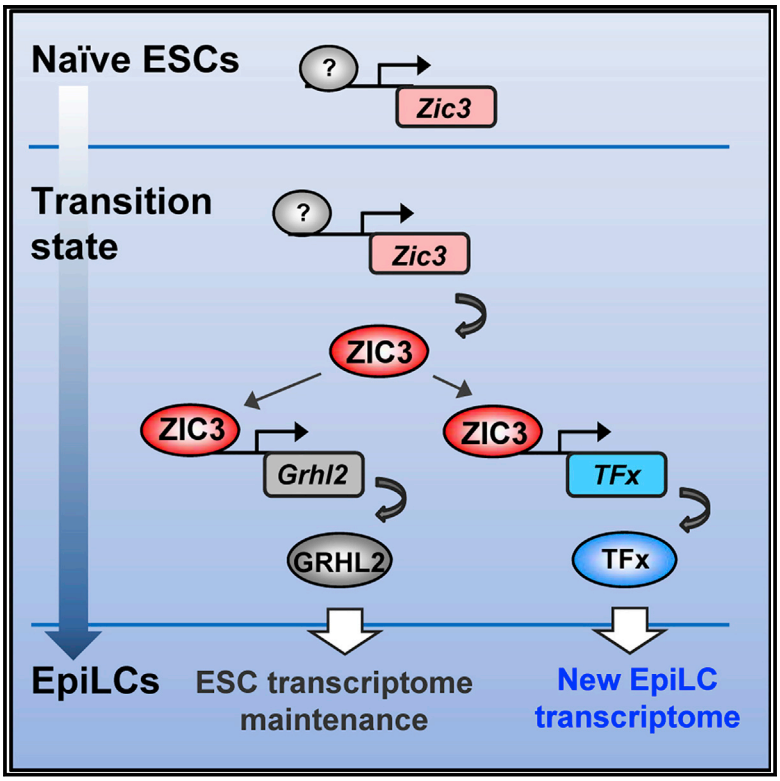
Article

Cell Reports

ZIC3 Controls the Transition from Naive to Primed Pluripotency

Graphical Abstract



Authors

Shen-Hsi Yang, Munazah Andrabi, Rebecca Biss, Syed Murtuza Baker, Mudassar Iqbal, Andrew D. Sharrocks

Correspondence

s.yang@manchester.ac.uk (S.-H.Y.), andrew.d.sharrocks@manchester.ac.uk (A.D.S.)

In Brief

We know relatively little about how the early steps of differentiation from naive embryonic stem cells (ESCs) are controlled. Here, Yang et al. profile the chromatin accessibility changes during the transition of ESCs to epiblast-like cells and identify the transcription factor ZIC3 as an important regulator of this transition.

Highlights

- Transcription factor ZIC3 regulates gene expression during the ESC to EpiLC transition
- Extensive changes occur in the open chromatin landscape as ESCs progress to EpiLCs
- ZIC3 activates the expression of a network of transcription factors
- ZIC3-activated genes in EpiLCs are upregulated in the post-implantation epiblast

Yang et al., 2019, Cell Reports *27*, 3215–3227
June 11, 2019
https://doi.org/10.1016/j.celrep.2019.05.026

CellPress

Cell Reports
Article

ZIC3 Controls the Transition from Naive to Primed Pluripotency

Shen-Hsi Yang,[1,*] Munazah Andrabi,[1] Rebecca Biss,[1] Syed Murtuza Baker,[1] Mudassar Iqbal,[1] and Andrew D. Sharrocks[1,2,*]
[1]Faculty of Biology, Medicine and Health, University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, UK
[2]Lead Contact
*Correspondence: s.yang@manchester.ac.uk (S.-H.Y.), andrew.d.sharrocks@manchester.ac.uk (A.D.S.)
https://doi.org/10.1016/j.celrep.2019.05.026

SUMMARY

Embryonic stem cells (ESCs) must transition through a series of intermediate cell states before becoming terminally differentiated. Here, we investigated the early events in this transition by determining the changes in the open chromatin landscape as naive mouse ESCs transition to epiblast-like cells (EpiLCs). Motif enrichment analysis of the newly opening regions coupled with expression analysis identified ZIC3 as a potential regulator of this cell fate transition. Chromatin binding and genome-wide transcriptional profiling following Zic3 depletion confirmed ZIC3 as an important regulatory transcription factor, and among its targets are genes encoding a number of transcription factors. Among these is GRHL2, which acts through enhancer switching to maintain the expression of a subset of genes from the ESC state. Our data therefore place ZIC3 upstream of a set of pro-differentiation transcriptional regulators and provide an important advance in our understanding of the regulatory factors governing the early steps in ESC differentiation.

INTRODUCTION

Early embryonic development involves the transition of pluripotent embryonic stem cells through intermediate cell states into the cell lineages that initiate subsequent development events. Using defined *in vitro* conditions, several different states have been identified for mouse embryonic stem cells (ESCs), starting from the naive ground state and progressing through epiblast-like cells (EpiLCs), to establish an epiblast stem cell (EpiSC) state (Hayashi et al., 2011; reviewed in Kalkan and Smith, 2014). Subsequently, EpiSCs can differentiate into the three germ layers: mesoderm, ectoderm, and endoderm. Mouse ESCs can be maintained in the naive ground state in defined media, which includes two kinase inhibitors (known as "2i") to block the MEK/ERK and GSK3 signaling pathways (Ying et al., 2008; reviewed in Wray et al., 2010). Withdrawal of 2i, allows the cells to progress to either EpiLCs or EpiSCs by altering culture conditions (Betschinger et al., 2013; Hayashi et al., 2011). The naive ESCs are thought to represent a model for the pre-implantation epiblast (embryonic 3.5 [E3.5]–4.5) whereas EpiLCs or EpiSCs cells are models for the post-implantation epiblast (E5.5) (Kalkan et al., 2017).

As ESCs progress from the naive ground state, large changes are observed in their chromatin landscapes and underlying gene expression programs (Marks et al., 2012; Factor et al., 2014; reviewed in Habibi and Stunnenberg, 2017). The pluripotent state is maintained through the action of a core set of transcription factors and chromatin regulators that include the well-studied NANOG, KLF4, SOX2, and OCT4 (reviewed in Young, 2011). However, comparatively less is known about the regulators controlling the transition to EpiLCs and EpiSCs. Recently, OTX2 was identified as a key transcription factor driving this transition, partly through cooperative interactions with OCT4/POU5F1 (Acampora et al., 2013; Buecker et al., 2014; Yang et al., 2014). Proteomics analysis also identified ZIC2/3 and OCT6/POU3F1 as interacting proteins for OCT4, specifically in EpiLCs (Buecker et al., 2014), suggesting a potential co-regulatory role for these transcription factors in this context. Further changes occur during the transition to EpiLCs, and in addition to transcriptional regulators, other proteins have been shown to play an important role during this transition such as the extracellular signaling protein, Cripto, which controls metabolic reprogramming (Fiorenzano et al., 2016).

To further our understanding of the regulatory networks controlling the transition from the naive ESC state to EpiLCs, we examined the chromatin accessibility changes accompanying this early transition in mouse ESCs. We focused on areas of dynamic chromatin opening and through DNA binding motif enrichment and associated gene expression data analysis, we identified the transcription factor ZIC3 as an important regulatory transcription factor in this context. ZIC3 controls the expression of EpiLC marker genes such as *Fgf5* and many of the ZIC3 target genes encode transcriptional regulators such as GRHL2, which has an important role in enhancer formation in the transition to EpiLCs. ZIC3 therefore is immediately upstream of a set of pro-differentiation regulators that work together to establish the EpiLC state.

RESULTS

Identification of Transcription Factors Involved in the Transition to EpiLCs through Open Chromatin Profiling

Cell state transitions are accompanied by changes to the underlying regulatory chromatin landscape (Stergachis et al., 2013).

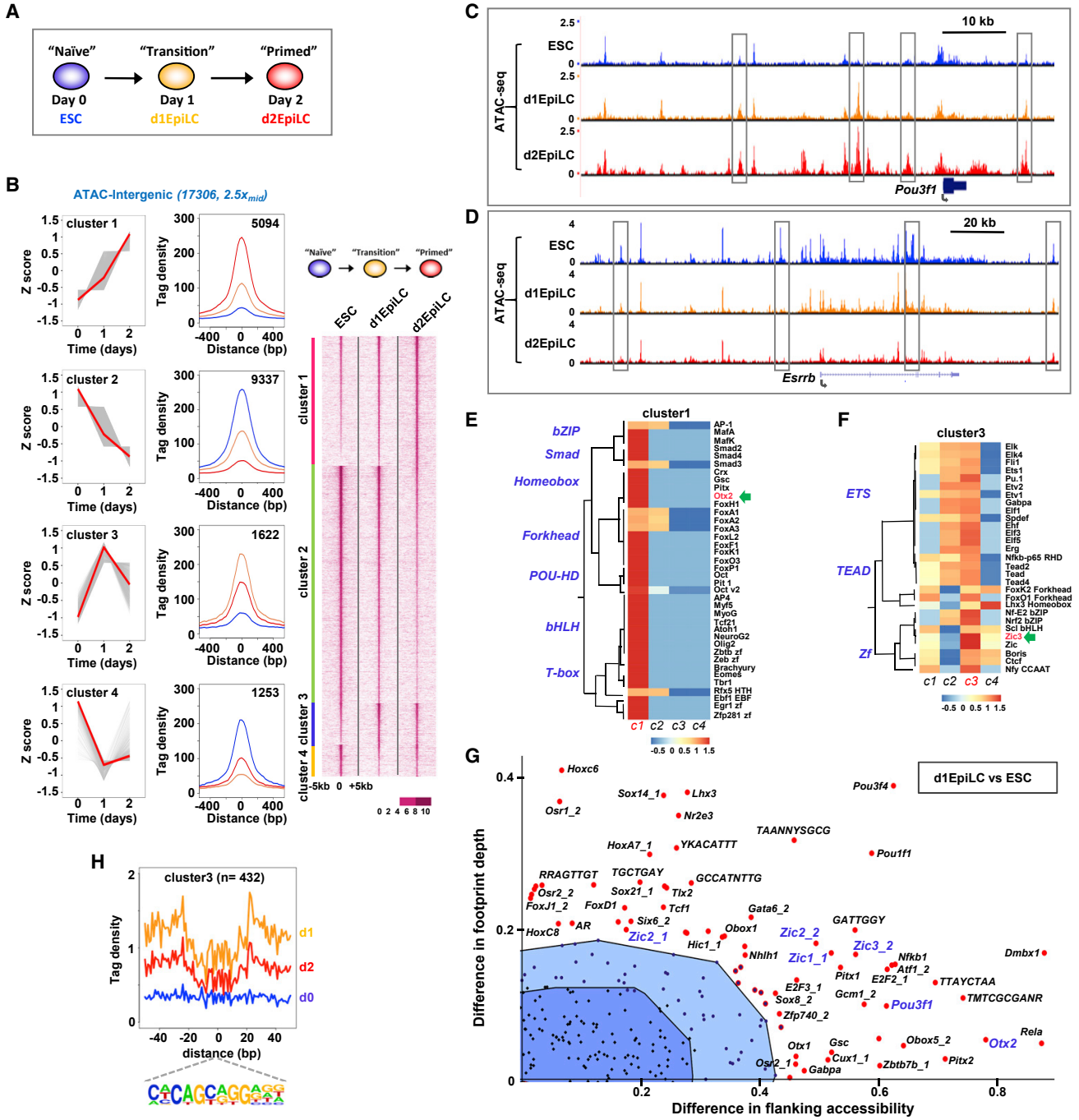


Figure 1. Identification of Transcriptional Regulators of the ESC to EpiLC Transition by Open Chromatin Profiling

(A) Schematic of the experimental time course of the naive ESC to EpiLC transition.

(B) Heatmap of the ATAC-seq profiles across a 10-kb window of intergenic regions showing > 2.5-fold change in accessibility between any two conditions (right). Average tag densities of each of four identified clusters (middle; blue = ESC, orange = d1EpiLCs, red = d2EpilCs) and average tag density profiles (z scored) are shown across the time course (left). Medians (red) and data for individual peaks (gray) are indicated.

(C and D) University of Santa Cruz (UCSC) genome browser views of the ATAC-seq profiles around the *Pou3f1* (C) and *Esrrb* (D) loci. Dynamically changing peaks are boxed.

(E and F) Heatmap showing the enrichment of transcription factor binding motifs across each of the open chromatin cluster profiles (z-normalized p values) for motifs enriched in cluster 1 (E) or cluster 3 (F).

(legend continued on next page)

These changes can then be used to infer the potential roles of upstream transcription factors (Sung et al., 2014). To begin to understand the regulatory events occurring during the conversion of naive mouse ESCs to EpiLCs, we therefore profiled the accessible chromatin landscape of mouse ESCs as they transition to EpiLCs over a 2-day period (Figure 1A) using assay for transposase-accessible chromatin using sequencing (ATAC-seq). Gene expression changes at matched time points were also profiled using single-cell RNA sequencing (scRNA-seq) from 816 cells. Biological replicates for ATAC-seq analysis were obtained for each time point, which showed high concordance (Figure S1A) and were therefore merged before peak calling for further analysis. Open chromatin regions were then identified at each time point and the resulting peaks consolidated into a single reference dataset (238,236 peaks in total). These peaks were then partitioned between promoter proximal (−2 to +0.5 kb), intragenic, and intergenic regions to examine whether genomic location affected the overall changes in chromatin accessibility. We then identified regions that showed differential accessibility between any two conditions, giving 3,041 (promoter), 16,510 (intragenic), and 17,306 (intergenic) differentially accessible regions. These regions were then clustered into four broad patterns based on their chromatin opening dynamics (Figures 1B, S1B, and S1C): regions that increased accessibility at day 1 and became further accessible at day 2 (cluster 1), regions that decreased accessibility at day 1 and became even more inaccessible at day 2 (cluster 2), or regions that transiently opened or closed at day 1 (clusters 3 and 4).

We next recovered the genes associated with each promoter and matched intergenic peaks to their likely associated genes by using the nearest-gene model. Then, we compared the changes in expression relative to the changes in open chromatin associated with each gene. By focusing on the differentially changing ATAC-seq peaks and gene expression changes, we observed a good concordance between chromatin opening and gene expression changes (Figures S1D and S1E). These changes become more marked after 2 days, as cells acquire the EpiLC state. The changes we observed in chromatin accessibility profiling therefore generally report on the activity status of associated genes. To further examine whether these different accessibility profiles reflected the underlying changes in gene expression, we clustered the gene expression changes into four similar patterns. We then took the open chromatin peaks in each cluster, associated them with all genes located at different peak-to-gene distances, and calculated the enrichment of the resulting set of genes among the equivalent clusters derived from scRNA-seq data. Overall, there was an excellent correlation between the two datasets, with the best matches occurring between similar cluster patterns (Figures S1F and S1G). This was particularly marked for intergenic regions in cluster 3, where there was very little enrichment with any other gene expression cluster other than cluster 3, irrespective of the distance to the transcription start site (TSS). Furthermore, each of the ATAC-seq clusters was associated with groups of genes exhibiting a unique set of Gene Ontology (GO) terms (Figures S2A–S2D). For example, the cluster 1 and 2 regions are associated with various developmental terms, as might be expected by their sequential changes in the transition to EpiLCs. The regulatory regions of genes encoding transcription factors associated with the two cell states show expected changes during the transition to EpiLCs; several peaks within the *Pou3f1* locus (EpiLC transcription factor) show sequential opening and are found in cluster 1 (Figure 1C). Conversely, peaks in the *Essrb* locus (ESC transcription factor) show progressive closing and are found in cluster 2 (Figure 1D). In contrast, cluster 4 genes are associated with various stem cell processes, consistent with transient regulatory region closing as illustrated by the *Nodal* locus (Figure S2F).

Having established the relevance of open chromatin profiling to gene expression changes, we next wanted to identify the relevant regulators. To that end, we searched the differentially accessible regions for over-represented transcription factor binding motifs. Each of the accessibility clusters has a different repertoire of motifs (Figures 1E, 1F, and S3A–S3F) with ZEB1, KLF4, ZIC3, and TCF3 being the most enriched binding motifs for clusters 1–4, respectively. Interestingly, OTX2 binding motifs were identified in cluster 1 regions, which is consistent with the fact that these regions become sequentially more open in the transition to EpiLCs and the known role for OTX2 in driving early ESC fate decisions (Yang et al., 2014; Buecker et al., 2014). We were particularly interested in cluster 3, as these regions are characterized by transient opening at day 1, suggesting an important role in the transition toward EpiLCs. To further interrogate the underlying transcription factor networks in this cluster, we used Bivariate Genomic Footprinting (BaGFoot) (Baek et al., 2017) to identify transcription factor motifs that exhibit increased footprint depth (and hence occupancy) and/or local DNA accessibility. Numerous motifs were identified, including those for several homeodomain and ZIC proteins (Figures 1G and S4A). ZIC binding motifs had previously been associated with OCT4/POU5F1 binding regions in EpiLCs (Buecker et al., 2014); therefore, we focused on this binding site and more closely examined the chromatin accessibility surrounding this motif at the three differentiation time points. Clear footprints were observed in open chromatin clusters 1 and 3 around this motif in day 1 (d1) EpiLCs and the depth and local accessibility mirrored the general accessibility profiles of these clusters across different time points (Figures 1H and S4B). We also compared the open chromatin of d2EpiLCs to ESCs and applied a similar analysis. Multiple motifs were again identified as becoming more accessible and potentially more occupied, including those for OTX2 and the ZIC transcription factors (Figure S5).

(G) BaGFoot analysis of the open chromatin regions in ESCs and d1EpiLCs (using all dynamic intergenic peaks from clusters 1 to 4 in B). The top-right quadrant from the whole plot (see Figure S4) is shown. Motifs showing significant increases in either local accessibility and/or footprint depth are labeled. Non-significant regions are shown in the dark (bag) or light (fence) blue-shaded regions.

(H) Average tag densities in a 100-bp window surrounding the ZIC3 binding motif (bottom) in ESCs (blue), d1EpiLCs (orange), or d2EpilCs (red) are shown for ATAC-seq peaks from cluster 3.

See also Figures S1–S5.

