association of *ZIC*-bound regions with regions that are active enhancers in primed hESCs but not naïve hESCs (Supplementary Fig. S3a–c).

We ran ChromHMM on five published[6,12,36,37] histone ChIP-seq datasets (H3K4me1, H3K4me3, H3K27ac, H3K27me3 and H3K9me3) in naïve and primed hESCs to unbiasedly annotate the chromatin state of ZIC2/3-bound regions in primed hESCs and their corresponding states in naïve hESCs. Owing to ZIC2/3 occupying disparate chromatin landscapes, we subcategorised ZIC2/3-binding sites using a nearest-neighbour clustering approach (Fig. 1g–j and Supplementary Data 2), on published H3K27ac ChIP, H3K27me3 ChIP and ATAC-seq datasets in naïve and primed hESCs[6,12,36,37]. We identified four clusters: (C1) regions open and highly enriched for the active chromatin mark H3K27ac in both naïve and primed hESCs; (C2) regions with some increase in chromatin accessibility in primed hESCs with no drastic changes in H3K27 post-translational modifications relative to naïve hESCs; (C3) regions with a dramatic increase in chromatin openness and H3K27ac in primed hESC relative to naïve hESCs, along with strong enrichment for the pluripotency factors OCT4, SOX2 and NANOG; (C4) regions with high levels of repressive H3K27me3 and H2AK119ub in primed hESCs (Fig. 1k and Supplementary Fig. S3d). *ZIC* motifs were enriched in all four clusters and motifs for pluripotency factors, particularly *SOX2*, were specifically enriched in C3 regions (Supplementary Fig. S3e). C3 regions were also distinctive in being least enriched for H3K4me3 and consisting almost entirely of distal regulatory elements (Supplementary Fig. S3d, f). Consistent with the distinct chromatin states, we observed high expression of genes proximal to C1 peaks, low expression of genes proximal to C4 peaks, and primed-specific upregulation of genes proximal to C3 peaks (Fig. 1l, m).

### ZIC2 and ZIC3 act in tandem to maintain primed pluripotency

Despite the apparent non-viability of *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ primed hESCs, we could achieve highly efficient mutation of *ZIC2* on a *ZIC3*$^{-/-}$ background via electroporation of Cas9 RNP targeting *ZIC2* (Fig. 2a, b). Although the targeted cells were progressively lost from culture by 15 days post-electroporation, we could study the effects of combined *ZIC2* and *ZIC3* deficiency 5 days after targeting without significant loss of *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ primed hESCs (Fig. 2b and Supplementary Fig. S4a). Acute ablation of *ZIC2* and *ZIC3* resulted in spontaneous differentiation characterised by increased expression of CD56/CD184[1], and apoptosis indicated by high annexin-V staining (Fig. 2c, d and Supplementary Fig. S4a–c).

Individual *ZIC* mutants, or compound mutants in which either *ZIC2* or *ZIC3* were intact, showed minimal changes in their transcriptome (Supplementary Fig. S4d and Supplementary Data 3). In contrast, *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ hESCs and *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$*ZIC5*$^{-/-}$ hESCs showed drastic changes in gene expression compared to wild-type control primed hESCs (Supplementary Fig. S4d). Upon hierarchical clustering of all differentially expressed genes (DEGs) in single and compound *ZIC2/3/5* deletion mutants, we observed that the transcriptomes of *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ and *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$*ZIC5*$^{-/-}$ hESCs were similar to each other, highlighting a lesser role for ZIC5 (Fig. 2e). The analysis yielded only two overarching genesets—genes significantly de-repressed upon loss of *ZIC2* and *ZIC3* (geneset 1) and genes that were downregulated upon loss of *ZIC2* and *ZIC3* (geneset 2) (Supplementary Data 4). Geneset 1 consisted mostly of genes that were upregulated upon differentiation towards the mesendoderm lineage (Fig. 2f–h), suggesting that loss of both ZIC2 and ZIC3 pushes primed hESCs developmentally forward towards differentiation. However, we also saw upregulation of some naïve-specific genes (Fig. 2f, g) as well as elevated expression of the naïve-specific LTR5Hs transposons[11,13] and LTR5Hs-proximal genes (Supplementary Fig. S5). Geneset 2 was enriched for genes

https://doi.org/10.1038/s41467-024-52431-1

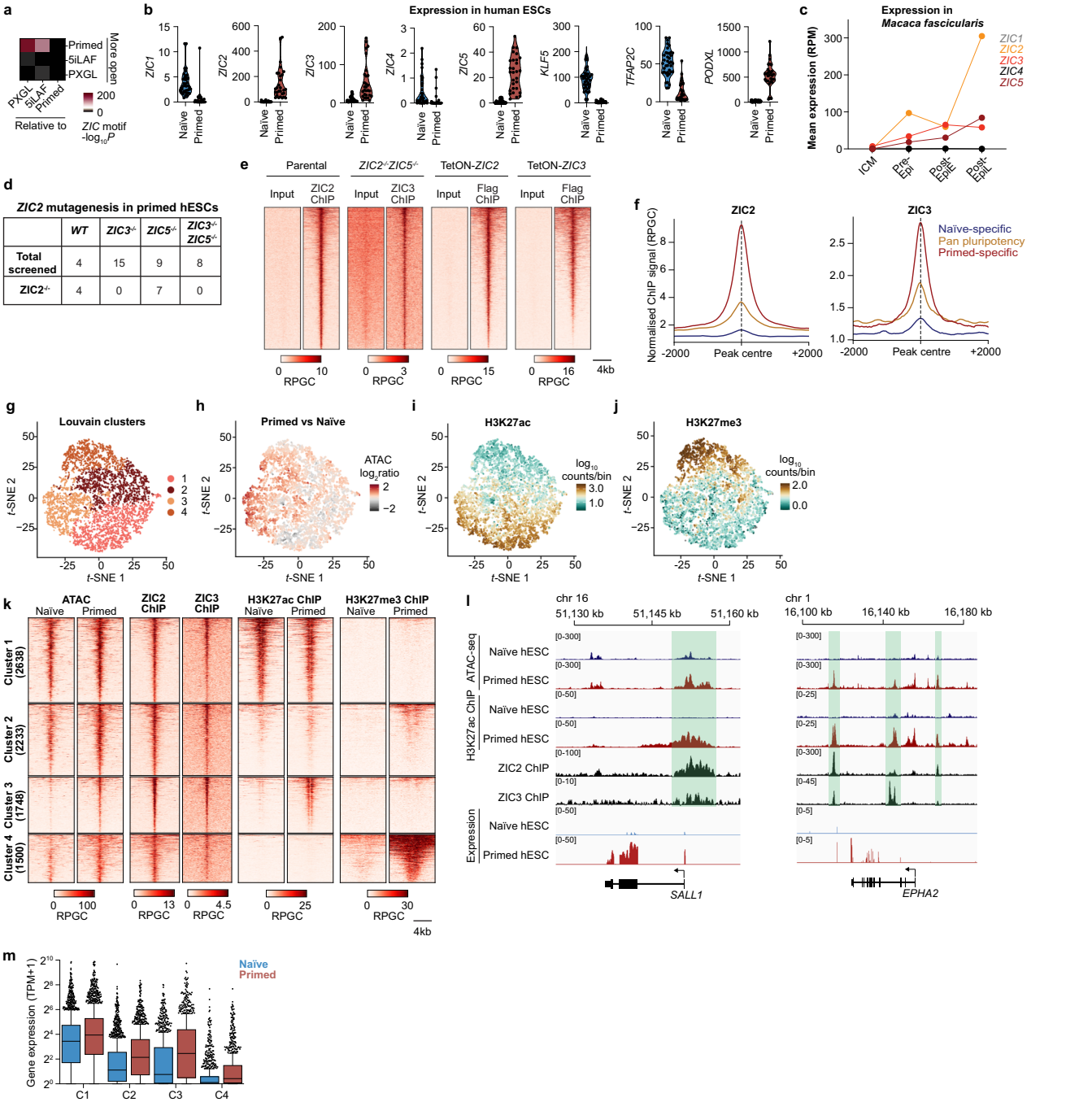


**Fig. 1 | Characterisation of ZIC2 and ZIC3 chromatin occupancy in naïve and primed hESCs. a** Enrichment of *ZIC* motif from pairwise comparisons of regions open specifically, in naïve (cultured in two conditions: PXGL and 5iLAF) and primed hESCs. *P* values calculated from a one-tailed Binomial test and were statistically significant for primed-specific ATAC peaks after adjusting for multiple comparisons by the Benjamini–Hochberg procedure. **b** Normalised expression of *ZIC* family transcription factors in naïve and primed hESCs. Expression of naïve pluripotency markers *KLF5* and *TFAP2C*, and primed pluripotency marker *PODXL*, are shown as reference. Values are expressed as TPM. **c** Mean expression of *ZIC1-5* in the inner cell mass and epiblast of cynomolgus macaques. RPM reads per million mapped reads. **d** Record of success in generating *ZIC2*$^{-/-}$ clonal lines in primed hESCs with genetic background indicated. *ZIC2*$^{-/-}$ hESCs were successfully generated on WT and *ZIC5*$^{-/-}$ background but not on *ZIC3*$^{-/-}$ background. **e** Heatmap showing enrichment of ZIC2, ZIC3, Flag-ZIC2 and Flag-ZIC3 over ZIC2 ChIP-seq peaks. **f** Normalised ChIP-seq coverage of ZIC2 and ZIC3 over primed-specific, naïve-specific and pan pluripotency ATAC peaks. Louvain clusters of ZIC2-bound chromatin projected on a *t*-SNE plot—each dot represents a 2 kb bin around a ZIC2 peak (**g**), change in chromatin accessibility of each ZIC2 peak in primed hESCs relative to naïve hESCs (**h**), enrichment of H3K27ac ChIP-seq reads (**i**) and H3K27me3 ChIP-seq reads (**j**) in primed hESCs. **k** Heatmap showing read density of ATAC, ZIC2, ZIC3 and H3K27 post-translational modifications within Louvain clusters of ZIC2-bound chromatin in naïve and primed hESCs. **e**, **f**, **k** RPGC reads per 1x genome coverage. **l** Gene browser tracks around two representative C3 loci. Values are reported in RPGC. **m** Expression of genes in naïve and primed hESCs within 50 kb of a ZIC2-bound locus as defined in (**g**). Mean values for each gene from four biological replicates were calculated and plotted in the appropriate set. Boxplots show median and inter-quartile range. Whiskers are calculated using the Tukey method. **b**, **m** TPM transcripts per million. Source data are provided as a Source Data file.

