nature cell biology

ARTICLES

https://doi.org/10.1038/s41556-019-0453-8

Maternal factor NELFA drives a 2C-like state in mouse embryonic stem cells

Zhenhua Hu[1,8], Dennis Eng Kiat Tan[1,8], Gloryn Chia[2,8], Haihan Tan[1], Hwei Fen Leong[1], Benjamin Jieming Chen[1], Mei Sheng Lau[1], Kelly Yu Sing Tan[3], Xuezhi Bi[3,4], Dongxiao Yang[3], Ying Swan Ho[3], Baojiang Wu[5], Siqin Bao[5], Esther Sook Miin Wong[6] and Wee-Wei Tee[1,7]*

Mouse embryonic stem cells (ESCs) sporadically transit into an early embryonic-like state characterized by the expression of 2-cell (2C) stage-restricted transcripts. Here, we identify a maternal factor—negative elongation factor A (NELFA)—whose heterogeneous expression in mouse ESCs is coupled to 2C gene upregulation and expanded developmental potential in vivo. We show that NELFA partners with Top2a in an interaction specific to the 2C-like state, and that it drives the expression of *Dux*—a key 2C regulator. Accordingly, loss of NELFA and/or Top2a suppressed *Dux* activation. Further characterization of 2C-like cells uncovered reduced glycolytic activity; remarkably, mere chemical suppression of glycolysis was sufficient to promote a 2C-like fate, obviating the need for genetic manipulation. Global chromatin state analysis on NELFA-induced cells revealed decommissioning of ESC-specific enhancers, suggesting ESC-state impediments to 2C reversion. Our study positions NELFA as one of the earliest drivers of the 2C-like state and illuminates factors and processes that govern this transition.

During embryonic development, maternal factors facilitate the reprogramming of epigenetically distinct parental genomes towards totipotency and initiate zygotic gene activation (ZGA)[1–3]. In the mouse, ZGA peaks at the 2-cell (2C) stage[4] and is characterized by the activation of 2C-specific genes such as *Zscan4* (ref. [5]), *Dux*[6,7] and the endogenous retrovirus *MERVL*[8–10]. The identification of pivotal maternal factors is thus instrumental for understanding the molecular foundations of totipotency, which is thought to be exclusive to the zygote and early cleavage-stage embryos[11]. However, recent studies have shown that mouse embryonic stem cells (mESCs) are metastable and can give rise to a rare subpopulation of early embryonic-like cells with totipotent-like characteristics.

The discovery of these 2C-like mESCs thus provides an opportunity to investigate the molecular control of totipotency and ZGA outside of the early embryos. Although 2C marker genes are well defined, the upstream drivers remain elusive[12]. In this study, we identify negative elongation factor A (NELFA)—a member of the NELF complex that regulates RNA polymerase II pausing[13]—as a maternal factor central for 2C gene expression. NELFA acts independent of the NELF complex and drives 2C reversion by activating *Dux* through interaction with DNA topoisomerase 2a (Top2a). We further discover that suppression of glycolysis can promote 2C reversion, and that Prdm14 serves as a barrier in this cell-state transition. In summary, our study reveals an instructive link between metabolism and induction of the 2C-like state, and uncovers molecular players that govern the pluripotent-to-2C state reversion.

Results

The maternal factor NELFA is heterogeneously expressed in mESCs. To identify upstream drivers of the 2C program, we performed hierarchical clustering on published transcriptomes of mouse pre-implantation embryos from various developmental stages[14], including mESCs, and obtained seven clusters with the most significant variations (Fig. 1a). We identified cohorts of genes that exhibited stage-specific expression patterns and developmental functions, such as *Nanog/Klf4* (cluster 7; ESC specific) and *Dux/Zscan4* (cluster 3; 2C specific) (Supplementary Table 1). We validated our methodology and recapitulated the stage-specific expression profile obtained using an independent transcriptome dataset[15] (Extended Data Fig. 1a).

Cluster 1 genes ($n=2,072$) were of particular interest as this cluster contains genes whose expression is restricted to the oocyte and early 2C stage that may contribute to ZGA. Notably, transcriptional and chromatin-based processes, such as 'chromatin binding' (Gene Ontology term 0003682) and 'transcription from RNA polymerase II promoter' (Gene Ontology term 0006366), were enriched in this cluster, consistent with molecular events that occur in early embryos for reprogramming towards totipotency[14,16–19]. We further intersected genes from cluster 1 with those that were significantly upregulated in other cellular models of the 2C-like state (zinc finger and SCAN domain containing 4 (Zscan4)-positive genes[20] and those causing *Dux* overexpression[7] and *CAF-1* depletion[7,16]) and narrowed our selection down to eight common genes (Fig. 1b). Of these, only *NELFA* is a known transcriptional regulator that is best studied as part of the NELF complex[13].

Immunofluorescence confirmed that NELFA is a maternally supplied factor present in the oocyte and early embryos as opposed to blastocysts (Fig. 1c). NELFA also showed a distinct expression profile compared with the other NELF subunits (Extended Data Fig. 1b). In mESC, only a rare subpopulation of cells express NELFA

[1]Chromatin Dynamics and Disease Epigenetics Group, Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore. [2]Department of Pharmacy, National University of Singapore, Singapore, Singapore. [3]Bioprocessing Technology Institute, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore. [4]Duke-NUS Medical School, Singapore, Singapore. [5]Research Centre for Animal Genetic Resources of Mongolia Plateau, College of Life Sciences, Inner Mongolia University, Hohhot, China. [6]Laboratory of Developmental and Regenerative Biology, Institute of Medical Biology, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore. [7]Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [8]These authors contributed equally: Zhenhua Hu, Dennis Eng Kiat Tan, Gloryn Chia. *e-mail: wwtee@imcb.a-star.edu.sg

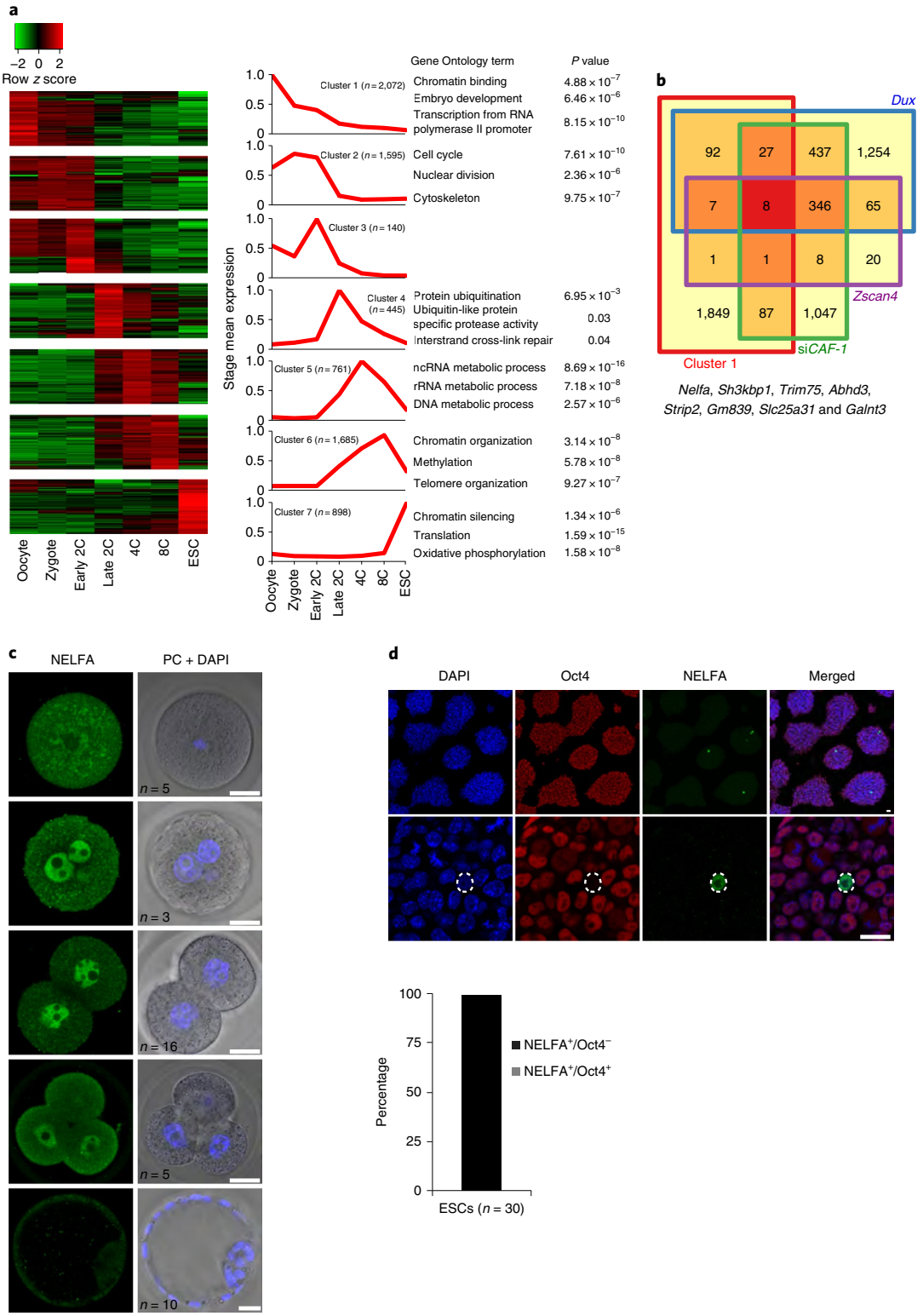


Fig. 1 | The maternal factor NELFA is heterogeneously expressed in mESCs. a, Clusters of genes showing distinct stage-specific expression patterns during mouse pre-implantation development (including mESCs). Representative enriched Gene Ontology terms and their corresponding *P* values are shown on the right. *P* values were calculated using hypergeometric distribution with the function enrichGO in the package clusterProfiler. 4C, 4-cell; 8C, 8-cell; ncRNA, non-coding RNA; rRNA, ribosomal RNA. **b**, Venn diagram indicating that eight genes from cluster 1 were also upregulated in various models of 2C-like cells (namely, Zscan4$^+$ mESCs[20], *Dux*-overexpressing mESCs[7] and *CAF-1* knockdown mESCs[16]). **c**, Immunofluorescence for NELFA (green) in the mouse oocyte, zygote, 2C embryo, four-cell embryo and blastocyst (top to bottom). A single focal plane is shown for each. *n* indicates the total number of embryos analysed in at least two independent experiments. DAPI was used for nuclear staining (blue). Scale bars: 20 μm. PC, phase contrast. **d**, Top: immunofluorescence for NELFA (green) and Oct4 (red) in mESCs revealed a rare NELFA$^+$ subpopulation. Dashed outlines highlight an ESC that is positive for NELFA and negative for Oct4. Three independent experiments were performed. Scale bar: 20 μm. Bottom: quantification of the cell population was performed based on the immunofluorescence data ($n = 30$ ESCs).

but lack the pluripotency marker Oct4 (Fig. 1d). This heterogeneous expression is specific to NELFA but not for other NELF complex subunits (Extended Data Fig. 1c), suggesting that NELFA might serve a complex independent function.

NELFA^{high} mESCs mark a 2C-like state. To better characterize the NELFA^{high} mESCs, we generated a NELFA-Strep-HA-P2A-EGFP reporter (NELFA fused to a streptavidin-HA epitope tag and connected to enhanced green fluorescent protein (EGFP) that is separated by a P2A self-cleaving peptide). We recapitulated the heterogeneous expression of NELFA and determined that ~0.1–0.3% of mESCs upregulate NELFA under conventional serum-containing mESC culture conditions (Fig. 2a). Next, using flow cytometry, we isolated NELFA^{high} and NELFA^{low} mESCs for transcriptome profiling (Supplementary Table 2) and observed that many 2C-restricted transcripts, including *MERVL*, were upregulated (Fig. 2b and Extended Data Fig. 2a). Further analysis revealed that the *MERVL* family (for example, *mt2_mm* and *Mervl-int*) and major satellite repeats (*gsat_mm*) were most strongly activated among other repetitive elements (Fig. 2b). Importantly, we confirmed by immunostaining that Zscan4 and MERVL (murine endogenous retrovirus-like MuERV-L) proteins were co-expressed in NELFA^{high} mESCs (Fig. 2c). Reciprocally, high NELFA expression was detected in Zscan4^{+} cells in an independent Zscan4-Emerald reporter mESC line (Extended Data Fig. 2b)[21]. Critically, comparison of NELFA^{high}-upregulated genes against our previously analysed seven transcriptome clusters (Fig. 1a) confirmed that NELFA^{high}-upregulated genes are most highly expressed at the 2C stage, while NELFA^{high}-downregulated genes are highly expressed at later stages of embryonic development and in mESCs (Fig. 2d). These observations thus suggested that NELFA^{high} cells might reside in a 2C-like state. In excellent support, NELFA^{high} transcriptome showed a strong positive correlation with the transcriptomes of other 2C cellular models (Extended Data Fig. 2c).

MERVL-expressing 2C-like cells were previously shown to exhibit bipotent chimeric competency in embryos[10]. To assess the developmental potential of NELFA^{high} mESCs in vivo, we injected mCherry-labelled NELFA^{high} or NELFA^{low} mESCs into late mouse morulae to generate chimeric blastocysts. We found that NELFA^{high} cells were able to contribute to both embryonic inner cell mass and extraembryonic trophectoderm in ~16% of chimeric blastocysts. In comparison, NELFA^{low} cells could only contribute exclusively to the inner cell mass in all blastocysts analysed (Fig. 2e). In conclusion, NELFA^{high} cells express 2C-specific genes and exhibit expanded cell fate potency, consistent with the hallmarks of reported 2C-like cells[10,16,22].

NELFA is a driver of the 2C-like state. Next, we asked whether overexpression of *NELFA* would be sufficient to induce a 2C-like state. For this, we generated a doxycycline (Dox)-inducible NELFA-EGFP mESC and enriched for NELFA-induced cells by flow cytometry for RNA sequencing (RNA-Seq). Indeed, we observed robust upregulation of key 2C genes and endogenous retroviruses (Fig. 3a and Supplementary Table 3). The upregulated genes were highly expressed at 2C stages relative to all other developmental stages (Fig. 3b) and were similarly upregulated in other reported 2C-like cells (Extended Data Fig. 3a). Using an independent Dox-inducible NELFA line, we confirmed that >80% of NELFA-induced cells showed concomitant upregulation of Zscan4 and MERVL proteins (Fig. 3c). To further demonstrate the causal effect of NELFA in promoting the expression of 2C genes, we depleted *NELFA* in Zscan4-Emerald reporter mESCs and observed a pronounced reduction in the population of ESCs expressing the Emerald marker driven by the *Zscan4* promoter (Zscan4^{Em+}; Fig. 3d). This phenomenon was specific to NELFA since the abrogation of *NELFB*—the central subunit of the NELF complex[23]—did not affect the proportion of Zscan4^{Em+} cells (Fig. 3d). To unequivocally demonstrate a 2C-specific effect of NELFA, we transiently overexpressed NELFA as a function of *NELFB* and *NELFE* knockdown and showed that 2C genes could still be induced (Extended Data Fig. 3b).

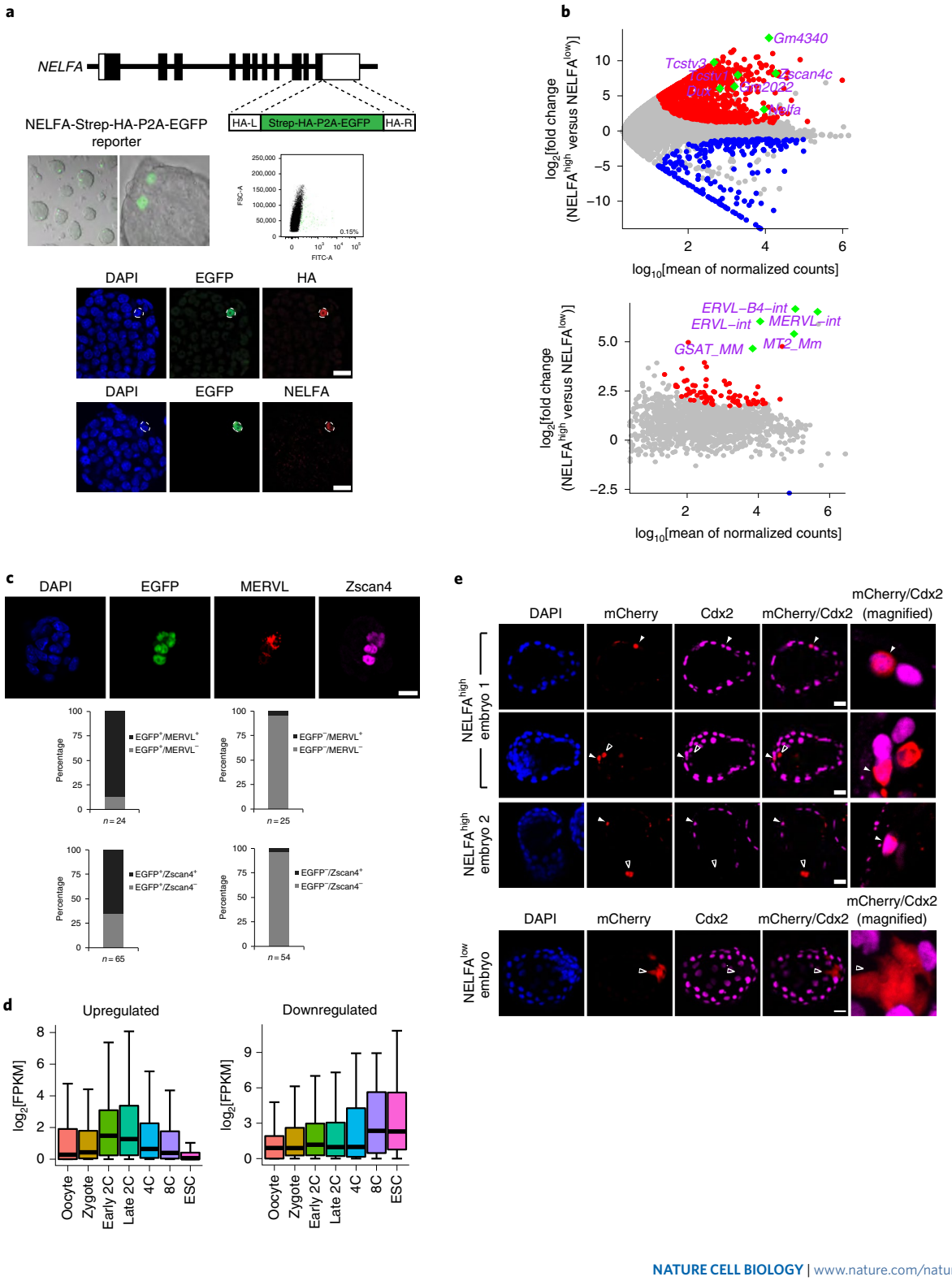
Next, we profiled the chromatin accessibility of NELFA-induced cells by assay for transposase-accessible chromatin (ATAC) sequencing (ATAC-Seq) and detected 11,306 regions that gained ATAC signals and 18,630 regions that lost ATAC signals. We confirmed that genes upregulated in NELFA-induced and NELFA^{high} cells exhibited increased chromatin accessibility (Fig. 3e). Using GREAT[24], we determined that the upregulated 2C genes from other 2C-transcriptome datasets were likewise enriched within ATAC-gained regions and depleted in the ATAC-lost regions (Extended Data Fig. 3c). Comparison against mouse embryo ATAC-Seq data[14] revealed that the ATAC-gained signals resembled those in 2C embryos, whereas ATAC-lost signals were more similar to those of late embryos and ESCs (Fig. 3f). Taken together, we conclude that NELFA is a previously unappreciated driver of the 2C-like state.