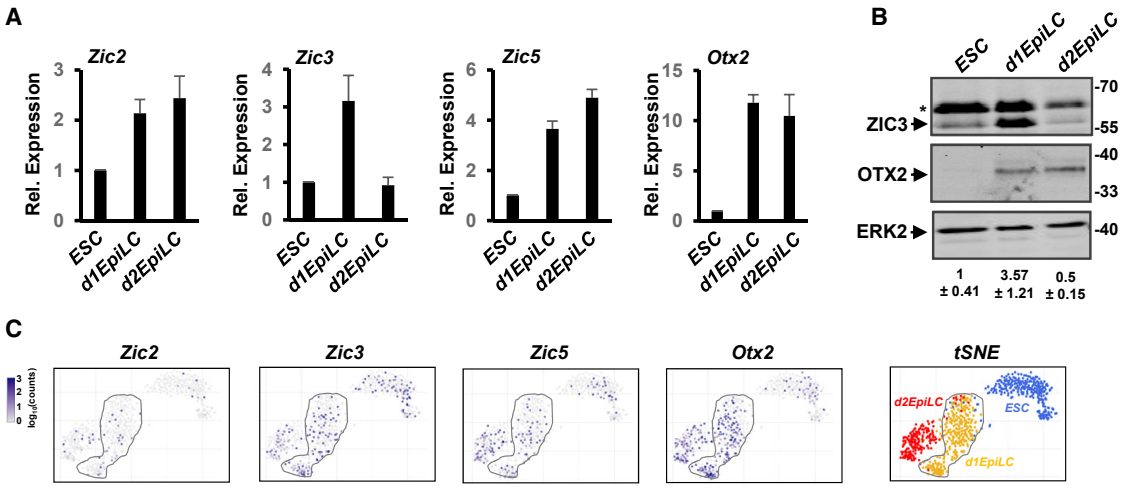


Figure 2. Expression Profiles of Zic Transcription Factors in ESCs and EpiLCs

(A) qRT-PCR analysis of *Zic2*, *Zic3*, *Zic5*, and *Otx2* expression in the indicated cell states (n = 3).

(B) Western blot analysis of ZIC3, OTX2, and ERK2 expression. The asterisk marks a non-specific band. Quantification of ZIC3 protein levels is shown below each lane (n = 3).

(C) The scRNA-seq analysis of *Zic2*, *Zic3*, *Zic5*, and *Otx2* expression. t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis of the entire scRNA-seq dataset is shown on the right, with the originating cell types color coded. d1EpiLCs are circled.

See also Figure S6.

Together these results establish the dynamics of chromatin accessibility changes accompanying the transition from ESCs to EpiLCs and identify ZIC transcription factors as likely important players in controlling gene regulation during this transition.

ZIC3 Is Transiently Upregulated during the Transition to EpiLCs

There are multiple members of the ZIC transcription factor family; therefore, we determined their relative expression levels during the transition to EpiLCs. *Zic1* and *Zic4* are not expressed to appreciable levels, whereas *Zic2* and *Zic5* show progressively increased expression at d1 and d2 of the differentiation time course (Figure 2A). However, *Zic3* shows a transient increase in expression at d1, which is even more pronounced at the protein level (Figure 2B). These findings are supported by scRNA-seq analysis, where *Zic3* expression is enriched in the d1EpiLCs (Figure 2C). Importantly, although *Zic3* RNA expression is heterogenous at the single cell level, ZIC3 is expressed at the protein level in all d1EpiLCs (Figure S6). In contrast, OTX2 expression shows fewer dynamic changes and is increased at d1 and remains at a stable level in d2EpiLCs (Figures 2A–2C). The ZIC transcription factors therefore show dynamic changes in their expression that accompany the transitions to EpiLCs, and the transient expression kinetics of ZIC3 in particular indicates that this is a likely candidate for controlling the transition phase. Interestingly, ZIC3 has previously been implicated in the maintenance of pluripotency in ESCs, suggesting that it may play a dual role (Lim et al., 2007).

Determination of the ZIC3 Cistrome

Next, we focused on ZIC3. As a first step in determining its regulatory potential, we identified its genome-wide binding profile using chromatin immunoprecipitation sequencing (ChIP-seq). Initially, we focused on the transition state on d1 and identified 4,724 high confidence ZIC3 bound regions (Figure 3A; Table S1). The majority of these are located in inter- and intra-genic regions, and the ZIC3 binding regions are associated with 5,216 target genes based on the nearest-neighbor model. Consistent with a role for ZIC3 in cell fate changes, these target genes are enriched in GO terms for many differentiation processes, and signaling pathways such as the BMP and STAT pathways (Figure 3B). As expected, the ZIC binding motif is highly enriched in these regions along with more moderate enrichment for several other transcription factors, including ESSRB and SOX proteins, which have previously been implicated in regulatory activities in stem cells (Figure 3C).

To uncover ZIC3 binding dynamics and link these to the changing chromatin accessibility profiles, we performed additional ChIP-seq experiments for ZIC3 in ESCs and d2EpiLCs. Replicate experiments showed good concordance (Figure S7A) and clustered together in principal-component analysis (PCA) (Figure S7B). Overall, the binding dynamics showed transiently increased occupancy of ZIC3 at d1, which was reduced back to a lower level at d2 (Figures S7C and S7D). These occupancy changes were accompanied by a transient increase in chromatin opening across the binding regions at d1 (Figures S7C and S7D). This transient opening could be observed in more detail when analyzing the cut frequencies around the ZIC3 binding motifs (Figure S7E). This was particularly marked when considering the ZIC3 binding regions, which were not already occupied in ESCs (d1 unique peaks; Figure S7E, right). To gain further insight into the relationship between binding and chromatin accessibility dynamics as ESCs transition to EpiLCs, we focused on the chromatin regions showing changes in accessibility between any two cell conditions (Figures 1B, S1C, and S1D). ZIC3 binding across these regions generally

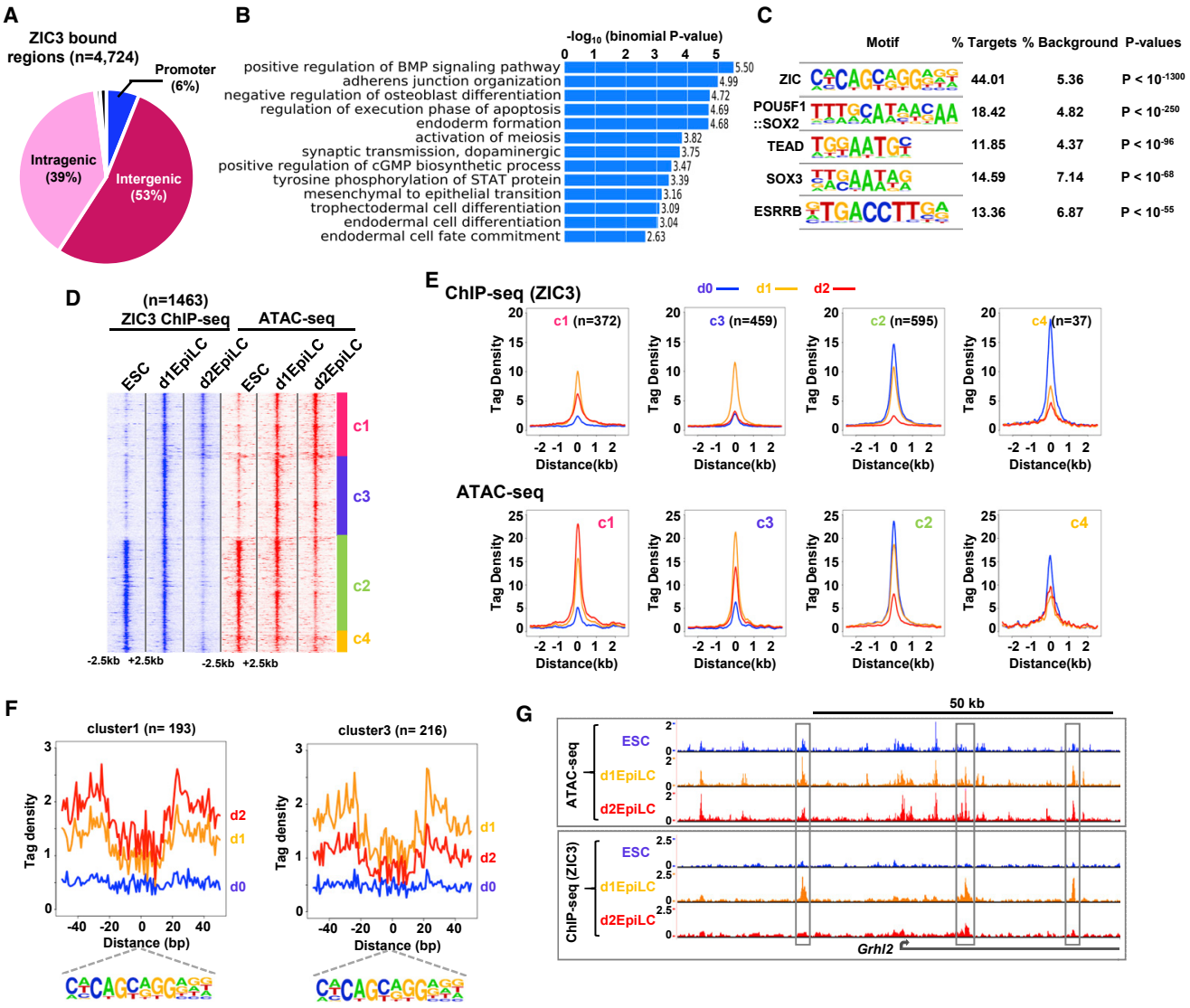


Figure 3. ChIP-Seq Analysis of ZIC3 Genomic Binding

(A) Genome-wide distribution of ZIC3 binding sites in d1EpiLCs. Promoter is defined as −2.5 to + 0.5 kb.

(B) Gene ontology analysis of ZIC3-associated genes (biological process).

(C) Top five enriched motifs found in the ZIC3 binding regions.

(D) Heatmap of the ZIC3 ChIP-seq profiles across a 5-kb window of all inducible ATAC-seq peaks (> 2.5-fold change for intra- or inter-genic peaks and > 2-fold change for promoter peaks) (left). The corresponding ATAC-seq signals at each ZIC3 binding region are shown on the right. Data are clustered (clusters c1-c4) according to ATAC-seq signals.

(E) Average tag densities of ZIC3 binding peaks from each of four identified clusters in each cell population (blue = ESC, orange = d1EpiLCs, red = d2EpilCs) for ZIC3 ChIP-seq signal (top) or ATAC-seq signal (bottom).

(F) Average ATAC-seq tag densities in an 80-bp window surrounding the ZIC3 motif (bottom) in cluster c1 (left), or cluster c3 (right). Data from ESCs (blue), d1EpiLCs (orange), and d2EpilCs (red) are shown.

(G) UCSC genome browser views of the ATAC-seq (top) and ZIC3 ChIP-seq (bottom) profiles around the *Grhl2* locus. Dynamically changing ZIC3 binding peaks are boxed.

See also Figure S7 and Table S1.

mirrors the changes in chromatin accessibility (Figures 3D and 3E). For example, in cluster c3, ZIC3 binding is strongly enhanced at d1, as chromatin accessibility increases, and is lost again at d2 as chromatin accessibility decreases again. However, in cluster c1, ZIC3 binding becomes reduced at d2 (consistent with its decreased expression) but chromatin accessibility increases, indicating a disconnection between ZIC3 binding kinetics and chromatin accessibility in these regions. This may reflect other factors acting to maintain or enhance the chromatin accessibility status at these regions.

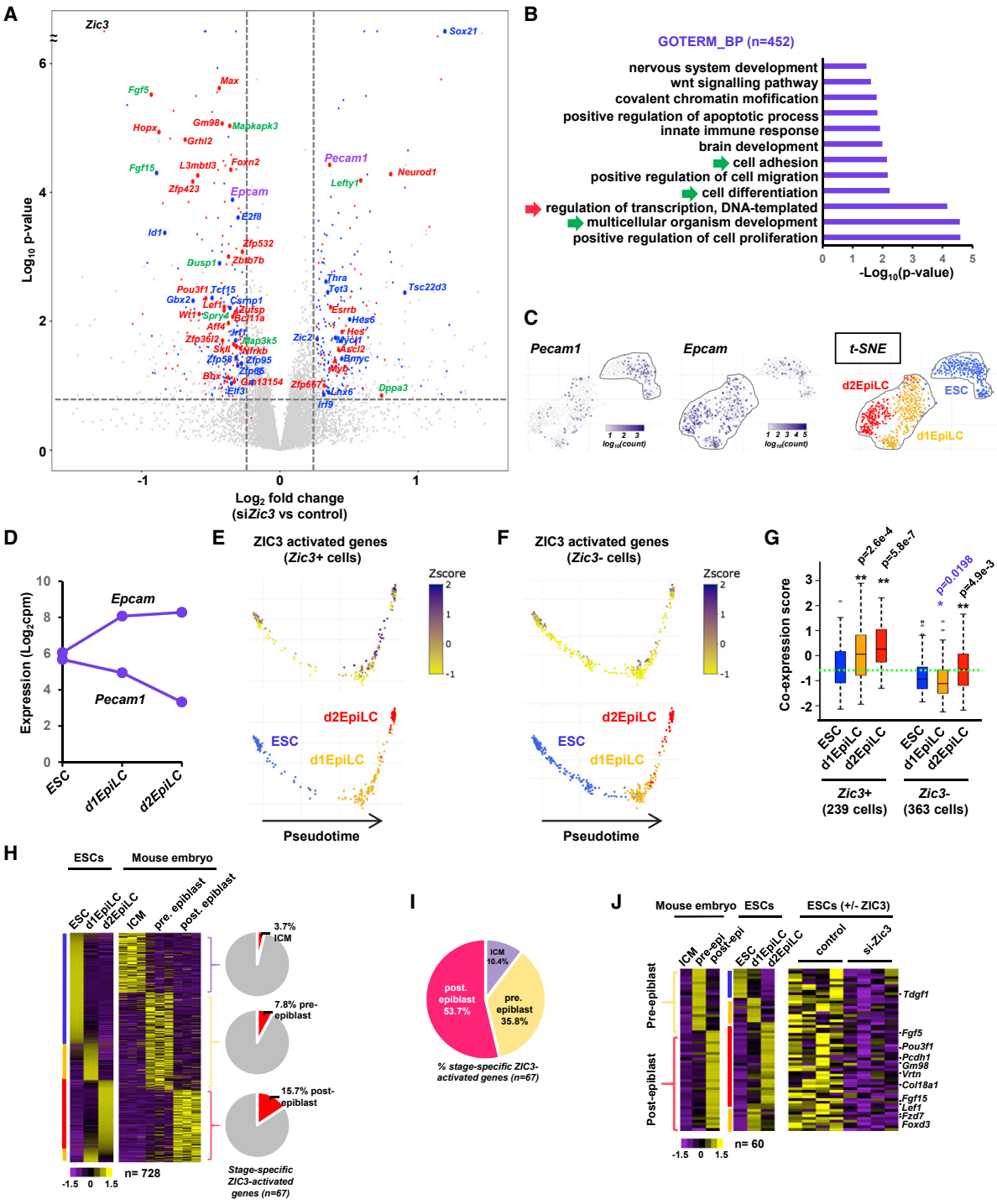


Figure 4. Identification of Zic3-Regulated Genes

(A) Volcano plot showing changes in gene expression in d1EpiLCs following *Zic3* knockdown by small interfering RNA (siRNA). Gene names are color coded for transcription factors (red for direct and blue for indirect targets) and signaling molecules (green). *Pecam1* and *Epcam* are shown in purple.

(B) Gene ontology analysis of all ZIC3-regulated genes for the biological process (BP) category.

(C) Expression of *Pecam1* and *Epcam* at single cell level. Data are plotted as $\log_{10}$ counts per cell, and superimposed on t-SNE analysis of the entire RNA-seq dataset for three time points (right).

(D) Expression of *Epcam* and *Pecam1* from aggregated scRNA-seq analysis in the indicated cell populations and shown as $\log_2$ counts per million base pairs (cpm).

(legend continued on next page)

The changes in chromatin accessibility were also revealed by focusing on the cleavage events around the ZIC3 motifs located in the ZIC binding regions (Figures 3F, S7F, and S7G). When considering all ZIC3 binding regions, both the depth and local accessibility are transiently enhanced in d1EpiLCs (Figure S7F). By focusing on different subclusters, different patterns could be discerned (Figures 3F and S7G). For example, footprint depth and local accessibility in regions belonging to cluster 3 show the highest levels in d1EpiLCs, as expected from the increased ChIP-seq signals in these regions. This behavior is exemplified by the *Grhl2* locus where several ZIC3 peaks are maximally present in d1EpiLCs, and this transient increase is accompanied by chromatin opening at the same loci (Figures 3G and S7H).

Collectively, these data reveal a dynamically changing ZIC3 cistrome during the transition from ESCs to EpiLCs. These dynamic changes are accompanied by underlying changes to the open chromatin landscape surrounding their sites. However, it should be noted that a large number of ZIC3 binding regions exhibit little change in chromatin accessibility, raising the possibility that ZIC3 has a role in ESCs in addition to its function during the transition to EpiLCs.

ZIC3-Dependent Gene Regulatory Events

Having established the ZIC3 cistrome and its dynamic changes, we next asked whether ZIC3 influences gene expression through these dynamic binding events. We depleted *Zic3* (Figures S8A and S8B) and determined the changes in transcriptome at the d1EpiLC transition state. ZIC3 protein levels were reduced by more than 80%. A total of 452 genes changed expression (> 1.2-fold), with 53% showing reduced expression following *Zic3* depletion, which is consistent with a potential activator role for ZIC3 (Figure 4A; Table S2). GO term analysis of the ZIC3-regulated genes, revealed enrichment of categories including cell adhesion alongside several developmental terms, various signaling pathways, and "regulation of transcription" (Figures 4B and S8C). Indeed, among these genes, there are numerous transcription factors and signaling pathway components (Figure 4A), indicating large changes in the regulatory systems in the cells. Two additional notable examples are the genes encoding the cell surface proteins, EPCAM and PECAM1, which show reciprocal changes in expression following *Zic3* depletion (Figure 4A, highlighted in purple) and are usually expressed at distinct times during the differentiation process, with PECAM1 being an ESC marker and EPCAM an EpiLC marker (Figures 4C and 4D). We next examined the expression of the ZIC3-activated genes across single cells that had been ordered by pseudotime analysis (Trapnell et al., 2014) of scRNA-seq data. Each cell was scored for expression of each target gene in a binary manner and the overall fraction of genes expressed per cell determined. There is a clear increase in expression of the ZIC3 regulon as cells progress toward d1EpiLCs and beyond (Figure 4E). However, this is not apparent in cells that do not co-express *Zic3* (Figure 4F). This association is further reflected by the increases in co-expression levels of the ZIC3-activated genes in d1- and d2EpliCs, which is not observed in cells with low *Zic3* expression (Figure 4G).