https://doi.org/10.1038/s41467-024-52431-1

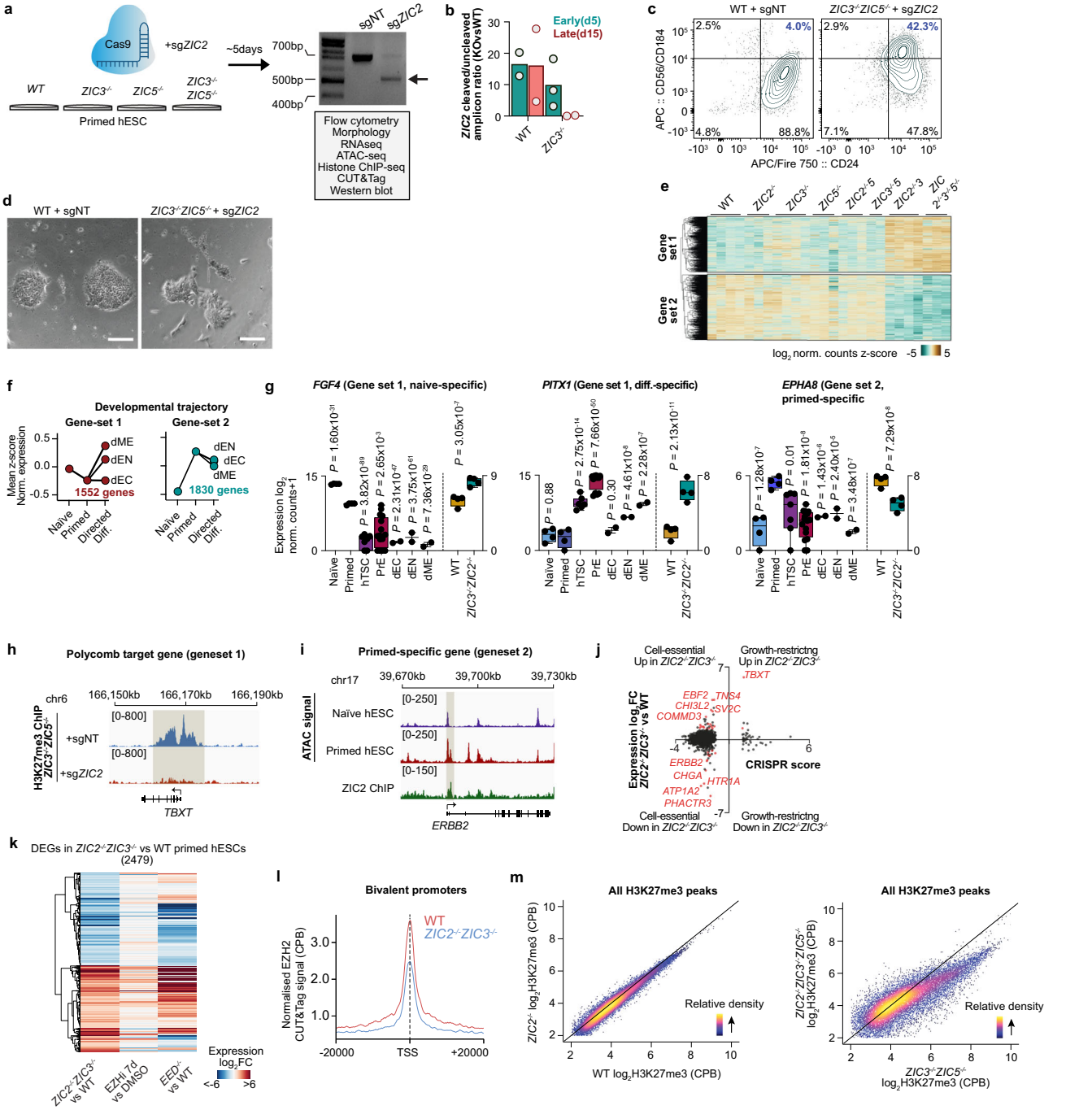


**Fig. 2 | Role of ZIC2 and ZIC3 in maintaining integrity of primed hESCs.** **a** Schematic of experiments involving transient deletion of *ZIC2* in primed hESCs. Representative image of genotyping showing successful excision 5 days after electroporation (excised allele indicated with arrow). **b** Ratio of KO to WT amplicon over time after targeting of *ZIC2* in hESCs of genotypes indicated. **c** Flow cytometry for CD56/CD184 and CD24 of WT and *ZIC3*$^{-/-}$*ZIC5*$^{-/-}$ primed hESCs depleted of *ZIC2*. **d** Brightfield images of WT and *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$*ZIC5*$^{-/-}$ primed hESCs. Scale bars 150 µm. Data are representative of three biological replicates. **e** Heatmap with scaled expression of all DEGs identified from pairwise comparison of WT primed hESCs and primed hESCs depleted of *ZIC2*, *ZIC3* and *ZIC5* as single- or combinatorial deletions. Two genesets were identified from hierarchical clustering. **f** Expression dynamics of genesets identified from DEGs in hESC-derived in vitro cultured lineages of the epiblast. **g** Scaled expression of select genes in the indicated lineages, and WT and *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ primed hESCs. Number of dots represents the number of biological replicates used to represent each lineage. Median values are shown, and boxplots represent the inter-quartile range. Whiskers indicate minimum and maximum expression values of each gene within the dataset. *P* values were calculated from a two-tailed Wald test and adjusted for multiple comparisons by the Benjamini–Hochberg procedure. Genome browser tracks of ZIC2 ChIP-seq, H3K27me3 ChIP-seq (**h**) or ATAC-seq (**i**) around genes representative of genesets 1 and 2 in primed hESCs upon concomitant loss of ZIC2 and ZIC3. **j** Scatterplot of genes that are identified to be growth-limiting or essential in primed hESCs, with fold change of expression in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ hESCs vs WT primed hESCs indicated. **k** Heatmap indicating change in expression of genes upregulated or downregulated in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$, showing expression changes of same genes in hESCs treated for 7 days with EZH2i or in *EED*$^{-/-}$ hESCs. **l** Normalised EZH2 CUT&Tag signal over annotated bivalent promoters in primed hESCs, in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ and WT primed hESCs. **m** Normalised H3K27me3 signal in 5 kb bins among H3K27me3-enriched regions upon loss of ZIC2 compared with WT primed hESC (left) and loss of ZIC2/3/5 compared to loss of ZIC3/5 in primed hESCs (right). Source data are provided as a Source Data file.

upregulated in primed hESCs compared to naïve hESCs (Fig. 2f, g, i), suggesting an important role for ZIC2 and ZIC3 in maintaining high expression of genes characteristic of primed pluripotency.

Cross-analysis of DEGs identified in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ primed hESCs with data from a dropout CRISPR screen in primed hESCs did not yield an obvious explanation for the lethality in the absence of ZIC2 and ZIC3 (Fig. 2j)[38]. However, a handful of essential genes show reduced expression and *TBXT*, a driver of primitive streak formation[39] and a growth-restricting gene, is de-repressed in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ hESCs (Fig. 2h, j). Overall, however, there is a clear loss of survival and self-renewal in primed hESCs deficient for both ZIC2 and ZIC3, highlighting their central, yet highly overlapping role, in safeguarding primed pluripotency.

### Concomitant deletion of ZIC2 and ZIC3 is linked to upregulation of PRC2 targets and loss of concentrated H3K27me3

We noted that genes upregulated upon loss of ZIC2 and ZIC3 (geneset 1) predominantly included canonical Polycomb targets. We observed strong overlap between genes de-repressed in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ hESCs and those de-repressed in *EED*$^{-/-}$ hESCs[40] or in hESCs treated with EZH2i for 7 days[8] (Fig. 2k and Supplementary Fig. S6a–c). Correspondingly, we observed a drop in EZH2 occupancy over annotated bivalent promoters[41] in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ hESCs (Fig. 2l). Additionally, *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ hESCs showed selectively reduced H3K27me3 enrichment over the regions that showed high H3K27me3 density in control primed hESCs (Fig. 2h, m). Changes in distribution of EZH2, H3K27me3 and H2AK119ub were highly correlated upon loss of ZIC2/3 (Supplementary Fig. S6d–f), and modest gain of H3K27ac and H3K4me3 was also observed over loci that lost polycomb marks (Supplementary Fig. S6g,h).

A number of genes de-repressed upon differentiation of primed hESCs towards the mesoderm or endoderm lineages are regulated by bivalent promoters[40] and are silenced by H3K27me3, whose enrichment over these loci is subsequently reduced upon exit from pluripotency (Supplementary Fig. S6i)[1]. However, primed hESCs acutely depleted of ZIC2 and ZIC3 are distinguishable from hESCs undergoing directed differentiation in that they still maintain expression of core pluripotency factors (Supplementary Fig. S6j)[1,4,8]. This upregulation of genes linked to differentiation without loss of pluripotency factors is similar to what is observed from acute inhibition of PRC2 activity (Supplementary Fig. S6j), suggesting that the upregulation of PRC2-repressed genes results from a specific loss of PRC2-mediated silencing rather than as a secondary consequence of differentiation.

### ZIC2/3 maintain accessibility over BRG1-dependent primed-specific enhancers

Given the strong enrichment of ZIC2/3 over enhancers selectively active in primed hESCs, and downregulation of primed-specific genes in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ hESCs, we investigated the importance of *ZIC* factors in maintaining accessibility at these enhancers. We thus conducted ATAC-seq in *ZIC*-mutant and WT primed hESCs. In *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ hESCs, we predominantly observed loss in accessibility of loci with enrichment of the *ZIC* motif (Fig. 3a, b, Supplementary Fig. S7a and Supplementary Data 5–7). C3 regions, which consist of *ZIC*-bound primed-specific enhancers, underwent the greatest loss in accessibility (Fig. 3c, d). Likewise, we observe a general drop in accessibility of primed hESC-specific open chromatin regions—a phenomenon most striking over loci that had motifs corresponding to *ZIC* factors (Fig. 3e). Consistent with a regulatory role for these open chromatin sites, we found that genes proximal to loci that lose accessibility in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ primed hESCs, or proximal to *ZIC*-bound C3 regions, are disproportionately likely to be downregulated upon loss of ZIC2/3 in primed hESCs (Fig. 3f). Collectively, these results indicate that ZIC2 and ZIC3 positively regulate transcription of proximal genes by maintaining an open chromatin state.

Since the number of differentially accessible regions (DARs) we identified was modest in number despite the strong phenotype in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ hESCs, we employed *k*-means clustering on consolidated ATAC-seq data from all experiments. We identified three clusters corresponding to regions that show variation in openness, the largest of which (cluster C) corresponded to regions that show reduced accessibility in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ hESCs (Fig. 3g and Supplementary Fig. S7b). GREAT analysis of cluster C showed that proximal genes are associated with the PI3K signalling pathway and patterning of the neural ectoderm (Fig. 3g)[42]. This is in accordance with spontaneous differentiation towards the mesendoderm lineage observed in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ hESCs. Appropriately, a preponderance of cluster C genomic regions corresponded to ZIC2-bound C3 regions (Supplementary Fig. S7c, d), confirming an essential role of ZIC2/3 in maintaining accessibility over primed-specific enhancers.

We next asked how ZIC2/3 function to open chromatin. *SMARCA4*/BRG1 constitutes the catalytic subunit of the SWI/SNF chromatin remodelling complex, and given the low expression of *SMARCA2*/BRM in primed hESCs (Supplementary Fig. S8a), is selectively important for the opening of primed-specific enhancers[35]. We found association of ZIC2 with BRG1 by co-immunoprecipitation (Fig. 3h,i and Supplementary Fig. S8b, c). Comparison with published data[35] indicates high enrichment of BRG1 over C3 loci and striking loss in accessibility over these regions in *SMARCA4*$^{-/-}$ naïve hESCs cultured to a primed-like state (Fig. 3j and Supplementary Fig. S8d). The *ZIC* motif was highly enriched among BRG1-bound loci in primed hESCs (Supplementary Fig. S8e) and we observed concordant overlap between genes deregulated in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ hESCs and *SMARCA4* knockdown primed hESCs (Supplementary Fig. S8f), which have a mild inability to maintain pluripotency[35]. We conducted CUT&Tag of BRG1 in WT and *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ hESCs and observed an impressive correspondence between loss of BRG1 occupancy and loss of chromatin accessibility over C3 loci upon ZIC2/3 deletion (Fig. 3k–m). These results indicate that *ZIC* factors maintain primed-specific enhancer activity through BRG1 during progression of human pluripotency.

We observed downregulation of a developmentally critical putative *ZIC* target, *DNMT3B* in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ hESCs (Supplementary Fig. 9a). *DNMT3B* is upregulated in primed hESCs compared to naïve hESCs and is the predominant driver for de novo DNA methylation in the primed epiblast[43] (Supplementary Fig. S9b). We identified three regions proximal to the *DNMT3B* promoter that showed enrichment for ZIC2 in primed hESCs and that are selectively open in primed hESCs. Among these, two candidate enhancers showed reduced BRG1 occupancy in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ hESCs (Supplementary Fig. S9c). These regions showed enhancer activity in a reporter assay, and CRISPR interference-mediated silencing of these enhancers resulted in reduced *DNMT3B* expression, indicating a role for ZIC2/3 in positive regulation of *DNMT3B* in primed pluripotency (Supplementary Fig S9d–g).

### ZIC2/3 facilitate exit from naïve pluripotency and restrict plasticity of naïve hESCs

Given that *ZIC2* and *ZIC3* are upregulated during developmental progression of the epiblast and transition to primed pluripotency (Fig. 1b, c and Supplementary Fig. S10a), we sought to determine whether ZIC2 and ZIC3 are essential for acquiring a primed-like state. Naïve hESCs deficient for ZIC2 and ZIC3 failed to undergo re-priming (restoration of primed-state upon culture in primed media). This effect was characterised by dramatically reduced cell numbers and greater retention of cells positive for the naïve marker SUSD2[44] (Supplementary Fig. S10b–e). To determine whether *ZIC* TFs could drive exit from naïve pluripotency, we used primed hESCs with a construct that ectopically expresses Flag-tagged *ZIC2* upon treatment with doxycycline (Fig. 4a and Supplementary Fig. S10f, g). Inducible expression of *ZIC2* blocks primed hESCs from undergoing naïve reversion (Supplementary Fig. S10h, i), while impairing self-renewal of steady state