NELFA activates 2C genes through Dux. The transcription factor Dux was identified as a key effector of the 2C program[6,7]. However, *Dux* is only expressed during ZGA, whereas NELFA is maternally

Fig. 2 | NELFA^{high} mESCs mark a 2C-like state. a, Generation of a NELFA-Strep-HA-P2A-EGFP reporter mESC line (HA-L and HA-R represent left and right homology arms, respectively). Reporter mESCs showed restricted expression of EGFP, as determined by fluorescence live imaging (middle panel), confocal microscopy for HA, NELFA and EGFP (bottom panel) and FACS. Representative FACS and immunofluorescence data from three independent experiments are shown. FSC-A (forward scatter area) and FITC-A (fluorescein isothiocyanate filter). **b**, MA plots ('M' refers to log-intensity differences (M-values), and 'A' refers to log-intensity averages (A-value)) showing changes in gene (top) and repetitive element expression (bottom) between NELFA^{high} and NELFA^{low} mESCs. Red and blue data points represent genes and repetitive elements that are significantly up- and downregulated in NELFA^{high} cells, respectively. Selected 2C genes and repetitive elements (green diamonds) are highlighted. Two RNA-Seq replicates were generated for differential gene expression analyses by DESeq2 (FDR ≤ 0.01; fold change ≥ 2). **c**, Top: immunofluorescence for various 2C-like markers (as indicated) in NELFA-Strep-HA-P2A-EGFP reporter mESCs. Representative images are shown from two independent experiments. Bottom: quantification of the expression of the markers in individual cells was performed based on the immunofluorescence data. Numbers of NELFA reporter ESCs are shown below each graph. **d**, Box plots showing that NELFA^{high}-upregulated genes ($n = 1,086$) are most highly expressed in 2C-stage embryos (left), whereas NELFA^{high}-downregulated genes ($n = 246$) tend to be expressed at later stages of pre-implantation development and in mESCs (right). Centre lines show median values, box limits represent the upper and lower quartiles, and whiskers show 1.5× the interquartile range. Two-tailed Student's *t*-tests were used for all comparisons. For NELFA^{high}-upregulated and –downregulated genes, the respective *P* values for other stages against the early and late 2C stages are provided in the source data. FPKM, fragments per kilobase of transcript per million mapped reads. **e**, Representative images showing immunostaining of chimeric blastocysts injected with mCherry-labelled NELFA^{high} and NELFA^{low} mESCs. Different confocal *z* sections are presented for NELFA^{high} embryo 1 to illustrate the dual contribution of injected cells to both inner cell mass (ICM; open arrows) and trophectoderm (filled arrows). $n = 4$ independent experiments were performed. A total of 73 embryos were injected with NELFA^{high} cells and 74 embryos were injected with NELFA^{low} cells. Of the 73 NELFA^{high} chimeric blastocysts, 12 (16%) showed integration of injected cells in both the ICM and trophectoderm. All NELFA^{low} chimeric blastocysts showed that injected cells contributed exclusively to the ICM. Trophectoderm is marked by Cdx2 (magenta). Scale bars: 20 μm (**a**, **c** and **e**).

supplied, preceding *Dux* and 2C genes in vivo[14] (Fig. 1c and Extended Data Fig. 1b), hinting at a pro-2C functionality of NELFA upstream of *Dux* activity. Further investigations revealed significant overlap between our NELFA ATAC-Seq data and published Dux ATAC-Seq data[7] (Fig. 4a), and an over-representation of the Dux-binding motif in genes upregulated in NELFA-induced cells (Fig. 4b). Taken together, these data strongly suggested that NELFA might promote 2C gene expression with Dux as an effector.

To test this hypothesis, we overexpressed NELFA in both wild-type and *Dux* knockout mESCs[6]. Consistent with our earlier results, many prototypic Dux targets were upregulated upon NELFA overexpression in wild-type mESCs (Fig. 4c and Supplementary Table 4), but this upregulation was abolished in *Dux* knockout cells (Fig. 4c and Supplementary Table 5). In contrast, 2C genes were still robustly induced upon NELFA depletion in the Dux-overexpressing cell line (Fig. 4d). These results supported a model in which NELFA



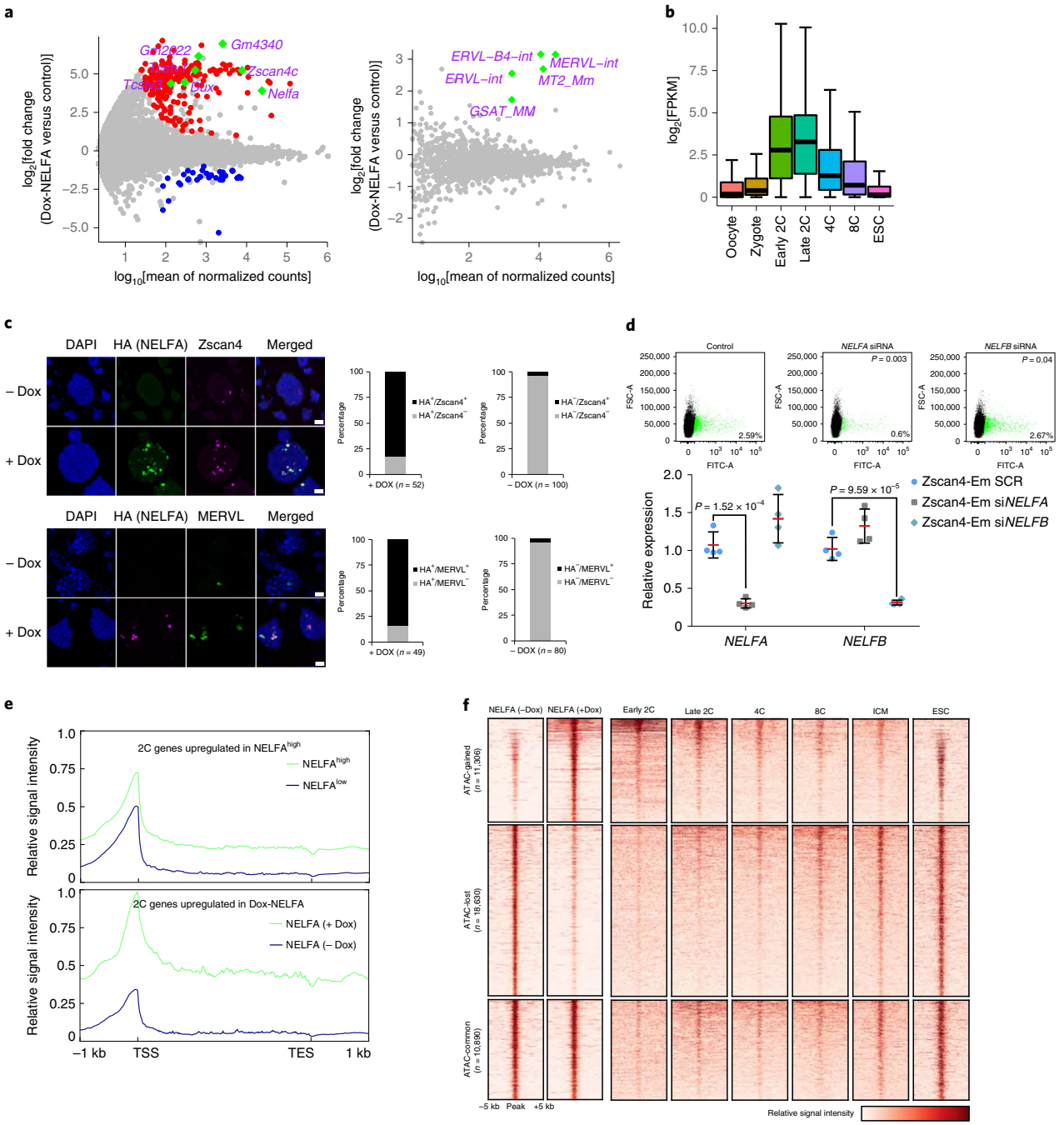


Fig. 3 | NELFA is a driver of the 2C-like state. a, MA plots ('M' refers to log-intensity differences (M-values) and 'A' refers to log-intensity averages (A-values)) showing expression changes in genes (left) and repetitive elements (right) between Dox-inducible NELFA-Strep-HA-P2A-EGFP 2C-like cells and control mESCs. Red and blue data points represent genes and repetitive elements that are significantly up- and downregulated, respectively. Selected 2C genes and repetitive elements are denoted by green diamonds. Three RNA-Seq replicates were generated for differential gene expression analyses by DESeq2 (FDR ≤ 0.01; fold change ≥ 2). **b**, Box plot showing that genes upregulated in NELFA-induced 2C-like cells ($n = 229$) are enriched at the 2C stage during pre-implantation development. Centre lines show median values, box limits represent the upper and lower quartiles, and whiskers show 1.5× the interquartile range. Two-tailed Student's *t*-tests were used for all comparisons. Respective *P* values for other stages against early and late 2C stages in NELFA^{high}-upregulated genes were: 6.1×10^{-16} and $<2.22\times10^{-16}$ (oocyte), 1.5×10^{-1}, and 3.0×10^{-15} (zygote), 1.2×10^{-3} and 1.4×10^{-4} (4C), 3.3×10^{-6} and 2.2×10^{-7} (8C) and $<2.22\times10^{-16}$ and $<2.22\times10^{-16}$ (mESC). **c**, Left: immunofluorescence for various 2C-like markers (as indicated) in Dox inducible NELFA-StrepHA ESCs. Representative images are shown from three independent experiments. Right: quantification of the expression of the markers in individual cells was performed based on the immunofluorescence data. Numbers of ESCs are indicated below each graph. Scale bars: 40 μm. **d**, Flow cytometry analysis of Zscan4-Emerald mESCs showing the effects of different NELF subunit knockdowns on the percentage of Zscan4^{Em+} cells ($n = 4$ independent experiments). In the bottom graph, red bars show mean values, while error bars represent s.e.m. The efficiency of knockdowns was assessed by quantitative reverse-transcription PCR (RT-qPCR). *P* values were determined by two-tailed Student's *t*-test. SCR, scramble siRNA control. **e**, Metagene analysis of chromatin accessibility, as determined by ATAC-Seq. The relative signal intensity is plotted across gene bodies of genes that are upregulated in NELFA^{high} (top) and Dox-induced NELFA (bottom) 2C-like mESCs. TES, transcription end site; TSS, transcription start site. **f**, ATAC-Seq heatmap comparing the landscapes of accessible chromatin in NELFA-induced 2C-like cells and cells from various stages of development. The embryo ATAC-Seq data were obtained from GEO accession GSE66390 (ref. [14]).

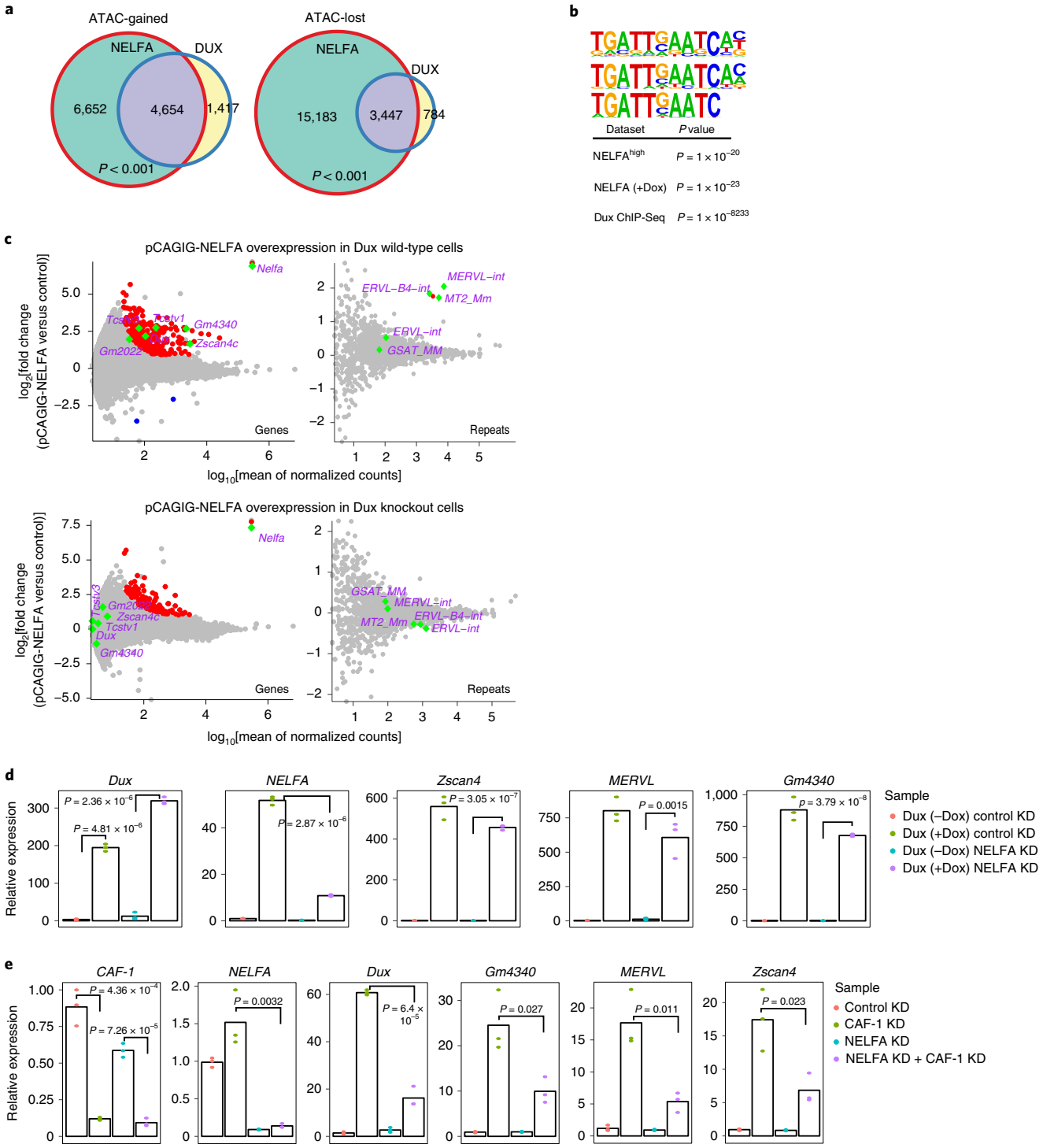


Fig. 4 | NELFA activates 2C genes through *Dux*. a, Venn diagrams showing significant overlap between ATAC-gained and ATAC-lost regions in Dox-induced NELFA and Dux-overexpressing mESCs. The *P* values were calculated using the enrichPeakOverlap function in the ChIPseeker R package based on 1,000 shuffles. ATAC-gained and ATAC-lost regions in Dux-overexpressing cells were obtained from published data[7]. **b**, Motif enrichment analysis of upregulated gene promoters from NELFA^{high} ($n = 1,086$ genes) and NELFA-induced transcriptomes, and of Dux-bound peaks from published Dux ChIP-Seq data[7], showing over-represented DUX transcription factor binding motifs. *P* values were calculated from cumulative hypergeometric distribution using the Homer2 package. **c**, MA plots ('M' refers to log-intensity differences (M-values), and 'A' refers to log-intensity averages (A-values)) showing expression changes in genes (left) and repetitive elements (right) between pCAGIG-NELFA overexpression and controls in *Dux* wild-type cells (top) and *Dux* knockout cells (bottom) (adjusted *P* value ≤ 0.01; fold change ≥ 2). Red and blue data points represent genes and repetitive elements that were significantly up- and downregulated, respectively. Selected 2C genes and repetitive elements are denoted by green diamonds. Two RNA-Seq replicates were generated for differential gene expression analyses by DESeq2 (FDR ≤ 0.01; fold change ≥ 2). pCAGIG-NELFA refers to a mammalian expression plasmid expressing NELFA in addition to IRES-GFP. **d**,**e**, RT-qPCR analysis for 2C genes in Dux-overexpressing mESCs following *NELFA* knockdown (**d**) and in *NELFA* knockdown mESCs as a function of *CAF-1* knockdown (**e**). *P* values were determined by two-sided Student's *t*-test. Means are represented as bar graphs, as calculated from $n = 3$ independent experiments. KD, knockdown.

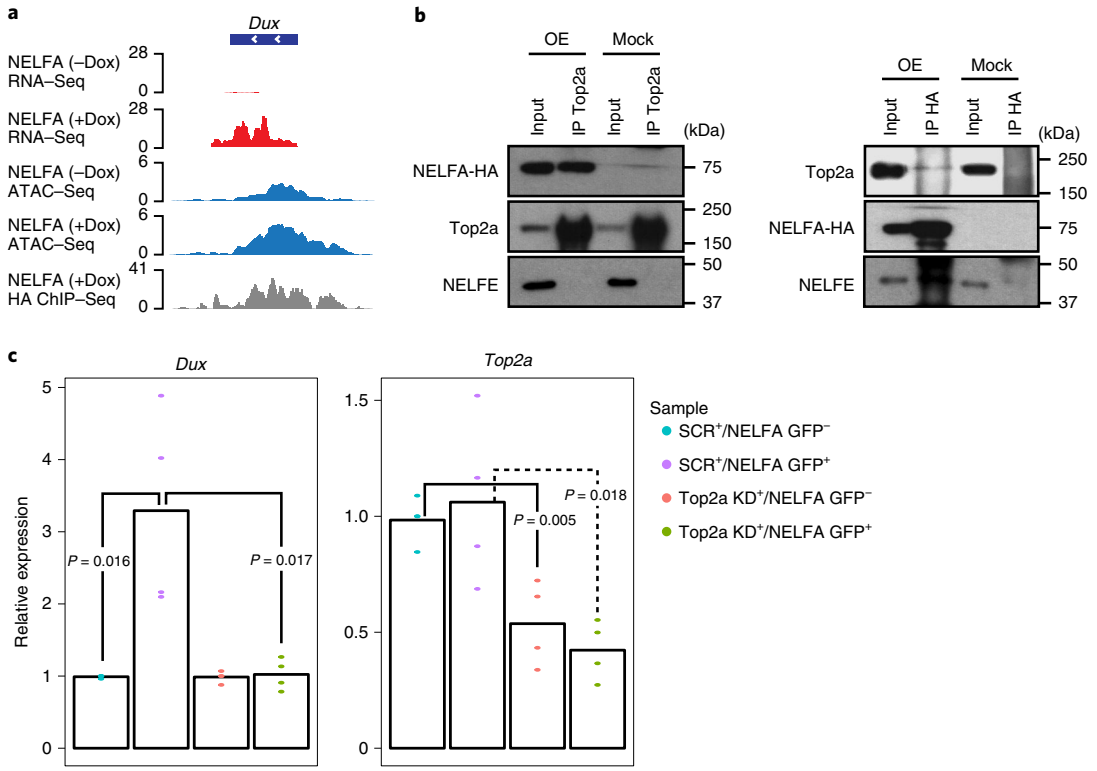


Fig. 5 | NELFA cooperates with TOP2A to activate *Dux*. a, Genome browser tracks showing that NELFA induction promotes the transcription of *Dux* messenger RNA, accompanied by an increase in chromatin accessibility at the *Dux* locus, and that NELFA (HA) binds to the *Dux* locus. **b**, Western blot analysis of Top2a, NELFA (HA), and NELFE following Top2a or NELFA immunoprecipitation (IP). IPs were performed on nuclear extracts from mESCs transfected with either NELFA-HA-expressing or mock expression plasmid. Blots are representative of two independent experiments with similar results. OE, overexpression. **c**, RT-qPCR analysis of *Dux* expression levels following transient *Top2a* knockdown in NELFA-overexpressing cells. *P* values were determined by two-sided Student's *t*-test. Means are represented by bar graphs, as calculated from $n = 4$ independent experiments. Unprocessed western blots are shown in the source data.

acts through Dux to promote the 2C-like state transition. *CAF-1* depletion is known to also induce a 2C state in a *Dux*-dependent manner[7]. Notably, in this independent 2C model, the loss of *NELFA* similarly impaired expression of *Dux* and canonical 2C transcripts, cogently demonstrating the importance of NELFA in activating *Dux* (Fig. 4e). Taken together, we provide compelling evidence that NELFA acts via *Dux* to promote 2C gene expression.

NELFA cooperates with Top2a to activate *Dux*. Next, we asked whether NELFA promotes 2C gene expression by the direct activation of *Dux*. Indeed, upon NELFA induction, we observed increased NELFA occupancy at the *Dux* locus that corresponded to chromatin opening and transcription of *Dux* (Fig. 5a and Extended Data Fig. 4), suggesting that *Dux* is a direct transcriptional target of NELFA. To gain mechanistic insights, we performed affinity purification followed by mass spectrometry and uncovered numerous interactors of NELFA upon its overexpression in mESCs, noting with interest that Top2a was the top interactor in our replicate analysis (Supplementary Table 6). Top2a was recently shown to be involved in ZGA by mediating chromatin decompaction and RNA polymerase II transcription activation in *Caenorhabditis elegans*[25]. To validate our results from the affinity purification followed by mass spectrometry, we performed reciprocal co-immunoprecipitation and confirmed that the pull-down of each protein recovered the other (Fig. 5b). Additionally, we established that NELFA's interaction with Top2a was probably subunit specific, as NELFE was not recovered in Top2a pull-down (Fig. 5b), corroborating our earlier findings that NELFA exerted a function independent of the NELF complex (Fig. 3d and Extended Data Fig. 3b). We subsequently demonstrated that the NELFA–Top2a interaction is also enhanced in an independent, chemically induced 2C-like state (Extended Data Fig. 6a; see later), lending support for a specific role of NELFA–Top2a partnership in 2C transition. Importantly, knockdown of *Top2a* in NELFA-overexpressing cells led to impaired upregulation of *Dux* (Fig. 5c), illustrating that NELFA required Top2a as a partner to accomplish *Dux* activation, and this was also consistent with the independent chemically induced 2C-like model (Extended Data Fig. 6b; see below). In conclusion, we show that NELFA is a driver of the 2C-like state, activating *Dux* in partnership with Top2a.