Finally, we asked whether the ZIC3-regulated genes are relevant in the context of early mouse embryonic development. We analyzed the clusters of genes that exhibit peak expression levels at each stage of embryonic development (Boroviak et al., 2015) and first compared the data to our own RNA-seq data from aggregated single cell analysis of ESCs, d1EpiLCs, and d2EpiLCs. Overall, there is good concordance between the datasets with ESCs being most similar to the inner cell mass (ICM), d2EpiLCs resembling the post-implantation epiblast, and d1EpiLCs representing an intermediate state (Figure 4H). Next, we asked whether ZIC3 is involved in regulating the expression of any of these genes that act as markers of early embryonic development. Importantly, when we superimposed our *Zic3* depletion dataset on top of these clusters, there was a sequential increase in the number of ZIC3-activated genes among the marker genes expressed maximally at each stage of embryonic development (Figure 4H, right; Figure 4I). Among these genes are known markers and regulators of differentiation in the pre-epiblast and post-epiblast stages such as *Foxd3* and *Fgf5* (Figure 4J) (Hanna et al., 2002; Khoa le et al., 2016). Thus, ZIC3 activates the expression of a large number of marker genes that are characteristic of the mouse pre- and post-implantation epiblast.

Next, we identified the direct target genes for ZIC3. To achieve this, and gain further insight into the likely direct roles of ZIC3, we took our ChIP-seq data and associated ZIC3 ChIP peaks with nearby genes. By intersecting this with our RNA-seq data, we uncovered a total of 207 directly regulated target genes for ZIC3 (Table S2). The majority of these are activated by ZIC3 (65%) (Figure S8D), suggesting a role for ZIC3 in upregulating gene expression during the transition to EpiLCs. Indeed, the directly activated ZIC3 target genes show an overall increase in expression during the transition to EpiLCs and this is

(E and F) Pseudotime analysis of ESCs, d1EpiLCs, and d2EpilCs based on the entire scRNA-seq dataset (bottom). The binarized expression (transformed z score) of the ZIC3 activated genes (n = 240) is plotted on top of these profiles (top) in cells that either show *Zic3* expression (E) or lack the expression of *Zic3* and *Otx2* (F). Pseudotime analysis was initially performed with all cells but in each case only cells exhibiting the *Zic3* expression characteristics are shown.

(G) Boxplot showing the co-expression scores for the ZIC3-activated genes in ESCs, d1EPiLCs, and d2EpilCs. Cells are split according to whether *Zic3* is expressed or not. Horizontal lines represent the median score, and the dotted green line is the median score in ESCs.

(H) Heatmaps showing the expression levels of genes categorized as uniquely expressed in ICM, pre-epiblast, or post-epiblast (right; Boroviak et al., 2015) and the corresponding expression levels in the aggregated scRNA-seq from ESCs, d1EPiLCs, and d2EpilCs (left). The heatmap is sorted based on the scRNA-seq data from the ESC-derived cells at each of the expression clusters from mouse embryos. Data are row z normalized for each dataset. The pie charts show the proportions of each of the stage-specific gene sets that are activated by ZIC3.

(I) Pie chart showing the proportions of ZIC3-activated lineage-specific genes from each stage of embryonic development.

(J) Heatmap showing the effect of *Zic3* depletion on the pre- and post-epiblast stage-specific genes in d1EpiLCs (right) and the heatmaps for the corresponding gene expression levels in early embryonic developmental stages (left) or ESCs, d1EpiLCs, and d2EpilCs (center).

Genes shown are from the red quadrants of the bottom two pie charts in (H). All heatmaps are individually z normalized.

See also Figure S8 and Table S2.

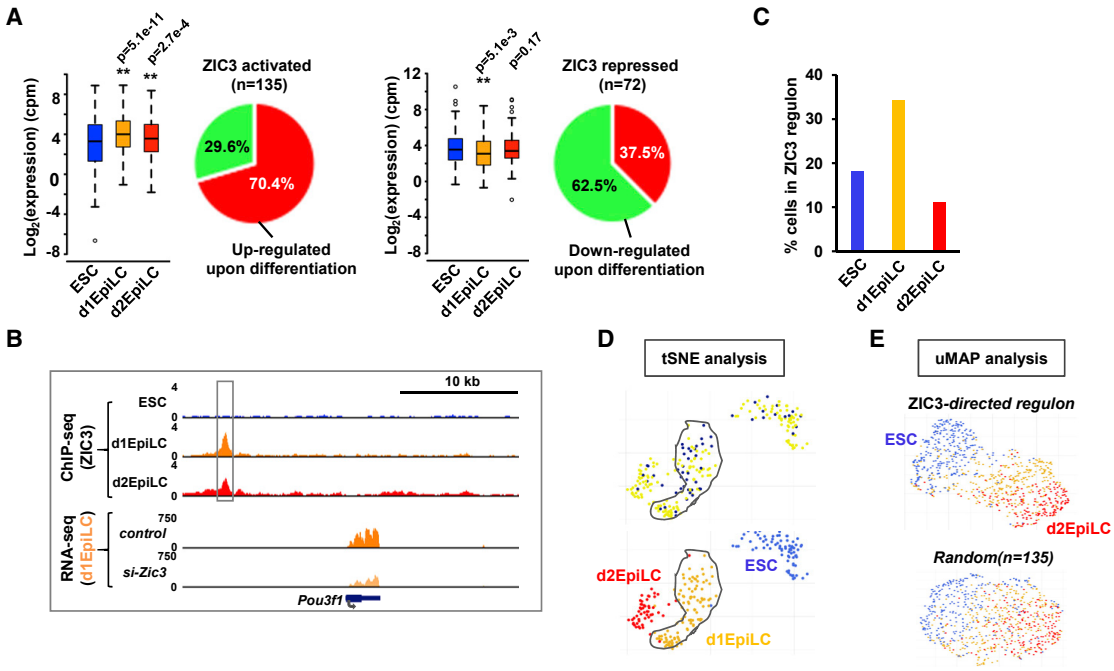


Figure 5. The Direct ZIC3 Target Gene Network

(A) Boxplots of the expression of directly regulated ZIC3 target genes (i.e., bound by ZIC3) in ESCs, d1EpiLCs, and d2EpilCs for activated (top) or repressed (bottom) genes. The proportions of direct ZIC3 target genes increasing and decreasing expression in d1EpiLCs upon differentiation from ESCs are shown in the pie charts on the right.

(B) UCSC genome browser views of the ZIC3 ChIP-seq (top) and RNA-seq (bottom) profiles around the *Pou3f1* locus. The major ZIC3 binding peak is boxed.

(C) AUCell analysis of the expression of the directly activated ZIC3 target gene regulon in ESCs, d1EPiLCs, and d2EpilCs. The percentage of cells is shown from each stage of differentiation that exhibits co-expression of the ZIC3 regulon.

(D) An scRNA-seq analysis of the ZIC3 regulon expression. Data are mapped (blue marked cells; top) on top of tSNE analysis of the entire RNA-seq dataset (bottom), with the originating cell types color coded. The d1EpiLCs are circled. Only *Zic3* positive cells are shown.

(E) Uniform manifold approximation and projection (uMAP) analysis of the scRNA-seq ESCs, d1EPiLCs, and d2EpilCs using either the 135 ZIC3-activated direct target genes in the ZIC3 regulon (top) or 135 randomly selected genes (bottom) to drive the clustering. Cells are color coded according to their known origins (blue = ESCs, orange = d1EpiLCs, and red = d2EpiLCs).

See also Figure S8 and Table S2.

particularly marked in d1EpiLCs, with over 70% of these genes showing upregulation (Figure 5A, left). Two notable directly activated target genes are *Pou3f1*, which is an important player in ESC differentiation (Iwafuchi-Doi et al., 2012) (Figure 5B), and *Wt1*, which encodes a bifunctional splicing factor and sequence-specific transcription factor that is known to function post-transcriptionally to regulate developmental RNAs in mouse ESCs (Bharathavikru et al., 2017) (Figure S8E). Reciprocally, we observe the opposite for the "directly repressed" genes, albeit to a lower level of significance (Figure 5A, right), leaving open the possibility that ZIC3 may be a bifunctional transcription factor. We also examined co-expression of ZIC3 directly activated genes using AUCell analysis, which is specifically designed to identify co-expression across single cells (Aibar et al., 2017). More frequent expression of the ZIC3 regulon was observed in the d1EpiLCs (Figure 5D), consistent with a role for ZIC3 in controlling gene expression at this transition point. Finally, we asked whether the ZIC3 regulon has predictive potential in determining cell types from scRNA-seq data and showed that the 135 directly activated ZIC3 target genes are not only able to separate ESCs from EpiLCs (Figure 5E) but are also able to segregate cells from different embryonic stages using scRNA-seq data derived from mouse embryos (Mohammed et al., 2017), with E4.5 epiblast cells forming a distinct compact cluster (Figure S8F).

ZIC3 is therefore involved in directly controlling the expression of a set of target genes that are generally upregulated in the transition from ESCs to EpiLCs and in the pre- and post-epiblast stages in the developing embryo.

ZIC3 Triggers a Complex Downstream Transcriptional Regulatory Network

To further understand the mechanisms through which ZIC3 affects downstream transcriptional programs that result in the EpiLC phenotype, we examined the functions of several of its target genes. Many of the ZIC3-regulated genes encode transcription factors, suggesting that ZIC3 acts mechanistically to trigger subsequent waves of changes in the transcriptome, mediated by these intermediary transcription factors. A large number of these are direct targets, including *Wt1*, *Lef1*, *Grhl2*, and *Pou3f1*. Further analysis of the scRNA-seq data demonstrates that a subset of these transcription factors is strongly co-expressed in d1EpiLCs (Figures 6A, S9A, and S9B). This

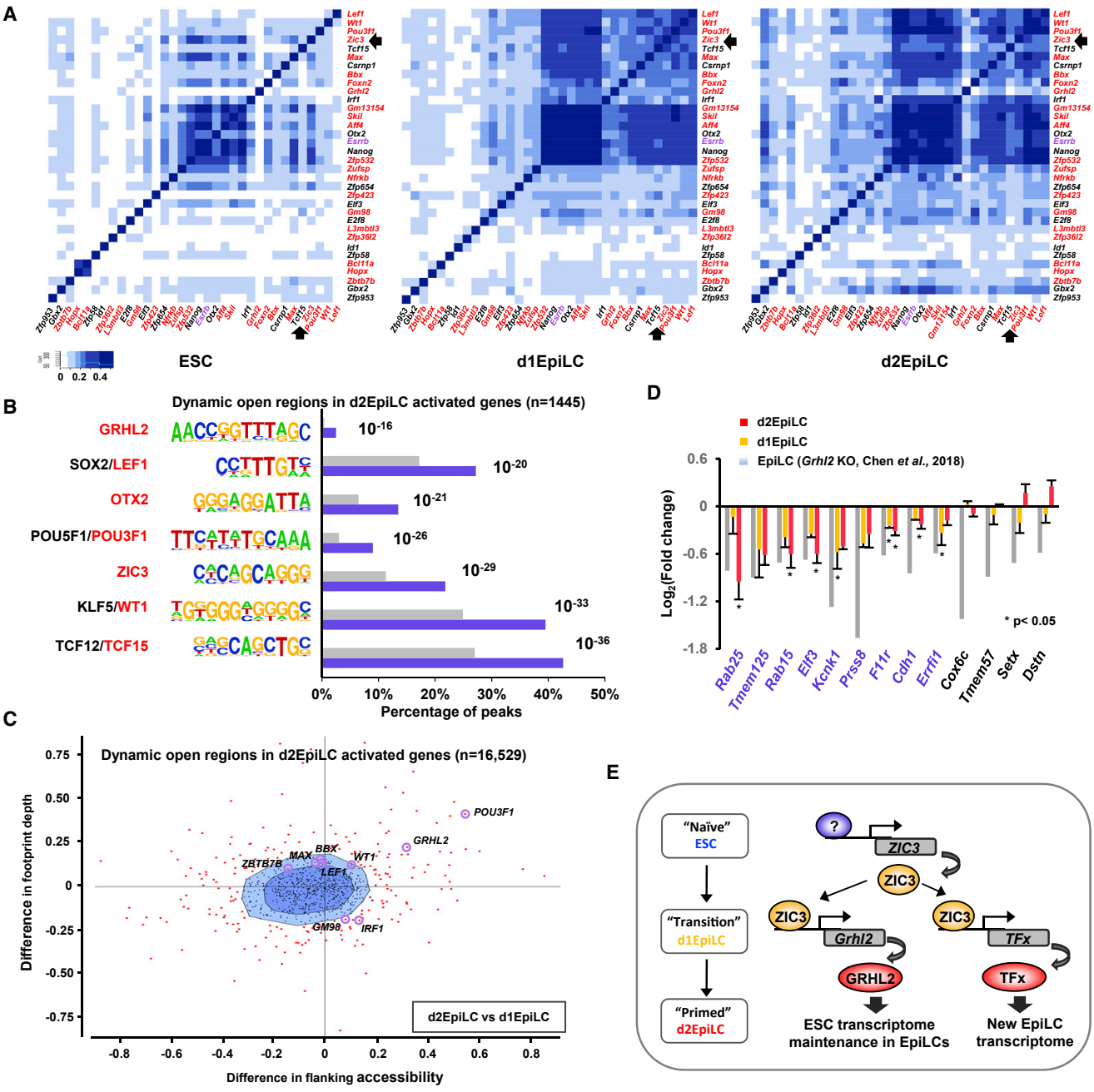


Figure 6. ZIC3-Regulated Transcription Factors Control Downstream Gene Expression Programs

(A) Jaccard's similarity plots of the co-expression of the indicated transcription factor-encoding ZIC3-activated genes in ESCs (left), d1EPiLCs (middle), and d2EpilCs (right). Direct ZIC3 targets are shown in red and the location of *Zic3* is highlighted with an arrow.

(B) Enrichment of DNA motifs within open chromatin regions associated with genes that change expression > 2-fold from d1EpiLCs to d2EpiLCs. Only dynamically opening inter- and intra-genic ATAC-seq peaks (i.e., from clusters 1 and 4; Figures 1B and S1C) were analyzed (n = 1,445). The percentages of each motif present in each set of peaks (blue) and the genomic background (gray) are indicated, and the p values are shown next to each of the columns.

(C) BaGFoot analysis of all of the open chromatin regions in d2EpiLCs that are associated (125 to +125 kb; n = 16,529) with genes that increase in expression (> 2-fold) in going from d1EpiLCs to d2EpiLCs. Motifs corresponding to binding sites for transcription factors encoded by ZIC3-activated genes are labeled.

(D) The expression of the indicated direct GRHL2 target genes (Chen et al., 2018) following depletion of *Grhl2* (gray bars) or *Zic3* in d1EpiLCs (orange bars) and d2EpiLCs (red bars). Data are shown based on the fold changes seen in RNA-seq data. Asterisks show significantly changing expression levels (p value < 0.05) and standard deviations are indicated (n ≥ 3).

(E) Model showing the transcriptional events centered on ZIC3 during the transition from ESCs to EpiLCs. GRHL2 maintains the ESC transcriptome (Chen et al., 2018) whereas other ZIC3-regulated transcription factors likely contribute to the newly established EpiLC transcriptome.

See also Figure S9.

transcription factor network is largely absent in ESCs but is maintained in d2EpiLCs. To begin to understand the impact of these transcription factors on downstream gene expression patterns, we asked whether we could find evidence for their DNA binding motifs in open chromatin regions associated with genes that show elevated expression in d2EpiLCs. We focused on regions that show increases in accessibility during the transition to EpiLCs (i.e., clusters 1 and 4 in Figure 1B) and found that a number of motifs are over-represented in these regions, including those for ZIC3 and OTX2 (Figure 6B). Importantly, motifs are also over-represented for a number of transcription factors encoded by ZIC3-activated genes, such as TCF15, WT1, LEF1, GRHL2, and POU3F1 (Figure 6B), consistent with a potential role for these transcription factors in enhancing the expression of these genes in EpiLCs. As an alternative approach, we used BaGFoot on all of the open regions that are associated with the genes showing enhanced expression in the transition from d1EpiLCs to d2EpiLCs and found that GRHL2 and POU3F1 binding motifs were among the motifs showing strong evidence for increased footprint depth and localized chromatin opening in EpiLCs (Figure 6C). Given the strong presence of binding motifs in the regulatory regions of potential target genes, we next sought evidence for regulatory activity of the corresponding transcription factors. We focused on GRHL2 as this has recently been shown to play an important role in switching enhancer usage during the transition of ESCs to EpiLCs (Chen et al., 2018). GRHL2 is encoded by a direct ZIC3 target gene (see Figure 3G), suggesting a potential functional hierarchy with ZIC3 acting upstream of GRHL2 in a transcriptional cascade. This hierarchy predicts that ZIC3 depletion should have a similar effect on downstream gene expression profiles as depletion of GRHL2. We therefore focused on a set of directly activated GRHL2 target genes (i.e., bound by GRHL2), which showed the largest decreases in expression in EpiLCs following loss of GRHL2 expression (Chen et al., 2018). Importantly, the majority of these were downregulated upon depletion of ZIC3 in d1EpiLCs and/or d2EpiLCs (Figure 6D), consistent with a transcription factor relay network whereby ZIC3 controls GRHL2 expression to subsequently influence downstream gene expression. More generally, ZIC3 controls the expression of a set of transcription factors that are able to generate a cascade effect on gene expression at later stages of ESC differentiation. Indeed, depletion of *Zic3* in ESCs cultured in EpiLC differentiation media leads to increased alkaline phosphatase positive colony formation when returned to 2i/leukemia inhibitory factor (LIF) media, consistent with inefficient differentiation toward EpiLCs and the retention of ESC-like properties (Figures S9C and S9D).

DISCUSSION

The transcription factor networks controlling maintenance of the pluripotent state in ESCs are relatively well understood. However, it is less clear how naive ESCs begin differentiation by transitioning through the EpiLC state. The transcription factor OTX2 was previously shown to control enhancer activation during the transition from naive ESCs toward EpiLCs (Acampora et al., 2013; Yang et al., 2014; Buecker et al., 2014). Here, we took an unbiased approach using ATAC-seq to uncover novel transcriptional regulators of this transition. We focused on ZIC3, which exhibits transient expression kinetics and chromatin binding as cells change fate to EpiLCs. ZIC3 plays a key role in controlling gene expression during differentiation to EpiLCs and, in particular, a large number of genes encoding signaling molecules and transcription factors. Through activating the expression of transcription factor encoding genes, ZIC3 acts at a pivotal point in a transcriptional cascade, which determines the EpiLC phenotype (Figure 6E). For example, the ZIC3-regulated transcription factor GRHL2 has been shown to play an important role in maintaining a stem cell-specific gene expression program as cells progress to the EpiLC state through an enhancer switching mechanism (Chen et al., 2018). Other ZIC3-regulated transcription factors play a role in controlling other gene expression programs during differentiation such as TCF15, which has previously been shown to be important in priming EpiSCs for differentiation (Davies et al., 2013).