https://doi.org/10.1038/s41467-024-52431-1

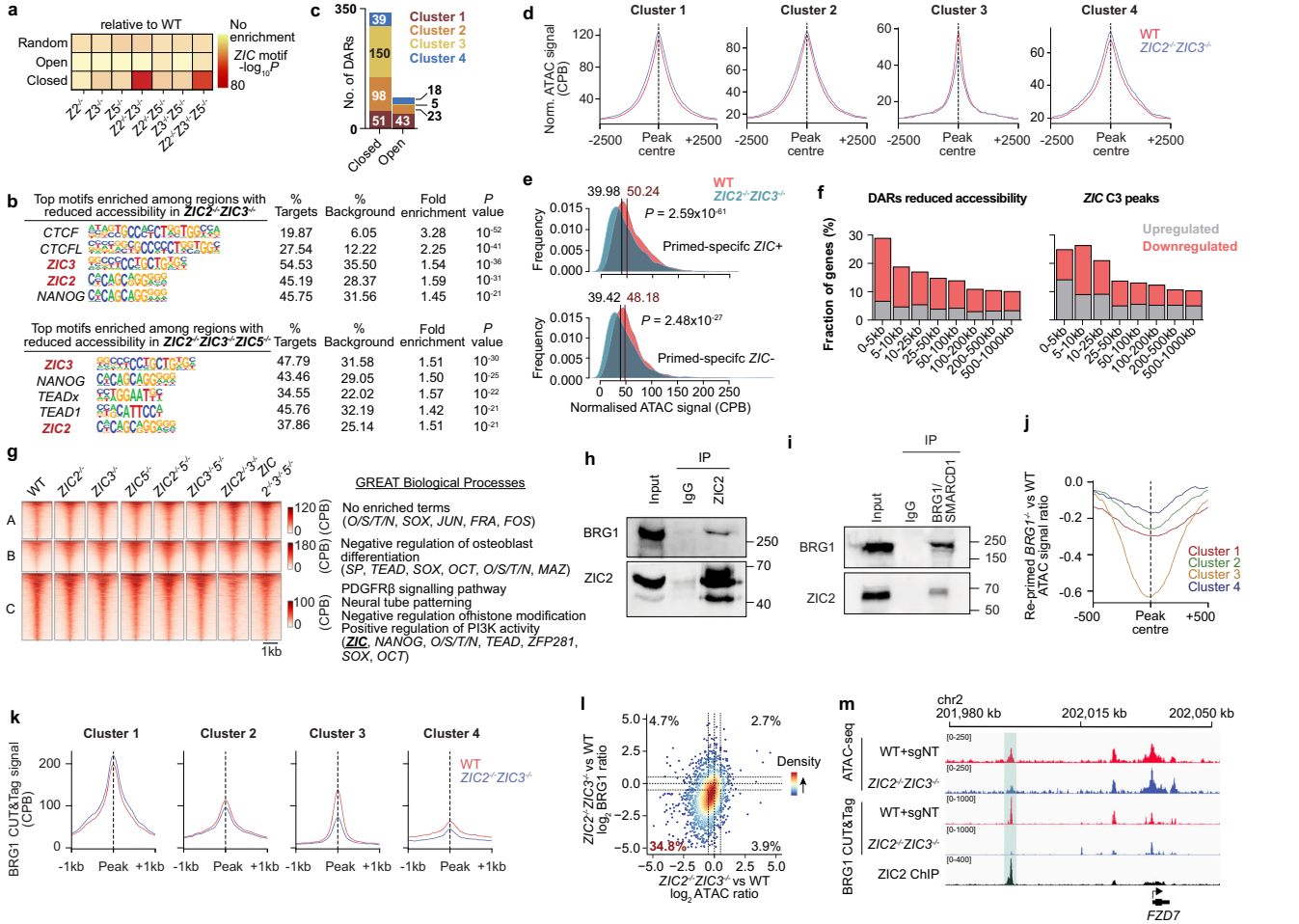


**Fig. 3 | ZIC2 and ZIC3 maintain BRG1-dependent chromatin accessibility over primed hESC-specific enhancers. a** Enrichment of motifs corresponding to *ZIC* factors in differentially accessible regions (DARs) in single- and compound *ZIC2/3/5* mutants. **b** Top motifs enriched among loci that lose accessibility in $ZIC2^{-/-}ZIC3^{-/-}$ and $ZIC2^{-/-}ZIC3^{-/-}ZIC5^{-/-}$ compared to WT primed hESCs. **a**, **b** *P* values calculated from a one-tailed Binomial test, and were statistically significant after adjusting for multiple comparisons by the Benjamini–Hochberg procedure. **c** Number of DARs in $ZIC2^{-/-}ZIC3^{-/-}$ hESCs that overlap with ZIC2-bound loci, grouped by Louvain clusters. **d** Composite plots showing ATAC-seq signal in WT and $ZIC2^{-/-}ZIC3^{-/-}$ hESCs over ZIC2-bound loci, grouped by Louvain clusters. **e** Kernel density estimate plots of ATAC-seq signal over primed-specific enhancers, stratified by whether each loci contains motifs corresponding to *ZIC* factors or not. Median values are shown and *P* values were calculated from a two-tailed paired Wilcoxon signed-rank test. **f** Correspondence of regions that lose accessibility in $ZIC2^{-/-}ZIC3^{-/-}$ hESCs and C3 loci to differentially expressed genes, grouped by distance. **g** Heatmap of scaled ATAC-seq signal, reported in CPB, over all DARs identified across all mutants. Three sets of DARs were identified by *k*-means clustering. Enriched biological process terms identified from GREAT analysis and top motifs enriched among each cluster are shown. **h** Western blot of BRG1 upon immunoprecipation of ZIC2 in primed hESCs. **i** Western blot of ZIC2 upon simultaneous immunoprecipitation of BRG1 and SMARCD1. **h**, **i** Images are representative of two biological replicates. Size markers (in kilodaltons) indicated on right. **j** Composite plot showing ATAC-seq signal ratio of $SMARCA4^{-/-}$ hESCs compared to WT upon culture of naïve hESCs in primed hESC media conditions. **k** Composite plots showing normalised BRG1 CUT&Tag signal over ZIC2-bound peaks grouped by Louvain clusters, in WT and $ZIC2^{-/-}ZIC3^{-/-}$ hESCs. **d**, **e**, **g**, **k** CPB = counts per bin. **l** Scatterplot of BRG1 CUT&Tag and ATAC-seq signal ratios over 200 bp bins in the centre of C3 regions, in WT hESCs relative to $ZIC2^{-/-}ZIC3^{-/-}$ hESCs. Dotted lines show origin and gating. Numbers in each quadrant represent percentage of all regions considered within each gate. **m** Genome browser tracks of ATAC-seq and BRG1 CUT&Tag in WT and $ZIC2^{-/-}ZIC3^{-/-}$ primed hESCs around a ZIC2/3-regulated gene. Source data are provided as a Source Data file.

primed hESCs to a lesser extent (Supplementary Fig. S10j). Upon ectopic *ZIC2* expression in naïve hESCs, there was a flattening of cell colonies, progressive loss of the naïve $CD75^{hi}CD24^{low}$ population, and gradual emergence of $CD75^{low}CD24^{hi}$ cells—characteristic of primed pluripotency (Fig. 4b, c)[34]. At the molecular level, induction of *ZIC2* in naïve hESCs, primarily resulted in upregulation of genes and a shift towards primed-specific gene expression (Fig. 4d, e and Supplementary Fig. S10k–m).

Our own and published data indicate that naïve hESCs have modest residual expression of *ZIC3* (Supplementary Fig. S11a). Consistent with a recent CRISPR screen[45] that identified *ZIC3* as a gene that impedes naïve pluripotency, we observed enhanced efficiency of naïve reversion from $ZIC3^{-/-}$ primed hESCs (Supplementary Fig. S11b–e). It is known that while naïve hESCs can differentiate readily towards the trophoblast lineage, primed hESCs have a greater propensity to differentiate to amnion-like cells (AMLC) when transitioned to hTSC culture media[33,46,47]. Loss of ZIC2/3 in naïve hESCs prior to differentiation resulted in improved conversion to trophoblast and reduced expression of AMLC markers (Supplementary Fig. S11f–j). Conversely, naïve hESCs induced to express *ZIC2* or *ZIC3* for short periods prior to hTSC derivation had a propensity to express AMLC markers (Fig. 4f, g and Supplementary Fig. S11k, l), and expression of *ZIC2* after culture in hTSC medium strongly antagonised trophoblast formation (Supplementary Fig. S11m). This suggests that *ZIC2/3* expression causes hESCs to exit naïve pluripotency to a more developmentally advanced state that is more likely to give rise to AMLCs than hTSCs and that lineage potency is affected by *ZIC2/3* expression levels in naïve hESCs[15,48].

https://doi.org/10.1038/s41467-024-52431-1

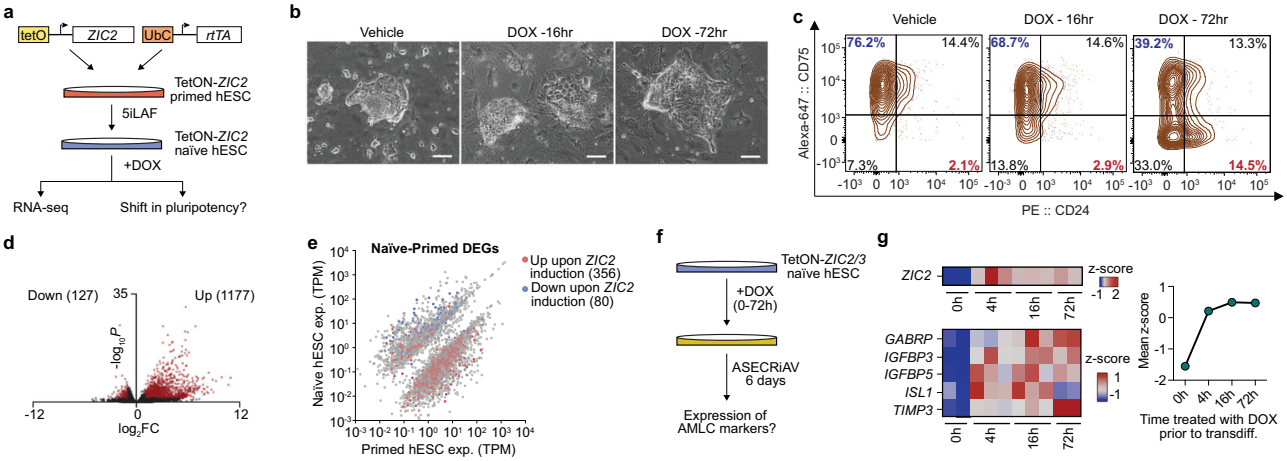


**Fig. 4 | ZIC2 facilitates exit from naïve pluripotency. a** Schematic of experiments designed to determine where 3xFlag-ZIC2 localises to on chromatin and how it affects naïve pluripotency. **b** Brightfield images of TetON-*ZIC2* naïve hESCs upon treatment with doxycycline compared to DMSO control. Scale bar 100 µm. Data are representative of two biological replicates. **c** Contour plots showing expression of CD75 and CD24 in TetON-*ZIC2* naïve hESCs upon treatment with doxycycline compared to DMSO control. **d** Volcano plot showing number of DEGs in naïve hESCs upon ectopic expression of *ZIC2*. *P* values were calculated from a two-tailed Wald test and adjusted for multiple comparisons by the Benjamini–Hochberg procedure. **e** 2D plot showing overlap of genes that become differentially expressed upon ectopic expression of *ZIC2* in naïve hESCs, with genes that are differentially expressed in naïve and primed hESCs. TPM transcripts per million. **f** Experimental setup to test lineage potency of naïve hESCs in response to ASECRiAV medium, upon ectopic expression of *ZIC2* or *ZIC3*. **g** (Top) Expression of *ZIC2* in naïve hESCs upon treatment with doxycycline for different timepoints. (Bottom) Expression of AMLC markers in naïve hESCs pre-treated with doxycycline for different timepoints and then cultured in ASECRiAV medium. Gene expression trends are shown. Source data are provided as a Source Data file.

## SWI/SNF acts downstream of ZIC2/3 to remodel chromatin in naïve hESCs towards primed pluripotency

To determine how *ZIC* factors facilitate exit from naïve pluripotency, we induced expression of Flag-*ZIC2* in naïve hESCs and conducted ATAC-seq and ChIP-seq for Flag, BRG1 and H3K27me3 (Fig. 5a). We did not observe gain of H3K27me3 at C4 (bound by ZIC2 and enriched for H3K27me3 in primed hESCs) regions upon *ZIC2* expression in naïve hESCs (Supplementary Fig. S12a), arguing against ZIC2/3 playing a direct role in PRC2 recruitment. Instead, we found extensive evidence that ZIC2/3 is directly responsible for opening chromatin and promoting gene expression.

When *ZIC2* was ectopically expressed in naïve hESCs, we observed that it preferentially homed to regions open specifically in primed, but not in naïve hESCs (Fig. 5b). There was a predominant gain rather than loss of DARs of chromatin, with a strong relationship between the binding of Flag-ZIC2 and chromatin opening (Fig. 5c, d and Supplementary Data 8). We did, however, observe decreased accessibility over LTR5Hs elements (Supplementary Fig. S12b), which are uniquely open and active in naïve hESCs, indicating exit from naïve pluripotency. Far more genes were upregulated than downregulated upon *ZIC2* induction (Fig. 4d) and there was a strong correspondence between increased accessibility at de novo ZIC2-bound loci and upregulation of proximal genes (Fig. 5e), along with a general transition towards a primed-like chromatin accessibility landscape and transcriptome (Supplementary Fig. S12c, d). Overexpression of *ZIC3* produced a similar pattern of chromatin opening as ZIC2, consistent with their overlapping roles (Supplementary Fig. S12e). Notably, there was a strong correspondence between gain of ATAC-seq signal upon *ZIC2* induction in naïve hESCs and loss of ATAC-seq signal in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ primed hESCs (Fig. 5f).