Suppression of glycolysis by 2-deoxy-D-glucose (2-DG) induces the 2C transcriptional program in mESCs. To uncover biological processes affected by NELFA upregulation, we performed gene set enrichment analysis (GSEA) on NELFA^{high} and NELFA-induced transcriptomes. Using the Kyoto Encyclopedia of Genes and Genomes database, we found that metabolism (metabolic pathways; mmu01100) emerged as the most altered process in both cell types (NELFA^{high} cells: normalized enrichment score (NES) = −2.24; false discovery rate (FDR) = 0.001; NELFA-induced cells: NES = −1.41; FDR = 0.001). Our equivalent analysis of Zscan4-Emerald mESCs[20] converged to a similar conclusion (NES = −2.03; FDR = 1.76×10^{-4}; Extended Data Fig. 5a and Supplementary Table 7). We note that the glycolysis pathway was downregulated in various 2C-like cells (Extended Data Fig. 5b). In NELFA-induced cells, we observed a decrease in chromatin accessibility on glycolysis-associated genes and reduced expression of glucose transporters (Fig. 6a and Extended Data Fig. 5b). Further validation confirmed the downregulation of key glycolytic enzymes in various 2C-like cells, congruous

with a decrease in chromatin accessibility (Extended Data Fig. 5c). As an independent assessment of the metabolic state, we utilized a green fluorescent glucose analogue, 2-deoxy-2-[(7-nitro-2,1,3-benzoxadiazol-4-yl)amino]-D-glucose (2-NBDG), to measure glucose uptake directly in an independent Zscan4-mCherry 2C-reporter cell line. Consistent with our transcriptomic data, we observed reduced 2-NBDG uptake predominantly in Zscan4$^+$ cells (Extended Data Fig. 5d). Our data collectively demonstrate that suppressed glycolysis is a general feature of the 2C-like state.

Hence, we asked whether pharmacological suppression of glycolysis might promote the emergence of 2C-like mESCs in vitro. We treated the NELFA reporter mESCs with a glycolysis inhibitor (2-DG) and observed an eightfold increase in the NELFA^{high} subpopulation following 4 d of treatment (Fig. 6b). A similar effect was also observed for Zscan4-Emerald reporter mESCs, indicating that suppression of glycolysis can activate the 2C program (Extended Data Fig. 5e). Although cell proliferation was affected, we did not detect significant cell death in this culture condition (Extended Data Fig. 5f), implying that the increase in the 2C-like population was not caused by selective elimination of naïve mESCs. In contrast, the inhibition of oxidative phosphorylation by rotenone did not increase the 2C-like population and caused elevated cell death (Extended Data Fig. 5g), consistent with previous studies[26]. Hence, we conclude that suppressing glycolysis with 2-DG supplementation is a robust and reproducible method to boost 2C-like cells in culture.

RNA-Seq of unsorted 2-DG-treated NELFA reporter mESCs uncovered 175 upregulated genes, including prototypic 2C genes (Fig. 6c and Supplementary Table 8). We observed a strong overlap of these upregulated genes with genes highly expressed in early and late 2C-stage embryos (Fig. 6d), in reported 2C-like mESCs (Extended Data Fig. 5h) and in NELFA^{high} and NELFA-induced cells (Extended Data Fig. 5h). Concordantly, immunofluorescence also showed upregulation of MERVL and Zscan4 in the 2-DG-treated NELFA$^+$ mESCs (Fig. 6e). Additional unsupervised clustering of transcriptomes from pre-implantation embryos, 2C-like mESCs and other totipotent-like cells confirmed that 2-DG-treated and NELFA-induced cells were most similar to 2C-like mESCs and the 2C embryo (Fig. 6f). We remark that even with metabolic manipulation, NELFA function remained critical; knockdown of *NELFA* in combination with 2-DG treatment still impaired the upregulation of key 2C genes (Extended Data Fig. 5i). Finally, we assessed the developmental potential of 2-DG-treated NELFA^{high} mESCs in vivo and confirmed that these cells showed increased potency in mouse chimera assays (Fig. 6g).

We reported earlier that NELFA partners with Top2a to activate *Dux* (Fig. 5). Notably, this interaction was specifically enhanced in 2-DG-treated cells (Extended Data Fig. 6a) and *Top2a* knockdown similarly impaired *Dux* activation in this chemically induced state (Extended Data Fig. 6b), fortifying the role of the NELFA–Top2a interaction in promoting *Dux* expression. In conclusion, our study uncovers metabolic insights governing the transition between pluripotent and 2C-like states, and demonstrates a facile method for inducing a 2C-like state without the need for genetic manipulation.

Exit from naïve pluripotency is required for robust activation of 2C gene expression in mESCs. To further dissect molecular changes during this pluripotency-to-2C-like transition, we analysed our NELFA-induced ATAC-Seq data with additional data generated from 2-DG-treated cells. In addition to increased chromatin accessibility at 2C-associated MERVL elements, we observed a consistent reduction at ESC-specific enhancers and super-enhancers (Fig. 7a and Extended Data Fig. 7a). Equivalent analysis of published *Dux* ATAC-Seq data[7] arrived at a similar conclusion (Extended Data Fig. 7a). The loss of chromatin accessibility was more pronounced at enhancers than corresponding promoters, indicating that enhancer remodelling preceded promoter changes for ESC pluripotency genes (Fig. 7b). This suggests that targeted decommissioning of ESC-specific enhancers may be a key feature associated with the pluripotency-to-2C-like conversion—a phenomenon also observed during transcription factor-induced reprogramming of somatic cells into induced pluripotent cells[27].

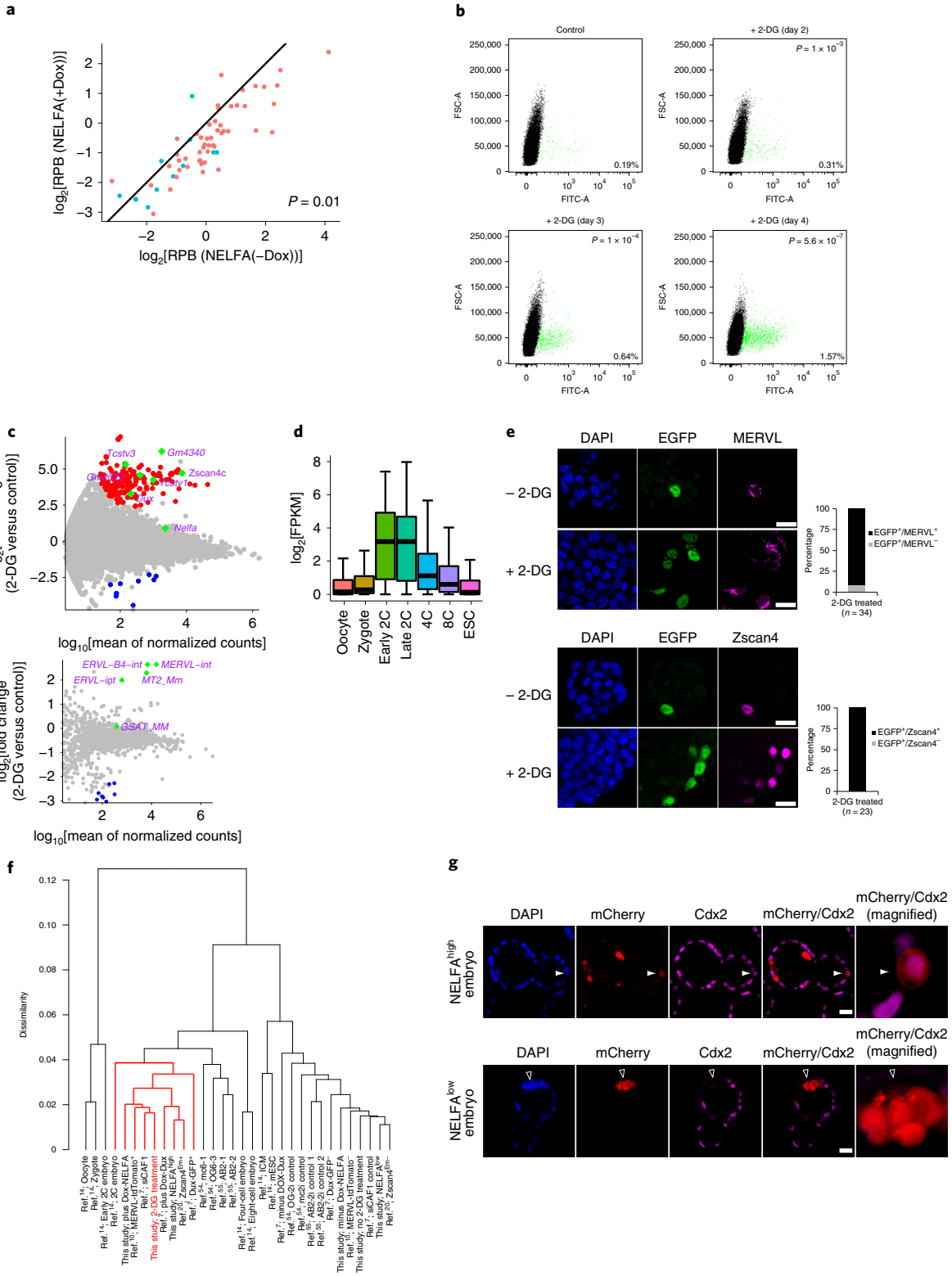
The results reported above suggest a model in which destabilization of the naïve pluripotent network may be necessary for ESCs to enter a 2C-like state. To investigate further, we cultured our NELFA reporter mESCs in a naïve ESC maintenance medium (N2B27/2i/leukemia inhibitory factor (LIF)) that promotes homogenous expression of pluripotency factors[28] and noted near-complete elimination of the NELFA^{high} subpopulation (Extended Data Fig. 7b). Further attempts to induce the 2C state via chromatin assembly factor-1 (CAF-1) depletion or 2-DG treatment in this naïve culture condition proved ineffectual (Fig. 7c). These data thus indicate that exit from naïve pluripotency is necessary for induction of the 2C program. In support, we noted from our ATAC-Seq that the *Prdm14* super-enhancer was one of the most affected loci (Fig. 7d and Supplementary Table 9). We thus re-examined the transcriptome data from *Prdm14* knockdown mESCs[29] and observed an upregulation of 2C genes that had originally been unremarked upon (Fig. 7e). Prdm14 is a transcriptional regulator that safeguards naïve

Fig. 6 | Suppression of glycolysis by 2-DG induces the 2C-like transcriptional program. a, Scatter plot displaying changes (log_2-normalized reads per billion (RPB)) in the chromatin accessibility of glycolysis genes ($n = 63$ genes) upon NELFA induction by Dox. Each point represents a gene in the glycolysis pathway. Red represents a decrease in expression in NELFA-induced cells compared with uninduced control cells. Blue represents non-downregulated glycolysis genes. **b**, FACS analysis of NELFA-StrepHA-P2A-EGFP reporter mESCs at various time points following 2-DG treatment. Representative data from seven independent experiments are shown. **c**, MA plots ('M' refers to log-intensity differences (M-values), and 'A' refers to log-intensity averages (A-values)) showing expression changes in genes (top) and repetitive elements (bottom) in NELFA reporter mESCs after 2-DG treatment. Red and blue data points represent up- and downregulated genes and repetitive elements. Selected 2C genes and repetitive elements are denoted by green diamonds. Two RNA-Seq replicates were generated for differential gene expression analyses by DESeq2 (FDR ≤ 0.01; fold change ≥ 2). **d**, Box plot showing that genes upregulated in NELFA reporter mESCs after 2-DG treatment ($n = 175$ genes) are most highly expressed in 2C embryos. Centre lines show median values, box limits represent the upper and lower quartiles, and whiskers show 1.5× the interquartile range. *P* values are provided in the source data. **e**, Left: immunofluorescence and quantification for EGFP (marking NELFA; green), MERVL (top; magenta) and Zscan4 (bottom; magenta) in NELFA reporter mESCs with and without 2-DG treatment. Two independent experiments were performed. Right: quantification of the immunofluorescence data. Numbers of ESCs are provided below each graph. **f**, Unsupervised clustering of transcriptomes from various cells, including NELFA reporter cells, NELFA-induced cells, 2-DG-treated cells and others[7,10,14,20,54,55]. **g**, Representative images showing immunostaining of chimeric blastocysts injected with mCherry-labelled NELFA^{high} and NELFA^{low} mESCs following 2-DG treatment. Two independent experiments were performed. A total of 41 embryos were injected with NELFA^{high} cells and 50 embryos were injected with NELFA^{low} cells. Of the NELFA^{high} chimeric blastocysts, 12% showed integration of injected cells in both the ICM (open arrows) and trophectoderm (filled arrows). All NELFA^{low} chimeric blastocysts showed that injected cells contributed exclusively to the ICM. Trophectoderm is marked by Cdx2 (magenta). DAPI nuclear staining is shown in blue. Scale bars: 20 μm (**e** and **g**). *P* values were determined by two-tailed Student's *t*-test (**a**, **b** and **d**).

pluripotency[29–31], and the depletion of this factor causes downregulation of pluripotency markers and misexpression of lineage makers, including extraembryonic endoderm (XEN) genes[29]. Notably, these effects were less evident in naïve 2i culture conditions[30]. In summary, our unbiased analysis of the chromatin landscape in 2C-like cells led us to uncover Prdm14 as an impediment to the ESC-to-2C-like transition.

Discussion

In this study, we identified the maternal factor NELFA as one of the earliest drivers of the 2C-like state, acting upstream of previously identified marker genes, such as *Dux*, *Zscan4* and *MERVL*, that are upregulated during ZGA. Mechanistically, NELFA partners with Top2a in an interaction specific to the 2C-like state to promote *Dux* activation. The *C. elegans* paralogue of *Top2a* has recently



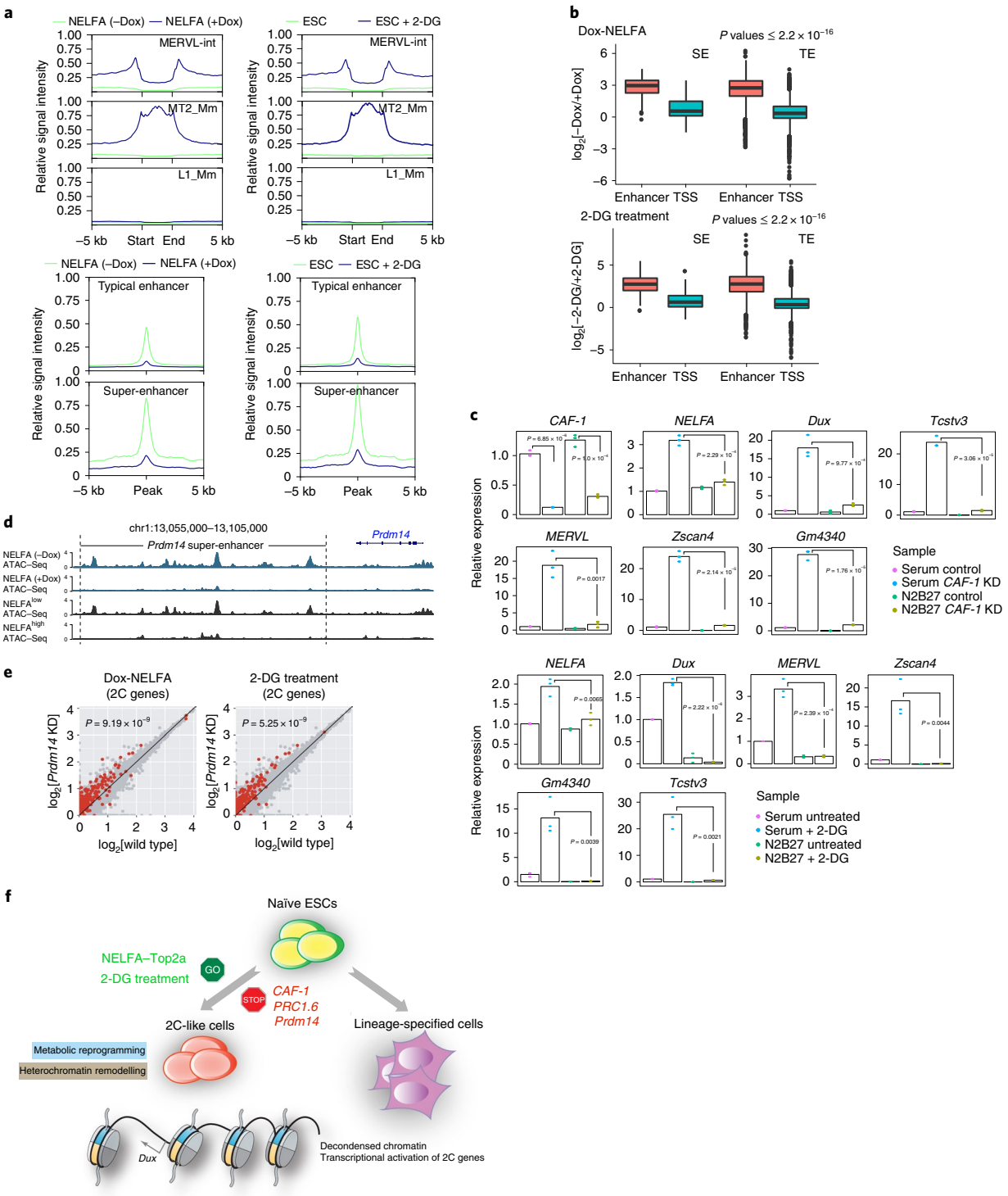


Fig. 7 | Exit from naïve pluripotency is required for activation of the 2C-like state. a, Metagene analysis of ATAC-Seq signals in NELFA-induced and 2-DG-treated NELFA reporter mESCs across specific repetitive elements (top) and ESC-specific enhancers[56] (bottom). **b**, Box plots showing the distribution of the $\log_2$ ratio of the normalized number of reads between enhancers and their nearest TSS in NELFA-induced and 2-DG-treated mESCs. Centre lines show median values, box limits represent the upper and lower quartiles, and whiskers show 1.5× the interquartile range. *P* values (all $\leq 2.2 \times 10^{-16}$) were determined for all comparisons by two-tailed Student's *t*-test. Super-enhancer (SE): $n = 236$); typical enhancer (TE): $n = 5{,}741$. **c**, RT-qPCR analysis of 2C genes in cells grown under conventional serum-containing culture conditions versus naïve N2B27/2i/LIF conditions, as a function of *CAF-1* knockdown ($n = 3$ independent experiments; top) and 2-DG treatment ($n = 3$ independent experiments; bottom). *P* values were determined by two-tailed Student's *t*-test. Bars represent mean expression. **d**, Genome browser tracks displaying ATAC-Seq signals at the *Prdm14* super-enhancer and promoter regions in NELFA reporter and NELFA-induced mESCs. **e**, Scatter plots depicting changes in gene expression in *Prdm14* knockdown cells[29]. Red data points denote 2C genes that are upregulated in NELFA-induced cells ($n = 229$ genes; left) or 2-DG-treated mESCs ($n = 175$ genes; right). Two-tailed Student's *t*-tests were used to derive the *P* values. **f**, Model of 2C activation. NELFA induction and 2-DG treatment promote 2C gene expression and the ESC-to-2C-like transition, characterized by extensive chromatin state changes and metabolic reprogramming. In particular, NELFA partners with Top2a to promote *Dux* activation, leading to expression of other downstream 2C genes. An exit from naïve pluripotency via ESC-specific enhancer decommissioning (for example, *Prdm14* enhancer) is also necessary for this transition.

been shown to enable germline ZGA in vivo, in part through triggering chromatin decompaction[25]. Consistent with this observation, induction of NELFA also led to chromatin decondensation, which we note is a well-known feature of 2C-like cells in general (Extended Data Fig. 8). In support of its possible role in ZGA, *Top2a* knockout mouse embryos are embryonically lethal as early as the four- to eight-cell stage[32].

Clearly, an important future direction is to determine whether the maternal factor NELFA is responsible for ZGA in vivo, and whether it evokes molecular events similar to those we have described above. To address this, the generation of maternal/zygotic *NELFA* knockout animals will be critical. In relation to this, we note a recent report in which *Dux* knockout mice were shown to develop past the 2C stage[33], unlike *Dux* knockout mESCs that cannot upregulate many canonical 2C genes. Such incongruity implies that our current molecular understanding of in vitro 2C-like cells may be oversimplified, and that yet-unknown differences exist between 2C embryonic and 2C-like ESC states. Likewise, it is also possible that *Dux*-independent and/or redundant mechanisms may exist in vivo. This knowledge gap was alluded to in our initial exploration of early 2C embryonic regulators, where we observed that the in vivo set of early upregulated genes only showed limited overlap with genes upregulated in existing 2C-like mESC models (Fig. 1b). Additionally, two factors—Dppa2 and Dppa4—were recently identified as positive regulators of ZGA[34,35]. However, in our analysis of NELFA-specific ATAC peaks, we did not recover Dppa2/4 motifs.