Previous studies suggested that ZIC transcription factors are likely involved in early developmental decisions in naive ESCs. ZIC2 and ZIC3 were identified as interactors of OCT4 in EpiLCs, suggesting a role in cooperative transcriptional regulation as the OCT4 cistrome is remodeled in the transition from ESCs (Buecker et al., 2014). Indeed, we provide further support for this model as we identified an enrichment of OCT4-like binding motifs in ZIC3 binding regions (Figure 3C). However, other motifs are enriched in the ZIC3 bound regions, suggesting a broader cooperativity with a wide range of transcription factors. Interestingly, ZIC3 has also been implicated in the maintenance of pluripotency in ESCs (Lim et al., 2007). However, in the latter study, the ESCs were maintained in the presence of serum and LIF, conditions that do not fully recapitulate the naive state. Our data are therefore generally consistent with a role for ZIC3 in ESCs, but point to ZIC3 acting at an early stage during the transition from naive ESCs. ZIC3 knockout mice exhibit early embryonic developmental defects prior to gastrulation, leading to defects in left-right patterning (Ware et al., 2006). Moreover, mutation of ZIC3 in humans causes a syndrome known as X-linked heterotaxy, where similar patterning defects are observed (reviewed in Bellchambers and Ware, 2018). It is possible that these developmental defects arise due to the changes we observe at the earliest cell fate transitions from naive ESCs. Indeed, consistent with this early role, many of the ZIC3-regulated target genes show peak expression during the transition from the ICM through to the pre-and post-implantation epiblast in the embryo (see Figure 4H).

ZIC transcription factors can bind similar DNA binding motifs (Badis et al., 2009; reviewed in Hatayama and Aruga, 2018) and hence there is the potential for further cross talk between these factors at the level of chromatin binding. As ZIC3 exhibits transient activation kinetics, and ZIC2/5 expression is maintained at later stages, it is possible that some of the functions of ZIC3 are maintained and/or expanded by these factors as EpiLCs differentiate further. Indeed, ZIC binding motifs still show increased occupancy and opening in d2EpiLCs (Figure S5) and together these observations suggest a pivotal role for ZIC transcription factors in maintaining the EpiLC state. ZIC2 was previously shown to act in concert with the Mbd3/NuRD complex in ESCs to cause transcriptional

repression and loss of ZIC2 affected subsequent ESC differentiation (Luo et al., 2015). However, the ESCs in this study were cultured in the presence of serum and LIF conditions that only partially recapitulate the naive ESC state. Further work in ESCs grown under these conditions also implicated ZIC2 as an important player in maintaining the transcriptional regulatory circuits in these cells (Iwafuchi-Doi et al., 2012; Matsuda et al., 2017). Moreover, ZIC1 and ZIC2 have been shown to play a role much later in development in the context of the neuronal gene expression program in cerebellar granule neurones (Frank et al., 2015). Future studies will be needed to unravel whether ZIC3 functions in the early ESC transitions are modified by other ZIC family members later in the differentiation and development program.

It is unclear whether ZIC transcription factors are transcriptional activator or repressor proteins. Work on ZIC2 suggests a repressive role (Luo et al., 2015). In contrast, the majority of the directly regulated ZIC3 target genes (65%) are activated by ZIC3, although a substantive proportion are repressed. It is possible that like many transcription factors, ZIC3 can adopt different roles at different regulatory regions, and in this context may poise genes in the transition state for subsequent activation in EpiLCs. Nevertheless, through its transcriptional regulatory activities, ZIC3 plays an important role in controlling the transition from naive ESCs to the more advanced EpiLC state. Through activating genes encoding transcription factors such as GRHL2 it contributes to maintaining a plastic state that retains stem-cell properties but is poised for subsequent differentiation. It is highly likely that other ZIC3-regulated transcription factors play equally important roles in creating this flexible regulatory environment.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Lines and Culture Conditions
- METHOD DETAILS
  - Real-Time RT-qPCR
  - Western Blot Analysis
  - ATAC-Seq Assays
  - ChIPmentation Assays
  - RNA-Seq Assays
  - Single-Cell (sc) RNA-Seq Assays
  - Immunofluorescence Assays
  - ESC Clonogenicity Assay
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Bioinformatics and Statistical Analysis
  - ATAC-Seq Data Analysis
  - ChIPmentation Data Analysis
  - RNA-Seq Data Analysis
  - Single-Cell Transcriptomics
  - Statistical Analysis of the ESC Clonogenicity Assay
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.05.026.

ACKNOWLEDGMENTS

We thank Karren Palmer and Mairi Challinor for excellent technical assistance; Peter March in the Bioimaging facility, as well as staff in the Genomic Technologies and Bioinformatics facilities; and Catherine Millar, Hilary Ashe, and members of our laboratories for comments on the manuscript and stimulating discussions. This work was funded by the BBSRC (BB/M000630/1) and the Wellcome Trust (103857/Z/14/Z; A.D.S. and S.M.B.) and MRC grants MR/M008908/1 and MR/M012174/1 (M.I.).

AUTHOR CONTRIBUTIONS

The conception or design of the work was performed by S.-H.Y. and A.D.S.; the acquisition, analysis, or interpretation of data by S.-H.Y., M.A., R.B., S.M.B., and M.I.; and the drafting of the work or revising it critically by S.-H.Y., M.A., S.M.B., M.I., and A.D.S. All authors approved the manuscript and are accountable for all aspects of the work.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 4, 2018
Revised: March 14, 2019
Accepted: May 6, 2019
Published: June 11, 2019

REFERENCES

Acampora, D., Di Giovannantonio, L.G., and Simeone, A. (2013). Otx2 is an intrinsic determinant of the embryonic stem cell state and is required for transition to a stable epiblast stem cell condition. Development *140*, 43–55.

Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086.

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

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X., et al. (2009). Diversity and complexity in DNA recognition by transcription factors. Science *324*, 1720–1723.

Baek, S., Goldstein, I., and Hager, G.L. (2017). Bivariate Genomic Footprinting Detects Changes in Transcription Factor Activity. Cell Rep. *19*, 1710–1722.

Bellchambers, H.M., and Ware, S.M. (2018). ZIC3 in Heterotaxy. Adv. Exp. Med. Biol. *1046*, 301–327.

Betschinger, J., Nichols, J., Dietmann, S., Corrin, P.D., Paddison, P.J., and Smith, A. (2013). Exit from pluripotency is gated by intracellular redistribution of the bHLH transcription factor Tfe3. Cell *153*, 335–347.

Bharathavikru, R., Dudnakova, T., Aitken, S., Slight, J., Artibani, M., Hohenstein, P., Tollervey, D., and Hastie, N. (2017). Transcription factor Wilms' tumor 1 regulates developmental RNAs through 3′ UTR interaction. Genes Dev. *31*, 347–352.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Boroviak, T., Loos, R., Lombard, P., Okahara, J., Behr, R., Sasaki, E., Nichols, J., Smith, A., and Bertone, P. (2015). Lineage-Specific Profiling Delineates the

Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. Dev. Cell *35*, 366–382.

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell *14*, 838–853.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr. Protoc. Mol. Biol. *109*, 21.29.1–21.29.9.

Chen, A.F., Liu, A.J., Krishnakumar, R., Freimer, J.W., DeVeale, B., and Blelloch, R. (2018). GRHL2-Dependent Enhancer Switching Maintains a Pluripotent Stem Cell Transcriptional Subnetwork after Exit from Naive Pluripotency. Cell Stem Cell *23*, 226–238.e4.

Davies, O.R., Lin, C.Y., Radzisheuskaya, A., Zhou, X., Taube, J., Blin, G., Waterhouse, A., Smith, A.J., and Lowell, S. (2013). Tcf15 primes pluripotent cells for differentiation. Cell Rep. *3*, 472–484.

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

Dray, S., and Dufour, A.B. (2007). The ade4 package: implementing the duality diagram for ecologists. J. Stat. Softw. *22*, 1–20.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA *95*, 14863–14868.

Factor, D.C., Corradin, O., Zentner, G.E., Saiakhova, A., Song, L., Chenoweth, J.G., McKay, R.D., Crawford, G.E., Scacheri, P.C., and Tesar, P.J. (2014). Epigenomic comparison reveals activation of "seed" enhancers during transition from naive to primed pluripotency. Cell Stem Cell *14*, 854–863.

Fiorenzano, A., Pascale, E., D'Aniello, C., Acampora, D., Bassalert, C., Russo, F., Andolfi, G., Biffoni, M., Francescangeli, F., Zeuner, A., et al. (2016). Cripto is essential to capture mouse epiblast stem cell and human embryonic stem cell pluripotency. Nat. Commun. *7*, 12589.

Frank, C.L., Liu, F., Wijayatunge, R., Song, L., Biegler, M.T., Yang, M.G., Vockley, C.M., Safi, A., Gersbach, C.A., Crawford, G.E., and West, A.E. (2015). Regulation of chromatin accessibility and Zic binding at enhancers in the developing cerebellum. Nat. Neurosci. *18*, 647–656.

Goldstein, L.D., Chen, Y.J., Dunne, J., Mir, A., Hubschle, H., Guillory, J., Yuan, W., Zhang, J., Stinson, J., Jaiswal, B., et al. (2017). Massively parallel nanowell-based single-cell gene expression profiling. BMC Genomics *18*, 519.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.

Habibi, E., and Stunnenberg, H.G. (2017). Transcriptional and epigenetic control in mouse pluripotency: lessons from in vivo and in vitro studies. Curr. Opin. Genet. Dev. *46*, 114–122.

Hanna, L.A., Foreman, R.K., Tarasenko, I.A., Kessler, D.S., and Labosky, P.A. (2002). Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo. Genes Dev. *16*, 2650–2661.

Hatayama, M., and Aruga, J. (2018). Role of Zic Family Proteins in Transcriptional Regulation and Chromatin Remodeling. Adv. Exp. Med. Biol. *1046*, 353–380.

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. Cell *146*, 519–532.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Iwafuchi-Doi, M., Matsuda, K., Murakami, K., Niwa, H., Tesar, P.J., Aruga, J., Matsuo, I., and Kondoh, H. (2012). Transcriptional regulatory networks in epiblast cells and during anterior neural plate development as modeled in epiblast stem cells. Development *139*, 3926–3937.

Kalkan, T., and Smith, A. (2014). Mapping the route from naive pluripotency to lineage specification. Philos. Trans. R. Soc. Lond. B Biol. Sci. *369*, 20130540.

Kalkan, T., Olova, N., Roode, M., Mulas, C., Lee, H.J., Nett, I., Marks, H., Walker, R., Stunnenberg, H.G., Lilley, K.S., et al. (2017). Tracking the embryonic stem cell transition from ground state pluripotency. Development *144*, 1221–1234.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S.R., Tan, G., et al. (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. *46* (*D1*), D260–D266.

Khoa le, T.P., Azami, T., Tsukiyama, T., Matsushita, J., Tsukiyama-Fujii, S., Takahashi, S., and Ema, M. (2016). Visualization of the epiblast and visceral endodermal cells using Fgf5-P2A-Venus BAC transgenic mice and Epiblast Stem Cells. PLoS ONE *11*, e0159246.

Kumar, L., and Futschik, M.E. (2007). Mfuzz: a software package for soft clustering of microarray data. Bioinformation *2*, 5–7.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Lim, L.S., Loh, Y.H., Zhang, W., Li, Y., Chen, X., Wang, Y., Bakre, M., Ng, H.H., and Stanton, L.W. (2007). Zic3 is required for maintenance of pluripotency in embryonic stem cells. Mol. Biol. Cell *18*, 1348–1358.

Luo, Z., Gao, X., Lin, C., Smith, E.R., Marshall, S.A., Swanson, S.K., Florens, L., Washburn, M.P., and Shilatifard, A. (2015). Zic2 is an enhancer-binding factor required for embryonic stem cell specification. Mol. Cell *57*, 685–694.

Mahony, S., and Benos, P.V. (2007). STAMP: a web tool for exploring DNA-binding motif similarities. Nucleic Acids Res. *35*, W253–W258.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., and Stunnenberg, H.G. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. Cell *149*, 590–604.

Matsuda, K., Mikami, T., Oki, S., Iida, H., Andrabi, M., Boss, J.M., Yamaguchi, K., Shigenobu, S., and Kondoh, H. (2017). ChIP-seq analysis of genomic binding regions of five major transcription factors highlights a central role for ZIC2 in the mouse epiblast stem cell gene regulatory network. Development *144*, 1948–1958.

McCarthy, D.J., Campbell, K.R., Lun, A.T., and Wills, Q.F. (2017). Scater: preprocessing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics *33*, 1179–1186.

McInnes L., Healy J., and Melville, J., (2018) UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv, arXiv:1802.03426, https://arxiv.org/abs/1802.03426.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., et al. (2017). Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. Cell Rep. *20*, 1215–1228.

O'Donnell, A., Yang, S.H., and Sharrocks, A.D. (2008). MAP kinase-mediated c-fos regulation relies on a histone acetylation relay switch. Mol. Cell *29*, 780–785.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods *14*, 979–982.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Schmidl, C., Rendeiro, A.F., Sheffield, N.C., and Bock, C. (2015). ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat. Methods *12*, 963–965.

Stergachis, A.B., Neph, S., Reynolds, A., Humbert, R., Miller, B., Paige, S.L., Vernot, B., Cheng, J.B., Thurman, R.E., Sandstrom, R., et al. (2013). Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. Cell *154*, 888–903.

Sung, M.H., Guertin, M.J., Baek, S., and Hager, G.L. (2014). DNase footprint signatures are dictated by factor dynamics and DNA sequence. Mol. Cell *56*, 275–285.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

van der Maaten, L.J.P., and Hinton, G.E. (2008). Visualizing Data Using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. Bioinformatics *28*, 2184–2185.

Ware, S.M., Harutyunyan, K.G., and Belmont, J.W. (2006). Zic3 is critical for early embryonic patterning during gastrulation. Dev. Dyn. *235*, 776–785.

Wei, T., and Simko, V. (2017). corrplot: Visualization of a Correlation Matrix (R Foundation for Statistical Computing).

Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. Biochem. Soc. Trans. *38*, 1027–1032.

Yang, S.H., Kalkan, T., Morrisroe, C., Smith, A., and Sharrocks, A.D. (2012). A genome-wide RNAi screen reveals MAP kinase phosphatases as key ERK pathway regulators during embryonic stem cell differentiation. PLoS Genet. *8*, e1003112.

Yang, S.H., Kalkan, T., Morissroe, C., Marks, H., Stunnenberg, H., Smith, A., and Sharrocks, A.D. (2014). Otx2 and Oct4 drive early enhancer activation during embryonic stem cell transition from naive pluripotency. Cell Rep. *7*, 1968–1981.

Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.

Young, R.A. (2011). Control of the embryonic stem cell state. Cell *144*, 940–954.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit monoclonal anti-p44/42 MAPK (Erk1/2)(137F5)	Cell Signaling Technology	Cat# 4695; RRID:AB_390779
Rabbit polyclonal anti-OTX2	Proteintech	Cat# 13497-1-AP; RRID:AB_2157176
Rabbit polyclonal anti-ZIC3	Abcam	Cat# ab222124; RRID:AB_2801559
IRDye 800CW Donkey anti-Mouse IgG	LI-COR	Cat# 926-32212; RRID:AB_621847
IRDye 680LT Donkey anti-Rabbit IgG	LI-COR	Cat# 926-68023; RRID:AB_10706167
Goat anti-Rabbit IgG (H+L) cross-adsorbed secondary antibody, Alexa Fluor 488	ThermoFisher	Cat# A-11008; RRID:AB_143165
Chemicals, Peptides, and Recombinant Proteins		
NDiff 227 media	Takara Bio Europe SAS	Cat# Y40002
DMEM/F-12(1:1)	ThermoFisher	Cat# 31330-038
BSA fraction V (7.5%)	ThermoFisher	Cat# 15260-037
KnockOut Serum Replacement	ThermoFisher	Cat# 10828010
CHIR99021	Miltenyl Biotech.	Cat# 130-106-539
PD0325901	Miltenyl Biotech.	Cat# 130-106-549
bFGF	R&D system	Cat# 233-FB-025
Activing A	Peprotech	Cat# 120-14E-10
Gelatin	Millipore	Cat# ES-006-B
Bovine plasma fibronectin	Sigma	Cat# F1141
Accutase	Sigma	Cat# A6964
Nuclei EZ lysis buffer	Sigma	Cat# N3408
Pierce protease inhibitor, EDTA-free	ThermoFisher	Cat# A32965
Proteinase K	Ambion	Cat# AM2546
ProLong™ Gold Antifade Mountant with DAPI	ThermoFisher	Cat# P36941
Critical Commercial Assays		
QIAshredder RNA Extraction Column	QIAGEN	Cat# 79654
RNeasy Plus Mini Kit	QIAGEN	Cat# 74134
RNase-free DNase set	QIAGEN	Cat# 79254
miniElute Reaction Cleanup	QIAGEN	Cat# 28204
TruSeq RNA library kit v2	Illumina	Cat# RS-122-2001
Nextera DNA library Prep kit	Illumina	Cat# FC-121-1031
Nextera Index kit	Illumina	Cat# FC-121-1012
NEBNext high fidelity 2x PCR master mix	NEB	Cat# M0541
Ampure XP beads	Beckman Coulter Agencourt	Cat# A63881
Leukocyte Alkaline Phosphatase kit	Sigma	Cat# 86R-1KT
Deposited Data		
ATAC-seq	ArrayExpress	ArrayExpress: E-MTAB-7207
ChIPmentation-seq	ArrayExpress	ArrayExpress: E-MTAB-7208
RNA-seq	ArrayExpress	ArrayExpress: E-MTAB-7206
Single cell (sc) RNA-seq	ArrayExpress	ArrayExpress: E-MTAB-7211
Experimental Models: Cell Lines		
Mouse Rex1GFPd2 ES cells	Yang et al., 2012	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
See Table S3		N/A
ON-TARGETplus Mouse Zic3 siRNA	Horizon, Dharmacon	Cat# L-045667-00-0020
ON-TARGETplus Non-targeting Pool	Horizon, Dharmacon	Cat# D-001810-10
Software and Algorithms		
Bowtie2 v2.3.0	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
SAMtools v1.3.1	Li et al., 2009	http://samtools.sourceforge.net/
Trimmomatic v0.32	Bolger et al., 2014	http://www.usadellab.org/cms/?page=trimmomatic
HOMER	Heinz et al., 2010	http://homer.ucsd.edu/homer/
MACS2	Zhang et al., 2008	https://github.com/taoliu/MACS
BEDTools	Quinlan and Hall, 2010	https://bedtools.readthedocs.io/en/latest/
Java treeview	Eisen et al., 1998	http://jtreeview.sourceforge.net/docs/overview.html
edgeR	Robinson et al., 2010	https://bioconductor.org/packages/release/bioc/html/edgeR.html
R Mfuzz package (Fuzzy cMeans clustering)	Kumar and Futschik, 2007	https://bioconductor.org/packages/release/bioc/html/Mfuzz.html
STAMP tool	Mahony and Benos, 2007	https://bio.tools/stamp
BaGFoot	Baek et al., 2017	https://sourceforge.net/projects/bagfootr/files/
JASPAR	Khan et al., 2018	http://jaspar.binf.ku.dk
FIMO	Grant et al., 2011	http://meme-suite.org/doc/fimo.html
GREAT	McLean et al., 2010	http://great.stanford.edu/public/html/
STAR	Dobin et al., 2013	https://github.com/alexdobin/STAR
HTSeq	Anders et al., 2015	https://htseq.readthedocs.io/en/release_0.11.1/
DESeq2 v1.18.1	Anders and Huber, 2010	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
DAVID	Huang et al., 2009	https://david.ncifcrf.gov
scater package (calculateCPM function)	McCarthy et al., 2017	https://bioconductor.org/packages/release/bioc/html/scater.html
t-SNE plot	van der Maaten and Hinton, 2008	https://lvdmaaten.github.io/tsne/
UMAP plot	McInnes et al., 2018	https://github.com/ropenscilabs/umapr
monocole (v2) package	Trapnell et al., 2014	https://github.com/cole-trapnell-lab/monocle-release
R package *ade4* (Jaccard Similarity Index)	Dray and Dufour, 2007	https://github.com/cran/ade4
R package *Beeswarm*	N/A	https://rdrr.io/cran/beeswarm
SCENIC R package	Aibar et al., 2017	https://github.com/aertslab/SCENIC
Fiji-ImageJ	National Institute of Health	https://imagej.net/Fiji/Downloads