Since ZIC2/3 are implicated in recruiting BRG1 to C3 loci in primed pluripotency, and ectopic expression of *ZIC2* in naïve hESCs resulted in increased accessibility over these regions, we queried the role of BRG1 in facilitating this process. Within 8 h of DOX-mediated *ZIC2* induction in naïve hESCs, BRG1 occupancy increases at loci that become bound by Flag-ZIC2 (Fig. 5g and Supplementary Data 9), especially at regions that become more accessible (Fig. 5h–j and Supplementary Fig. S12f). Correspondingly, an increase in BRG1 at C3 loci occurs upon *ZIC2* expression (Supplementary Fig. S12g). We found this trend to be congruent with ATAC-seq data upon ectopic *ZIC2* expression in naïve hESCs, and with BRG1 CUT&Tag data in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ primed hESCs (Fig. 5k, l). We used ACBI1, a PROTAC degrader of several SWI/SNF complex components including BRG1[49] (Fig. 5m and Supplementary Fig. S8c), to determine if these changes in accessibility occur in absence of their chromatin remodelling activity. Indeed, treatment with ACBI1 during the duration of doxycycline treatment in naïve hESCs abolished the increase in accessibility mediated by ectopic Flag-ZIC2 over C3 loci and ZIC2-bound primed-specific open chromatin sites (Fig. 5n–r) without broadly disrupting genome-wide chromatin accessibility (Fig. 5o, p and Supplementary Fig. S12h, i). In line with this, the absence of SWI/SNF prevented ectopic *ZIC2* in naïve hESCs from causing the formation of CD24$^{hi}$CD75$^{low}$ primed-like hESCs (Supplementary Fig. S12j) highlighting SWI/SNF as a key mediator of *ZIC*-dependent progression towards primed pluripotency.

## Transcriptional machinery consolidates ZIC2 enrichment over primed-specific enhancers

SWI/SNF complexes depend on transcriptional initiation to remain associated with chromatin[50] and we considered whether transcription itself is likewise important for ZIC2 function. Notably, all classes of regions bound by ZIC2 were strongly enriched for RNAPII and Ser5p-RNAPII[51], as well as the transactivator BRD4[52,53], indicating association with sites of transcriptional initiation (Fig. 6a). To determine the importance of transcriptional initiation for ZIC2 binding we treated primed hESCs with triptolide, which prevents transcriptional initiation[54], flavopiridol, which pauses transcription[55], and STM2457, which prevents methylation of RNA[56] (Fig. 6b). At 4 h of treatment, triptolide had a strong, selective negative effect on ZIC2 binding specifically at C3 regions, while other inhibitors had no considerable effect (Fig. 6c–g). These results indicate that transcriptional initiation is critical for ZIC2 binding specifically at enhancers, distinguishing regulation at these loci from the dynamics observed over ZIC2-bound promoters.

https://doi.org/10.1038/s41467-024-52431-1

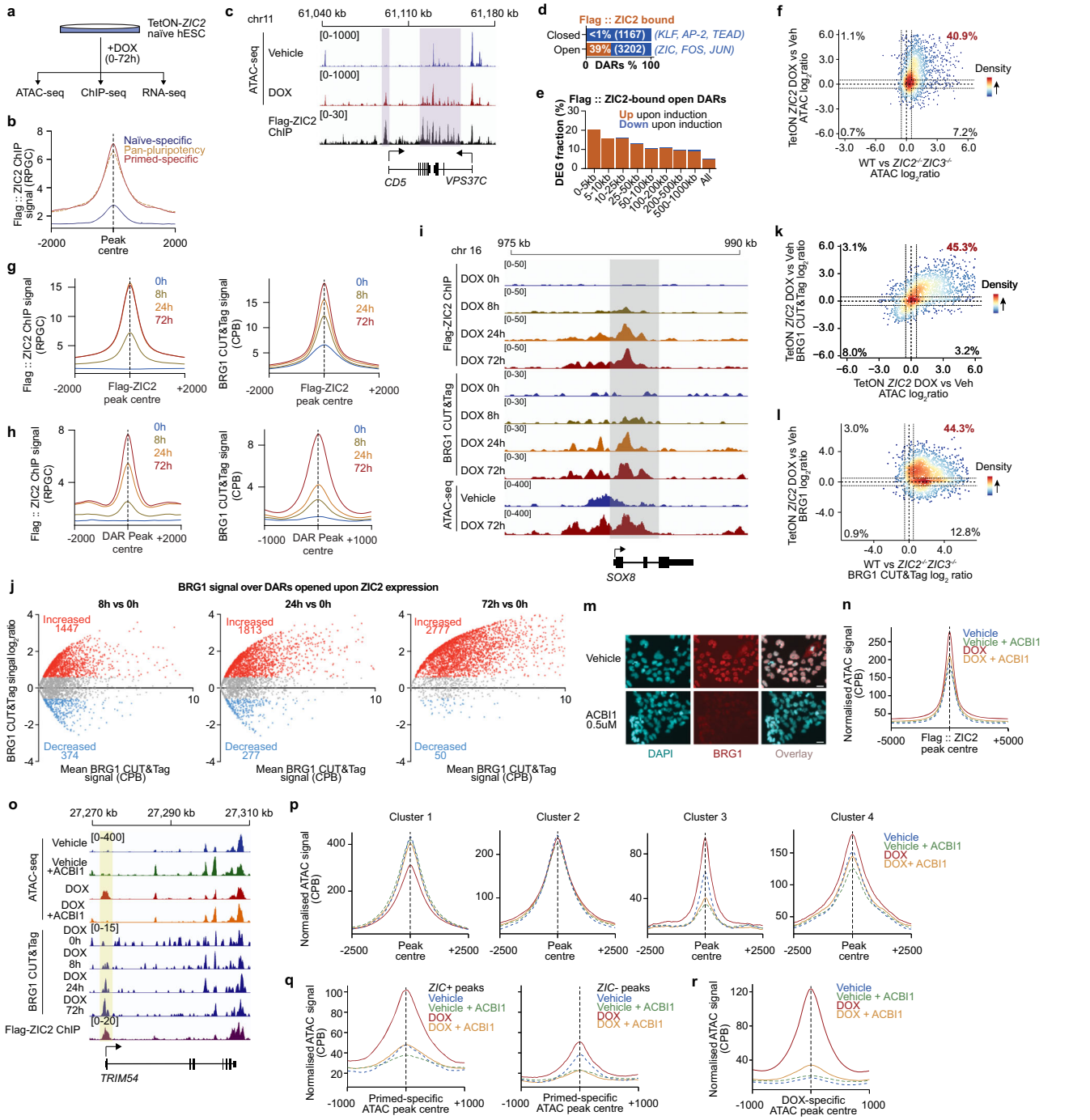


**Fig. 5 | SWI/SNF acts downstream of *ZIC* factors in opening primed hESC-specific enhancers. a** Schematic of experiments designed to test the effect of ectopic *ZIC2/3* expression in naïve hESCs. **b** Enrichment of 3xFlag-ZIC2 over ATAC-seq peak sets indicated. **c** Genome browser tracks showing correspondence between 3xFlag-ZIC2 occupancy and increased chromatin accessibility. **d** Fraction of DARs that overlap with Flag-ZIC2-bound loci. Top motifs enriched over DARs that are not bound by 3xFlag-ZIC2 are shown. **e** Fraction of genes that show differential expression upon *ZIC2*-induction (DEGs) as a function of distance from ZIC2-bound loci that become open upon induction. **f** Scatterplot over 200 bp bins around centre of C3 regions showing congruence between loss of accessibility in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ hESCs and gain in accessibility upon 3xFlag-*ZIC2* expression in naïve hESCs. Composite plots showing Flag-ZIC2 ChIP signal (left) and BRG1 CUT&Tag signal (right) over Flag-ZIC2 bound sites (**g**) or DARs that are opened (**h**) upon *ZIC2* induction. **b**, **g**, **h** RPGC reads per 1x genome coverage. **i** Genome browser tracks showing ZIC2 and BRG1 occupancy over the *SOX8* locus upon *ZIC2* induction. **j** Means-average plots showing BRG1 occupancy over regions that become accessible upon Flag-ZIC2 induction in naïve hESCs. **k** Scatterplot over 200 bp bins around centre of C3 regions showing congruence between changes in accessibility and BRG1 occupancy upon ectopic expression of ZIC2 in naïve hESCs. **l** Scatterplot over 200 bp bins around centre of C3 regions, showing congruence between gain in BRG1 occupancy upon ectopic expression of ZIC2 in naïve hESCs and loss in BRG1 occupancy in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ primed hESCs. **m** Immunofluorescence showing degradation of BRG1 in primed hESCs upon treatment with ACBI1 for 2 days. Scale bar 50 µm. Images are representative of two biological replicates. **n** Normalised ATAC signal over regions bound by ectopic Flag-ZIC2. **o** Genome browser tracks showing ZIC2 and BRG1 occupancy over the *TRIM54* locus upon Flag-ZIC2 induction and failure to open chromatin upon treatment with ACBI1. Normalised ATAC signal over distinct classes of ZIC2-bound regions (**p**), primed-specific ATAC peaks with and without ZIC motifs (**q**) and regions that become accessible upon ectopic expression of Flag-ZIC2 (**r**) in naïve hESCs in the presence or absence of ACBI1. **n**–**r** Treatments were carried out for 3 days. **j**, **n**, **p**–**r** CPB counts per bin. Source data are provided as a Source Data file.

https://doi.org/10.1038/s41467-024-52431-1

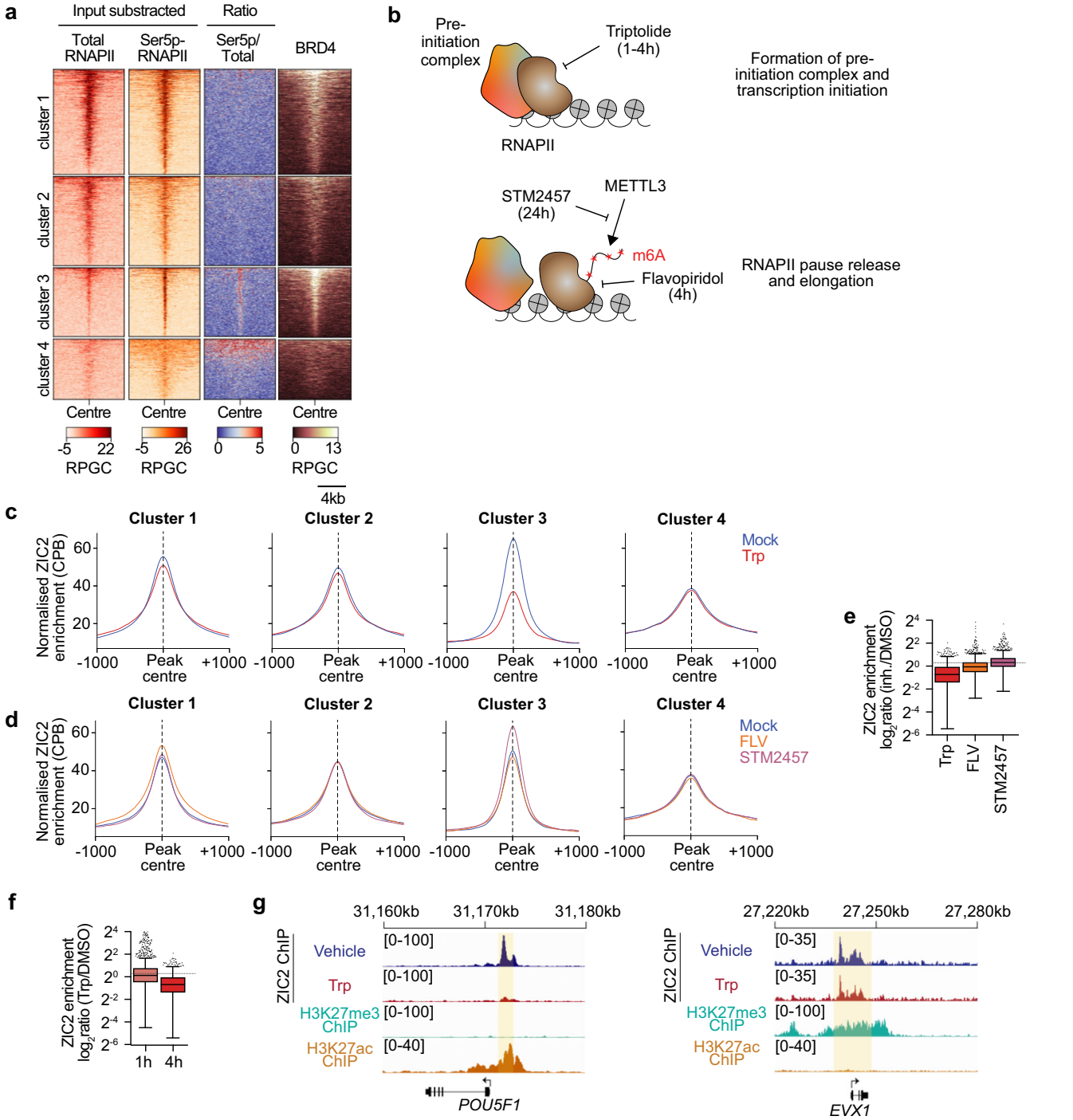


**Fig. 6 | Transcriptional machinery consolidates recruitment of ZIC2 over primed-specific enhancers. a** Enrichment of total RNAPII, Ser5p-RNAPII and BRD4 over ZIC2-bound regions. RPGC reads per 1x genome coverage. Enrichment of Ser5p-RNAPII to total RNAPII is also shown. **b** Schematic of experiments designed to test the effect of inhibiting various components of transcription and its machinery on recruitment of ZIC2. **c** Effect of treatment with triptolide (Trp) for 4 h on ZIC2 occupancy in primed hESCs. **d** Effect of treatment with flavopiridol (FLV) and STM2457 for 4 and 24 h, respectively, on ZIC2 occupancy in primed hESCs. **c**, **d** CPB = normalized counts per bin. **e** Comparison of 4 h FLV and 24 h STM2457 treatment to 4 h Trp treatment on ZIC2 occupancy over C3 loci. Results were computed from two independent biological replicates. **f** Comparison of Trp treatment for 1 and 4 h on ZIC2 occupancy over C3 loci. ChIP-seq data for the treatment of primed hESCs with triptolide for 4 h was computed from two independent biological replicates, and data for treatment with triptolide for 1 h was computed from one replicate. **e**, **f** Median values are shown, and boxplots represent the inter-quartile range. Whiskers are calculated using the Tukey method. **g** Genome browser tracks showing differential effects of ZIC2 occupancy over primed-specific enhancers and bivalent regions upon treatment with triptolide in primed hESCs. Source data are provided as a Source Data file.