Our finding that different 2C-like mESC models exhibited suppressed glycolysis dovetails neatly with decades-old embryological studies, in which early cleavage-stage embryos were shown to be metabolically less active, consuming less oxygen and glucose compared with blastocysts and ESCs[36–39]. In fact, the deliberate blockade of glycolysis is required for development of the totipotent early embryo[40]. In excellent agreement, we demonstrated that 2-DG-induced glycolytic inhibition promoted the emergence of 2C-like mESCs (Fig. 6b and Extended Data Fig. 5). However, although 2-DG treatment could enhance the proportion of 2C-like cells, this mode of metabolic manipulation is still not sufficient to promote stable propagation of these cells in vitro (Extended Data Fig. 9a). In this regard, it is pertinent to remark that other metabolic pathways may be important for energy production during early embryonic development, such as lipid metabolism[41]. How these various metabolic pathways could be combinatorically manipulated to robustly promote the maintenance and propagation of totipotent-like cells is an important thrust for follow-up investigations.

A recent study demonstrated that pyruvate is essential for ZGA to occur in vivo[42]. In agreement, we noted that NELFA^{high} 2C-like mESCs could indeed persist in a glucose-deprived, pyruvate-supplemented medium (Extended Data Fig. 9b). Related to this, select mitochondrial tricarboxylic acid cycle metabolic enzymes, such as Idh3a and Aco2, are known to transiently translocate to the nucleus during ZGA[42]. However, in the in vitro setting, we did not observe any obvious changes in the localization of either enzyme in the NELFA-positive 2C-like cells (Extended Data Fig. 9c), highlighting fundamental differences between the nature of the 2C embryo and 2C-like ESC state over and above the gene expression changes discussed earlier.

Previous studies on 2C gene regulation have largely centred on chromatin regulation, implicating a number of epigenetic repressors in negative correlation with the 2C-like state[16,43–46]. Here, we show that metabolic manipulation could also impact the 2C-like state. Our results thus illustrate the intimate relationship between metabolism and chromatin structure, and how this regulatory axis may be manipulated to alter cell fate[47–49]. In the case of 2-DG treatment, chromatin immunoprecipitations (ChIPs) for the methylated histone modification H3K27me3 indicated reduced levels of this mark at 2C genes (Extended Data Fig. 10a), demonstrating that metabolic manipulation could promote a 2C-like state, in part through chromatin modulation. Yet, regardless of metabolic or chromatin interventions for 2C-like-state induction, we emphasize that NELFA (and its ensuing pathways) remains key to this transition. This is underpinned by the requirement for *NELFA* in activating *Dux* and other 2C genes in 2-DG-treated and *CAF-1* knockdown cells (Fig. 4e and Extended Data Fig. 5i). Interestingly, upon NELFA induction, CAF-1 is downregulated at the protein level, which may provide another means for *Dux* activation (Extended Data Fig. 10b). We further observed that NELFA itself is upregulated upon *CAF-1* knockdown and/or *Dux* overexpression (Fig. 4d,e), highlighting a potential positive feedback mechanism to ensure robust 2C state transition.

The observation of suppressed glycolysis in 2C-like cells could be a consequence of the positive association between glycolysis and naïve pluripotency. Indeed, pluripotency factors such as Oct4 are known to directly regulate the expression of glycolytic regulators[50], and our observation of Oct4 loss in 2C-like cells (Fig. 1d) is consistent with a suppressed glycolytic state. Reciprocally from our ATAC-Seq analysis, we also demonstrated that the 2-DG-induced 2C-like state correlates with ESC enhancer decommissioning, signalling exit from naïve pluripotency (Fig. 7a,d). Enhancers play a cardinal role in driving cell-state-specific gene expression[51]. Our data indicate that ESC enhancer decommissioning is an early epigenetic event in response to 2C fate induction, and further highlight that regulated dissolution of the naïve pluripotency network may be necessary for entry into the 2C-like state. This is corroborated by curtailed expression of 2C genes in naïve ESC culture conditions (Fig. 7c) and the observed downregulation of pluripotency factors in 2C-like cells[7,10,43]. Indeed, our data are consistent with the findings from a recent single-cell study showing that downregulation of pluripotency genes precedes the activation of 2C genes[52]. Through our enhancer analysis, we also defined Prdm14—a stabilizer of naïve pluripotency[29–31]—as restrictive for the transition of ESCs into early embryonic states (Fig. 7e), concordant with previous findings that *Prdm14*-depleted ESCs upregulated other lineage markers, including XEN genes, under serum-containing ESC culture conditions[29]. Interestingly, a recent study highlighted a molecular link between XEN-like and 2C-like states[53]; understanding how Prdm14 and other pluripotency factors act at the nexus of the ESC-to-early embryonic transition should reveal additional insights into an expanding molecular roadmap for establishing the 2C-like state[43].

In conclusion, we hereby establish the maternal factor NELFA as a driver of the 2C-like mESC state, and describe distinctive metabolic alterations that accompany the transition to a developmentally antecedent cell state. Additionally, our finding that metabolic manipulation by small-molecule supplementation is instructive for mESC reversion to an earlier 2C-like state now provides a refreshingly different dimension to 2C gene regulation, which may potentially be translated into cellular reprogramming strategies (Fig. 7f).

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41556-019-0453-8.

Received: 27 February 2019; Accepted: 9 December 2019;
Published online: 13 January 2020

References

1. Ishiuchi, T. & Torres-Padilla, M. E. Towards an understanding of the regulatory mechanisms of totipotency. *Curr. Opin. Genet. Dev.* **23**, 512–518 (2013).
2. Wu, G. & Scholer, H. R. Lineage segregation in the totipotent embryo. *Curr. Topics Dev. Biol.* **117**, 301–317 (2016).

3. Zhou, L. Q. & Dean, J. Reprogramming the genome to totipotency in mouse embryos. *Trends Cell Biol.* **25**, 82–91 (2015).
4. Hamatani, T., Carter, M. G., Sharov, A. A. & Ko, M. S. Dynamics of global gene expression changes during mouse preimplantation development. *Dev. Cell* **6**, 117–131 (2004).
5. Falco, G. et al. *Zscan4*: a novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. *Dev. Biol.* **307**, 539–550 (2007).
6. De Iaco, A. et al. DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* **49**, 941–945 (2017).
7. Hendrickson, P. G. et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934 (2017).
8. Fadloun, A. et al. Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat. Struct. Mol. Biol.* **20**, 332–338 (2013).
9. Peaston, A. E. et al. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* **7**, 597–606 (2004).
10. Macfarlan, T. S. et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).
11. Beddington, R. S. & Robertson, E. J. An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development* **105**, 733–737 (1989).
12. Svoboda, P. Mammalian zygotic genome activation. *Semin. Cell Dev. Biol.* **84**, 118–126 (2018).
13. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).
14. Wu, J. et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652–657 (2016).
15. Deng, Q., Ramskold, D., Reinius, B. & Sandberg, R. Single-cell RNA-Seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196 (2014).
16. Ishiuchi, T. et al. Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* **22**, 662–671 (2015).
17. Ancelin, K. et al. Maternal LSD1/KDM1A is an essential regulator of chromatin and transcription landscapes during zygotic genome activation. *eLife* **5**, e08851 (2016).
18. Hatanaka, Y. et al. Histone chaperone CAF-1 mediates repressive histone modifications to protect preimplantation mouse embryos from endogenous retrotransposons. *Proc. Natl Acad. Sci. USA* **112**, 14641–14646 (2015).
19. Liu, X. et al. Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature* **537**, 558–562 (2016).
20. Akiyama, T. et al. Transient bursts of Zscan4 expression are accompanied by the rapid derepression of heterochromatin in mouse embryonic stem cells. *DNA Res.* **22**, 307–318 (2015).
21. Zalzman, M. et al. *Zscan4* regulates telomere elongation and genomic stability in ES cells. *Nature* **464**, 858–863 (2010).
22. Choi, Y. J. et al. Deficiency of microRNA miR-34a expands cell fate potential in pluripotent stem cells. *Science* **355**, eaag1927 (2017).
23. Williams, L. H. et al. Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. *Mol. Cell* **58**, 311–322 (2015).
24. McLean, C. Y. et al. GREAT improves functional interpretation of *cis*-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
25. Wong, M. M., Belew, M. D., Kwieraga, A., Nhan, J. D. & Michael, W. M. Programmed DNA breaks activate the germline genome in *Caenorhabditis elegans*. *Dev. Cell* **46**, 302–315.e5 (2018).
26. Carbognin, E., Betto, R. M., Soriano, M. E., Smith, A. G. & Martello, G. Stat3 promotes mitochondrial transcription and oxidative respiration during maintenance and induction of naive pluripotency. *EMBO J.* **35**, 618–634 (2016).
27. Koche, R. P. et al. Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell* **8**, 96–105 (2011).
28. Ying, Q.-L. et al. The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).
29. Ma, Z., Swigut, T., Valouev, A., Rada-Iglesias, A. & Wysocka, J. Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. *Nat. Struct. Mol. Biol.* **18**, 120–127 (2011).
30. Yamaji, M. et al. PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells. *Cell Stem Cell* **12**, 368–382 (2013).
31. Grabole, N. et al. *Prdm14* promotes germline fate and naive pluripotency by repressing FGF signalling and DNA methylation. *EMBO Rep.* **14**, 629–637 (2013).
32. Akimitsu, N. et al. Enforced cytokinesis without complete nuclear division in embryonic cells depleting the activity of DNA topoisomerase IIα. *Genes Cells* **8**, 393–402 (2003).
33. Chen, Z. & Zhang, Y. Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. *Nat. Genet.* **1**, 947–951 (2019).
34. Eckersley-Maslin, M. et al. Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. *Genes Dev.* **33**, 194–208 (2019).
35. De Iaco, A., Coudray, A., Duc, J. & Trono, D. DPPA2 and DPPA4 are necessary to establish a 2C-like state in mouse embryonic stem cells. *EMBO Rep* **20**, e47382 (2019).
36. Houghton, F. D., Thompson, J. G., Kennedy, C. J. & Leese, H. J. Oxygen consumption and energy metabolism of the early mouse embryo. *Mol. Reprod. Dev.* **44**, 476–485 (1996).
37. Brinster, R. L. & Troike, D. E. Requirements for blastocyst development in vitro. *J. Anim. Sci.* **49**, 26–34 (1979).
38. Baumann, C. G., Morris, D. G., Sreenan, J. M. & Leese, H. J. The quiet embryo hypothesis: molecular characteristics favoring viability. *Mol. Reprod. Dev.* **74**, 1345–1353 (2007).
39. Leese, H. J. Metabolic control during preimplantation mammalian development. *Hum. Reprod. Update* **1**, 63–72 (1995).
40. Barbehenn, E. K., Wales, R. G. & Lowry, O. H. The explanation for the blockade of glycolysis in early mouse embryos. *Proc. Natl Acad. Sci. USA* **71**, 1056–1060 (1974).
41. Sturmey, R., Reis, A., Leese, H. & McEvoy, T. Role of fatty acids in energy provision during oocyte maturation and early embryo development. *Reprod. Domest. Anim.* **44**, 50–58 (2009).
42. Nagaraj, R. et al. Nuclear localization of mitochondrial TCA cycle enzymes as a critical step in mammalian zygotic genome activation. *Cell* **168**, 210–223.e11 (2017).
43. Rodriguez-Terrones, D. et al. A molecular roadmap for the emergence of early-embryonic-like cells in culture. *Nat. Genet.* **50**, 106–119 (2018).
44. Macfarlan, T. S. et al. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev.* **25**, 594–607 (2011).
45. Maksakova, I. A. et al. Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. *Epigenetics Chromatin* **6**, 15 (2013).
46. Suzuki, A. et al. Loss of MAX results in meiotic entry in mouse embryonic and germline stem cells. *Nat. Commun.* **7**, 11056 (2016).
47. Reid, M. A., Dai, Z. & Locasale, J. W. The impact of cellular metabolism on chromatin dynamics and epigenetics. *Nat. Cell Biol.* **19**, 1298–1306 (2017).
48. Cluntun, A. A. et al. The rate of glycolysis quantitatively mediates specific histone acetylation sites. *Cancer Metab.* **3**, 10 (2015).
49. Mentch, S. J. et al. Histone methylation dynamics and gene regulation occur through the sensing of one-carbon metabolism. *Cell Metab.* **22**, 861–873 (2015).
50. Kim, H. et al. Core pluripotency factors directly regulate metabolism in embryonic stem cell to maintain pluripotency. *Stem Cells* **33**, 2699–2711 (2015).
51. Hu, Z. & Tee, W.-W. Enhancers and chromatin structures: regulatory hubs in gene expression and diseases. *Biosci. Rep.* **37**, BSR20160183 (2017).
52. Fu, X., Wu, X., Djekidel, M. N. & Zhang, Y. Myc and Dnmt1 impede the pluripotent to totipotent state transition in embryonic stem cells. *Nat. Cell Biol.* **21**, 835–844 (2019).
53. Zhao, T. et al. Single-cell RNA-Seq reveals dynamic early embryonic-like programs during chemical reprogramming. *Cell Stem Cell* **23**, 31–45.e7 (2018).
54. Yang, J. et al. Establishment of mouse expanded potential stem cells. *Nature* **550**, 393–397 (2017).
55. Yang, Y. et al. Derivation of pluripotent stem cells with in vivo embryonic and extraembryonic potency. *Cell* **169**, 243–257.e25 (2017).
56. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).

Methods

Cell culture. Unless otherwise stated, all mESC lines were cultured in serum-containing ESC medium containing KnockOut DMEM, 15% foetal calf serum, L-glutamine, non-essential amino acids, penicillin/streptomycin and 2-mercaptoethanol and supplemented with LIF, 1 μM PD0325901 and 3 μM CHIR99021 (2i). For the serum-free naïve culture condition, mESCs were grown in N2B27 medium supplemented with LIF and 2i. For the 2-DG experiments, ESCs were grown in serum-containing ESC medium with 4 mM 2-DG (Sigma–Aldrich; catalogue number D8375) for up to 4 d.

Generation of NELFA reporter and Dox-inducible mESC lines. For the NELFA-Strep-HA-P2A-EGFP reporter ESC line, a donor vector harbouring a Strep-HA-P2A-EGFP cassette flanked by NELFA left and right homology arms (826 and 924 base pairs, respectively) was co-transfected with the CRISPR–Cas9 single-guide RNA cloning vector pX458 containing the *NELFA* single-guide RNA sequence (GCTGACCCTCATCAGACCAG). Positive transfectants were GFP sorted, and targeted single clones were validated by genotyping. For generation of the Dox-inducible NELFA overexpression ESC lines, the Tet-On 3G inducible expression system (Clontech) was used. Briefly, both NELFA-Strep-HA and NELFA-Strep-HA-P2A-EGFP constructs were cloned into the pTRE3G vector, linearized with ScaI and co-transfected with pCAG-IRES-puromycin into E14 mESCs stably expressing EF1α-Tet3G. ESCs were kept under constant puromycin (1.2 μg ml^{-1}) and G418 (250 μg ml^{-1}) selection. A total of 0.4 μg ml^{-1} Dox was used for induction of NELFA expression. For generation of Dux-overexpressing mESCs, E14 mESCs were transfected with the mouse codon-optimized Tet-inducible *Dux* plasmid pCW57.1-mDux-CA-3HA (modified from the original pCW57.1-mDux-CA plasmid (Addgene plasmid 99284)) that had been linearized with ScaI. Puromycin selection (2 μg ml^{-1}) was carried out for 1 week to enrich for stable transfectants before single-colony sub-cloning. Selected sub-clones were kept under constant puromycin selection to prevent transgene silencing. For the generation of Zscan4-mCherry reporter mESCs, E14 mESCs were transfected with a Zscan4-mCherry plasmid (a gift from M. E. Torres-Padilla). Transfected cells were selected with hygromycin (250 μg ml^{-1}) for 1 week before single-colony sub-cloning. Selected sub-clones were kept under constant hygromycin selection to prevent transgene silencing.

Embryo injections. mCherry-labelled NELFA^{high} and NELFA^{low} mESCs were isolated by fluorescence-activated cell sorting (FACS) or manual picking and injected into either eight-cell-stage embryos or E3.25 blastocysts (from ICR or C57BL/6N strains). Five to seven cells were injected into each host embryo. Chimeric blastocysts were cultured in KSOM medium at 37 °C under a 5% CO_2 atmosphere until late E4.5 and fixed in 4% paraformaldehyde (PFA) for immunofluorescence staining. The female mice used to generate the embryos were 8–10 weeks old and males were 12–20 weeks old. The study was compliant with all relevant ethical regulations regarding animal research.

Small interfering RNA (siRNA) transfection. siRNAs were purchased from GE Healthcare Dharmacon and Qiagen and transfected into mESCs using Lipofectamine RNAiMAX (Thermo Fisher Scientific; catalogue number 13778150) according to the manufacturer's instructions. Both forward and reverse transfections were carried out. The final concentration of siRNA used was 40 nM per transfection, and the effect of knockdown was analysed at 72 h. The siRNAs are listed in Supplementary Table 12.

Immunofluorescence. Briefly, mESCs were grown on gelatin-coated imaging dishes, fixed in 4% PFA for 10 min at room temperature, permeabilized and blocked in blocking buffer (0.1% (*v/v*) Triton X-100, 1% (*w/v*) BSA and 1× phosphate buffered saline (PBS)) for 30 min. Cells were incubated in primary antibodies overnight at 4 °C, followed by three washes in blocking buffer. Secondary antibodies were added for 1 h, followed by two washes in blocking buffer and a final wash in PBS. 4′,6-diamidino-2-phenylindole (DAPI) was used as a nuclear counterstain. Mouse embryos were fixed in 4% PFA for 15 min at room temperature, permeabilized and blocked in blocking buffer for 30 min. Primary antibodies were added overnight and washed three times in blocking buffer, then secondary antibodies containing DAPI were added for 2 h. The antibodies used are listed in Supplementary Table 13.

Quantitative reverse-transcription PCR. Between 0.5 and 1 μg of total RNA was used as input for complementary DNA (cDNA) preparation with the SensiFAST cDNA Synthesis Kit (Bioline; 65054). Reverse transcription was carried out with the standard cycling condition per the manufacturer's instructions. cDNA was diluted 40-fold and quantitative PCR (qPCR) was conducted using the 2× PowerUp SYBR Green master mix (Thermo Fisher Scientific; catalogue number A25742). The relative gene expression fold change was calculated using the ΔΔCt method, and statistical significance assessed using Student's *t*-test. The primers used for qPCR are listed in Supplementary Table 11.

FACS. A FACSAria Cell Sorter (BD Biosciences) was used to quantify the proportion of EGFP$^+$ and EGFP$^-$ cells from reporter mESC lines. The MoFlo cell sorter (Beckman Coulter) was used to sort EGFP$^+$ and EGFP$^-$ cells for the RNA-Seq and ATAC-Seq studies. The gating strategy is detailed in Extended Data Fig. 7c.

PE Annexin V staining. NELFA reporter mESCs with or without 2-DG or rotenone treatment were harvested and 5×10^5 cells from each condition were used for Annexin V staining. Briefly, cells were first washed with PBS (Gibco; Thermo Fisher Scientific) before a second wash with sterile Annexin V binding buffer (10 mM HEPES (pH 7.4), 140 mM NaCl and 2.5 mM $CaCl_2$). Washed cells were resuspended and incubated with PE Annexin V stain (BD Biosciences; catalogue number 556421) diluted 40-fold in Annexin V binding buffer on ice for 15 min, protected from light. Finally, stained cells were washed once with Annexin V binding buffer before FACS analysis.

2-NBDG treatment. Zscan4-mCherry reporter mESCs were first seeded on imaging slides and starved overnight in ESC medium lacking glucose before treatment. The following day, the reporter cells were incubated with fresh media containing 200 μM 2-NBDG (Thermo Fisher Scientific; catalogue number N13195) for 30 min before live fluorescence imaging.

Cytotoxic and cell proliferation assay. Cytotoxic and cell proliferation assessment of 2-DG and rotenone treatment was performed using the CellTox Green Cytotoxicity and CellTiter-Glo Luminescent commercial kits (Promega), respectively. Each assay was set up in 12-well plates at a density of 1,000 cells per well with either treatment with 2-DG (Sigma–Aldrich) or rotenone (Tocris Bioscience) following manufacturer's instructions. Fluorescence readings were obtained at various time points with a fluorescence plate reader (Cytation 3; BioTek).