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Andrew Sharrocks (a.d.sharrocks@manchester.ac.uk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines and Culture Conditions

Mouse *Rex1GFP*d2 ES cells were maintained as described previously in NDiff 227 media (Takara Bio Europe SAS, Y40002) containing 2i inhibitors (CHIR99021 and PD0325901; Miltenyi Biotec, 130-106-539 and 130-106-549) on dishes coated with gelatin (Millipore, ES-006-B) (Yang et al., 2012). The d1EpiLCs were created by plating 2.5x10^4 cells/cm^2 on dishes coated with bovine plasma fibronectin (5 μg/ml; Sigma, F1141) and then growing for 1 day in NDiff N2B27 media containing bFGF (12ng/ml; R&D system, 233-FB-025), activin A (20 ng/ml; Peprotech,120-14E-10), and KnockOut Serum Replacement (1%; ThermoFisher, 10828010) (Hayashi

et al., 2011). For d2EpiLCs, the half volume of medium was removed and replenished with freshly prepared medium, and growth continued for a further day. RNAi was performed as described previously (Yang et al., 2012).

METHOD DETAILS

Real-Time RT-qPCR

Real time RT-qPCR was carried out as described previously (O'Donnell et al., 2008). Data were normalized for the geometric mean expression of the control genes *hmbs* and *ppia*. The primer-pairs used for RT-PCR are listed in Table S3.

Western Blot Analysis

Western blotting was carried out with the primary antibodies; Erk2 (137F5; Cell Signaling, 4695), Otx2 (ProteinTech., 13497-1-AP) and ZIC3 (Abcam, ab222124). All experiments were carried out in 12-well plates. The lysates were directly harvested in 2xSDS sample buffer (100 mM Tris.Cl pH 6.8, 4% SDS, 20% glycerol, 200 mM DTT and 0.2% bromophenol blue) followed by sonication (Bioruptor, Diagenode). The proteins were detected using a LI-COR Odyssey Infrared Imager as described previously (Yang et al., 2012).

ATAC-Seq Assays

The cells were dissociated from the plates with Accutase (Sigma, A6964) for 3 minutes at 37°C. ATAC-seq samples and libraries were generated as described previously (Buenrostro et al., 2015) except the nuclei were prepared using 100 μl of ice cold Nuclei EZ lysis buffer (Sigma, N3408). The nuclei pellets were resuspended in 10 μl H_2O and nuclei were counted. The tagmentation reaction was performed with 50 thousand nuclei and 2.5 μl of Tn5 transposase (0.5 μM) in 25 μl reaction volumes for 30 mins at 750 rpm at 37°C. The tagmented genomic DNA was purified by using miniElute Reaction Cleanup kits (QIAGEN, 28204) and eluted in 10.5 μl. The libraries were generated by 9 cycles of PCR reaction using adaptor primers (Nextera Index kit; Illumina, FC-121-1012) and NEBNext high fidelity 2x PCR master mix (NEB, M0541), followed by two-sided size selection by Ampure XP beads purification (0.4x reaction volume then 1.2x reaction volume; Beckman Coulter Agencourt, A63881). The typical yield is between 150-300 ng. The sequencing was performed on an Illumina Next-seq genome analyzer according to the manufacturer's protocols.

ChIPmentation Assays

For ChIP-seq using the ChIPmentation method (Schmidl et al., 2015), the cells (1.4-2 × 10^7 cells sufficient for 5 ChIPmentation experiments) were dissociated with Accutase (Sigma, A6964) for 3 mins at 37°C and fixed in 1% formaldehyde in 0.03%BSA/F12 for 10 min at room temperature. After quenching with 0.125 M glycine, cells were pelleted and the pellets were washed with 0.03%BSA/PBS. The nuclei were lysed in lysis buffer (10 mM Tris-Cl pH 8.0, 10 mM NaCl, 0.2% NP40 and 1 tablet of Complete protease inhibitor cocktail (Thermo Scientific) per 50 mL for 20 mins at 4°C. The nuclei were counted, snap frozen in liquid N_2 and stored at −80°C. Prior to ChIPmentaion, nuclei pellets were resuspended in H_2O and topped up with 0.25% SDS (25x10^6 cells/ml; 130 μl/aliquot). The nuclei solution was then sonicated 3 times, 10 cycles (30 s on/off) at 4°C (Bioruptor, Diagenode). The IP solution was prepared through sequential dilution by adding 1.5 volumes of equilibration buffer (10 mM Tris.Cl pH8.0, 140 mM NaCl, 0.6 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate and 0.1% SDS) and 0.92 volumes of TopUp buffer (10 mM Tris-Cl pH8.0, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate and 0.05% SDS).

ChIPmentation assays were performed essentially as described previously (Schmidl et al., 2015). Briefly, 100 μl of ZIC3 antibody solution (1 μg; Abcam, ab222124) was cross-linked to the protein A beads were incubated with 100 μl IP solution (1.2x10^6 nuclei) at 4°C overnight. The beads were sequentially washed twice with 250 μl of low salt buffer, high salt buffer, LiCl wash buffer and once in 150 μl of 10mM Tris-Cl pH 8.0. Next, the tagmentation reactions (25 μl) were performed with 1 μl of Nextera Tn5 transposase (Nextera kit; Illumina, FC-121-1030) in tagmentation buffer (33 mM Tris-OAc pH 7.8, 66 mM potassium-OAc, 10 mM Mg-OAc and 16% DMF) at 1200 rpm at 37°C for 10 mins. Ice cold low salt buffer (150 μl) was immediately added to stop the enzymatic reaction on ice for 5 mins. The beads were then washed twice with 150 μl of low salt buffer and TE. The tagmented ChIPed samples were then resuspended in 48 μl of ChIPmentation elution buffer (10 mM Tris.Cl pH8.0, 300 mM NaCl, 5 mM EDTA and 0.4% SDS) and incubated with 2 μl of proteinase K (20 mg/ml; Ambion, AM2546) at 55°C for 1 hr and then at 65°C overnight. The beads were further incubated for 1hr at 55°C with proteinase K (1 μl 20 μg/ml)/ChIPmentation elution buffer (19 μl). The combined eluates were topped up with 300 μl of ERC and ChIPed DNA purified by a mini Elute kit (QIAGEN). ChIPed DNA was eluted in 10.5 μl of EB buffer. The sequencing libraries were generated by 12 cycles of PCR reaction using adaptor primers (Nextera Index kit; Illumina, FC-121-1012) and NEBNext high fidelity 2x PCR master mix (NEB, M0541), followed by two-sided size selection by Ampure XP beads purification (0.65x reaction volume then 1.2x reaction volume; Beckman Coulter Agencourt, A63881). The typical yield is between 150-300 ng. The sequencing was performed on an Illumina Hi-seq 4000 genome analyzer according to the manufacturer's protocols.

RNA-Seq Assays

Total RNA was prepared using RNAeasy Plus Mini kit (RNase-free DNase set; QIAGEN, 74134) according to the manufacturer's protocols except extra "in column" DNase digestion was performed (QIAGEN, 79254). Libraries for RNA-seq were generated using the Illumina TruSeq RNA library prep kit v2 (Illumina, RS-122-2001) and sequencing was performed on an Illumina Hi-seq 4000 genome analyzer according to the manufacturer's protocols (ArrayExpress accession: E-MTAB-7206).

Single-Cell (sc) RNA-Seq Assays

Single cell RNA-seq was performed on the ICELL8 single-cell RNA-seq system as described previously (Goldstein et al., 2017) except that cryogenically frozen cells were used (ArrayExpress: E-MTAB-7211).

Immunofluorescence Assays

Cells were seeded on 24 well-plate acid-treated glass slides at $2.5x10^4/cm^2$ in EpiLC media for 1 and 2 days (d1EpiLC, d2EpiLC). Cells were washed twice with PBS and fixed with 4% paraformaldehyde treatment for 10 min at room temperature (RT). Following washing three times with PBS, the cells were permeablised with 0.1% Triton X-100/PBS 20 min at RT then washed three times with PBS. The cells were then blocked with 5% normal goat serum/PBS for 45 min at RT, and then stained with diluted primary ZIC3 antibody (2 μg/ml final) in blocking solution at 4°C overnight. The cells were then washed four times with PBS, and then incubated with a secondary antibody (1/500 dilution, Alexa Fluor 488 goat anti-rabbit) in blocking solution at RT for 1 hr. The cells were then washed four times with PBS, and mounted in Prolong Gold antifade reagent with DAPI (Invitrogen P36941).

ESC Clonogenicity Assay

To identify alkaline phosphatase positive ESC-like clones, we used the method described previously (Kalkan et al., 2017). Naive ESCs were reverse-transfected with siRNA at 100 nM in fibronectin-coated 12 well-plates in EpiLC media. Cells were re-seeded 15 hr after transfection at 800 cells per well in gelatin-coated 6-well plates in 2i/LIF media. Six days later, plates were fixed and stained for AP (Sigma, 86R-1KT). Images were taken by microscope and colonies were counted using FiJi-ImageJ.

QUANTIFICATION AND STATISTICAL ANALYSIS

Bioinformatics and Statistical Analysis

All software was run with default settings, unless otherwise indicated. Raw sequencing reads (76-nt length; paired end) were trimmed and filtered using Trimmomatic v0.32 with paired-end mode to remove adapters, truncated reads (3′) and reads with < 25 nucleotides (TRAILING:5 SLIDINGWINDOW:4:15 MINLEN:25; Bolger et al., 2014). Filtered reads were mapped against National Center for Biotechnology Information build 37/mm9 of mouse genome using Bowtie2 v2.3.0 (allow up to two mismatches, *-X 2000 and –dovetail*; Langmead et al., 2009). Unmapped pairs (- F 4) were discarded using SAMtools v1.3.1 (Li et al., 2009). Reads were then de- duplicated using the *MarkDuplicates* function of the Picard tools (http://broadinstitute.github.io/picard/). Only reads that were uniquely mapped to the genome were preserved (MAPQ ≥ 30). The reads mapped to the mitochondrial genome (sed '/chrM/d') and overlapping with mm9 blacklist regions (intersectBed –v) were removed. The normalized tag density profiles were generated using HOMER (annotatePeak.pl; Heinz et al., 2010) and were plotted using a customised R script. Heatmaps were generated using Java treeview (Eisen et al., 1998). The UCSC tracks were generated by genomeCoverageBed for ATAC-seq normalized to total reads in peaks (RIPs) (BEDtools; Quinlan and Hall, 2010), MACS2 for ChIPmentation normalized to total tags (Zhang et al., 2008) or RSeQC for RNA-seq (Wang et al., 2012).

ATAC-Seq Data Analysis

Data from two biological replicates were first compared to check for concordance ($R^2 > 0.96$; Figure S1), and then merged into a single read file for each time point. ATAC-seq peaks (open-chromatin regions) were then called using MACS2 (Zhang et al., 2008) on individual replicates with the following parameters: -q 0.01–nomodel–shift −75–extsize 150. The high confidence peak sets were selected from biological replicates using the *intersectBed* function from BEDTools (Quinlan and Hall, 2010) with parameters -f 0.50, -r. This ensures a reciprocal overlap of > 50% between the two peaks being selected. To get a union set of peaks from all three conditions (ESC, d1EpiLC and d2EpiLC), high confidence MACS peaks from each condition were merged using *mergePeaks* module from HOMER (d = 100; Heinz et al., 2010) so only a single peak was retained when two or more peaks from different conditions had peak to peak distance < 100 bp). All downstream analysis was based on this union set of 238,236 peaks.

For identifying differentially accessible regions and fuzzy cMeans clustering, the union set of peaks was divided into promoter (−2 kb to +0.5 kb), intragenic- (defined by peaks located within mm9 protein coding regions) and intergenic-regions (all remaining peaks). Read counts for all peaks in the union set were obtained using the *annotatePeaks* module of HOMER package (Heinz et al., 2010) and were quantified using edgeR (Robinson et al., 2010). Fuzzy cMeans clustering using the R Mfuzz package (Kumar and Futschik, 2007) was then performed on each set of ATAC-accessible peaks identified in the Promoter, Intergenic and Intragenic regions, respectively. Initially, the Fuzzy cMeans clustering was performed to classify peaks into 12 clusters, which were subsequently merged into 4 clusters upon manual inspection. The final differentially accessible peaks were filtered based on EdgeR analysis (minimal CPM ≥ 4 in any of the three conditions), q < 0.05 and fold change ≥ 2 (promoter peaks) ≥ 2.5 (intergenic- and intragenic-peaks) (on at least one pairwise comparison between conditions).

To determine motif enrichment in clustered regions, over represented transcription factor motifs in each of the four clusters were identified using findMotifsGenome module of the HOMER package (Heinz et al., 2010). Motifs were then clustered using Fuzzy cMeans clustering and were also assigned to their respective families using the STAMP tool (Mahony and Benos, 2007). The relative enrichment scores of the clustered motifs were then transformed to Z-scores and plotted as heatmaps using the R *pheatmap* package.

To identify the transcription factors undergoing substantial changes in occupancy levels and chromatin accessibility around their binding sites between the transition states we used BaGFoot (Baek et al., 2017) software on the clustered ATAC-seq peaks. ATAC-seq peaks from all 4 clusters were merged for a reliable detection of footprint depth with robust statistical significance. We collected transcription factors from the JASPAR (Khan et al., 2018) database, which were manually curated to exclude transcription factors from non-vertebrate species, giving us 872 transcription factor motifs. The mouse genome (mm9) was scanned for motif occurrences of these transcription factors using Find Individual Motif Occurrences (*FIMO*) (Grant et al., 2011) as recommended by the software (1.5 M motif threshold count). We performed pairwise comparisons for the transition states (d1EpiLC versus ESC and d2EpiLC versus ESC) and calculated the changes in accessibility and footprint-depth. Results are displayed as bagplots.

ChIPmentation Data Analysis

ChIPmentation data was compared to input chromatin and peaks were called on each replicate using MACS2 v2.1.1 using parameters: *–keep-dup all -q 0.01 -g mm -f BAMPE -B–SPMR–call-summits* (Zhang et al., 2008). The high confident peak set (peaks identified in both biological replicates) was selected using *mergePeaks* module from HOMER (d = 400, peak summit distance = 400; Heinz et al., 2010). Similarly, the *mergePeaks* (d = 250, peak summit distance = 250) was used to subset peaks that overlapped with differential accessible ATAC-peaks.

Motif discovery and the significance of discovered motifs was performed by HOMER (findMotifsGenome.pl; Heinz et al., 2010) using the sequences within ± 100 bp around the binding region summits, using the default background setting i.e., sequences randomly selected from the genome with the same GC content as the target sequences.

Nearest genes were assigned to peaks and the Gene Ontologies (GO) were analyzed using GREAT (McLean et al., 2010). Genomic distributions were determined using HOMER (Heinz et al., 2010).

RNA-Seq Data Analysis

A manually curated gtf file was built for expression quantification of all datasets. Briefly, the gtf (vM1) file for mm9 from the GENCODE website was downloaded and genes specified by transcript_type (protein_coding, lincRNA and antisense) were retained. In addition, genes missing from the GENCODE gtf file but in ENSEMBL gtf file were added to our list. After manual filtering and inspection, the gtf file comprises of 25,875 unique ensembl id's and 25,753 unique gene symbols.

Filtered paired-end reads were mapped to the mouse genome (mm9 assembly) using STAR v2.5.3a (Dobin et al., 2013) with the manually curated mm9-gtf file and default parameters. Ribosomal RNA (rRNA) reads were removed from the mapped files. Read counts for each sample were quantified using HTSeq v0.9.1 (Anders et al., 2015), which estimates number of reads mapped to each gene. The raw read counts from the HTSeq were subsequently used to quantify the differential expression levels for genes using DESeq2 v1.18.1 (Anders and Huber, 2010). Data were taken as significant if a fold change of > 1.2 and p value < 0.05 was obtained. Additional genes were included in our analysis above this p value threshold if they changed expression in a consistent direction in paired samples and also exhibited a mean fold change > 1.2. For volcano plots, $\log_2$ fold changes (FC) of differentially expressed genes were plotted against their $\log_{10}$ p- values using the inbuilt function of the R statistical package. The Gene Ontology (GO) analyses were performed using DAVID (Huang et al., 2009).

Single-Cell Transcriptomics

Single cells from three samples (ECS, d1EpiLCs and d2EpiLCs) were captured and isolated using the ICELL8 single cell system. A custom script was used to perform assignment and error correction of cell barcodes/UMIs, low quality reads trimming and to run a cross species contamination checking. After the QCs, reads were aligned to a customized mouse reference genome of mm9 using STAR aligner (v2.4.2a). Reads aligning to genes were counted using HTSeq (v0.6.1.p1) with setting the stranded option to "yes." This count matrix was then used for the downstream analysis of the dataset using statistical computing programming language R.