## Discussion

All cells of the human body trace back to a single pluripotent cell type present in the first few days of the developing embryo. The primed epiblast which forms in the 2nd week of embryogenesis constitutes the gateway between an organism of near homo-cellular composition to one with thousands of different cell types. We provide compelling evidence that ZIC2 and ZIC3 act together as key regulators during this critical period of human development (Fig. 7).

https://doi.org/10.1038/s41467-024-52431-1

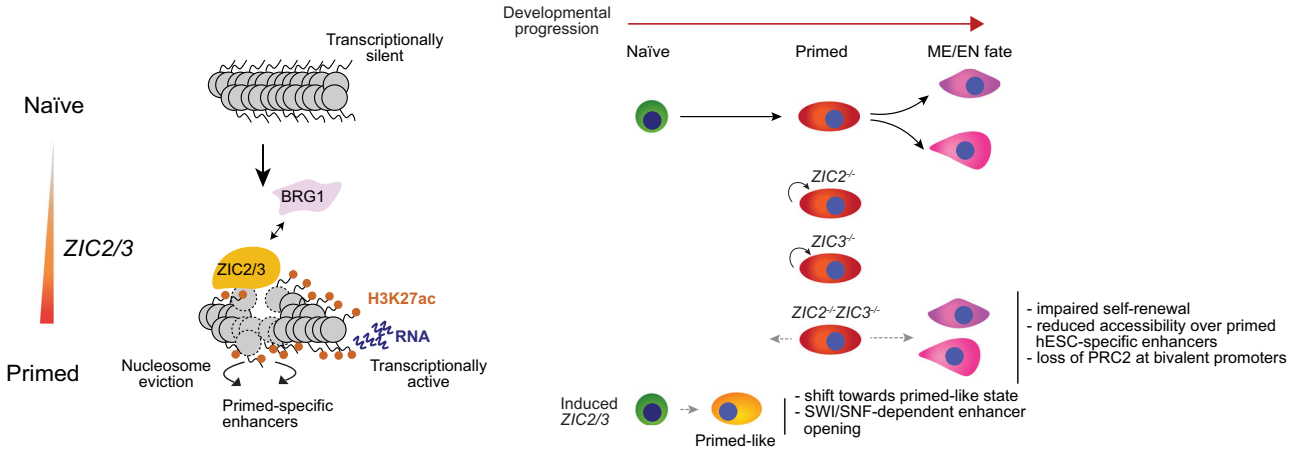


**Fig. 7 | Illustration of critical findings of this study.** ZIC2 and ZIC3 have largely overlapping roles and act together as crucial regulators of human primed pluripotency, facilitating recruitment of the SWI/SNF complex over primed hESC-specific enhancers. Their simultaneous ablation in primed hESCs results in selective closing of lineage-specific enhancers, a failure to self-renew or maintain primed pluripotency, and a gene expression pattern indicative of disrupted PRC2 activity coincident with aberrant expression of mesendoderm genes. Their ectopic expression in naïve hESCs facilitates progression towards primed pluripotency in a SWI/SNF-dependent manner.

There is extensive literature showing a role for *Zic2* and *Zic3* in embryonic development[23–26], however, our study uniquely shows essentiality in a cell type reflective of epiblast development. Some reports indicate a stronger phenotype from single *ZIC* knockouts than what we observed. RNAi depletion of *Zic3* in mESCs and hESCs is reported to result in increased expression of endodermal genes, a phenotype exacerbated in mESCs by simultaneous *Zic2* knockdown[57]. Depletion of *Zic2* in mESCs is reported to result in activation of hundreds of PRC2-target genes in mESCs[27]. While we did not observe a severe phenotype in *ZIC2* or *ZIC3* single knockouts, we did observe some upregulation of endoderm markers in *ZIC3*$^{-/-}$ hESCs (Supplementary Fig. 4d) and acute ablation of *ZIC2* did result in some spontaneous differentiation (Supplementary Fig. 4c). Thus, although more subtle in primed hESCs, loss of either ZIC2 or ZIC3 does have some destabilising effect on pluripotency, consistent with previous reports. Another report indicated that mESCs lacking functional *Zic2* and *Zic3*, cultured in 2i + LIF conditions are apparently viable[30]. Their ability to survive in 2i + LIF conditions may reflect the non-essentiality of *ZIC* factors in the naïve state. Intriguingly, the *Zic2*$^{-/-}$*Zic3*$^{-/-}$ mESCs were capable of modest contribution to chimeras, with a few percent of cells in the E9.5 embryo originating from *Zic2*$^{-/-}$*Zic3*$^{-/-}$ mESCs. It is possible that *Zic2*$^{-/-}$*Zic3*$^{-/-}$ mESCs have an ability to pass through formative and primed pluripotency unlike *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ hESCs, but it is also possible that they differentiate via a substantially different trajectory that bypasses the primed state.

In primed hESCs, we saw nearly equal gain and loss of gene expression upon *ZIC2/3* ablation, but induced expression of *ZIC2* in naïve hESCs showed a dominant trend towards transcriptional activation, with a distance-dependent correspondence between *ZIC*-binding and activation. While we find that ZIC2 and ZIC3 act in an overlapping manner, perturbation of *ZIC5* had little discernable impact on transcription. Correspondingly, in reporter assays, ZIC1–3 show far higher transactivation ability than ZIC4 and ZIC5[58]. ZIC2/3's role in the recruitment of BRG1 is consistent with a general function as an activator that opens chromatin at enhancers.

We observe both gain in BRG1 occupancy upon recruitment of ZIC2 to chromatin in naïve hESCs and loss of BRG1 occupancy to chromatin over primed-specific enhancers in the absence of both ZIC2 and ZIC3 in primed hESCs. This highlights a robust role of the functional interaction between *ZIC* factors and the SWI/SNF complex in dictating progression to and maintenance of primed pluripotency. Importantly, especially in naïve hESCs, the striking ability of ZIC2 to bind previously closed chromatin, recruit BRG1 and cause subsequent increases in chromatin accessibility over time raises the possibility that ZIC2/3 may act as pioneer factors[59]. The closing of chromatin we observe upon ZIC2/3 loss is slower and more subtle, but this is consistent with the observation that enhancers can remain open even if the pioneering factors that opened them are removed[60].

While the effect of ZIC2 in recruiting BRG1 was consistent in both naïve and primed hESCs, we found that the effect on H3K27me3 was context specific. The primed pluripotent state features a lower global level of H3K27me3 than the naïve state but a high density at promoters[7], and the loss of ZIC2/3 results in selective loss of EZH2 and H3K27me3 over these regions of high enrichment with concomitant de-repression of PRC2 targets. The mechanism by which this occurs is unclear. Loss of ZIC2/3 results in upregulation of *EPOP* and downregulation of *JARID2*, which are PRC2.1 and PRC2.2 core components, respectively. While it is non-obvious that a shift from PRC2.2 to PRC2.1 would result in loss of polycomb repression over concentrated regions, it is notable that naïve hESCs in which PRC2.1 is more prominent feature a more dispersed H3K27me3 landscape[7,8]. We also observe upregulation of *KDM6B* in the *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ primed hESCs. *KDM6B* is an H3K27 demethylase known to be important for mesoderm and endoderm differentiation[61]. However, none of these three genes are apparent direct *ZIC* targets. The mechanism by which H3K27me3 becomes concentrated at promoters during the naïve to primed transition, and the mechanism by which it is subsequently lost from promoters during differentiation, remains substantially unknown, as does ZIC2/3's link to this process.

We were intrigued by the question of whether the transcriptional machinery itself regulates association of ZIC2/3 with chromatin. Treatment with triptolide results in substantial detachment of ZIC2 from chromatin in 4 h, but only from enhancers. An experiment in K562 leukaemia cells showed that gross loss in chromatin accessibility of enhancers is not observed after 4 h of Trp treatment[62], suggesting that ZIC2 may not be lost because it loses direct access to DNA. However, an experiment in mESCs shows that BRG1 is almost uniformly lost within 2 h of Trp treatment[50] and a different experiment in mESCs shows that SWI/SNF inhibition causes rapidly reduced binding of pluripotency factors to their target motifs in enhancers[63]. The timing of ZIC2 loss over enhancers, occurring between 1 and 4 h after inhibition, is consistent with a model in which ZIC2 does not depend on transcriptional initiation per se but rather on the continued action of BRG1 to retain openness at enhancers in ESCs. However, even in

K562 cells in which Trp does not affect chromatin openness, Trp treatment has a much stronger effect on promoter–enhancer than promoter–promoter interactions[64], implying that the three-dimensional interactions normally present at enhancers are disproportionately perturbed. A similar result is observed upon acute RNAPII depletion in DLD-1 cancer cells[65]. We do not yet know if selective loss from enhancers upon RNA Pol II degradation is a general property of TFs or a selective property of *ZIC* TFs, but our finding raises intriguing questions about what factors beyond direct recognition of DNA affect binding of TFs to targets and how enhancers and promoters are differently regulated.

Cell fate decisions are dictated and shaped by their underlying epigenome. Simultaneous perturbation of ZIC2 and ZIC3 results in deregulation of two key aspects of gene regulation integral to primed pluripotency—closing of SWI/SNF-regulated enhancers that define it and de-repression of Polycomb target genes upregulated during the onset of gastrulation. Interestingly, the progressive opening of many but not all primed-specific enhancers upon ectopic expression of *ZIC2* indicates the role of *ZIC* factors as important limiting factors during progression towards primed pluripotency. This ability to both drive acquisition of primed pluripotency from a naïve state and maintain self-renewal capacity in the primed identifies ZIC2 and ZIC3 as important regulators of primed pluripotency.

## Methods

### Ethical permissions

All experiments were approved by the McGill University Faculty of Medicine Institutional Review Board and the CIHR Stem Cell Oversight Committee.

### Plasmid preparation and cloning

Vectors used in this study include pGL3-promoter (Promega), pRL-TK (Promega), FUW-tetO-loxP-hNANOG (Addgene 60849), FUW-M2rtTA (Addgene 20342), psPAX2 (Addgene 12260), pMD2.G (Addgene 12259), pUC57-Kan-3xFlag-hZIC2 and pUC57-Kan-3xFlag-hZIC3 (both custom synthesised by Azenta). All putative enhancers were PCR amplified with restriction site overhangs from human genomic DNA and cloned into pGL3-promoter using Acc651 and Xho1. 3xFlag-hZIC2 and 3xFlag-hZIC3 ORF were synthesised with Nhe1 and Age1 overhangs and cloned into custom vector and finally subcloned into Addgene 60849. All newly made constructs were validated by Sanger sequencing.

### Virus generation

Virus was generated from HEK 293T cells cultured in DMEM supplemented with 10% BCS. psPAX2 (250 ng), pMD2.G (250 ng) and lentiviral transfer plasmid of interest (500 ng) were transfected into 293Ts using Lipofectamine 3000 (Thermo Fisher) according to manufacturer's instructions. Media was changed to primed hESC media 12 h post-transfection. Conditioned media, containing virus particles were collected 48–60 h post-transfection and residual 293Ts were removed using a 0.45 μm filter. The filtrates were then stored as single-use aliquots at −80 °C, for up to 6 months, and thawed only prior to use.

### Cell culture

**Maintenance and reversion of hESCs.** UCLA1 primed hESCs[66] and AJC002 primed hiPSCs were cultured on hESC-qualified Matrigel (Corning) coated plates and media was refreshed daily with TeSR-E8 or mTESR-Plus (STEMCELL Technologies).