Nuclear extract preparation and co-immunoprecipitation. To prepare nuclear extract, ESCs were resuspended in ice-cold TMSD buffer (20 mM HEPES (pH 7.5), 5 mM $MgCl_2$ and 250 mM sucrose completed with 1 mM dithiothreitol (DTT), 5 mM sodium butyrate, 1 mM Aprotinin (A), 1 mM Pepstatin (P), 1 mM Leupeptin (L) and 2 mM phenylmethane sulfonyl fluoride (PMSF)) for 10 min and collected by centrifugation (800*g* at 4 °C). Next, the nuclei were released by resuspending and incubating the cell pellet with complete ice-cold TMSD buffer containing 0.1% (*v/v*) Nonidet P-40 on ice for 10 min. The released nuclei were pelleted by centrifugation (800*g* at 4 °C) and immediately lysed with ice-cold lysis buffer 1 (20 mM Tris-Cl (pH 7.9), 420 mM KCl, 1.5 mM $MgCl_2$ and 0.2 mM EDTA) completed with 10 mM sodium butyrate, 0.5 mM DTT, 1 mM A/L/P and 2 mM PMSF. Nuclear lysis was carried out at 4 °C with constant rotation (20 r.p.m.) and the extract was briefly sonicated in a Bioruptor (Diagenode) at a high setting for three cycles (30 s on; 30 s off). Thereafter, the nuclear lysate was centrifugated at maximum speed (20,000*g* at 4 °C) for 30 min and the supernatant (lysate 1) was transferred to a fresh microfuge tube. The insoluble pellet from the initial nuclear lysis was subjected to an additional round of extraction with ice-cold lysis buffer 2 (20 mM Tris-Cl (pH 7.9), 700 mM KCl, 1.5 mM $MgCl_2$ and 0.2 mM EDTA) completed with 10 mM sodium butyrate, 0.5 mM DTT, 1 mM A/L/P and 2 mM PMSF. This lysis step was similarly carried out at 4 °C with constant rotation (20 r.p.m.), and with five cycles of sonication (Bioruptor; Diagenode) at high settings (30 s on; 30 s off). The lysate was spun at maximum speed (20,000*g* at 4 °C) for 30 min and the supernatant (lysate 2) was transferred to a fresh microfuge tube. Both lysates were then dialysed in BC100 (50 mM Tris-Cl (pH 7.9), 2 mM EDTA, 10% glycerol, 100 mM KCl and 0.2 mM PMSF) and combined. Between 500 μg and 2 mg of lysates were used per immunoprecipitation. For each immunoprecipitation, 4 μg of either HA (Abcam; catalogue number ab9110), Top2a (Abcam; catalogue number ab52934), normal rabbit IgG (Cell Signaling; catalogue number 2729) or normal mouse IgG (Cell Signaling; catalogue number 5415) antibody was used per mg of nuclear extracts. Immunoprecipitation was carried out overnight with constant rotation (20 r.p.m.) at 4 °C. Then, 50 μl protein-G dynabeads (Thermo Fisher Scientific; catalogue number 10003D) was added to each immunoprecipitation reaction to capture the antibody–antigen complex. This was followed by extensive washes in BC200 (50 mM Tris-Cl (pH 7.9), 2 mM EDTA, 10% (*v/v*) glycerol, 200 mM KCl and 0.1% (*v/v*) NP-40). Immunoprecipitations were finally eluted with 2× Laemmli buffer and boiled at 99 °C for 10 min. Boiled samples were then placed on a magnetic stand and the eluates were transferred to a fresh tube before setting a standard western analysis, or for sample submission for mass spectrometry.

Mass spectrometry (nano-liquid chromatography with tandem mass spectrometry). The immunoprecipitation eluents were reduced in 25 mM DTT at 95 °C for 10 min and resolved by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (12% Mini-PROTEAN TGX precast gel) for 10 min at 200 V. Silver-stained gel bands were excised and in-gel digested with trypsin gold (mass spectrometry grade; Promega) using a previously described method[57] after reduction and alkylation. The resultant peptide mixture was extracted with 5% (*v/v*) formic acid in 50% (*v/v*) acetone-nitrile and vacuum dried. Liquid chromatography with tandem mass spectrometry analysis was performed using a nanoACQUITY UPLC system (Waters) coupled to an LTQ Orbitrap Elite ETD Mass Spectrometer (Thermo Fisher Scientific), as described previously described[57], with a top-15 data-dependent mass spectrometry survey scan at a resolution of 120,000 from 350–1,600 *m/z* and collision-induced dissociation tandem mass spectrometry with a normalized collision energy of 35%.

The mass spectrometry raw data file was searched against a *Mus musculus* database (downloaded on 17 March 2017 from SwissProt) by Proteome Discoverer version 1.4 (Thermo Fisher Scientific) using the SEQUEST-HT database search algorithm. The fixed modification was carbamidomethylation of cysteines and the dynamic modification was oxidation of methionine. The precursor and fragment mass tolerances were set to ±10 ppm and ±0.6 Da, respectively. An FDR at 1% was calculated using Percolator at 0.05 to select the best-scoring peptide-spectrum match. Only proteins with a minimum of two unique and high-confidence peptides identified were reported as identified.

RNA-Seq. Three RNA-Seq experiments were performed on mESCs under different conditions, each with two biological replicates. Total RNA was isolated from mESCs using the Zymo RNA Miniprep Kit (Zymo Research; catalogue number R2053) following the manufacturer's instructions. Ribosomal depletion was carried out using the NEBNext rRNA Depletion Kit (NEB; catalogue number E6310) and RNA-Seq libraries were constructed using the NEBNext Ultra II Directional RNA Library Kit (NEB; catalogue number E7760), per the manufacturer's instructions, and sequenced on a NextSeq 500 sequencer (Illumina). External RNA Controls Consortium spike-in controls (Thermo Fisher Scientific; catalogue number 4456740) were included in both RNA-Seq libraries generated from Dox-inducible NELFA-EGFP and 2-DG-treated NELFA reporter mESCs for downstream normalization purposes.

ATAC-Seq. ATAC libraries were generated following the Omni-ATAC-Seq protocol[58] as previously reported with slight modifications. Briefly, 50,000 cells were sorted using flow cytometry for each experimental condition. Sorted cells were first collected (500*g* for 5 min at 4 °C), the supernatant was discarded and cells were resuspended in 50 µl of RSB buffer containing 0.1% *v/v* NP-40, 0.1% *v/v* Tween 20 and 0.01% *v/v* Digitonin. Resuspended cells were incubated on ice for 5 min before adding 1 ml of RSB without NP-40 and Digitonin. The resulting nuclei were pelleted (500*g* for 10 min at 4 °C) and resuspended in 50 µl transposition mix (1× TD buffer, Illumina Tn5 transposase (100 nM final), 0.01% *v/v* Digitonin and 0.1% *v/v* Tween 20), before incubating at 37 °C for 30 min with constant agitation (1,000 r.p.m). Tagmented DNA was purified using the Zymo DNA Clean and Concentrator-5 kit (Zymo Research; catalogue number D4014) and libraries were amplified using Illumina Nextera indexes. Amplified libraries were size selected, pooled and sequenced on a NextSeq 500 sequencer.

ChIP sequencing (ChIP-Seq). Briefly, cells were first harvested with trypsin, quenched and resuspended in fresh media. Paraformaldehyde (Sigma–Aldrich, catalogue number F8775-25ML) was added to a final concentration of 1% (*v/v*) and cells were fixed at room temperature with constant rocking for 15 min. Thereafter, fixed cells were quenched with glycine at a final concentration of 125 mM for 5 min at room temperature. Fixed cells were then washed with ice-cold PBS. Fixed cell pellets were either snap frozen or lysed immediately. For ChIP, fixed cells were lysed with lysis buffer 1 (50 mM HEPES (pH 7.5 at 4 °C), 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40 and 0.25% Triton X) on ice. Released nuclei were centrifuged and washed once with lysis buffer 2 (10 mM Tris (pH 8 at 4 °C), 200 mM NaCl, 1 mM EDTA and 0.5 mM EGTA). Washed nuclei were then resuspended in lysis buffer 3 (10 mM Tris (pH 7.5 at 4 °C), 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and 0.5% *N*-lauroylsarcosine sodium salt) in a Bioruptor-compatible 15-ml tube. Sonication probes were attached before sonicating in a Bioruptor Plus (Diagenode; catalogue number B01020001) for a total of 15 cycles (30 s on; 30 s off) at high power. Insoluble debris were collected by centrifugation (20,000*g* for 30 min at 4 °C) and discarded. Soluble chromatin was then diluted twofold with ChIP dilution buffer (50 mM Tris (pH 7.5), 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and 1% (*v/v*) Triton) before pre-clearing with Protein A/G Dynabeads (Thermo Fisher Scientific; catalogue number 1003D). For ChIP, antibody-conjugated Dynabeads (protein A + G) were added to 0.5–2 mg chromatin for immunoprecipitation overnight with constant rotation. The following day, Dynabeads were washed with low-salt wash (20 mM Tris (pH 7.9), 2 mM EDTA, 150 mM NaCl, 1% (*v/v*) Triton X-100 and 0.1% (*v/v*) SDS), high-salt wash (20 mM Tris (pH 7.9), 2 mM EDTA, 500 mM NaCl, 1% (*v/v*) Triton X-100 and 0.1% (*v/v*) SDS) and LiCl wash (10 mM Tris (pH 7.9), 1 mM EDTA and 250 mM LiCl) before eluting in ChIP elution buffer (50 mM Tris (pH 8.0), 10 mM EDTA and 1% (*v/v*) SDS). Elution was done at 65 °C with constant agitation overnight. Immunoprecipitation DNA was reverse-crosslinked and purified with in-house SPRI beads, and sequencing libraries were constructed using the NEBNext Ultra II DNA kit (NEB; catalogue number E7645S) following the manufacturer's instructions

ChIP-qPCR. ChIP was carried out as described above. Between 2 and 4 µl of purified ChIP DNA was used for qPCR using the 2× PowerUp SYBR Green master mix (Thermo Fisher Scientific; catalogue number A25742). Fold enrichment was calculated based on the percent input method. The ChIP-qPCR primers used are listed in Supplementary Table 11.

Gene clustering to detect function groups. The transcriptome data for the seven stages of mouse pre-implantation embryos were downloaded from the Gene Expression Omnibus (GEO; accession number GSE66582 (ref. [14])).

Gene clustering was done in the R language (https://www.r-project.org/). Genes that are actively transcribed in at least one stage (fragments per kilobase of transcript per million mapped reads ≥ 5) were selected for clustering analysis. Gene-wise scaling was applied across all of the stages such that for a given gene, the highest stage expression was 1 and the lowest was 0 by the formula $x_{\text{scaled}} = \frac{x-\min(x)}{\max(x)-\min(x)}$, where x is a stage expression vector for a given gene. The gene distance matrix was then calculated based on the uncentred correlation similarity method using the dist function from the amap package (https://cran.r-project.org/web/packages/amap/index.html). Then, the hclust function from R standard stats package was applied to classify these genes into ten clusters using the hierarchical average linkage clustering method. Seven out of ten clusters that showed the most significant variations were presented.

RNA-Seq data analysis. Single- and paired-end raw sequencing reads were trimmed with Trim Galore (version 0.4.2_dev; https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with the following parameters: *-trim-n* (single-end, NELFA reporter mESCs) or *-paired* (paired-end, Dox-inducible NELFA-EGFP, 2-DG-treated NELFA reporter mESCs, and parental-matched wild-type and Dux knockout cells). Cleaned reads were then mapped to the mouse GRCm38 reference genome, guided by the vM9 gene model from the GENCODE project using the RSEM pipeline (version 1.1.11)[59]. For repetitive elements analysis, the repeat annotation in mm10 was obtained from the University of California Santa Cruz genome browser track repeatMasker. Reads mapping and quantification of individual repetitive element expression were carried out using the RepEnrich pipeline, which takes a two-step approach[60]. Briefly, reads were mapped to the mouse genome (mm10) by Bowtie (version 1.1.2)[61] and separated into multiple-mapped and uniquely mapped reads. Unique reads were assigned to the repetitive elements based on coordinate overlap, and multiple-mapped reads were subsequently mapped to repetitive element pseudogenomes generated from concatenating occurrences of individual repetitive elements. If a read maps to more than one repetitive element, each repetitive element receives a fraction of the count (1 divided by the number of repetitive elements aligned). DeSeq2 (version 1.16.1)[62] was applied to both differential gene and repetitive element expression analyses with default settings, except for the calculation of sample normalization factors, and using biological replicates as covariant. Specifically, in both differential gene and repetitive element expression analyses, sample normalization factors were calculated from External RNA Controls Consortium spike-ins for Dox-inducible NELFA-EGFP, 2-DG-treated NELFA reporter mESCs and parental-matched wild-type and Dux knockout cells, while sample normalization factors were calculated from the expression of all of the annotated genes for NELFA reporter mESCs. Genes and repetitive elements are considered to be differentially expressed if they show at least a twofold difference in expressions with an adjusted *P* value ≤ 0.01 after correcting for multiple testing by FDR (Benjamini and Hochberg FDR). To compare the expression profiles between the 2C-like cells generated by us and those established in other laboratories, RNA-Seq reads from Akiyama et al.[20] (Zscan4^{high}; GSE51682) and Hendrickson et al.[7] (si*CAF-1* and Dux overexpression; GSE85627) were downloaded from the National Center for Biotechnology Information GEO repository and processed as described above. Scatter and box plots were generated using ggplot2 in R.

Motif discovery and enrichment analysis. Homer2 (version 4.9.1) was used for motif discovery and enrichment analysis[63]. For motifs across promoters, the search space was defined as a 4 kilobase (kb) window centred at the transcription start site (*findMotifs.pl geneInput.txt mouse out/ -start -2000 -end 2000 -len 8,12 -p 10*). For motifs at ChIP-Seq binding sites, the search space was defined as a 50-base pair window centred at the peak summit (*findMotifsGenome.pl peakInput.bed mm10 out/ -size 50 -p 10*).

Functional enrichment analysis. We made use of the functions enrichGO and gseKEGG in the package clusterProfiler[64] to carry out Gene Ontology over-representation tests and Kyoto Encyclopedia of Genes and Genomes GSEA, respectively[65,66]. For Gene Ontology analysis, *P* values were calculated using hypergeometric distribution, and for GSEA analysis, *P* values were calculated based on one million permutations. For both types of analyses, pathways were considered as significant if the FDR-corrected *P* value was ≤0.05.

Sample clustering. Sample clustering was carried out to compare the 2C-like cells generated by this study (NELFA^{high}, NELFA-overexpressing and 2-DG-treated cells) against 2C-like cells from other reports (Zscan4^{high} (GEO accession GSE51682)[20], si*CAF-1* and *Dux* overexpression (GSE85627)[7] and MERVL-tdTomato^{high} (GSE33923)[10]). Additionally, pre-implantation mouse embryos of different developmental stages (GSE66582)[14] and extended pluripotency stem cells (ERP005641 (ref. [55]) and (GSE89303 (ref. [54])) were included for comparison. Read counts were obtained for each sample using the RSEM pipeline described above. Only genes that were expressed (read count ≥ 5) in at least 10% of investigated samples (three out of 32 samples in total) were included for analysis. A $\log_2$ transformation was applied after adding one pseudo-count (that is, $\log_2$[count + 1]). The ComBat function from the sva package (https://bioconductor.org/packages/release/bioc/html/sva.html) was applied on $\log_2$ expression values to correct for batch effects caused by different experiments and sequencing platforms.

A sample distance matrix was then calculated based on the correlation similarity method, using the Dist function from the amap package (https://cran.r-project.org/web/packages/amap/index.html), and the hclust function from the R standard statistical package was used for clustering by the hierarchical complete linkage clustering method.

ChIP-Seq data analysis. Paired-end raw sequencing reads were processed with Trim Galore to trim low-quality reads and remove adaptors with the parameters *-trim-n–paired*). Cleaned reads were then mapped to mm10 by Bowtie 2 (version 2.2.9)[67] with the parameters *-N 1 -L 25–no-mixed–no-discordant*. PCR duplicates were removed using SAMtools (version 1.4)[63]. Biological replicate alignment files were merged and peaks were called with the MACS2 callpeak function with the parameter –kep-dup all, to keep biological duplicate alignments[68]. Peaks with $P \leq 10^{-9}$ were kept for further analysis. Published ChIP-Seq data for DUX were downloaded from GEO accession GSE85632 (ref. [7]) and processed similarly.

ATAC-Seq data analysis. *ATAC-Seq peak calling and annotation.* Paired-end raw sequencing reads were trimmed with Trim Galore to remove adaptors and trim low-quality reads. Cleaned reads were then mapped to the mouse mm10 reference genome with Bowtie 2 (version 2.2.9)[67] with the following parameters: *-N 1–L 25–X 2000–no-mixed–no-discordant*. Mapped reads were sorted and PCR duplicates were removed using SAMtools (version 1.4)[63]. Differential ATAC peaks were called between corresponding conditions (NELFA Dox-induced cells versus uninduced; 2-DG-treated mESCs versus untreated), treating biological replicates separately. This was performed using the callpeak function of MACS2 to generate bedgraph files (with the parameters *-B–nomodel–shift -100–extsize 200*), on which the bdgdiff subcommand was applied (with the parameters *–l 500–g 250*) to call 'differential peaks'. Subsequently, high-quality overlapping peaks between biological replicates were selected using the bedtools intersect function (with the parameters *-u -f 0.5 –F 0.5*) for condition-specific peaks. In total, 18,630 and 11,306 regions lost and gained ATAC signals, respectively, in NELFA-induced cells; similarly, 18,046 and 14,811 regions were found to lose and gain ATAC signals, respectively, in mESCs treated with 2-DG. The software package GREAT[24] was used to link ATAC-gained and ATAC-lost regions in NELFA-induced cells to annotated genes (with the parameters *basal plus extension; proximal 5 kb upstream, 1 kb downstream, plus distal up to 15 kb*). ChIPSeeker[69] was used to determine overlap with genomic features; in addition, it was used for peak annotation to the nearest genes (for Supplementary Table 9), to guide the discovery of the most affected genes.

ATAC coverage bigwig file generation. Replicate-merged bedgraph files were generated using the MACS2 callpeak function (with the parameters *–B –SPMR –nomodel–shift -100–extsize 200 –t replicate1.bam replicate2.bam*). These bedgraph files were subsequently converted to bigwig format using the bedGraphToBigWig function from Kent informatics (http://hgdownload.soe.ucsc.edu/downloads.html#source_downloads). For comparisons with pre-implantation mouse embryo data, published ATAC data from Wu et al.[14] were retrieved from GEO accession GSE66390 and processed similarly. To detect chromatin accessibility across repetitive element regions, mapped reads were further filtered with SAMtools (with the parameters *view –q 10*) to remove low-quality multiple-mapping reads (for the top panel in Fig. 7a), and coverage bigwig files were generated as described above. Metagene heatmap and profile plots were generated using the deeptools suite[70].

Statistics and reproducibility. No statistical methods were used to pre-determine the sample size. Except when annotated otherwise, each experiment shown was carried out at least three times, with similar outcomes. Statistical analyses are described in detail for each panel. Briefly, statistics were performed using MACS2 for peak calling of the ChIP-Seq and ATAC-Seq experiments. Genes that were differentially expressed in the RNA-Seq were called using DESeq2 with the Wald test. The statistical analyses of ChIP-qPCR, quantitative reverse-transcription PCR, cell culture experiments and others were performed using the Prism GraphPad Software (version 7.03) or R (version 3.4.1 RC), and the *P* values were calculated either by two-tailed unpaired Student's *t*-test or two-tailed Wilcoxon tests.

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

Data availability

RNA-Seq, ChIP-Seq and ATAC-Seq data that support the findings of this study have been deposited in the GEO under accession code GSE113671. Previously published RNA-Seq data that were re-analysed here are available under accession codes GSE51682 (Zscan4^{high}), GSE85627 (si*CAF-1* and Dux overexpression), GSE33923 (MERVL-tdTomato^{high}), GSE66582 (pre-implantation mouse embryos of different developmental stages), and ERP005641 (European Nucleotide Archive: https://www.ebi.ac.uk/ena/data/view/PRJEB6168) and GSE89303 (extended pluripotency stem cells). Published ChIP-Seq data for DUX are available under accession code GSE85632. All other data supporting the findings of this study are available from the corresponding author upon reasonable request. Source data for Figs. 2–7 and Extended Data Figs. 1–7, 9 and 10 are available online.

Code availability

All of the codes used are available on request.

References

57. Bi, X. et al. Proteomic profiling of barley spent grains guides enzymatic solubilization of the remaining proteins. *Appl. Microbiol. Biotechnol.* **102**, 4159–4170 (2018).
58. Corces, M. R. et al. An improved ATAC-Seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
59. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
60. Criscione, S. W., Zhang, Y., Thompson, W., Sedivy, J. M. & Neretti, N. Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC Genomics* **15**, 583 (2014).
61. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
62. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
63. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
64. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**, 284–287 (2012).
65. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
66. Ogata, H. et al. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **27**, 29–34 (1999).
67. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
68. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
69. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
70. Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).