We implemented two measures of cell quality control (cell QC) based on library size and number of expressed genes. If the total read count of a cell is below 3x median absolute deviation (MAD) of the dataset then the cell was filtered out. Similarly, if the total number of genes expressed by a cell is lower than 3x MAD those cells were also filtered out. A further QC was done to filter out any cells that are outliers in terms of library size on the higher end as this could indicate doublets of cells. After all these filtering steps, a total of 816 cells (from 869) were left for downstream analysis.

Subsequently, the lowly expressed genes were filtered out if their average counts are less than 0.05 (raw counts) meaning a gene has to be expressed in at least 5% of cells with 1 read count or higher counts in smaller number of cells still accounting for 5% equivalent cells, which gave us 12,695 genes for downstream analysis. The normalized count data is represented as counts per million where the size factors are used to calculate the effective library size. These size factors were defined from the actual library size after centering to unity. We used the calculateCPM function from the scater package (McCarthy et al., 2017) to perform this normalization.

To identify the highly variable genes (HVGs) we first estimated the total variance in expression of each gene which is then decomposed into technical and biological components. We fitted a mean-variance trend to the expression of endogenous genes and then took those genes that have a larger biological variance component with an FDR value less than 0.05 as our HVGs. 283 genes were identified as HVGs for this dataset. These HVGs were then used to construct dimensionality reduction processes using PCA. For the t-SNE plot (van der Maaten and Hinton, 2008), 10 principal components from this PCA are given as input and the perplexity is set to

60. In addition, the theta (a parameter for speed/accuracy tradeoff) is set to 0.01 to increase the accuracy of the plot. For generating the Uniform Manifold Approximation and Projection (UMAP) plots (McInnes et al., 2018), 10 PCs are taken as input.

We used the monocole (v2) package to perform the pseudotime estimation (Trapnell et al., 2014). As we know the three stages of cells in our samples we use this information to identify the order of the genes. Genes that are differentially expressed between ESCs, d1EpiLCs and d2EpiLCs with a q-value less than 7.5e-08 were identified and subsequently used to order cells. We then apply DDRTree method to reduce the dimension of the dataset (Qiu et al., 2017). The pseudotime trajectory is visualized in the reduced dimension.

Co-expression scores across single cell RNA-seq data were calculated by first giving a binary score to the expression of each gene in each cell. These binary scores were summed for each cell, and then z transformed. The data are shown as boxplots for ESCs, d1EpiLCs and d2EpiLCs in Figure 4G after correcting for the numbers of cells at each condition.

To examine the Jaccard Similarity Index (JSI), Jaccard's distance was computed for the binary co-expression matrix of all ZIC3 activated TFs in each cell type (ESC, d1EpiLC, and d2EpiLC2) using the R package *ade4* (Dray and Dufour, 2007). The dissimilarity matrix was then converted to a similarity matrix by using the expression JSI = 1- $(JD)^2$ (where JD = Jaccard's distance) and the JSI based data were plotted as heatmaps. First, the d1EpiLC data was clustered on both row and columns using hierarchical clustering using heatmap.2 from R gplots package. The heatmaps generated for ESC and d2EpiLC data were plotted using the same gene ordering that was obtained from clustering of the d1EpiLC data.

The Matthew's Correlation Coefficient (MCC) for the co-expression data of ZIC3 activated genes that encode transcription factors, were computed using a custom R script. The correlation matrix for d1EpiLC was then ordered for the first principal component using R package *corrplot* (Wei and Simko, 2017) and plotted as a heatmap after clustering the rows and columns. As in the JSI plots, the gene order from the d1EpiLC heatmap was retained and the heatmaps for ESC and d2EpiLC data were plotted without further clustering.

Beeswarm plots for Jaccard Similarity Indices of ZIC3 activated target genes in each cell type were plotted using R package *Beeswarm* (https://rdrr.io/cran/beeswarm).

To interrogate the co-expression of genes in the ZIC3 regulon (ZIC3 targets) in our scRNA-seq data, the AUCell module from SCENIC R package was used (Aibar et al., 2017). AUCell calculates the enrichment of gene-sets (regulon) as area under the recovery curve (AUC) based on the rankings of all genes expressed in a particular cell. The AUC threshold was then determined and subsequently used to mark whether the cells contained an active- or inactive-regulon. This binary data was then visualized by superimposing onto t-SNE plots.

To verify the importance of the ZIC3 regulon during mouse embryogenesis, a scRNA-seq dataset generated during early mouse gastrulation was used (Mohammed et al., 2017). The raw reads were mapped to the same custom mm9 gft file and analyzed as described above. HVGs were used to generate the PCAs and 14 PCs were used as input for the t-SNE plot. For the t-SNE plot, the perplexity was set to 60 and theta to 0.01. In addition, 135 direct ZIC3-activated genes were selected as input to cluster cells using the t-SNE plot with the same perplexity and theta value. As a control, 135 randomly selected genes were used.

To generate pseudo-bulk RNA datasets from single cell data, the aggregated counts of each gene from each cell of the scRNA-seq were generated and quantified using the edgeR (Robinson et al., 2010). The Fuzzy cMeans clustering was performed to generate 4 broad expression clusters as described above. 8659 genes were selected for further downstream analysis (CMP $\geq$ 2, fold change > 1.5 at any pairwise comparison).

To correlate Fuzzy cMeans-generated clusters of ATAC-seq and RNA-seq data (Figures S2E and S2F), we first identified genes whose TSSs lie within a given genomic distance constraint from any peak within a given cluster of ATAC-seq peaks (Ai). Next we take the RNA-seq based gene clusters (Rj), and calculate the intersection of Rj with Ai (observed set). The expected set of genes was defined as all genes within a given genomic distance of an ATAC-seq cluster from a randomly selected number (according to comparator expression cluster size) of genes extracted from the background population (all mm9 genes). The p values for the enrichment were subsequently calculated using a hypergeometric test between the observed and expected datasets. This calculation is repeated for all to all combinations between ATAC-seq peak clusters and gene expression clusters, and for each of pre-defined set of genomic distances relative to the TSS (8 bins within the range of ± 10 kb to ± 250 kb from the center of the peak). The resulting p values are log-transformed (-$\log_{10}$(p values)) and shown as heatmaps in figures.

Statistical Analysis of the ESC Clonogenicity Assay

A Student's t test was done to compare the significance of the differences in colony numbers following treatment with control siRNA or *Zic3* siRNA under parameters with paired conditions and a 1-tailed test (n = 4).

DATA AND SOFTWARE AVAILABILITY

ATAC-seq data have been deposited in the ArrayExpress repository under accession number: ArrayExpress: E-MTAB-7207. ZIC3 ChIPmentation-seq has been deposited in the ArrayExpress repository under accession number: ArrayExpress: E-MTAB-7208. RNA-seq data have been deposited in the ArrayExpress repository under accession number: ArrayExpress: E-MTAB-7206. Single cell (sc) RNA-seq data ArrayExpress: E-MTAB-7211).

Cell Reports, Volume *27*

Supplemental Information

ZIC3 Controls the Transition from Naive to Primed Pluripotency

Shen-Hsi Yang, Munazah Andrabi, Rebecca Biss, Syed Murtuza Baker, Mudassar Iqbal, and Andrew D. Sharrocks

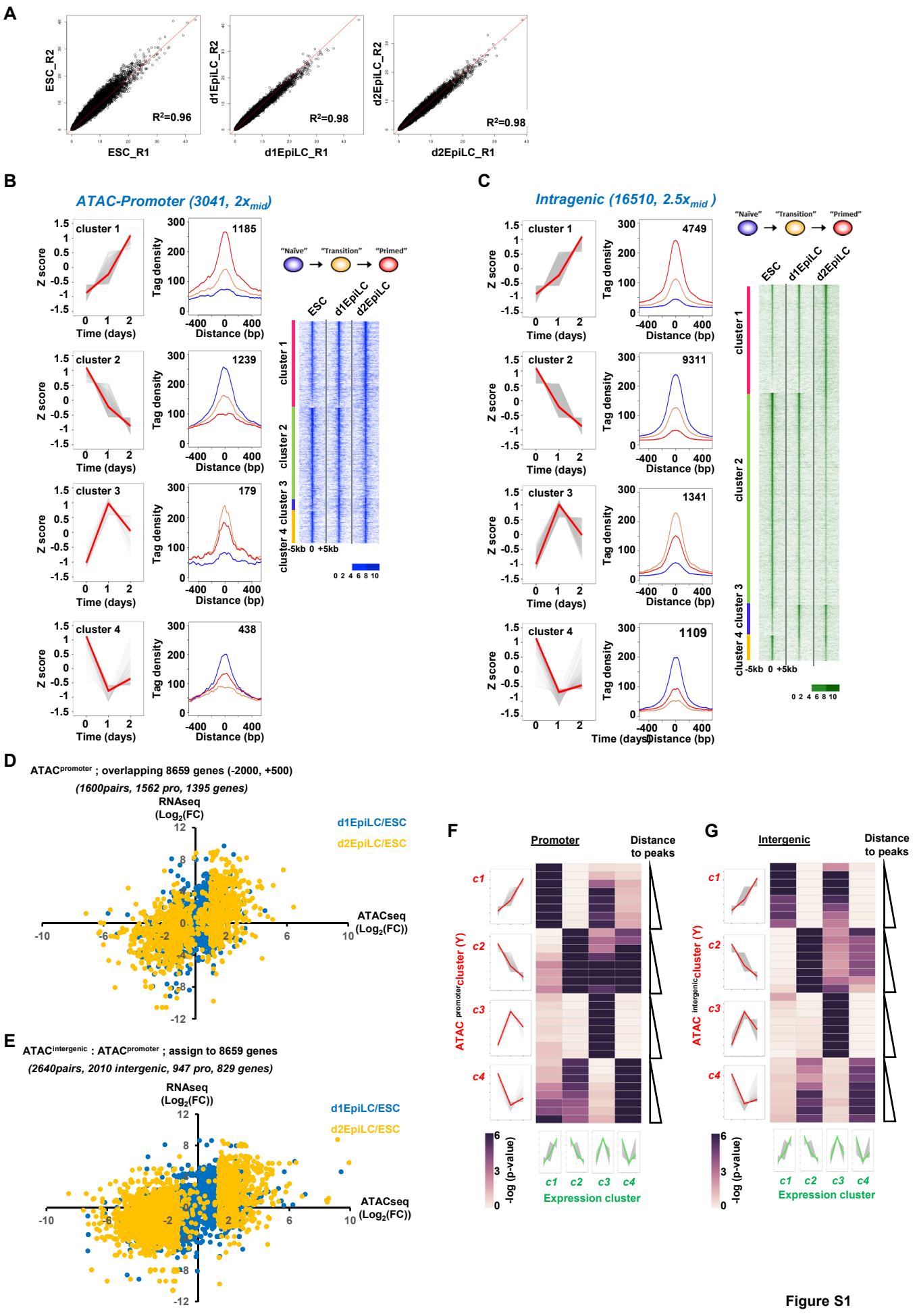

A
ESC_R2
ESC_R1
R²=0.96
d1EpiLC_R2
d1EpiLC_R1
R²=0.98
d2EpiLC_R2
d2EpiLC_R1
R²=0.98
B
ATAC-Promoter (3041, 2x mid)
cluster 1
cluster 2
cluster 3
cluster 4
Z score
Time (days)
Tag density
Distance (bp)
1185
1239
179
438
"Naïve"
"Transition"
"Primed"
ESC
d1EpiLC
d2EpiLC
-5kb 0 +5kb
C
Intragenic (16510, 2.5x mid)
4749
9311
1341
1109
D
ATACpromoter ; overlapping 8659 genes (-2000, +500)
(1600pairs, 1562 pro, 1395 genes)
RNAseq
(Log2(FC)
d1EpiLC/ESC
d2EpiLC/ESC
ATACseq
(Log2(FC))
E
ATACintergenic : ATACpromoter ; assign to 8659 genes
(2640pairs, 2010 intergenic, 947 pro, 829 genes)
F
Promoter
Distance to peaks
ATAC promoter cluster (Y)
-log (p-value)
Expression cluster
G
Intergenic
ATAC intergenic cluster (Y)

Figure S1

Figure S1. Open chromatin profiling during the ESC to EpiLC transition. Related to Figure 1. (A) Scatter plots of ATAC-seq data from two independent biology repeats (R1 and R2) from ESCs, d1EpiLCs and d2EpiLCs. The numbers of reads from the R1 and R2 repeats at each time point were determined in a consolidated peak set from all time points (362,008 peaks). R2 values are shown. (B and C) Heat maps of the ATAC-seq profiles across a 10 kb window of promoter regions (-2 to +0.5 kb relative to the TSS)(B) and intragenic regions (C) showing >2 fold (promoters) or >2.5 fold (intragenic regions) change in accessibility between any two conditions (right). Average tag densities of each of four identified clusters (middle; blue=ESC, orange=d1EpiLCs, red=d2EpilCs) and average tag density profiles (z scored and medians [red] and standard deviation [grey] shown) across the time course (left) are shown. (D and E) Scatter plots comparing ATAC-seq signals from individual peaks to the RNA-seq signals associated with their associated genes. Data are plotted as $\log_2$ fold change between two conditions (blue for d1EpiLCs versus ESCs and orange for d2EpiLCs versus ESCs). Promoter-located ATAC-seq peaks (D) or intergenic located ATAC-seq peaks (E) were analysed. ATAC-peaks showing differentially accessibility between any two time points were assigned to 8659 differentially expressed genes (defined as >1.5 fold change between any time point) by 1 bp overlap with their TSS (for promoter-located ATAC-seq peaks) or by first locating the closest differentially changing ATAC-seq peak at a promoter and then identifying the associated gene by 1 bp overlap with a TSS (for intergenic-located ATAC-seq peaks). The total number of ATAC-seq peak pairs to promoters (pro) and genes is indicated. Pearson's correlations are 0.428 (d1EpiLC) and 0.502(d2EpiLC) (D), 0.424(d1EpiLC) and 0.538(d2EpiLC) (E). (F and G) Enrichment of genes associated with each ATAC-seq peaks cluster, (F; promoter and G; intergenic)(illustrated on the left), and each gene expression based cluster (illustrated on the bottom). These enrichments are calculated at different peak-to-gene genomic distances (+/- 10, 25, 50, 100, 125, 150, 200 and 250 kb; increasing distances illustrated by a triangle). The colours represent the p-values calculated from a hypergeometric model.

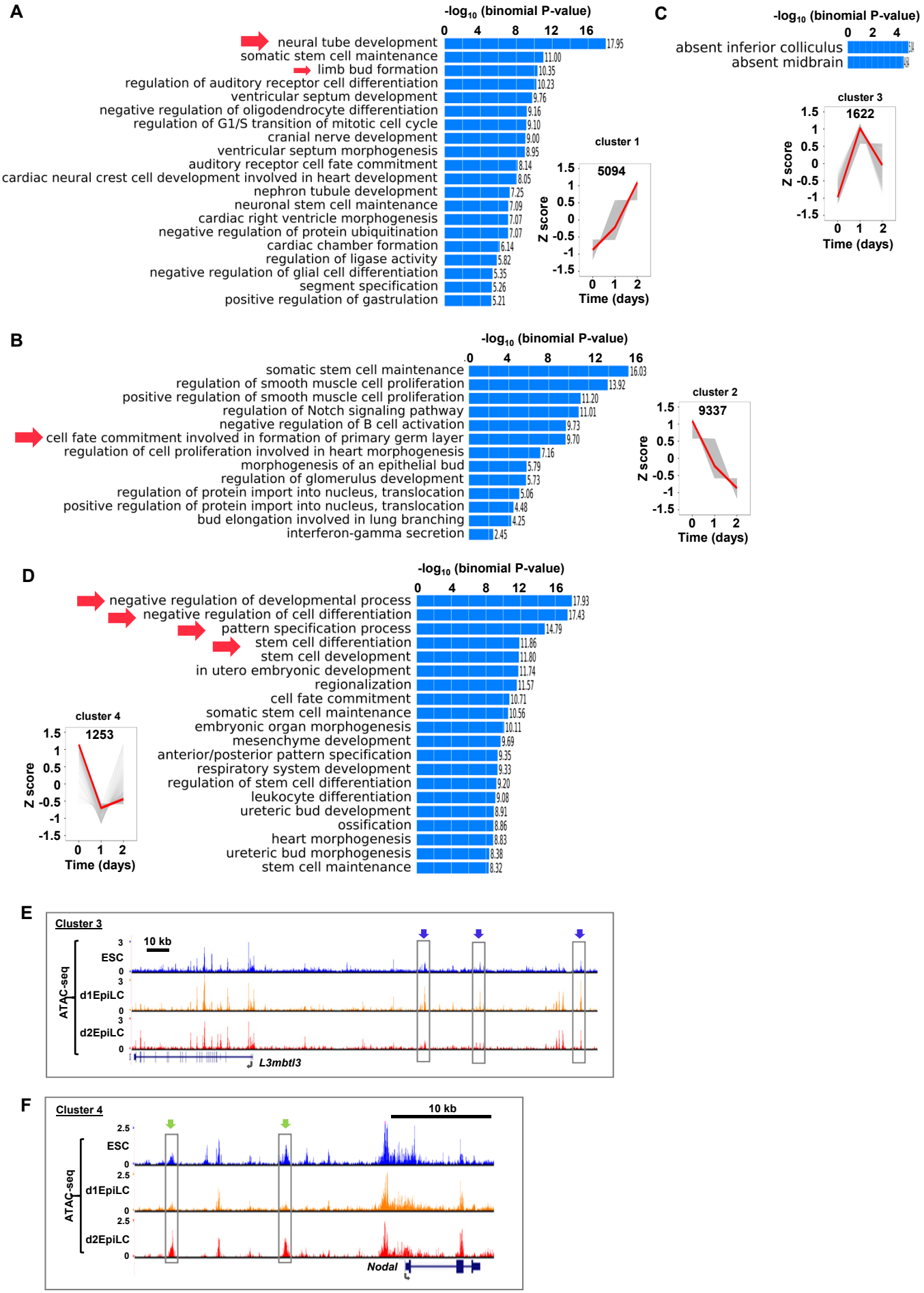

A
-log10 (binomial P-value)
0 4 8 12 16
neural tube development 17.95
somatic stem cell maintenance 11.00
limb bud formation 10.35
regulation of auditory receptor cell differentiation 10.23
ventricular septum development 9.76
negative regulation of oligodendrocyte differentiation 9.16
regulation of G1/S transition of mitotic cell cycle 9.10
cranial nerve development 9.00
ventricular septum morphogenesis 8.95
auditory receptor cell fate commitment 8.14
cardiac neural crest cell development involved in heart development 8.05
nephron tubule development 7.25
neuronal stem cell maintenance 7.09
cardiac right ventricle morphogenesis 7.07
negative regulation of protein ubiquitination 7.07
cardiac chamber formation 6.14
regulation of ligase activity 5.82
negative regulation of glial cell differentiation 5.35
segment specification 5.26
positive regulation of gastrulation 5.21
cluster 1
5094
Z score
Time (days)
C
-log10 (binomial P-value)
0 2 4
absent inferior colliculus
absent midbrain
cluster 3
1622
Z score
Time (days)
B
-log10 (binomial P-value)
0 4 8 12 16
somatic stem cell maintenance 16.03
regulation of smooth muscle cell proliferation 13.92
positive regulation of smooth muscle cell proliferation 11.20
regulation of Notch signaling pathway 11.01
negative regulation of B cell activation 9.73
cell fate commitment involved in formation of primary germ layer 9.70
regulation of cell proliferation involved in heart morphogenesis 7.16
morphogenesis of an epithelial bud 5.79
regulation of glomerulus development 5.73
regulation of protein import into nucleus, translocation 5.06
positive regulation of protein import into nucleus, translocation 4.48
bud elongation involved in lung branching 4.25
interferon-gamma secretion 2.45
cluster 2
9337
Z score
Time (days)
D
-log10 (binomial P-value)
0 4 8 12 16
negative regulation of developmental process 17.93
negative regulation of cell differentiation 17.43
pattern specification process 14.79
stem cell differentiation 11.86
stem cell development 11.80
in utero embryonic development 11.74
regionalization 11.57
cell fate commitment 10.71
somatic stem cell maintenance 10.56
embryonic organ morphogenesis 10.11
mesenchyme development 9.69
anterior/posterior pattern specification 9.35
respiratory system development 9.33
regulation of stem cell differentiation 9.20
leukocyte differentiation 9.08
ureteric bud development 8.91
ossification 8.86
heart morphogenesis 8.83
ureteric bud morphogenesis 8.38
stem cell maintenance 8.32
cluster 4
1253
Z score
Time (days)
E
Cluster 3
10 kb
ATAC-seq
ESC
d1EpiLC
d2EpiLC
L3mbtl3
F
Cluster 4
10 kb
ATAC-seq
ESC
d1EpiLC
d2EpiLC
Nodal