Naïve hESCs were reverted from primed hESCs and maintained on mitotically inactive PMEFs (Millipore)[4–6,12,46]. Briefly, 250,000 primed hESCs were dissociated and plated as single cells on MEFs. For the following 3 days, reverting cells were cultured in reset induction media (24.5 ml DMEM F-12, 24.5 ml Neurobasal media, 0.25 ml N2, 0.5 ml B27, 1x Glutamax, 150 μM L-ascorbic acid, 0.1 mM β-mercaptoethanol, 1 μM PD0325901, 10 μM Y-27632, 0.75 mM valproic acid and 10 ng/ml recombinant hLIF) and later, stably maintained in PXGL[15,44] (24.5 ml DMEM F-12, 24.5 ml Neurobasal media, 0.25 ml N2, 0.5 ml B27, 1x Glutamax, 150 μM l-ascorbic acid, 0.1 mM β-mercaptoethanol, 1 μM PD0325901, 10 μM Y-27632, 2 μM XAV939, 2 μM Gö6983 and 10 ng/ml recombinant hLIF). In some cases, PXGL-cultured naïve hESCs were purified for $CD75^{hi}$ population by FACS prior to experiments.

In other cases, we reverted to and cultured naïve hESCs in 5iLAF conditions[4,5,46]. Briefly, 100,000–200,000 primed hESCs were seeded as single cells on a mitotically inactive MEF feeder layer in each well of a 6-well plate in medium supplemented with 10 μM ROCK inhibitor Y-27632. Two days later, medium was changed to 5iLAF culture conditions (1x N2 supplement (Invitrogen; 17502048), 1x B27 supplement (Invitrogen; 17504044), 10 μg recombinant human LIF (STEMCELL Technologies), 1x Glutamax (Gibco), 1x nonessential amino acids (Gibco), 0.1 mM β-mercaptoethanol (Sigma), 1% penicillin-streptomycin (Gibco), 50 μg/ml BSA (Gibco Fraction V, 15260), and the following small molecules and cytokines: PD0325901 (Axon MedChem, 1 μM), IM-12 (Axon MedChem, 0.25–1 μM), SB590885 (Axon MedChem, 0.5 μM), WH4-023 (Axon MedChem, 1 μM), Y-27632 (Axon MedChem, 10 μM) and Activin A (Invitrogen, 10 ng/mL)). Naïve hESC colonies appeared after the first passage, which was carried out 8–12 days post-induction. Naïve hESCs were not cultured for more than five passages and were cultured in 5% $O_2$, 5% $CO_2$ at 37 °C. Medium was changed every other day. Primed hESCs were cultured in 20% $O_2$, 5% $CO_2$ at 37 °C. Primed hESC Media was refreshed daily, and cells were passaged every 5–7 days. All cell lines were routinely checked for mycoplasma contamination.

**Generation of CRISPR mutant lines.** All knockout lines were generated by electroporation of Cas9-sgRNA ribonucleoprotein complexes (RNPs). hESCs were cultured in media supplemented with 10 μM Y-27632 for 24 h prior and dissociated into single cells with 30% TrypLE. Meanwhile, 100 pmol sgRNA and 20 pmol Cas9 (all purchased from Synthego) were combined and pre-complexed for 10 min. Pre-complexed RNPs were combined with hESCs and 20 μl P3 solution (Lonza), transferred into a cuvette and electroporated in an Amaxa 4D nucleofector (Lonza) with pulse code CA137. Media was immediately added to electroporated cells and transferred into a well of a 24-well plate and cultured for 2 days in media supplemented with 10 μM Y-27632. To generate clonal lines, upon reaching confluency, primed hESCs were dissociated into single cells and plated at very low densities in a single well of a 6-well plate. After 7–14 days of culture, colonies arising from each single cell were manually picked and polyclonally expanded. Clonal lines were screened for homozygous deletion. Deletion was confirmed by genotyping PCR and loss of functionality was confirmed either by western blot or RNA-seq or qRT-PCR. In cases where bulk populations of cells were used instead of clonal lines, all experiments were carried out within 5–7 days of nucleofection, unless explicitly mentioned otherwise. $ZIC2^{-/-}ZIC3^{-/-}$ hESCs and $ZIC2^{-/-}ZIC3^{-/-}ZIC5^{-/-}$ described in text or figures refer to $ZIC3^{-/-}$ and $ZIC3^{-/-}ZIC5^{-/-}$ hESCs that have been electroporated with sgRNA targeting *ZIC2*. The only exceptions include H2AK119ub ChIP (Supplementary Fig. S6f), CUT&Tag (Fig. 3k–m and Supplementary Fig. S6g, h), re-priming (Supplementary Fig. S10c–e) and transdifferentiation (Supplementary Fig. S11f–j) experiments, in which case $ZIC2^{-/-}$ primed hESCs were electroporated with sgRNA targeting *ZIC3* to produce $ZIC2^{-/-}ZIC3^{-/-}$.

**Generation of stable expression lines.** TeSR-E8 media and virus-containing supernatant were combined in 1:1 v/v solution, supplemented with 10 μg/ml polybrene and 10 μM Y-27632—and added to a confluent well of primed hESCs. Transduction was carried out by spinfection—where the plate of cells with virus particles was spun at 700 × *g* for 1 h and the virus-containing media was aspirated

immediately after. Transduced cells were then cultured in media supplemented with 10 μM Y-27632 for the next 2 days. To generate stable lines, hESCs were dissociated into single cells and plated at very low densities in a single well of a 6-well plate. After 7–14 days of culture, colonies arising from each single cell were manually picked and polyclonally expanded. Homogeneity and stability of expression of each stable cell line were validated by western blot and immunofluorescence before any experiments were carried out. Unless otherwise mentioned, all inducible experiments were carried out in media supplemented with 100 ng/ml doxycycline hyclate or matched vehicle control.

**Re-priming of naïve hESCs (conversion from naïve to primed).** Naïve hESCs grown in 5iLAF culture conditions were dissociated into single cells and depleted of MEFs. Briefly, naïve hESCs were incubated with 30% TrypLE at 37 °C for 3 min. Human cells were separated from MEFs coarsely by manual pipetting; additionally, the cell suspension was plated on 0.1% gelatin-coated plates for 40 min. Since MEFs have a higher affinity of seeding on gelatin than hESCs, the supernatant enriched for human cells was collected and processed accordingly.

In total, 250,000 naïve hESCs were then seeded on hESC-qualified matrigel coated plates in mTESR-plus medium supplemented with 10 μM Y-27632. Media was changed daily until cells were collected for experiments.

**Transdifferentiation of naïve hESCs to hTSCs.** Naïve hESCs cultured in 5iLAF culture conditions were dissociated into single cells and depleted of MEFs as described above[4]. In total, 300,000 naïve hESCs were then either electroporated using Cas9 RNPs, or plated directly on matrigel coated plates in ASECRiAV medium. Media was changed every other days until cells were collected for experiments.

**Small molecule experiments.** Primed hESCs were treated with 500 nM triptolide (Tocris) for 1–4 h, 1 μM flavopiridol (SelleckChem) for 4 h and 25 μM STM2457 (SelleckChem) for 24 h, in their respective experiments. ACBI1, a kind gift from the lab of Dr. Sidong Huang, was used at 300 nM to treat naïve and primed hESCs, as indicated. In all cases, control cells were treated with the equivalent volume of DMSO for the same amount of time.

## ATAC-seq

**Library generation.** Cells were lysed in 50 μl ATAC resuspension buffer (10 mM Tris-HCL pH 7.4, 10 mM NaCl and 3 mM $MgCl_2$) supplemented with 0.1% Tween-20, 0.1% NP-40 and 0.01% digitonin, and incubated on ice for 3 min. Lysis buffer was then diluted with 1 ml ATAC resuspension buffer supplemented with 0.1% Tween-20 and tubes were inverted three times. Nuclei were pelleted by centrifugation at 500 × *g* for 10 min at 4 °C. Transposition reaction was carried out in 50 μl volume for each sample by resuspending nuclei in 25 μl 2x TD buffer (20 mM Tris-HCL pH 7.6, 10 mM $MgCl_2$ and 20% DMF), 2.5 μl transposase (100 nM final), 16.5 μl PBS, 0.5 μl 1% digitonin, 0.5 μl 10% Tween-20 and 5 μl $H_2O$. Reactions were incubated at 37 °C for 30 min in a thermomixer with 1000 RPM mixing. Reactions were quenched with 5× volume of buffer PB and DNA was purified using the MinElute PCR Purification Kit (Qiagen). Libraries were next amplified using Illumina-compatible primers. Naïve hESCs were purified from co-cultured MEFs as described above, before subsequent lysates were prepared.

**ATAC analysis.** Reads were trimmed to 50 bp using cutadapt and subsequently aligned against the GRCh38 reference genome, using bwa mem with default parameters. Uniquely mapped alignments ($q > 20$) were filtered using samtools. Picard was used to mark and remove duplicate reads and finally samtools was used to merge biological/technical replicates where appropriate, sort by chromosome number and create indices. Mitochondrial reads were removed prior to analysis. To identify DARs between two conditions, a custom script utilising csaw and edgeR was used. Motifs enriched in a given peakset were identified using the findmotifsgenome.pl command in Homer with the following notable parameters (-size given -mask). Occurrences of a given motif in a given peakset were identified using the annotatepeak.pl command in Homer. bamCoverage command in deepTools (version 2.5.3) was used to generate normalised bigWig and bedgraph files.

## Chromatin immunoprecipitation

**Library preparation.** ChIP was conducted as previously described with minor modifications[4,67]. Briefly, cells were fixed with 1% paraformaldehyde 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 1200 × *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 1 (10 mM Tris-HCL pH 8.0, 0.25% Triton X-100, 10 mM EDTA, 0.5 mM EGTA, 1x Protease inhibitors (Roche), 1 mM NaF, 1 mM $Na_2VO_3$ and 1 mM PMSF), then rotated for 15 min. Nuclei were pelleted by centrifugation at 1200 × *g* for 5 min at 4 °C. Nuclei were then resuspended with 1 ml lysis buffer 2 (10 mM Tris-HCL pH 8.0, 200 mM NaCl, 10 mM EDTA, 0.5 mM EGTA, 1x Protease inhibitor, 1 mM NaF, 1 mM $Na_2VO_3$ and 1 mM PMSF) and rotated for 10 min at 4 °C. Nuclei were then pelleted and resuspended in 900 μl lysis buffer 3 (10 mM Tris-HCL pH 8, 10 mM EDTA, 0.5 mM EGTA, 0.1% SDS, 1 mM NaF, 1 mM $Na_2VO_3$ and 1 mM PMSF) and transferred to 1 ml milliTube (Covaris). Sonication was performed with the Covaris M220 Focused-ultra sonicator (with the following settings: PIP = 75 W; CPB = 200; Duty Factor = 20%) for 10 min at 7 °C. The sonicated lysate was then centrifuged for 10 min at 15,000 × *g*, and the supernatant was retained.

Fifty microliters of Protein A/G Dynabeads (GE Healthcare) were washed three times with ChIP dilution buffer (16.7 mM Tris-HCL pH 8.0, 0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 167 mM NaCl), each wash consisting of addition of 1 ml of buffer and collection of the beads on a magnetic rack (BioRad). Beads were then resuspended in 900 μl ChIP dilution buffer and combined with the solubilised chromatin 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.

One percent of the supernatant was saved as input, the rest used for subsequent immunoprecipitations. The samples were then rotated overnight at 4 °C. Fifty microliters of Protein A/G Dynabeads were then 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% Triton X-100, 0.1% deoxycholate, 1 mM EDTA, 140 mM NaCl), 500 μl Wash Buffer B (50 mM HEPES pH 7.9, 0.1% SDS, 1% Triton X-100, 0.1% deoxycholate, 1 mM EDTA, 500 mM NaCl), and 500 μl TE buffer (10 mM Tris-HCL pH 8.0, 1 mM EDTA). DNA was eluted in 100 μl elution buffer (50 mM Tris-HCL pH 8.0, 1 mM EDTA, 1% SDS) after incubation at 65 °C for 10 min at 1400 rpm. The eluent 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. These 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, Thermo Fisher) was added, and the samples were incubated 30 min at 37 °C to degrade RNA. Hundred micrograms of Proteinase K was then added and the samples were incubated 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 M220 and libraries were generated according to manufacturers' instructions.

### ChIP-seq analyses

**Alignment and peak calling.** Fastq files were processed exactly as in ATAC-seq to generate BAM files, normalised bigwig files and identify ChIP peaks.

**Seurat-guided clustering of ZIC2 peaks.** H3K27me3, H3K27ac and ATAC reads in naïve and primed hESCs corresponding to 2 kb bins around ZIC2 peaks, were computed using bedtools multicov. A matrix was generated such that each column represented a ZIC2 peak, and each of six rows represented reads corresponding to a given ATAC/histone PTM signal in a cell type. Values were scaled and centred and the *k*-nearest neighbours approach was used to identify Louvain clusters, using pre-built functions from the Seurat package. All visualisations were made using the FeaturePlot function.