Acknowledgements

We thank M. Ko (Keio University School of Medicine) for the kind gift of Zscan4-Emerald reporter mESCs, M. E. Torres-Padilla (IES/Helmholtz Zentrum München) for the Zscan4-mCherry construct, D. Trono (École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland) for the Dux knockout and wild-type mESCs, G. Almouzni and J.-P. Quivy (CNRS/Institut Curie) for the CAF-1 antibodies, and D. A. Silva (BMSI, A*STAR), P. Hutchinson and T. Guo Hui (LSI, NUS) for excellent help with the flow cytometry. This research is supported by National Research Foundation (NRF) Singapore, under the NRF fellowship to W.-W.T. (NRF-NRFF2016-06) and Biomedical Research Council, Agency for Science, Technology and Research (1531C00144).

Author contributions

Z.H. performed all of the computational analyses and analysed the data. D.E.K.T. and G.C. performed the majority of the experiments and analysed the data. K.Y.S.T., X.B., Y.S.H., D.Y., H.F.L., B.J.C., H.T., M.S.L., B.W., S.B. and E.S.M.W. provided experimental support. W.-W.T. conceived of and designed the study, analysed the data and wrote the manuscript with contributions from all authors.

Competing interests

The authors declare no competing interests.

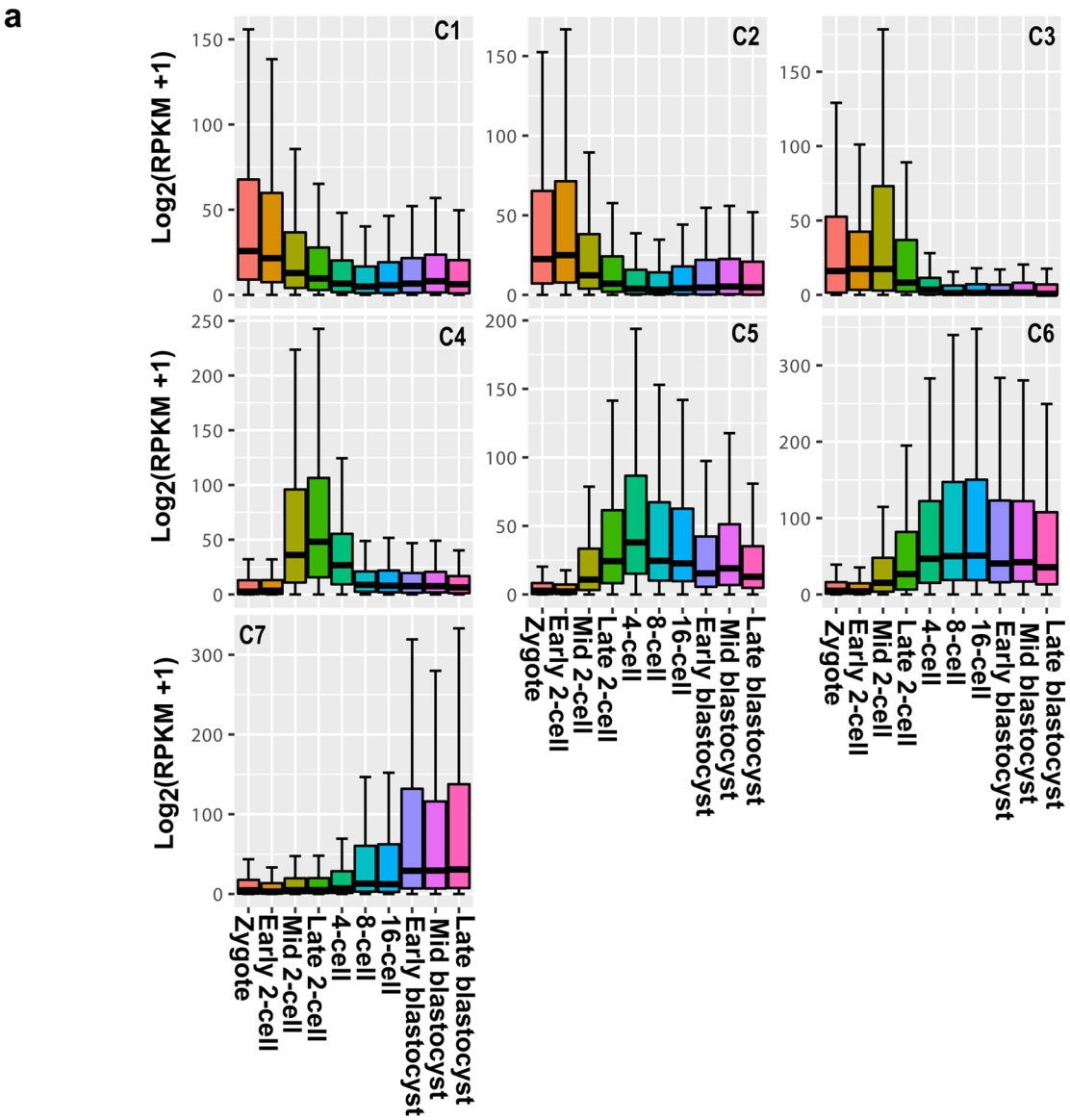
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41556-019-0453-8.

Supplementary information is available for this paper at https://doi.org/10.1038/s41556-019-0453-8.

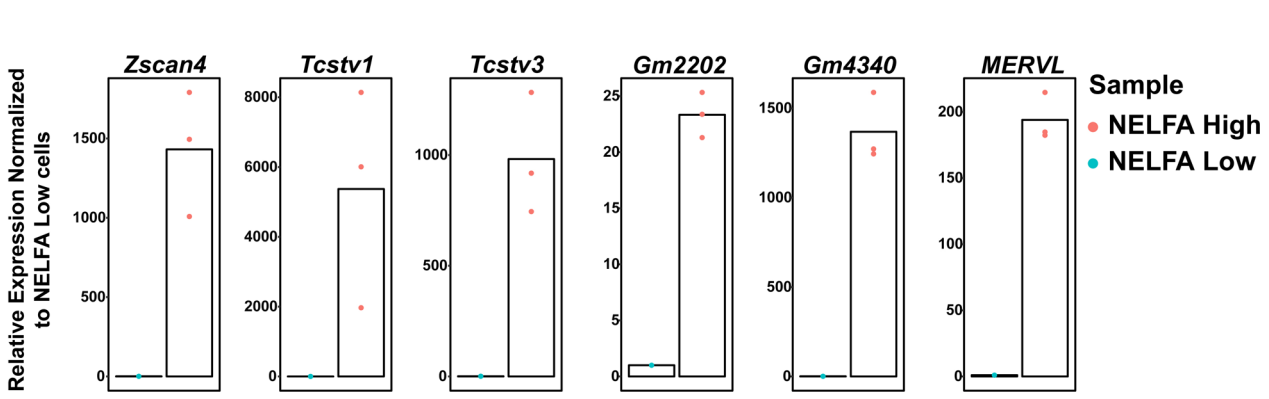
Correspondence and requests for materials should be addressed to W.-W.T.

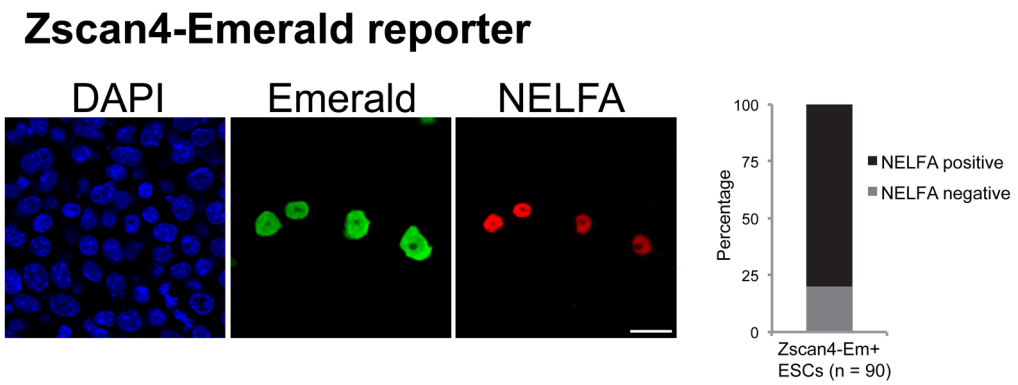
Reprints and permissions information is available at www.nature.com/reprints.

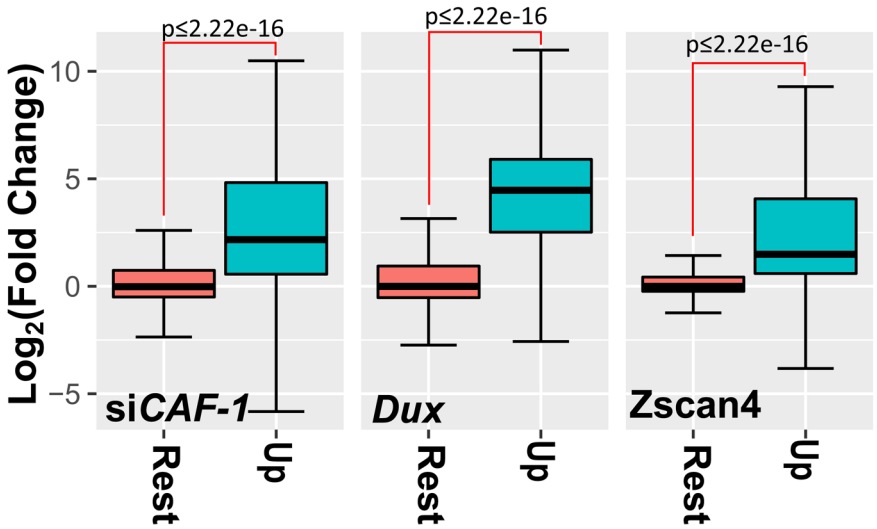


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Stage-specific expression profiles. a, Boxplots showing distinct stage-specific gene expression patterns during mouse pre-implantation development[15]. C1-C7, the seven clusters identified from Fig. 1a[14]. Number of genes (n) in C1 to C7 is 2072, 1595, 140, 445, 761, 1685, and 898 respectively. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range. **b**, *NELFA* showing distinct stage-specific expression in mouse pre-implantation embryos. Left panel: schematic depiction of the NELF (Negative Elongation Factor) complex, consisting of NELFA, NELFB, NELFC/D and NELFE subunits. Right panel: relative expression of NELF subunits in mouse pre-implantation embryos[14]. **c**, Immunofluorescence staining of Oct4 (red), NELFE (green upper panel) and NELFB (green lower panel) in mESCs, showing uniform expression of these factors. Scale bar = 20 μm.

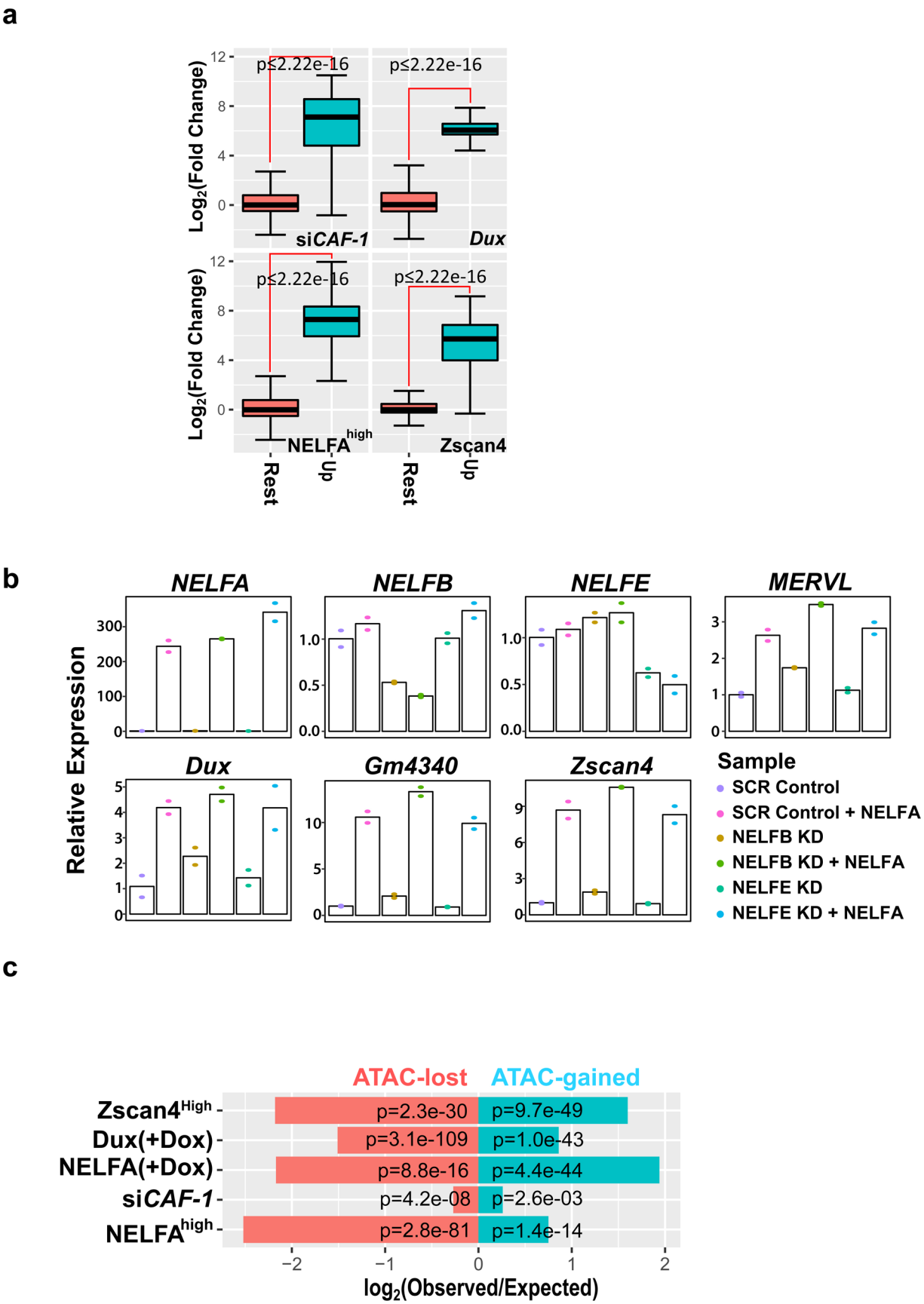






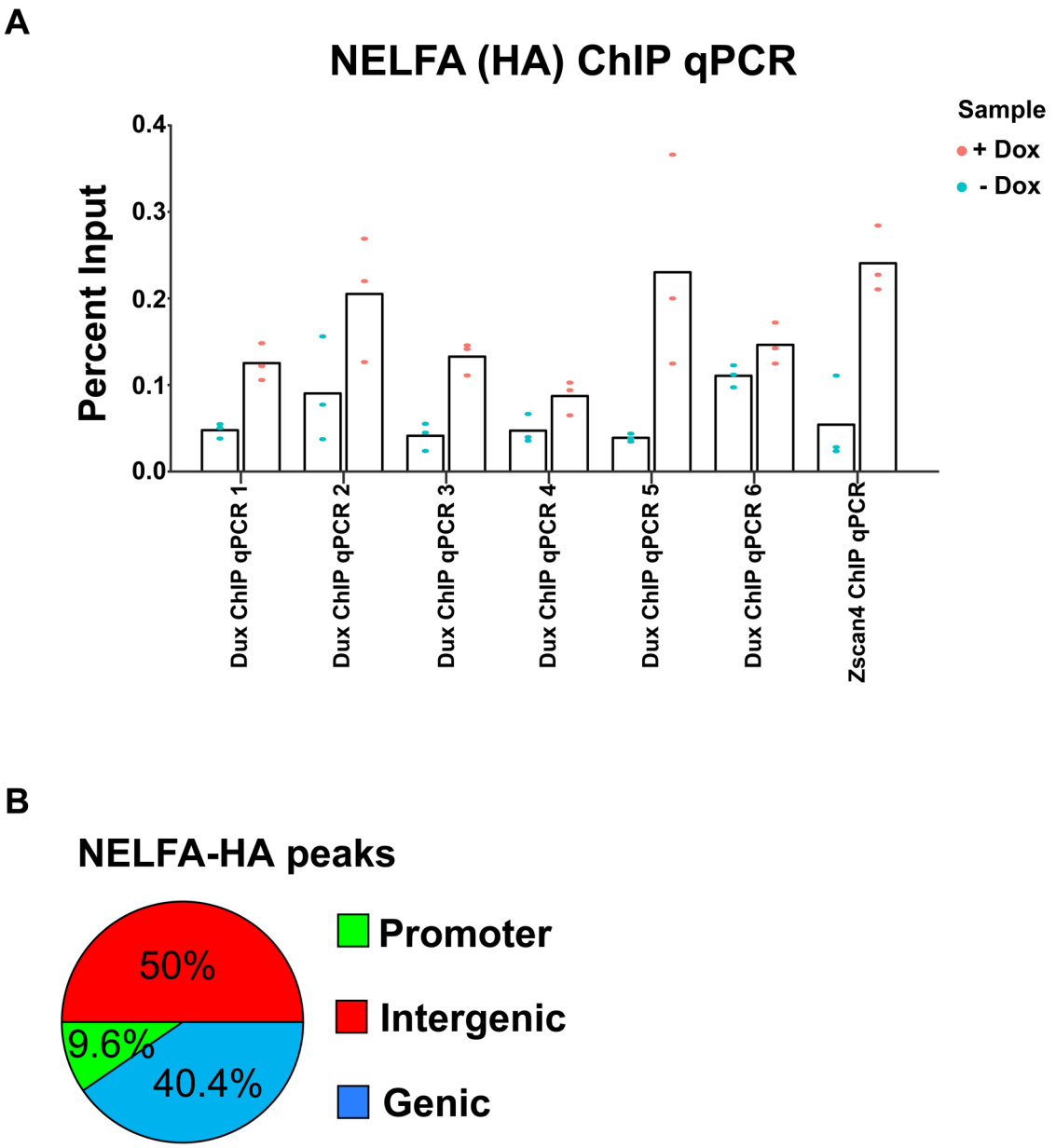
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | NELFA is a potent driver of the 2C state. a, RT-qPCR validation of selected 2C genes in NELFA reporter mESCs. The plot is generated with n=3 technical replicates from two independent experiments, with similar results obtained. **b**, Immunofluorescence for various 2C-like markers (as indicated) in Zscan4-Emerald reporter mESCs. Representative images are shown from two independent experiments. Quantification of expression of the markers in individual cells was performed based on the immunofluorescence data. Scale bar = 20 μm. **c**, NELFA^{high}-upregulated genes (n=1086) are specifically enriched in various 2C-like mESC transcriptomes. Boxplots showing the relative expression of NELFA^{high}-upregulated genes (Up) and other genes (Rest) in 2C-like cells induced by *CAF-1* inhibition (left panel), Dux overexpression (middle panel), and Zscan4-positive ESCs (right panel) respectively. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range. *p*-values determined by two-tailed Student's *T*-tests.