Figure S2

Figure S2. Gene ontology analysis of genes associated with clusters of dynamically changing ATAC-seq peaks. Related to Figure 1. (A-D) Gene ontology analysis of genes associated with the dynamically changing intergenic ATAC-seq peaks from clusters 1-4 (shown as insets; see Fig. 1B). Enriched terms associated with the biological process subcategory are shown. (E and F) UCSC genome browser views of the ATAC-seq profiles around representative gene loci associated with cluster 3 (*L3mbtl3*; E) and cluster 4 (Nodal; F) intergenic peaks. Dynamically changing peaks are boxed.

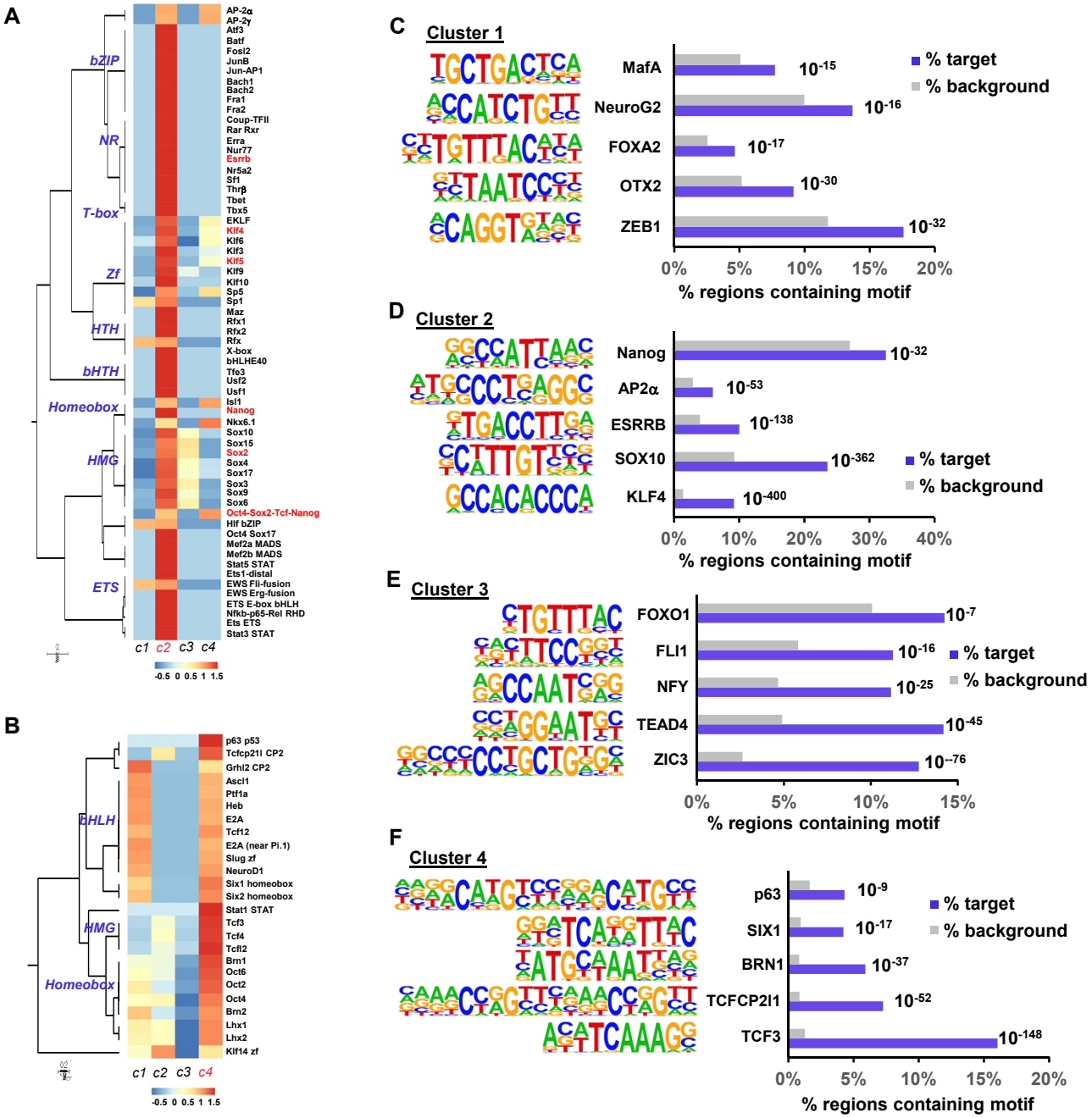


Figure S3. DNA motif enrichment in clusters of dynamically changing ATAC-seq peaks. Related to Figure 1. (A and B) Heat map showing the enrichment of transcription factor binding motifs across each of the open chromatin cluster profiles (z-normalised p-values) for motifs enriched in cluster 2 (A) or cluster 4 (B). (C-F) The top five enriched DNA motifs identified in each cluster of dynamically changing ATAC-seq peaks. The percentage of each motif present in each set of peaks and the genomic background is indicated, and the p-value is shown next to each of the columns.

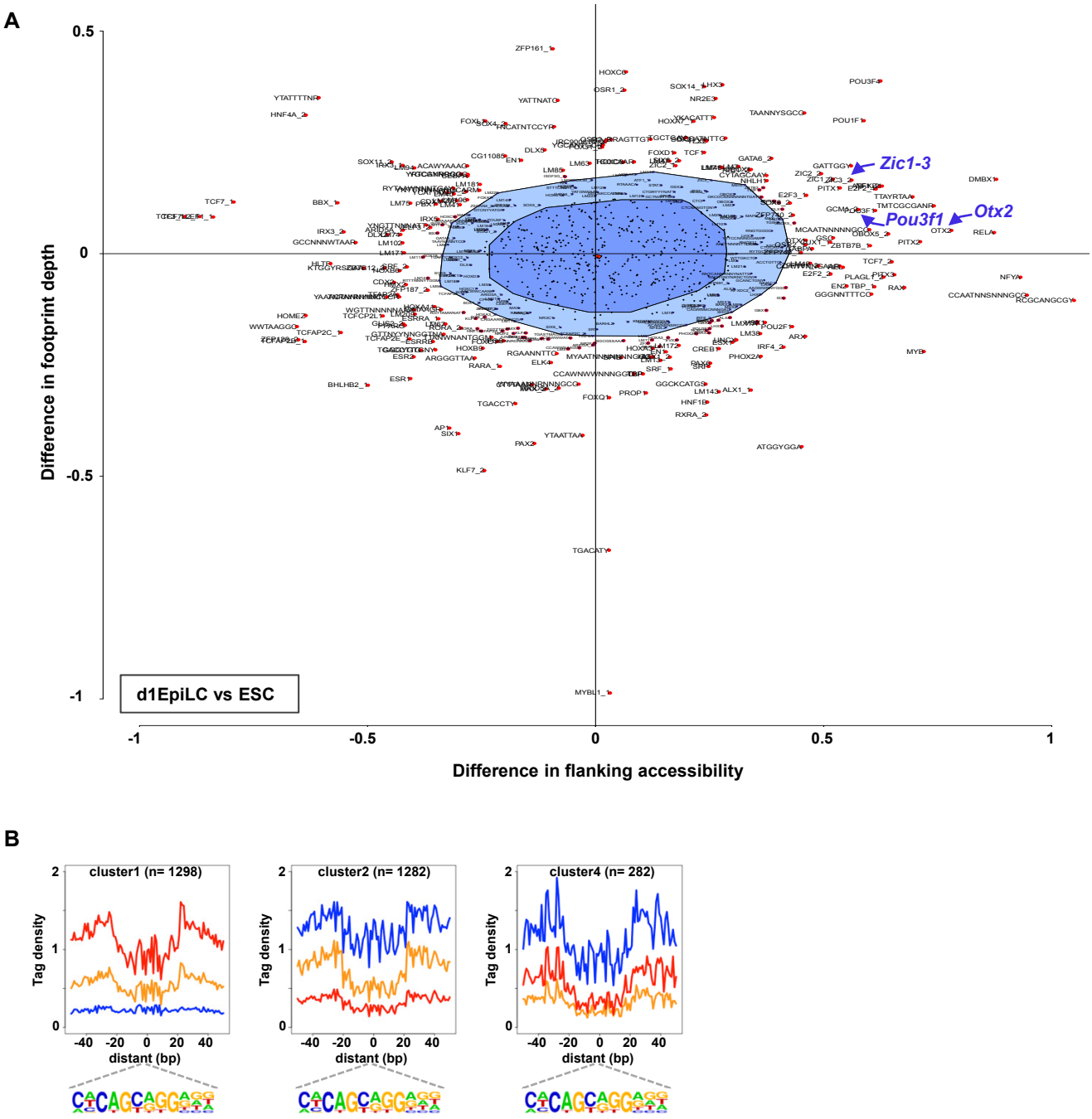


Figure S4. Dynamically changing chromatin accessibility around transcription factor binding motifs in d1EpiLCs. Related to Figure 1. (A) BagFoot analysis of the open chromatin regions in ESCs and d1EpiLCs. Motifs showing significant changes in either local accessibility and/or footprint depth are labelled. Non-significant regions are shown in the dark (bag) or light (fence) blue shaded regions. (B) Average tag densities in a 80 bp window surrounding the ZIC3 binding motif (bottom) in ESCs (blue), d1EpiLCs (orange) or d2EpilCs (red) are shown for the intergenic ATAC-seq peaks from each of the indicated clusters.

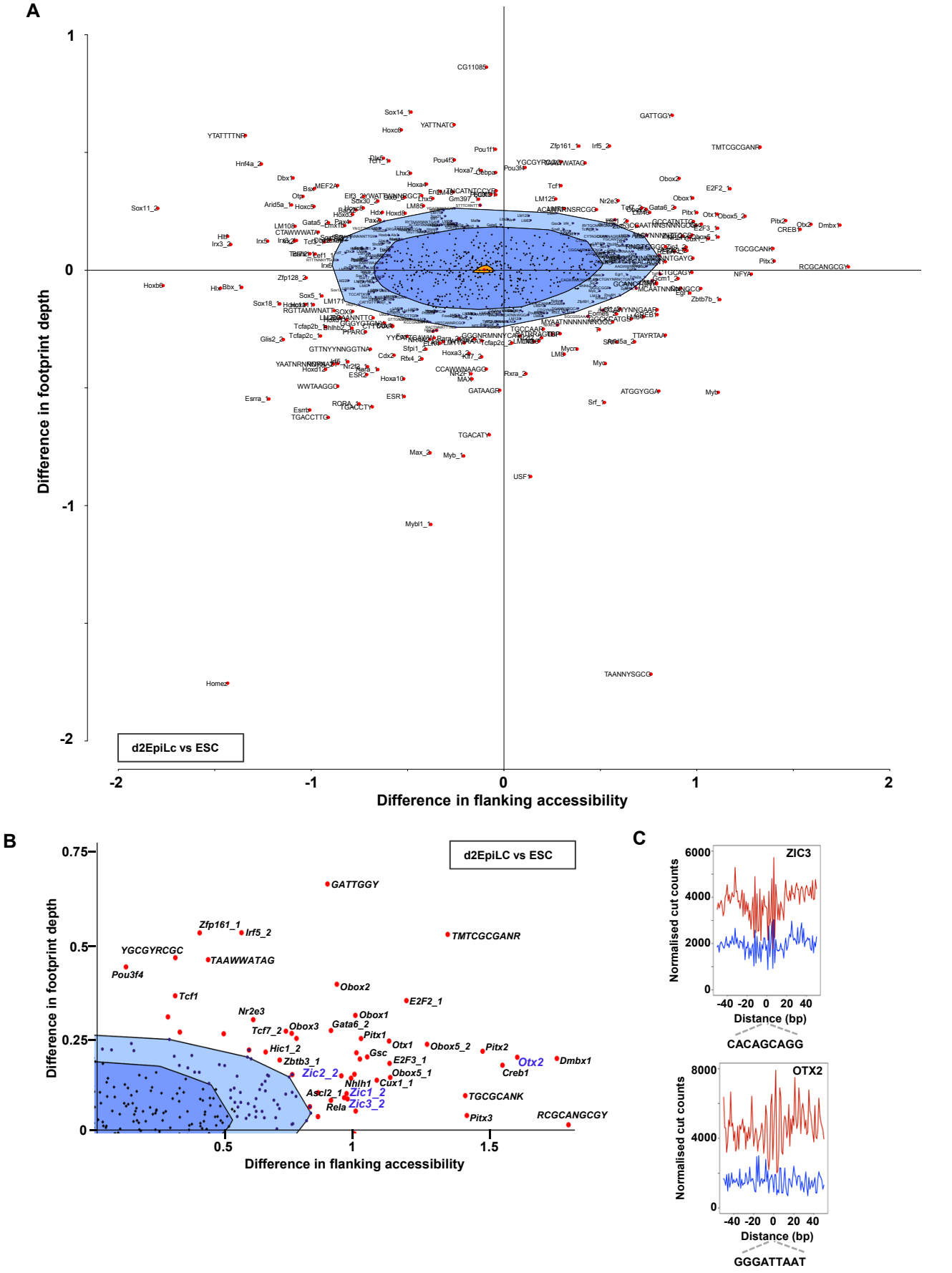


Figure S5. Dynamically changing chromatin accessibility around transcription factor binding motifs in d2EpiLCs. Related to Figure 1. (A) BagFoot analysis of the open intergenic chromatin regions in ESCs and d2EpiLCs. Motifs showing significant changes in either local accessibility and/or footprint depth are labelled. Non-significant regions are shown in the dark (bag) or light (fence) blue shaded regions. (B) Expanded version of the top right quadrant from (B). (C) Average tag densities in a 100 bp window surrounding the ZIC3 binding (top) or OTX2 (bottom) binding motifs in all peaks in ESCs (blue) or d2EpiLCs (red) are shown.

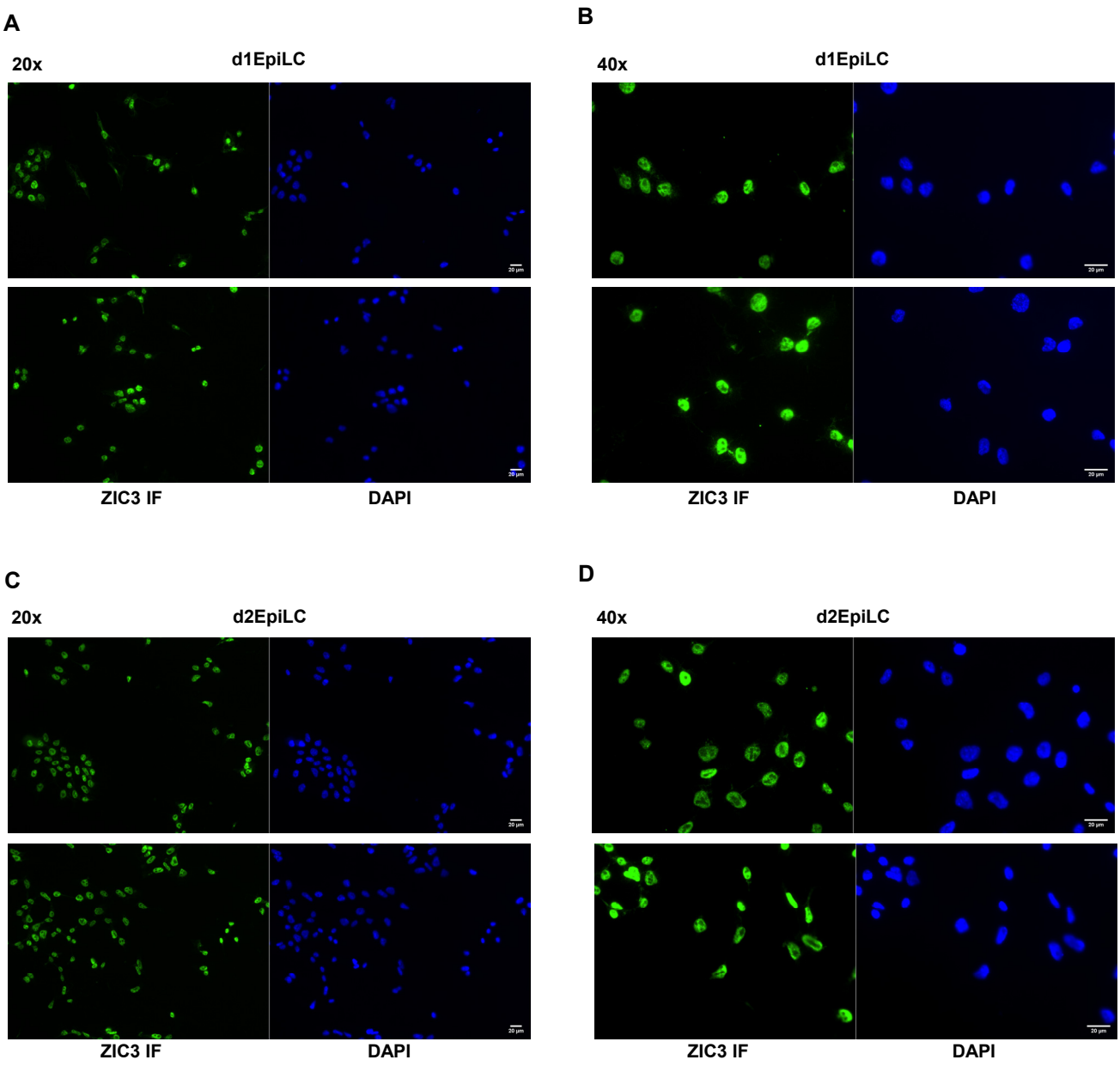


Figure S6. Immunofluorescence detection of ZIC3 in the nucleus. Related to Figure 2. ZIC3 protein was detected by immunofluorescence in d1EpiLCs (A and B) and d2 EpiLCs (C and D). Nuclei were identified by DAPI staining. Four representative fields are shown for each timepoint, with either 20x (A and C) or 40x (B and D) magnification. All cells contain nuclear ZIC3.