**Annotation of ZIC2 peaks to chromatin states.** H3K4me1, H3K4me3, H3K27ac, H3K9me3 and H3K27me3 ChIP-seq datasets, with their corresponding inputs, in naïve and primed hESCs from published literature were used to define chromatin states. BAM files were binarized using the BinarizeBam command in ChromHMM with a bin size of 200 bp and Poisson threshold of $10^{-8}$. Clustering of chromatin states in each cell type into 16 states was carried out concurrently with the LearnModel command. Next, the emissions file was used to collapse the 16 states into 8 biologically meaningful chromatin states. To assign ZIC2 peaks with a chromatin state in primed hESC, bedtools intersect was used—a minimum of 40% overlap was required for a ZIC2 peak to correspond to a chromatin state. In cases of redundancy, a hierarchy was maintained (in decreasing priority - bivalent promoter, active promoter, bivalent enhancer, active enhancer, polycomb repressed, constitutive heterochromatin, mixed signal, background), with bivalent promoter having the highest priority. The chromatin state of naïve hESCs which are bound by ZIC2 in primed hESCs was similarly annotated. The outputs were used with bedtools intersect to assign how naïve chromatin states transit as they go on to become bound by ZIC2 in primed hESCs.

### CUT&Tag

**Library preparation.** CUT&Tag was carried out on formaldehyde-fixed hESCs. Briefly, 100,000 cells were fixed in 0.1% paraformaldehyde for 10 min at room temperature and the reaction was quenched with 0.14 M glycine. Cell pellets were resuspended in 200 µl cold nuclear extraction (NE) buffer (20 mM HEPES pH 7.9, 10 mM KCl, 20% glycerol, 0.1% Triton X-100, 0.5 mM spermidine, 0.15 mM spermine and protease inhibitors) and incubated for 5 min on ice. Nuclei were resuspended in 50 µl NE buffer and incubated with activated Concanavalin A-coated magnetic beads for 10 min at room temperature. Supernatant was removed using a magnetic stand and nuclei resuspended in 25 µL antibody buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 0.15 mM spermine, 20% glycerol and 2 mM EDTA) containing 0.5 µg primary antibody. Samples were then incubated overnight at 4 °C on a nutator. To remove unbound antibody, bead-nuclei complexes were washed twice with 150 buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 0.15 mM spermine, 20% glycerol) using a magnetic stand and resuspended in 25 µL 150 buffer containing 0.5 µg anti-mouse or rabbit secondary antibody for CUTANA™ CUT&Tag workflows (EpiCypher). Samples were then incubated for 30 min at room temperature. After two washes with 150 buffer, bead/nuclei complexes were resuspended in 12 µl 1x pA/G-Tn5 pre-loaded adapter complex (EpiCypher, USA) made in 150 buffer and incubated for 1 h at room temperature on a nutator. Bead/nuclei complexes were washed twice in 150 buffer, resuspended in 50 µl cold CUTAC-hex buffer (10 mM TAPS, 5 mM $MgCl_2$, 10% 1,6-hexanediol, 20% glycerol and protease inhibitors) and incubated for 2 min at 4 °C followed by 20 min at 37 °C. To release tagmented DNA, bead/nuclei complexes were then resuspended in 5 µl 0.1% SDS and incubated for 1 h at 58 °C. To generate libraries, 15 µl of 0.67% Triton X-100 neutralisation solution was added to each sample followed by 2 µl of 10 µM universal or barcoded i5 primers, 2 µl of 10 µM uniquely barcoded i7 primers and 25 µl NEBNext High-Fidelity buffer (non-hot-start NEBNext formulation; NEB, USA). Samples were then placed in a thermocycler under the following cycling conditions: 5 min at 58 °C, 5 min at 72 °C, 45 s at 98 °C, 15 s at 98 °C (melting) and 10 s at 60 °C (annealing and extension). The last two steps were repeated for a total of 14–21 cycles followed by a 1-min final extension at 72 °C. To purify libraries, 25 µl AMPure XP beads (Beckman Coulter, Canada) were added to the reactions and samples were incubated for 10 min at room temperature on a nutator. DNA was washed twice with 70% ethanol using a magnetic stand and once with 100% ethanol. DNA was eluted in 15 µl in pre-warmed 10 mM Tris pH 8 buffer using a magnetic stand.

**Analysis.** Reads were trimmed and processed, as described above for ATAC-seq. In cases of high mapping to mitochondrial DNA, reads corresponding to chrM were removed prior to analysis.

### Transcriptome analyses

Cell pellets were lysed with 150 µl RNAzol RT (Sigma Aldrich) and total RNA was purified according to manufacturer's instructions.

**qRT-PCR.** Total RNA (500 ng) was used to generate cDNA using the SensiFast cDNA Synthesis Kit (Bioline); PCR conditions used were unchanged from the manufacturer's technical bulletin. The resulting cDNA was diluted fivefold using nuclease-free water. Each 3 µl qPCR reaction was carried out in duplicates using 1x PowerUp SYBR master mix (Thermo Fisher), gene-specific primers (Supplementary Data 10) and diluted cDNA solution and finally, quantified on the QuantStudio 5 instrument. *RAB7A* was used as an internal control unless mentioned otherwise.

**RNA-seq.** mRNA was enriched from 100–500 ng total RNA using NEBNext Poly(A) mRNA Magnetic Isolation Module Kit and libraries were generated using Swift RNA Library Kit or NEBNext Ultra Directional RNA library kit.

**RNA-seq analyses.** Fastq files were aligned against the GRCh38 reference genome with default parameters. Gencode v34 gtf file was used to define features to generate raw or normalised counts. HTSeq was used to calculate raw counts for each gene. DESeq2 was used to do differential analyses between different conditions. Normalised counts, reported as TPM (Supplementary Data 3), were calculated with StringTie. In cases, where *z*-scores were used, raw counts were normalised using DEseq2's median of ratios method and log transformed, where appropriate, prior to analyses.

### Flow cytometry and FACS

Cells were dissociated into single cells with 30% TrypLE as described elsewhere in this study and then incubated with fluorophore-conjugated antibodies, designed against cell surface markers, diluted in 1% KSR, for 30 min at 4 °C. Following incubation, cell suspensions were washed with PBS and finally resuspended with 1% KSR, supplemented with DAPI counterstain. Unless otherwise stated, all analyses were done on live, single cells and contaminating MEFs were removed prior to data acquisition. Flow cytometry measurements were recorded on the LSR Fortessa instrument and analysed on FlowJo v10. FACS isolation of cells was carried out on the BD FacsARIA Fusion instrument.

**Western blot.** Cells were harvested as described elsewhere in the study with 30% TrypLE. Cell pellets were lysed with ice-cold RIPA lysis buffer supplemented with 1x Halt Protease and Phosphatase

inhibitor (Thermo Fisher). To complete lysis, the resulting suspension was sonicated using the Covaris M220 sonicator (DF = 20%, CPB = 200, PIP = 75 W, temperature = 7 °C). Debris and chromatin were cleared by centrifugation and protein concentrations were determined by Bradford assay. Equal protein amounts across all samples were aliquoted and boiled in 1X Laemmli buffer at 95 °C, loaded onto and separated by SDS PAGE. Proteins were transferred onto Immobilon PVDF-FL membranes and blotted with required primary antibodies diluted in Odyssey Blocking buffer (LI-COR) overnight at 4 °C. Membranes were then blotted with fluorophore-conjugated secondary antibodies for 1 h. Membranes were then visualised using the Odyssey Imaging system. Details of antibodies used are described in Supplementary Data 10.

**Co-immunoprecipitation.** For experiments involving pulldown of ZIC2, 10–20 million cells were collected and washed once with cold PBS. Cells were then lysed with 1x packed cell volume (PCV) of co-IP lysis buffer (50 mM KCl, 20 mM HEPES pH 7.5, 0.5% CHAPS) supplemented with 1x complete protease inhibitor (Sigma), 1 mM sodium orthovanadate, 5 mM NaF, 5 mM beta glycerophosphate and 2 mM EDTA. The lysate was incubated on ice for an hour with mild shaking. 3x PCV PBS, supplemented with 1x complete protease inhibitor (Sigma), 1 mM sodium orthovanadate, 5 mM NaF, 5 mM beta glycerophosphate and 2 mM EDTA, was added and cell debris was removed by centrifugation at 16,000 × *g* for 5 min at 4 °C. Antibodies were pre-incubated with SureBeads (BioRad) in PBS for 20 min, and washed once with PBS, then with 2% BSA diluted in PBS for 30 min and finally washed twice with PBS, following which lysates were added to antibody-conjugated beads and incubated overnight, at 4 °C with end-to-end rotation. Beads were then washed three times with ice-cold wash buffer (2 M urea, 0.1% CHAPS diluted in PBS) and the immunoprecipitated material was eluted with 1x LDS sample buffer supplemented with 10% B-Me at 70 °C for 20 min. Western blot was carried out as described above. Blots were imaged with chemiluminescence using anti-mouse secondary antibody conjugated to HRP for BRG1 and VeriBlot for ZIC2. Reciprocal co-IP experiments were carried using the same protocol, but instead the SureBeads were crosslinked to primary antibodies with EGS to minimise elution of denatured antibody.

**Analysis of transposable elements.** For studies including transposable elements from short-read sequencing data, multimapping reads were not removed prior to analyses.

Briefly, for RNA-seq, default settings were used during mapping with STAR aligner with the following exceptions (--outFilterMultimapNmax 5000 --outMultimapperOrder Random --winAnchorMultimapNmax 5000). Raw counts corresponding to transposable elements were calculated using TEtranscripts[68], which was used to identify transposable elements with differential expression using DESeq2.

For ATAC-seq, default settings were used during mapping with BWA MEM. However, reads corresponding to multiple locations were not removed prior to analysis, during filtering.

**Immunofluorescence.** Cells plated on coverslips were incubated with 4% PFA for 20 min at room temperature. Coverslips were washed with PBS three times and later, permeabilized with 0.3% Triton X-100 for 10 min. Next, coverslips were rinsed with 0.8% glycine and blocked with IF buffer (0.2% Triton X-100, 0.05% Tween-20 and 0.5% BSA in PBS), followed by incubation in primary antibodies for 1 h at room temperature and secondary antibodies for 40 min. Washes were carried out before and after incubations with antibodies, with IF buffer. DAPI and phalloidin were used as counterstains where indicated. Finally, all coverslips were washed with MILIQ water and mounted onto glass sides. Images were captured using the EVOS M5000 imaging system unless otherwise stated. Details of antibodies used and dilutions are described in Supplementary Data 10.

**Luciferase reporter assay.** In total, 200,000 primed hESCs were seeded as single cells in media supplemented with 10 μM Y-27632 24 h before transfection. Cells were co-transfected with 400 ng pGL3-promoter and 100 ng pRL-TK plasmids using Lipofectamine Stem Transfection reagent (Thermo Fisher) according to manufacturer's instructions. Cells were harvested and lysed with 50 μl 1X Passive lysis buffer (Promega). Debris was removed from lysate by centrifugation at 15,000 × *g* for 10 min. For each sample, 100 μl firefly assay buffer was added to 20 μl lysate and readings for *Firefly* activity were taken using a luminometer. Subsequently, readings for *Renilla* activity were taken after the addition of 100 μl renilla assay buffer. Both assay buffers were made using a homebrew protocol.

**Gene set enrichment analysis (GSEA).** GSEA was conducted in *ZIC2*$^{-/-}$*ZIC3*$^{-/-}$ and WT control primed hESCs over genes de-repressed upon acute inhibition of EZH2 in primed hESCs[8], with default parameters, including number of permutations: 1000, type of permutation: phenotype, metric for ranking genes: Signal2Noise and enrichment statistic: weighted.

**Predictor of natural disordered regions (PONDR).** PONDR score was calculated by using the amino acid sequence of the mostly highly expressed isoform of *ZIC2*, *ZIC3* and *ZIC5*, using the VL-XT algorithm[69]. A PONDR score >0.5 was used as a cut-off for regions considered to be disordered.

**Homology of amino acid sequences.** Amino acid sequences of the most highly expressed isoform of each gene were used for analysis with the ClustalW program[70], with default parameters.

**Genome browser tracks.** All values used in comparative analysis are expressed as normalised counts per bin (CPB). Otherwise, reads per 1x genome coverage (RPGC) is used. RNA-seq values are expressed in a log scale.

**Data visualisation.** Most figures were generated using GraphPad Prism and ggplot2, unless otherwise stated. Metaplots were generated using deepTools. Sankey plots were generated using SankeyMATIC (https://sankeymatic.com/).

**Alteration to images.** Brightness and contrast were uniformly altered for microscopy images where necessary to improve image clarity.