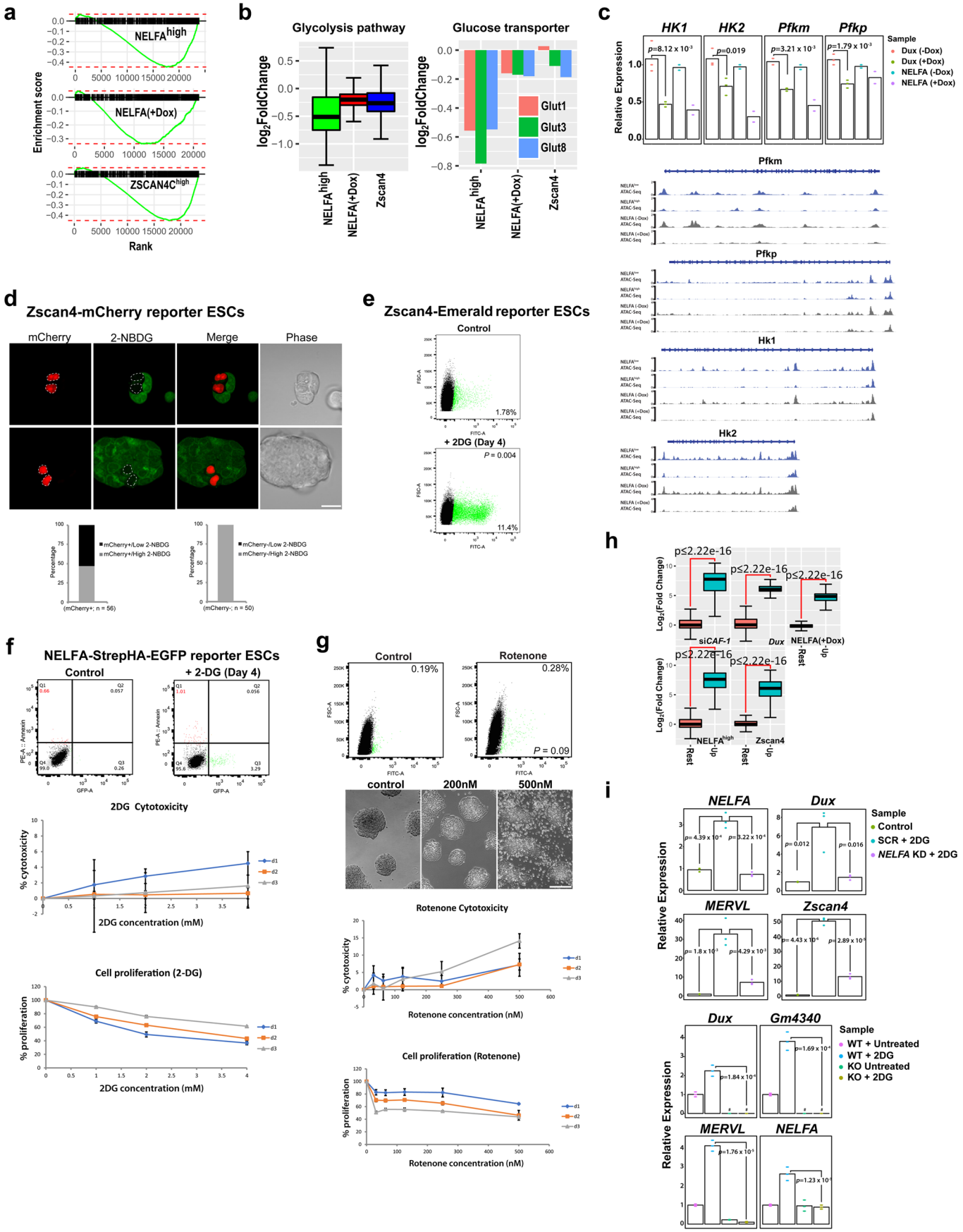


Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | NELFA-induced cells are 2C-like. a, NELFA-induced genes (n=229) are specifically upregulated in 2C-like transcriptomes. Boxplots showing the relative expression of NELFA-induced genes (Up) and non-induced genes (Rest) in various 2C-like cells, namely mESCs induced by *CAF-1* knockdown[7] (upper left panel) and Dux overexpression[7] (upper right), our NELFA^{high} cells (bottom-left), and Zscan4-positive mESCs[20] (bottom-right) respectively. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range. *p*-values determined by two-tailed Student's *T*-tests. **b**, RT-qPCR analysis of select 2C genes in mESCs following overexpression of NELFA, as function of *NELFB* and *NELFE* knockdown. Plot is representative of n=2 independent experiment with similar results obtained, with mean from two independent experiments represented as bar graph. **c**, 2C-like genes from various 2C-like cell models (NELFA^{high}, n=1086; NELFA(+Dox), n=229; Dux(+Dox), n=2236; Zscan4^{high}, n=456) are enriched within ATAC-gained region linked genes (n=4002), but depleted within ATAC-lost region linked genes (n=10332), in NELFA-induced 2C-like cells. ATAC regions were annotated to genes using GREAT[24], and *p*-values were determined by Chi-squared tests, assuming a background of 20,000 genes.

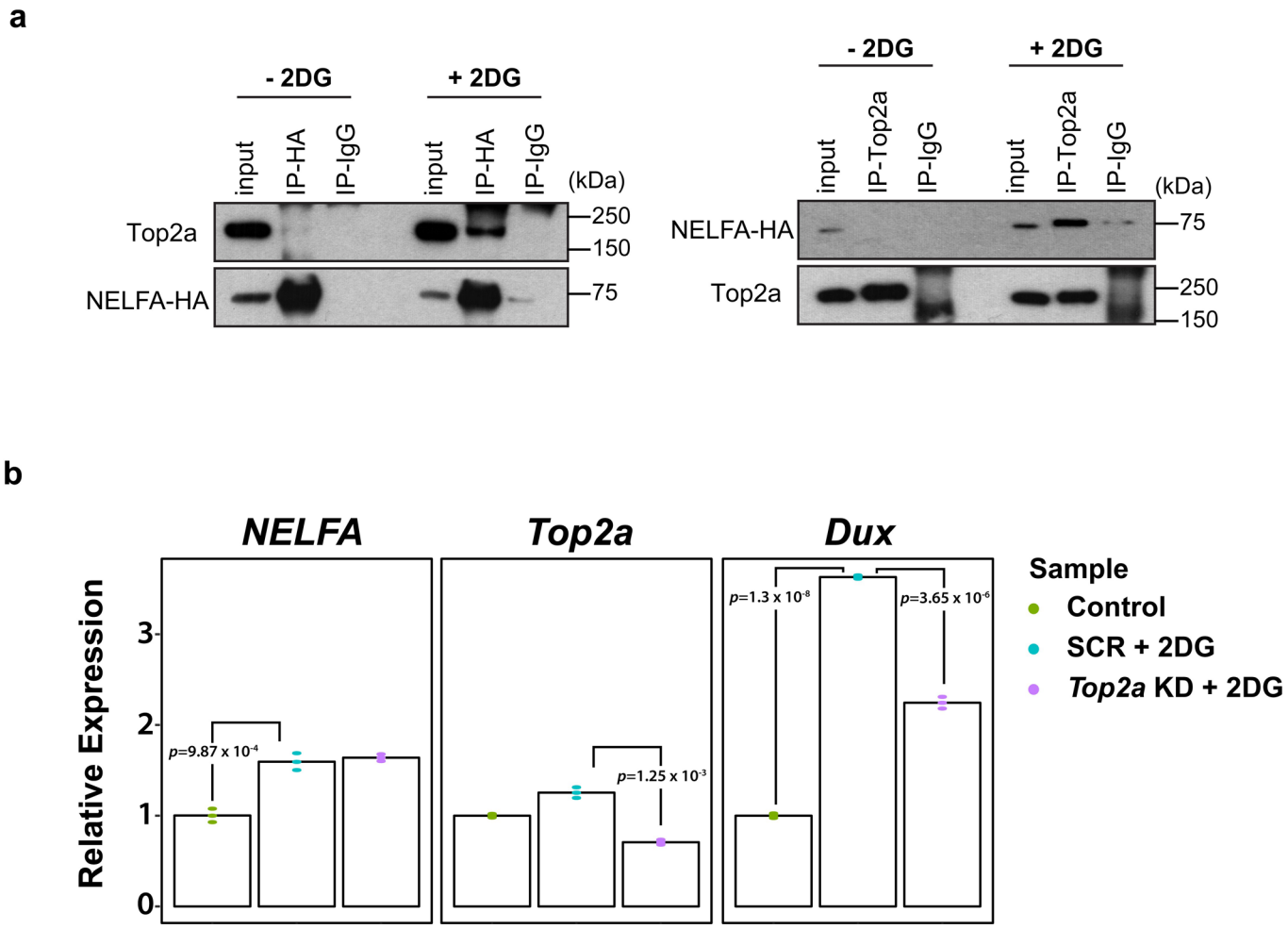


Extended Data Fig. 4 | NELFA (HA) ChIP-qPCR validation. a, NELFA (HA) ChIP-qPCR validation of NELFA occupancy at *Dux* and *Zscan4c* promoter regions. Percent input values shown are normalized against -Dox untreated control condition. Dots represent n=3 technical replicates from two independent experiments. Primers are listed in Supplementary Table 11. **b**, Pie chart detailing the genomic distribution of NELFA (HA) ChIP-seq peaks.

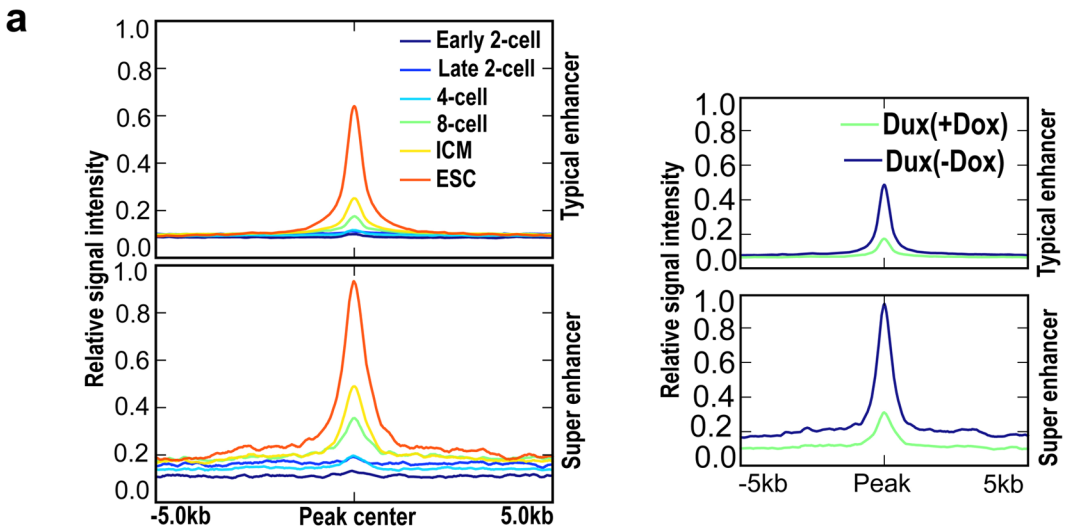


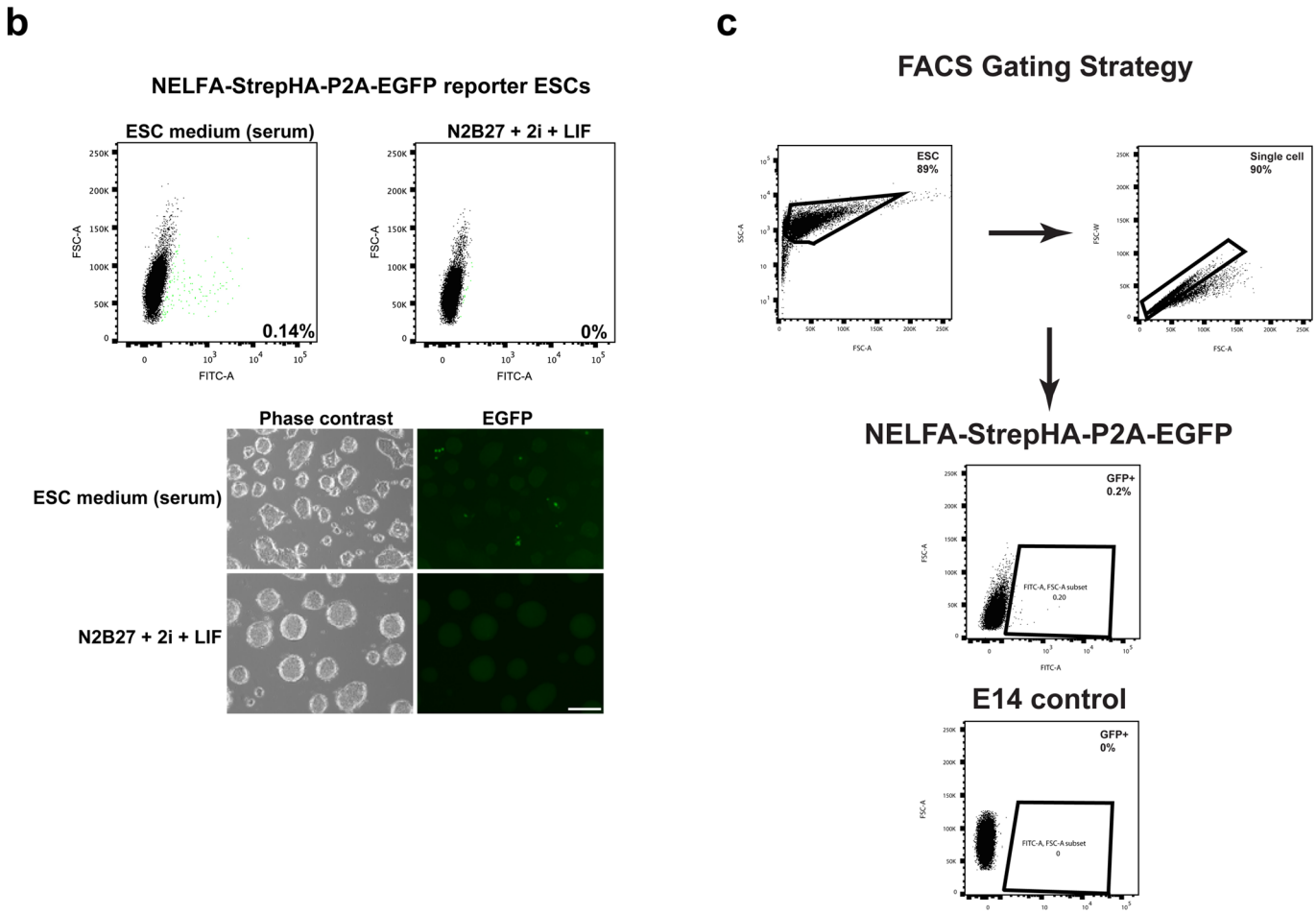
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Suppressed glycolysis is a feature of 2C-like cells. a, GSEA enrichment plot for metabolic pathways (mmu01100) in the respective cell lines. The leading gene list is in Supplementary Table 10**. b**, Left panel: box plots charting distribution of changes in glycolysis genes (n=65 genes; mmu00010) in the respective cell lines. Right panel: bar charts showing expression of glucose transporters genes in the respective cell lines. **c**, RT-qPCR analysis of key glycolysis genes (top panel) and ATAC-seq signals (bottom panel) in both NELFA- and Dux-induced 2C-like cells. n=3 and n=2 biological replicates for Dux and NELFA induction, respectively. **d**, 2-NBDG staining in Zscan4-mCherry reporter mESCs. Scale bar = 20 μm. Three independent experiments were performed. N numbers represent ESCs and are indicated on the graph. **e**, Representative FACS analysis of Zscan4-Em reporter mESCs after 2-DG treatment. 3 independent experiments. **f**, Representative FACS analysis of Annexin-V PE-stained NELFA reporter mESCs, with and without 2-DG treatment (upper panel). Assessment of cell cytotoxicity and proliferation rate as a function of 2-DG treatment (middle and lower panels), **g**, Representative FACS analysis of NELFA reporter cells following rotenone treatment (upper panel). Representative microscope images of mESCs following rotenone treatment (second panel from top). Scale bar = 200 μm. Assessment of cell cytotoxicity and proliferation rate following rotenone treatment (bottom 2 panels). **h**, 2-DG-induced genes (n=175) are specifically upregulated in 2C-like transcriptomes. Boxplots showing the relative expression of 2-DG-induced genes (Up) and non-induced genes (Rest) in various 2C-like cells. **i**, RT-qPCR analysis showing impaired activation of 2C genes upon 2-DG treatment and knockdown of *NELFA*, n=3 biological replicates. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range (**b**, **h**). Data presented as the mean ± s.e.m of $n = 3$ independent experiments (**f**, **g**). *p*-values determined by two-tailed Student's *T*-test (**c**, **e**, **g**, **h**, **i**).

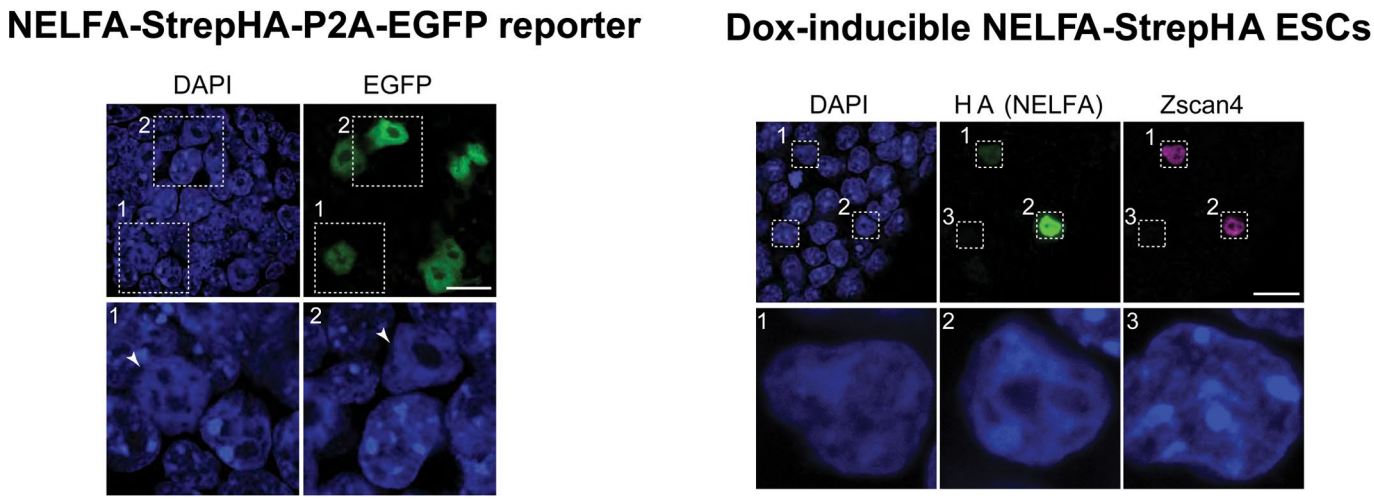


Extended Data Fig. 6 | Top2a is involved in the activation of Dux by NELFA. a, Reciprocal co-IP western analysis in 2-DG-treated mESCs, showing robust interaction between Top2a and NELFA specifically upon 2-DG treatment. Co-IP of NELFA (HA) (left panel) and Top2a (right panel) with and without 2-DG treatment. Two independent experiments were performed, with similar results. **b**, RT-qPCR analysis of *Dux* following transient *Top2a* knockdown and 2-DG treatment. Mean is represented as bar graph derived from n=3 independent experiments and *p*-values are determined using a two-tailed Student's *T*-tests. Unprocessed western blots are shown in Source Extended data Fig. 6.

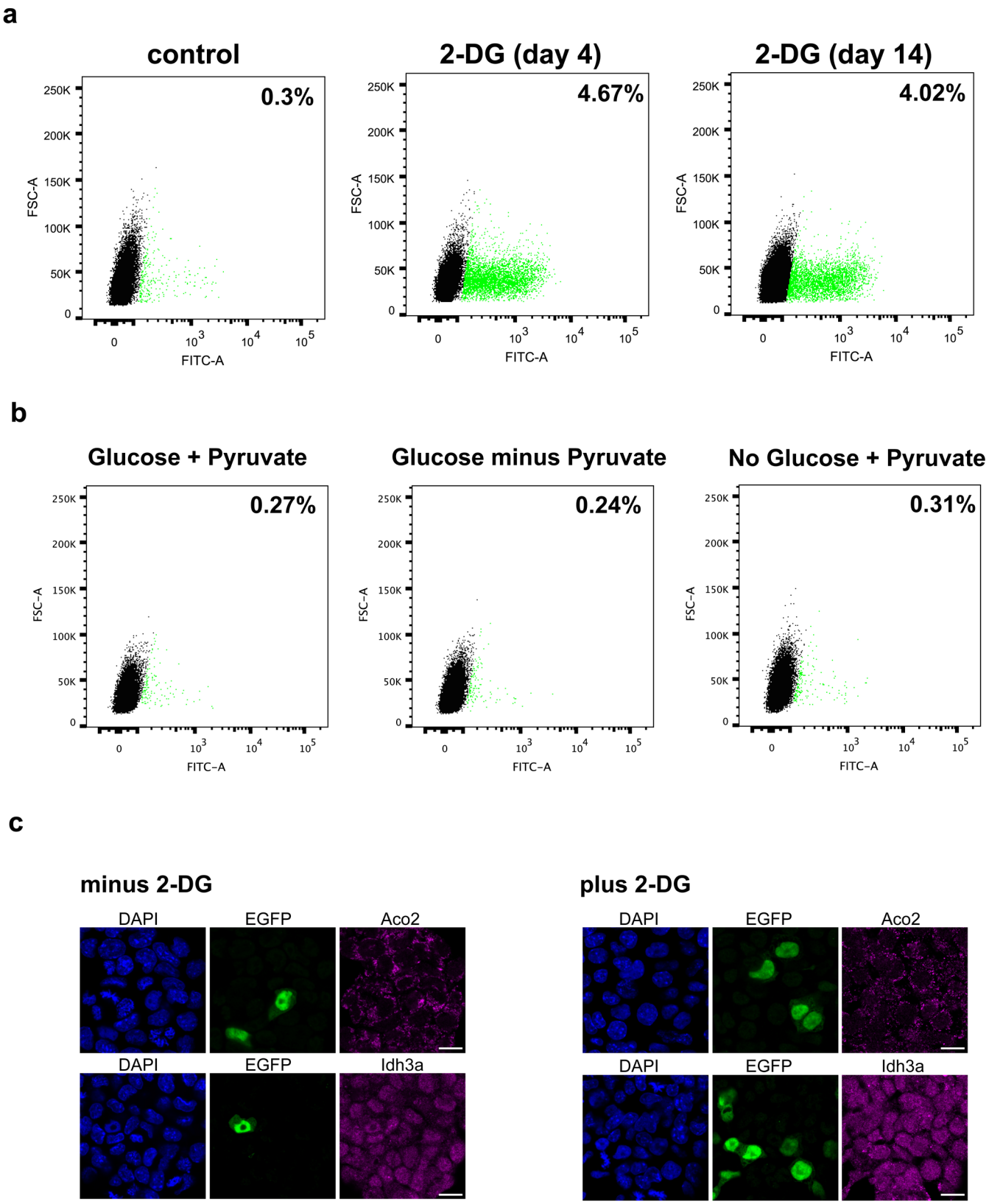




Extended Data Fig. 7 | Naïve mESC culture conditions restrict expression of 2C genes. a, Metagene analysis of ATAC-seq signals in pre-implantation mouse embryos of different developmental stages (left panel) and Dux-overexpressing cells (right panel) across ESC-specific enhancers, including super-enhancers and typical enhancers. Enhancer intervals were obtained from published data[56]. **b**, FACS analysis of NELFA-StrepHA-P2A-EGFP mESCs cultured in naïve, serum-free ESC conditions (N2B27/2i/LIF), showing the loss of the NELFA^{high} subpopulation under naïve conditions. Representative data from three independent experiments. Representative phase contrast and EGFP fluorescence images are also shown. Scale bar = 200 μm. **c**, General FACS gating strategy.