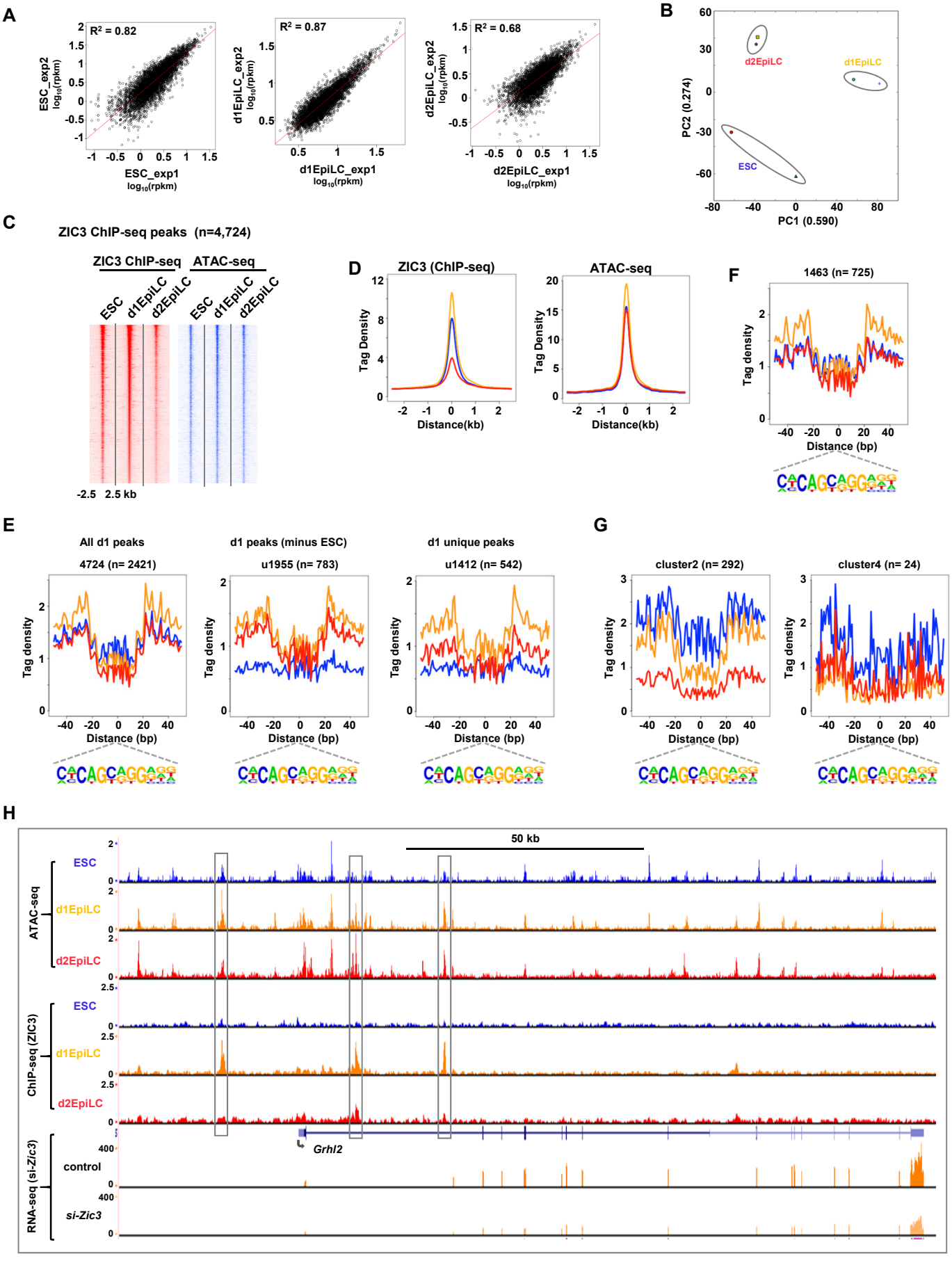

A
R² = 0.82
ESC_exp2
log10(rpkm)
ESC_exp1
log10(rpkm)
R² = 0.87
d1EpiLC_exp2
log10(rpkm)
d1EpiLC_exp1
log10(rpkm)
R² = 0.68
d2EpiLC_exp2
log10(rpkm)
d2EpiLC_exp1
log10(rpkm)
B
d2EpiLC
d1EpiLC
ESC
PC2 (0.274)
PC1 (0.590)
C
ZIC3 ChIP-seq peaks (n=4,724)
ZIC3 ChIP-seq
ATAC-seq
ESC
d1EpiLC
d2EpiLC
-2.5
2.5 kb
D
ZIC3 (ChIP-seq)
ATAC-seq
Tag Density
Distance(kb)
F
1463 (n= 725)
Tag density
Distance (bp)
E
All d1 peaks
4724 (n= 2421)
d1 peaks (minus ESC)
u1955 (n= 783)
d1 unique peaks
u1412 (n= 542)
Tag density
Distance (bp)
G
cluster2 (n= 292)
cluster4 (n= 24)
Tag density
Distance (bp)
H
50 kb
ATAC-seq
ESC
d1EpiLC
d2EpiLC
ChIP-seq (ZIC3)
ESC
d1EpiLC
d2EpiLC
Grhl2
RNA-seq (si-Zic3)
control
si-Zic3

Figure S7

Figure S7. ChIP-seq analysis of ZIC3 binding to chromatin in ESCs and d1EpiLCs. Related to Figure 3. (A) Scatter plots of the replicate ZIC3 ChIP-seq experiments performed in the indicated cell populations. Data are plotted based on the 4,724 high confidence ZIC3 binding regions in d1EpiLCs. Correlation coefficients (R^2) are indicated. (B) PCA analysis of the replicate ZIC3 ChIP-seq experiments performed in the indicated cell populations. (C) Heat map of the ZIC3 ChIP-seq profiles across a 5 kb window in each of the indicated cell populations (left). The corresponding ATAC-seq signals at each ZIC3 binding region are shown on the right. Data are ranked according to ZIC3 ChIP-seq signals in d1EpiLCs. (D) Average tag densities in a 5 kb window around all ZIC3 peaks (blue=ESC, orange=d1EpiLCs, red=d2EpilCs) for ZIC3 ChIP-seq signal (left) or ATAC-seq signal (right). (E-G) Average ATAC-seq tag densities in a 80 bp window surrounding the ZIC3 motif (bottom) in; (E) peaks partitioned according to their presence in d1EpiLCs (left), peaks found in d1EpiLCs but not ESCs (middle) or peaks found in d1ESCs but neither ESCs nor d2EpiLCs (right); (F) all ZIC3 ChIP-seq binding regions from clusters 1-4 in Fig. 3D; (G) ZIC3 binding regions from clusters c2 (left), or cluster c4 (right)(clusters defined in Fig. 3D). ATAC-seq tags from ESCs (blue), d1EpiLCs (orange) or d2EpilCs (red) are shown. (H) UCSC genome browser views of the ATAC-seq (top), ZIC3 ChIP-seq (middle) and RNA-seq (plus/minus *siZic3*; bottom) profiles around the entire *Grhl2* locus. Dynamically changing ZIC3 binding peaks are boxed.

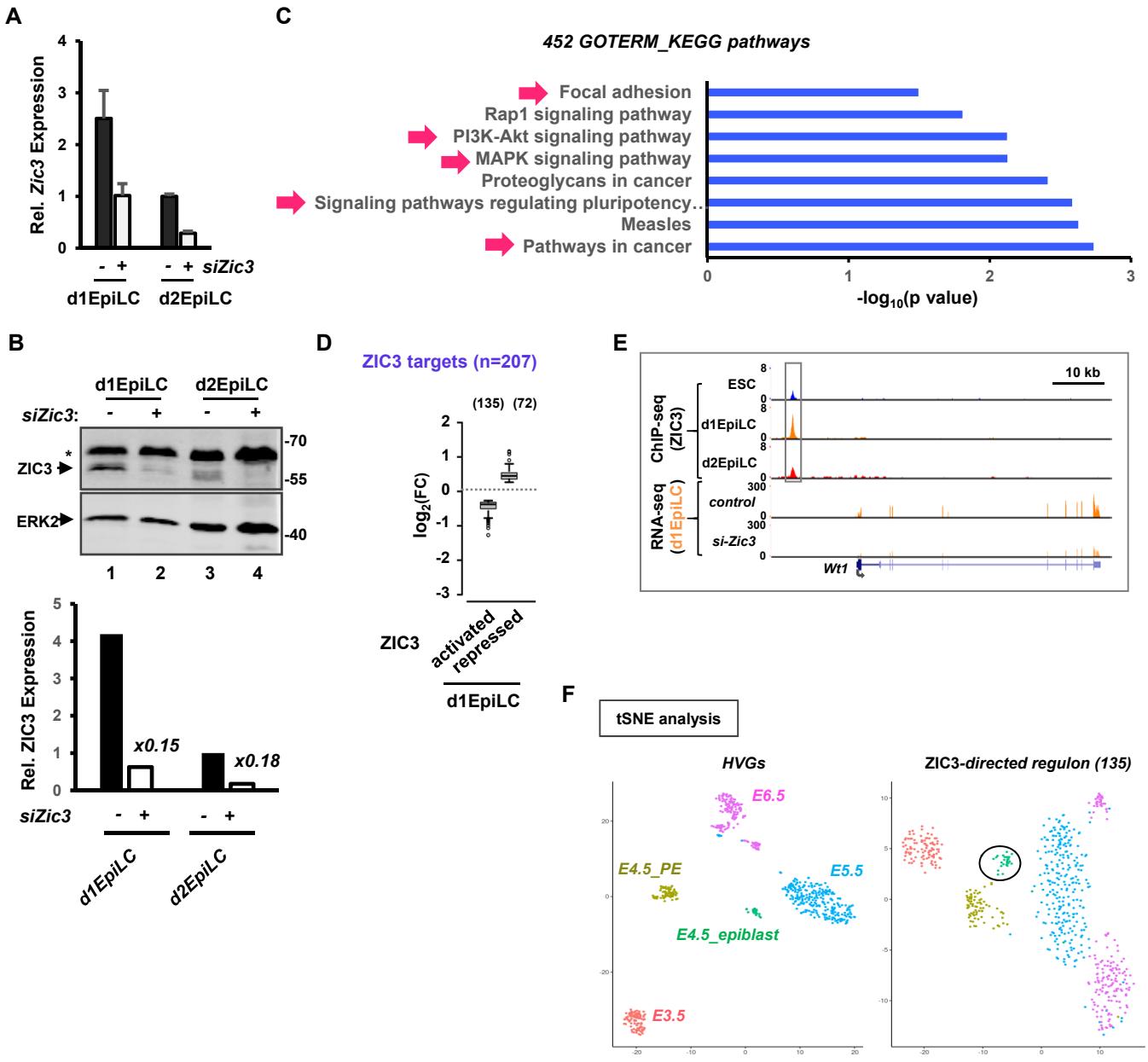


Figure S8. Properties of ZIC3-regulated genes. Related to Figures 4 and 5. (A and B) siRNA-mediated depletion of ZIC3. Levels of *Zic3* mRNA in d1EPiLCs and d2EpiLCs were measured by RT-qPCR (n=3) (A) and ZIC3 protein by western blot analysis (B). RT-qPCR data are shown relative to *Zic3* levels in d2EpiLCs treated with non-targeting control (-). ERK2 protein levels are shown as a loading control. The asterisk represents a cross reacting non-specific band. The protein expression data are quantified in the graph below and the fold change in expression caused by *Zic3* depletion is indicated. (C) Gene ontology analysis of ZIC3-regulated genes for the KEGG pathways. (D) Boxplot showing the overall changes in expression of genes directly bound by ZIC3 following *Zic3* depletion with siRNA in d1EpiLCs. The total number of genes going down (ie activated by ZIC3) and going up (ie repressed by ZIC3) is indicated above the graph. (E) UCSC genome browser views of the ZIC3 ChIP-seq (top) and RNA-seq (bottom) profiles around the *Wt1* locus. The major ZIC3 binding peak is boxed. (F) t-SNE analysis of scRNA-seq data from mouse embryos (Mohammed et al., 2017) using the HVGs from all of the scRNAseq data (left) or the 135 directly activated ZIC3 target genes (right). The colour coding from the HVG-derived t-SNE analysis and the cluster comprised of E4.5 cells is circled.

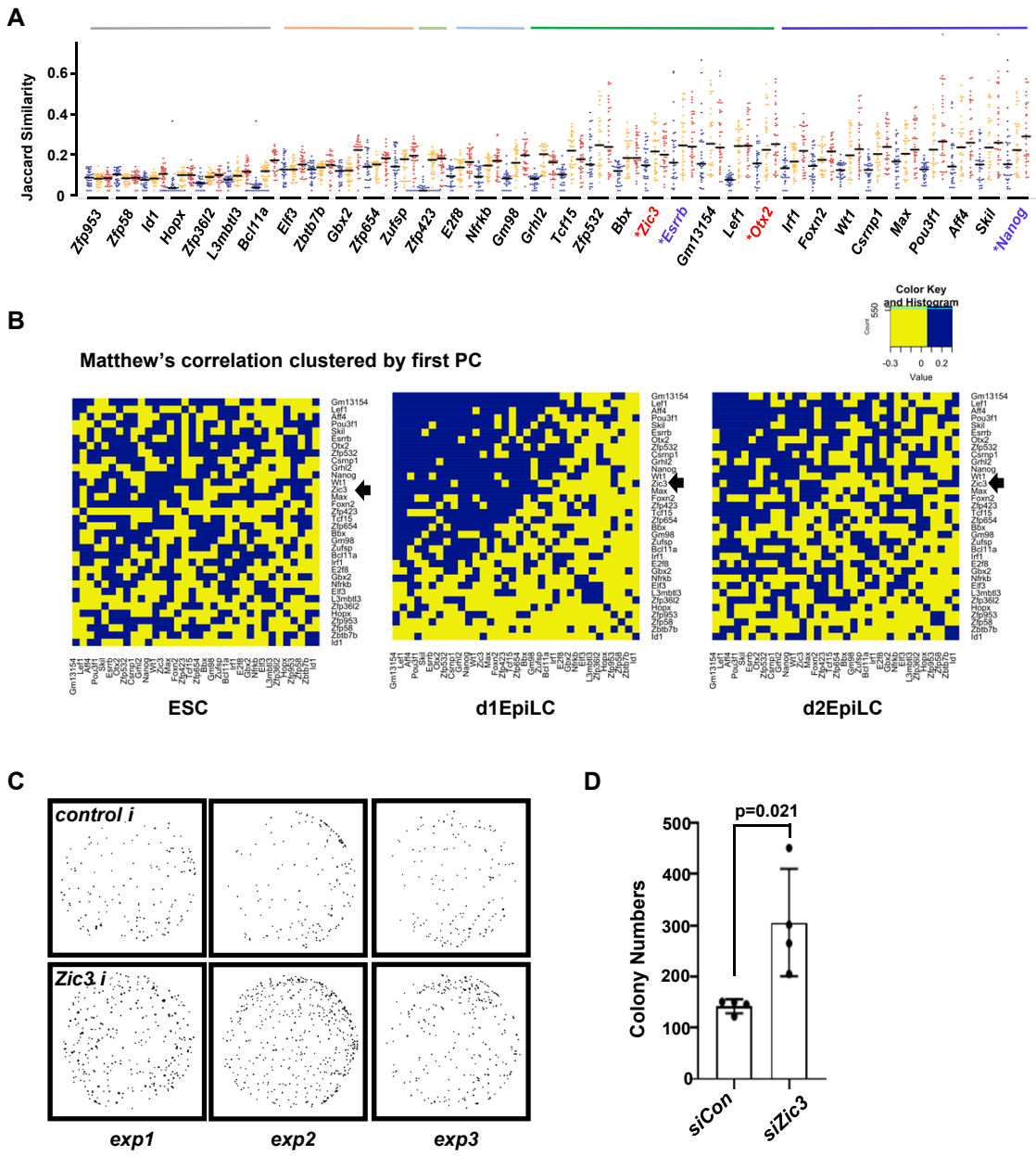


Figure S9. Co-expression analysis of ZIC3 activated transcription factors and phenotypic outcome of *Zic3* depletion. Related to Figure 6. (A) Jaccard's similarity score for each of the indicated ZIC3 activated target genes (in black font) to each other in ESCs (blue), d1EpiLCs (orange) and d2EpiLCs (Red). Black lines represent the median similarity scores. Only 31 target genes that encode transcription factors were included in this cross-comparison. The majority of genes exhibit higher similarity scores in d1EpiLCs which is often maintained in d2EpiLCs. (B) Matthew's correlation plots of the co-expression of the indicated transcription factor-encoding ZIC3 activated genes in ESCs (left), d1EpiLCs (middle) and d2EpilCs (right). The location of *Zic3* is highlighted with an arrow. (C and D) ESC clonogenicity assay. (C) Alkaline phosphatase-positive colonies were detected in ESCs grown in EpiLC differentiation media for 15 hours and returned to 2i/LIF medium. Cells were treated with either control siRNAs (top) or siRNAs targeting *Zic3* (bottom). (D) Quantification of the experiments showing individual data points from four independent experiments.