### Reporting summary

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

## Data availability

All RNA-seq, ChIP-seq, CUT&Tag and ATAC-seq data have been deposited in the Gene Expression Omnibus database under the accession number GSE270787. Mapping statistics for libraries generated in this manuscript are shown in Supplementary Data 12–15. Previously published datasets used in this study include GSE76970, GSE147750, GSE152101, GSE138012, GSE92625, GSE101074, GSE181242, GSE123055, GSE138012, GSE138688, GSE153212, GSE164786, GSE176172, GSE167979, GSE69646, GSE59435, GSE75868, GSE99627, GSE76626, GSE32465, GSE24447, GSE147751, GSE210859, GSE64758 and GSE17312. Use of each published dataset is described in Supplementary Data 11. The data supporting the findings of this study are available from the corresponding authors upon request. Source data for the figures and Supplementary Figs. are provided as a Source Data file. Source data are provided with this paper.

## References


1. Gifford, C. A. et al. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* **153**, 1149–1163 (2013).
2. Tsankov, A. M. et al. Transcription factor binding dynamics during human ES cell differentiation. *Nature* **518**, 344–349 (2015).
3. Creyghton, M. P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* **107**, 21931–21936 (2010).
4. Pastor, W. A. et al. TFAP2C regulates transcription in human naive pluripotency by opening enhancers. *Nat. Cell Biol.* **20**, 553–564 (2018).
5. 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).
6. Theunissen, T. W. et al. Molecular criteria for defining the naive human pluripotent state. *Cell Stem Cell* **19**, 502–515 (2016).
7. Zijlmans, D. W. et al. Integrated multi-omics reveal polycomb repressive complex 2 restricts human trophoblast induction. *Nat. Cell Biol.* **24**, 858–871 (2022).
8. Kumar, B. et al. Polycomb repressive complex 2 shields naive human pluripotent cells from trophectoderm differentiation. *Nat. Cell Biol.* **24**, 845–857 (2022).
9. Sahakyan, A. et al. Human naive pluripotent stem cells model X chromosome dampening and X inactivation. *Cell Stem Cell* **20**, 87–101 (2017).
10. Grow, E. J. et al. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522**, 221–225 (2015).
11. Pontis, J. et al. Hominoid-specific transposable elements and KZFPs facilitate human embryonic genome activation and control transcription in naive human ESCs. *Cell Stem Cell* **24**, 724–735 e725 (2019).
12. Theunissen, T. W. et al. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* **15**, 524–526 (2014).
13. Fuentes, D. R., Swigut, T. & Wysocka, J. Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. *Elife* **7**, e35989 (2018).
14. Seong, J. et al. Epiblast inducers capture mouse trophectoderm stem cells in vitro and pattern blastoids for implantation in utero. *Cell Stem Cell* **29**, 1102–1118.e1108 (2022).
15. Rostovskaya, M., Stirparo, G. G. & Smith, A. Capacitation of human naive pluripotent stem cells for multi-lineage differentiation. *Development* **146**, dev172916 (2019).
16. Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4**, 487–492 (2009).
17. Takashima, Y. et al. Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* **158**, 1254–1269 (2014).
18. Nakamura, T. et al. A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature* **537**, 57–62 (2016).
19. Xiang, L. et al. A developmental landscape of 3D-cultured human pre-gastrulation embryos. *Nature* **577**, 537–542 (2020).
20. Tyser, R. C. V. et al. Single-cell transcriptomic characterization of a gastrulating human embryo. *Nature* **600**, 285–289 (2021).
21. Zheng, Y. et al. Controlled modelling of human epiblast and amnion development using stem cells. *Nature* **573**, 421–425 (2019).
22. Warr, N. et al. Zic2-associated holoprosencephaly is caused by a transient defect in the organizer region during gastrulation. *Hum. Mol. Genet.* **17**, 2986–2996 (2008).
23. Dykes, I. M. et al. A requirement for Zic2 in the regulation of nodal expression underlies the establishment of left-sided identity. *Sci. Rep.* **8**, 10439 (2018).
24. Ware, S. M., Harutyunyan, K. G. & Belmont, J. W. Zic3 is critical for early embryonic patterning during gastrulation. *Dev. Dyn.* **235**, 776–785 (2006).
25. Bellchambers, H. M. & Ware, S. M. Loss of Zic3 impairs planar cell polarity leading to abnormal left-right signaling, heart defects and neural tube defects. *Hum. Mol. Genet.* **30**, 2402–2415 (2021).
26. Inoue, T. et al. Mouse Zic5 deficiency results in neural tube defects and hypoplasia of cephalic neural crest derivatives. *Dev. Biol.* **270**, 146–162 (2004).
27. Luo, Z. et al. Zic2 is an enhancer-binding factor required for embryonic stem cell specification. *Mol. Cell* **57**, 685–694 (2015).
28. Yang, S. H. et al. ZIC3 controls the transition from naive to primed pluripotency. *Cell Rep.* **27**, 3215–3227.e3216 (2019).
29. Xu, J. et al. Genome-wide CRISPR screen identifies ZIC2 as an essential gene that controls the cell fate of early mesodermal precursors to human heart progenitors. *Stem Cells* **38**, 741–755 (2020).
30. Lin, X. et al. Mesp1 controls the chromatin and enhancer landscapes essential for spatiotemporal patterning of early cardiovascular progenitors. *Nat. Cell Biol.* **24**, 1114–1128 (2022).
31. Koyabu, Y., Nakata, K., Mizugishi, K., Aruga, J. & Mikoshiba, K. Physical and functional interactions between Zic and Gli proteins. *J. Biol. Chem.* **276**, 6889–6892 (2001).
32. Linneberg-Agerholm, M. et al. Naive human pluripotent stem cells respond to Wnt, Nodal, and LIF signalling to produce expandable naive extra-embryonic endoderm. *Development* **146**, dev180620 (2019).
33. Dong, C. et al. Derivation of trophoblast stem cells from naive human pluripotent stem cells. *Elife* **9**, e52504 (2020).
34. Collier, A. J. et al. Comprehensive cell surface protein profiling identifies specific markers of human naive and primed pluripotent states. *Cell Stem Cell* **20**, 874–890.e877 (2017).
35. Huang, X. et al. OCT4 cooperates with distinct ATP-dependent chromatin remodelers in naive and primed pluripotent states in human. *Nat. Commun.* **12**, 5123 (2021).
36. Ji, X. et al. 3D chromosome regulatory landscape of human pluripotent cells. *Cell Stem Cell* **18**, 262–275 (2016).
37. Barakat, T. S. et al. Functional dissection of the enhancer repertoire in human embryonic stem cells. *Cell Stem Cell* **23**, 276–288.e278 (2018).
38. Yilmaz, A., Peretz, M., Aharony, A., Sagi, I. & Benvenisty, N. Defining essential genes for human pluripotent stem cells by CRISPR-Cas9 screening in haploid cells. *Nat. Cell Biol.* **20**, 610–619 (2018).
39. Martin, B. L. & Kimelman, D. Brachyury establishes the embryonic mesodermal progenitor niche. *Genes Dev.* **24**, 2778–2783 (2010).
40. Shan, Y. et al. PRC2 specifies ectoderm lineages and maintains pluripotency in primed but not naive ESCs. *Nat. Commun.* **8**, 672 (2017).
41. Court, F. & Arnaud, P. An annotated list of bivalent chromatin regions in human ES cells: a new tool for cancer epigenetic research. *Oncotarget* **8**, 4110–4124 (2017).
42. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
43. Dahlet, T. et al. Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity. *Nat. Commun.* **11**, 3153 (2020).
44. Bredenkamp, N., Stirparo, G. G., Nichols, J., Smith, A. & Guo, G. The cell-surface marker sushi containing domain 2 facilitates establishment of human naive pluripotent stem cells. *Stem Cell Rep.* **12**, 1212–1222 (2019).
45. Collier, A. J. et al. Genome-wide screening identifies Polycomb repressive complex 1.3 as an essential regulator of human naive pluripotent cell reprogramming. *Sci. Adv.* **8**, eabk0013 (2022).
46. Cinkornpumin, J. K. et al. Naive human embryonic stem cells can give rise to cells with a trophoblast-like transcriptome and methylome. *Stem Cell Rep.* **15**, 198–213 (2020).

47. Guo, G. et al. Human naive epiblast cells possess unrestricted lineage potential. *Cell Stem Cell* **28**, 1040–1056.e1046 (2021).
48. Kinoshita, M. et al. Capture of mouse and human stem cells with features of formative pluripotency. *Cell Stem Cell* **28**, 2180 (2021).
49. Farnaby, W. et al. BAF complex vulnerabilities in cancer demonstrated via structure-based PROTAC design. *Nat. Chem. Biol.* **15**, 672–680 (2019).
50. Brahma, S. & Henikoff, S. The BAF chromatin remodeler synergizes with RNA polymerase II and transcription factors to evict nucleosomes. *Nat. Genet.* **56**, 100–111 (2024).
51. Estaras, C., Benner, C. & Jones, K. A. SMADs and YAP compete to control elongation of beta-catenin:LEF-1-recruited RNAPII during hESC differentiation. *Mol. Cell* **58**, 780–793 (2015).
52. Dhalluin, C. et al. Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**, 491–496 (1999).
53. Sun, T. et al. Crosstalk between RNA m(6)A and DNA methylation regulates transposable element chromatin activation and cell fate in human pluripotent stem cells. *Nat. Genet.* **55**, 1324–1335 (2023).
54. Vispe, S. et al. Triptolide is an inhibitor of RNA polymerase I and II-dependent transcription leading predominantly to down-regulation of short-lived mRNA. *Mol. Cancer Ther.* **8**, 2780–2790 (2009).
55. Hsieh, T. S. et al. Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol. Cell* **78**, 539–553.e538 (2020).
56. Liu, C. et al. Absolute quantification of single-base m(6)A methylation in the mammalian transcriptome using GLORI. *Nat. Biotechnol.* **41**, 355–366 (2023).
57. Lim, L. S. et al. Zic3 is required for maintenance of pluripotency in embryonic stem cells. *Mol. Biol. Cell* **18**, 1348–1358 (2007).
58. Ahmed, J. N., Diamand, K. E. M., Bellchambers, H. M. & Arkell, R. M. Systematized reporter assays reveal ZIC protein regulatory abilities are subclass-specific and dependent upon transcription factor binding site context. *Sci. Rep.* **10**, 13130 (2020).
59. Zaret, K. S. & Mango, S. E. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr. Opin. Genet. Dev.* **37**, 76–81 (2016).
60. Mayran, A. et al. Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. *Nat. Genet.* **50**, 259–269 (2018).
61. Ohtani, K. et al. Jmjd3 controls mesodermal and cardiovascular differentiation of embryonic stem cells. *Circ. Res.* **113**, 856–862 (2013).
62. Wang, Z. et al. Prediction of histone post-translational modification patterns based on nascent transcription data. *Nat. Genet.* **54**, 295–305 (2022).
63. Iurlaro, M. et al. Mammalian SWI/SNF continuously restores local accessibility to chromatin. *Nat. Genet.* **53**, 279–287 (2021).
64. Barshad, G. et al. RNA polymerase II dynamics shape enhancer-promoter interactions. *Nat. Genet* **55**, 1370–1380 (2023).
65. Zhang, S., Ubelmesser, N., Barbieri, M. & Papantonis, A. Enhancer-promoter contact formation requires RNAPII and antagonizes loop extrusion. *Nat. Genet.* **55**, 832–840 (2023).
66. 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).
67. Graham-Paquin, A. L. et al. ZMYM2 is essential for methylation of germline genes and active transposons in embryonic development. *Nucleic Acids Res.* **51**, 7314–7329 (2023).
68. Jin, Y., Tam, O. H., Paniagua, E. & Hammell, M. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31**, 3593–3599 (2015).
69. Peng, K., Radivojac, P., Vucetic, S., Dunker, A. K. & Obradovic, Z. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinform.* **7**, 208 (2006).
70. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).

## Acknowledgements

The research was funded by the NSERC Grant RGPIN-2018-04856 and the CIHR Project Grant PJT-166169 and to W.A.P. and by the CIHR Project Grant PJT-173443 to J.A.L. I.H. was a recipient of the Graduate Excellence Fellowship and J.P. Collip Fellowship from McGill University. W.A.P. was supported by an FRQS Chercheurs-boursier. The authors would like to thank Dr. Amander Clark for sharing UCLA1 hESCs and Dr. Thomas Durcan for sharing AJC002 hiPSCs; Dr. Jessica K. Cinkornpumin for technical assistance on growing hTSCs; Anita Wang for preliminary work on this manuscript. We would like to thank the Centre for Applied Genomics at SickKids Hospital, Genome Quebec, Canada's Michael Smith Genome Sciences Centre at BC Cancer and Genomics Platform at IRIC for next-generation sequencing. We would like to acknowledge the Flow Cytometry Core Facility at McGill University for dedicated service.

## Author contributions

I.H., P.P., S.C.R, X.X.Z., L.C., J.D., C.C. and T.F. conducted experiments. I.H., S.C.R. and S.S. conducted bioinformatic analyses. I.H. and W.A.P. wrote the manuscript. J.A.L. and W.A.P. supervised the research.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-52431-1.

**Correspondence** and requests for materials should be addressed to William A. Pastor.

**Peer review information** *Nature Communications* thanks Peter Rugg-Gunn and the other, anonymous, reviewer for their contribution to the peer review of this work. A peer review file is available.