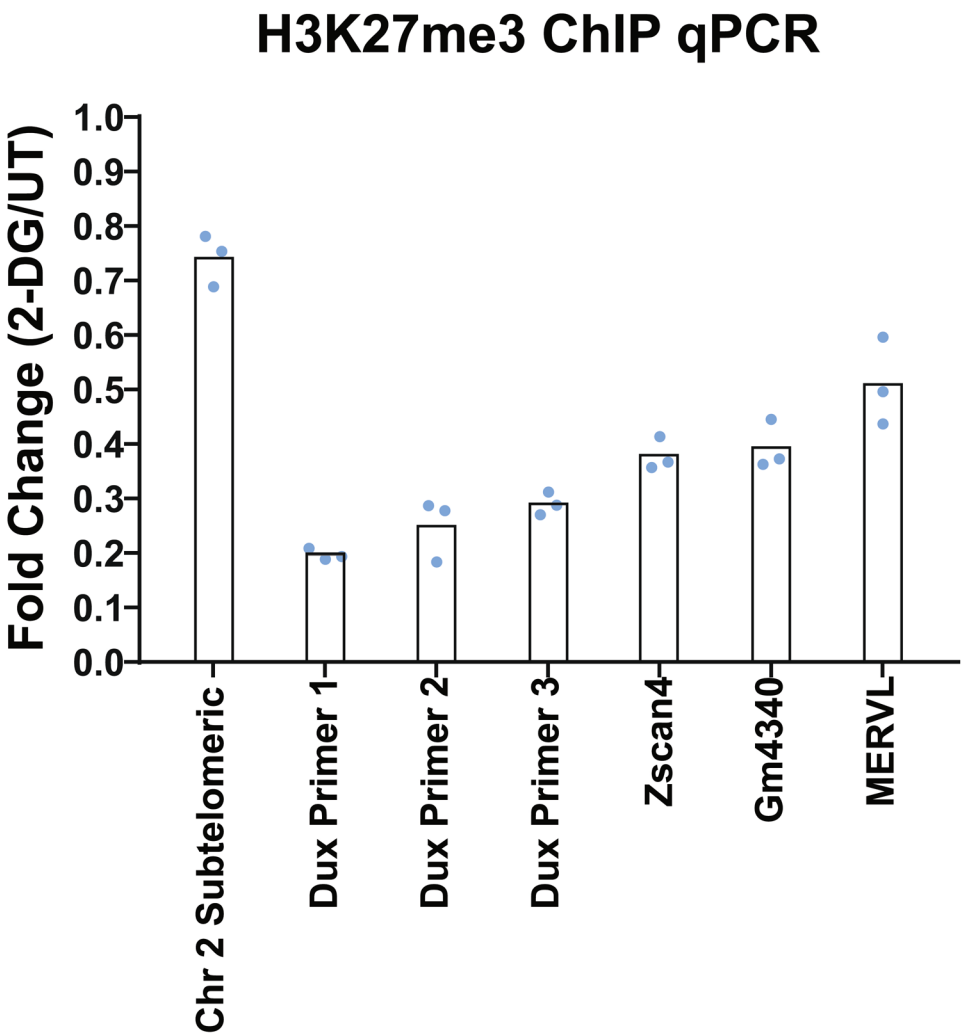


Extended Data Fig. 8 | NELFA induction triggers extensive chromatin remodeling. Left panel: Immunofluorescence for EGFP (marking NELFA; green) in NELFA reporter mESCs reveals structural changes in heterochromatin. 5 independent experiments were performed, with similar results. Right panel: Immunofluorescence for HA (marking NELFA; green) and Zscan4 (magenta) in Dox-induced NELFA-StrepHA-P2A-EGFP mESCs reveals structural changes in heterochromatin. DAPI (blue) staining depicts the different extents of chromatin decondensation. Cells demarcated by white dotted boxes are shown in magnification. Data is representative of 3 independent experiments. Scale bar = 20 μm.

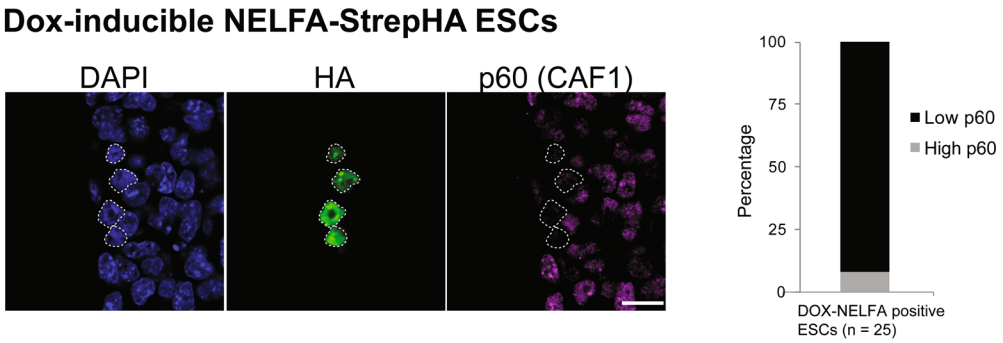


Extended Data Fig. 9 | Assessing metabolic requirements of 2C-like cells. a, FACS analysis of NELFA reporter cells following 4 and 14 days of 2-DG treatment. Two independent experiments were performed, with similar results. **b**, FACS analysis of NELFA reporter cells under different culture conditions as indicated. Two independent experiments were performed, with similar results. **c**, Confocal images of Aco2 and Idh3a in both NELFA reporter and NELFA-inducible mESCs. Representative images of two independent experiments are shown. Scale bar = 20 μm.

a



b



Extended Data Fig. 10 | H3K27me3 decreased across 2C genes following 2-DG treatment and CAF-1 is depleted in NELFA induced cells. a, ChIP-qPCR quantification of H3K27me3 in NELFA reporter mESCs, with and without 2-DG treatment. Fold change < 1 indicates loss of this histone modification at the tested locus. Chromosome 2 subtelomeric region serves as a negative control. Dots represent n=3 technical replicates from two independent experiments. The primers used are listed in Supplementary Table 11. **b**, Immunofluorescence for HA (marking NELFA; green) and the CAF-1 p60 subunit (magenta) in NELFA-induced mESCs. Quantification of expression of the markers in individual cells was performed based on the immunofluorescence data. N number represents ESCs and is indicated on the graph. Two independent experiments were performed, with similar results. Scale bar = 20 μm.

nature research

Corresponding author(s): Wee Wei Tee

Last updated by author(s): 29/11/19

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data collection for qPCR and ChIP qPCR were collected using Quantstudio design and analysis (version 1.4) by Applied biosystems. Confocal images were captured using Zeiss LSM 700 confocal microscope and were processed using Image J (Version 2.0.0-rc-46/1.50g), Adobe Photoshop and Adobe Illustrator CS4. Flow cytometry was performed on either a FACSAria II (SORP) or MoFlo XDP sorter, and data analyze collected using (Aria (FACSDiva, Version 7), MoFlo (Summit, Version 6)). FACS analysis were performed using BD LSR II and data was processed using (FlowJo, Version 9). Mass spectrometry samples were ran on a LTQ Orbitrap Elite ETD mass spectrometer (Thermo Fisher) and results were analyzed using proteome discoverer (Version 1.4, Thermo Fisher) using default settings.
Data analysis	All the analysis were done using custome R code. The code for data analysis is available upon reasonable request. Trim Galore (v0.4.2_dev); Bowtie (v1.1.2); DESeq2 (v1.16.1) ; Bowtie2 (v.2.2.9); SAMtools (v1.4); Homer2 (v4.9.1) ; BEDOPS (v2.4.2821); deeptools (v2.5.3); R (v3.4.1 RC) ; macs2 (v2.1.1);bedtools (v2.26.0); clusterProfiler(v3.10.1) ; sva(v3.30.1); ChIPseeker (v1.18.0); ggplot2 (v3.2.1); amap (v0.8-16); RepEnrich(v0.1, https://github.com/nskvir/RepEnrich); RSEM (v1.1.11, https://github.com/deweylab/RSEM); bedGraphToBigWig (v4, http://hgdownload.soe.ucsc.edu/downloads.html#source_downloads); GREAT (v3.0.O, http://great.stanford.edu/great/public-3.0.0/html/)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The datasets generated during the current study are deposited at GEO under the accession number GSE113671

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size calculation was not performed. The sample sizes were determined based on prior experience and standards within the field. All key experiments were repeated independently using different cells lines (e.g. 2-DG treatment using both NELFA-StrepHA-P2A-EGFP and Zscan4-Emerald cells; Fig 6b and Extended Fig 5e) or different techniques (e.g. Fig 6a, RNA-seq; Extended data 5c, qPCR; Extended data 5d, 2-NBDG labelling). For the immunofluoresence experiments, at least 2 independent experiments were performed with similar outcome. The the number of cells analyzed are clearly indicated in the figures. For the FACS analyses, multiple independent runs were performed (the exact number for each experiment is indicated in the figure legends) with at least 3 independent experiments per treatment condition unless otherwise stated. The sample sizes were determined based on prior experience and standards within the field. In all the ATAC-seq (Fig 3e and f; Fig 4a; Fig 7a and d; Extended data Fig 5c) and RNA-seq (Fig 2b; Fig 3a; Fig 4c; Fig 6c) analyses, each experiment was performed in two independent cell cultures. Sample size was not pre-determined, and experiments with this sample size are in accordance to conventions in this field. In the ChIP-seq analyses (Fig 5a), each experiment was performed in three independent cell cultures. Sample size was not pre-determined, and experiments with this sample size are in accordance to conventions in this field.
Data exclusions	No data was excluded.
Replication	All data are replicated at least thrice, and in some cases, twice. The exact numbers are clearly indicated in the figure legends. All attempts at replication were successful.
Randomization	The embryos used in Fig 2e and 6g were randomly assigned to two experimental groups. Randomization was not relevant for other experiments as the samples were treated or analyzed in the same manner.
Blinding	Blinding was not applicable to our stem cell associated assays (e.g 2-DG treatment, siRNA knockdown). The experiments were descriptive in nature.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

The following antibodies are used are listed in Extended Data Table 13:

Anti-NELFA, Bethyl (Cat# A301-910A, lot: A301-910A-1), 1 in 500 (IF)
Anti-NELFA, Sigma (Cat# SAB1406594, lot: 9057), 1 in 200 (IF)
Anti-NELFE, Sigma (Cat# HPA007187, lot: A106463), 1 in 250 (IF), 1 in 2000 (Western)
Anti-NELFB, Rockland (Cat# R01/200-301-D18, lot: 200-301-D18), 1 in 50 (IF), 1 in 1000 (western)
Anti-Aco2, Abcam (Cat# ab110321, lot: GR-254688-3), 1 in 500 (IF)
Anti-Idh3a, Abcam (Cat# ab58641, lot: GR302381-3), 1 in 200 (IF)
Anti-mCherry, Abcam (Cat# ab167453, lot: GR312817-9), 1 in 400 (IF)
Anti-p60 (CAF1), Gift from G.Almouzni, 1 in 1000 (IF)
Anti-Cdx2, Biogenex (Cat# MU392A-5UC, lot: MU392A0516), 1 in 250 (IF)
Anti-Top2a, Abcam (Cat# ab52934, lot: GR241638-21), 1 in 2000 (Western)
Anti-Oct-3/4, Santa Cruz (Cat# sc-8628, lot: I2115), 1 in 500 (IF)
Anti-GFP, Abcam (Cat# ab6673, lot: GR202861-4), 1 in 600 (IF)
Anti-MERVL-Gag, Epigentek (Cat# A-2801-100, lot: 607021), 1 in 400 (IF)
Anti-Zscan4, Abnova (Cat# H00201516-B01P, lot: G1211), 1 in 500 (IF)
Anti-HA, Biolegend, (Cat# 901501, lot: B220766), 1 in 250 (IF), 1:1000 (Western)
Anti-HA, Abcam (Cat# ab9110, lot:GR304617-2), 1 in 1000 (IF), 1 in 2000 (Western)
Anti-H3K27me3, Cell Signalling (Cat# C36B11, lot: Lot 8), 4ug per 1mg input for ChIP
Alexa Fluor 488 Donkey Anti-Rabbit IgG (H+L), Thermo-Fisher (Cat# A-21206, lot: 1723019), 1 in 500 (IF)
Alexa Fluor 488 Donkey Anti-Rat IgG (H+L), Thermo-Fisher (Cat# A-21208 , lot:1750277), 1 in 500 (IF)
Alexa Fluor 488 Donkey Anti-Goat IgG (H+L), Thermo-Fisher (Cat# A-11055 , lot:1737907), 1 in 500 (IF)
Alexa Fluor 594 Donkey Anti-Rabbit IgG (H+L), Thermo-Fisher (Cat# A-21207, lot: 1744751), 1 in 250 (IF)
Alexa Fluor 647 Rabbit Anti-Goat IgG (H+L), Thermo-Fisher (Cat# A-27018, lot: RA231460), 1 in 250 (IF)
Alexa Fluor 647 Donkey Anti-Rabbit IgG (H+L), Thermo-Fisher (Cat# A-31573, lot: 1786284), 1 in 250 (IF)
Alexa Fluor 647 Donkey Anti-Mouse IgG (H+L), Thermo-Fisher (Cat# A-31571, lot: 1757130), 1 in 250 (IF)

Validation

All antibodies used are commercially available and widely used by the scientific community.

NELFA: We have validated NELFA staining with 2 different antibodies and obtained identical results
NELFE: Correctly stained NELFE in a study of NELF complex pausing, as expected (Plos Genetics, e1004090)
Aco2: Correctly stained the metabolic enzyme for localization studies as expeced (Cell, 210-223 e11 (2017))
Idh3a: Same study as above (Cell, 210-223 e11 (2017))
mCherry: expressed as fusion protein to study localization, stained correctly as expected (doi: 10.7554/eLife.37949.)
p60 (CAF1): Used to detect CAF1 levels following manipulation, stained correctly as expected (doi: 10.1083/jcb.143.3.563)
Cdx2: Used for staining Cdx2 localization in embryo study, stained correctly as expected (doi: 10.1242/dev.141390)
Top2a: Used for immunostaining studies, stained correctly as expected (doi:https://doi.org/10.1016/j.cell.2017.06.034)
Oct4: Used for immunostaining studies, stained correctly as expected (https://doi.org/10.1073/pnas.1615540113)
MERVL: Used to study expression of MERVL in embryos, stained correctly as expected (doi: 10.15252/embr.201846240)
Zscan4: Used in study expression during 2C state transition in ES cells, stained correctly as expected (https://doi.org/10.1093/dnares/dsv013)
HA: used to detect HA-fused proteins for ChIP in a study of chromatin remodeling, captured and stained correctly as expected (https://doi.org/10.1093/nar/gky438)
H3K27me3: Used for ChIP-Seq profiling, worked as expected (doi: 10.1038/s41467-019-09624-w)

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	E14 mouse ES cells (from ATCC)
Authentication	E14 cells were purchased from ATCC and verified with transcriptome profiling and immunofluoresence experiments.
Mycoplasma contamination	Cells were tested routinely for mycoplasm and are free of contamination at the point of our experiments.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Mice (Mus musculus) were used to obtain mouse embryos for this study. The following strains were used: ICR and C57BL/6N. Female mice used were 8-10 weeks old and males 12-20 weeks old.
Wild animals	No wild animals used in this study
Field-collected samples	The study did not involve field-collected samples
Ethics oversight	IACUC approval was obtained

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	Raw and processed data associated with this study can be accessed under GSE113671
Files in database submission	ChIP-seq.NELFA_plusDox_rep1_HAantibody1_1.fastq.bz2 ChIP-seq.NELFA_plusDox_rep1_HAantibody1_2.fastq.bz2 ChIP-seq.NELFA_plusDox_rep1_input_1.fastq.bz2 ChIP-seq.NELFA_plusDox_rep1_input_2.fastq.bz2 ChIP-seq.NELFA_plusDox_rep2_HAantibody1_1.fastq.bz2 ChIP-seq.NELFA_plusDox_rep2_HAantibody1_2.fastq.bz2 ChIP-seq.NELFA_plusDox_rep2_input_1.fastq.bz2 ChIP-seq.NELFA_plusDox_rep2_input_2.fastq.bz2 ChIP-seq.NELFA_plusDox_rep3_HAantibody1_1.fastq.bz2 ChIP-seq.NELFA_plusDox_rep3_HAantibody1_2.fastq.bz2 ChIP-seq.NELFA_plusDox_rep3_HAantibody2_1.fastq.bz2 ChIP-seq.NELFA_plusDox_rep3_HAantibody2_2.fastq.bz2 ChIP-seq.NELFA_plusDox_rep3_input_1.fastq.bz2 ChIP-seq.NELFA_plusDox_rep3_input_2.fastq.bz2 GSE113671_mNELFA_DOXplus_ChIP_input.bw GSE113671_mNELFA_DOXplus_ChIP_IP.bw NELFA_ChIP_peaks.narrowPeaks.gz
Genome browser session (e.g. UCSC)	N/A

Methodology

Replicates	Total of seven samples with three replicates for IP and corresponding inputs. For replicate3, two antibodies have been used to ChIP for NELFA
Sequencing depth	Sample TotalReads CleanReads_TrimGalore UniquelyMapped ChIP-seq.NELFA_plusDox_rep1_HAantibody1 82,306,842 82,223,287 60,831,083 ChIP-seq.NELFA_plusDox_rep1_input 56,937,855 56,866,562 42,168,229 ChIP-seq.NELFA_plusDox_rep2_HAantibody1 79,302,552 79,205,493 58,217,787 ChIP-seq.NELFA_plusDox_rep2_input 58,241,856 58,156,534 43,106,961 ChIP-seq.NELFA_plusDox_rep3_HAantibody1 62,978,691 62,869,829 45,692,365 ChIP-seq.NELFA_plusDox_rep3_HAantibody2 65,429,240 65,359,184 47,842,272 ChIP-seq.NELFA_plusDox_rep3_input 72,208,108 72,123,076 50,467,858
Antibodies	Antibody information is listed in Extended Data Table 13
Peak calling parameters	macs2 callpeak -t NELFA_mergedIP.rmdup.bam -c NELFA_mergedInput.rmdup.bam -n NELFA_mm10 -g mm -f BAMPE --keep-dup all. "keep-dup all" is used because the PCR duplicates have already been removed before merging.
Data quality	Only NELFA peaks of high quality were kept by applying the p <= 10e-09. In addition, Peaks that overlap blacklist regions were ignored.
Software	Trim Galore (v0.4.2_dev); Bowtie2(v.2.2.9); SAMtools(v1.4); deeptools (v2.5.3); macs2(v2.1.1); bedtools (v2.26.0)

Flow Cytometry

Plots

Confirm that:

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

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

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

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

Methodology

Sample preparation	Mouse ES cells were dislodged with trypsin and was resuspended with FACS buffer (1X PBS, 2% Serum and 2X Pen-Strep). Finally, resuspended cells were passed through a cell strainer (0.7 um) into FACS tube. Samples were transported on ice to sort facility.
Instrument	For RNA-Seq samples, cells were sorted on either the FACSAria II (SORP) or MoFlo XDP machine (BD), while GFP population analysis was analyzed with the LSRII machine (BD).
Software	Flow cytometry was performed on either a FACSAria II (SORP) or MoFlo XDP sorter, and data analyze collected using (Aria (FACSDiva, Version 7), MoFlo (Summit, Version 6)). FACS analysis were performed using BD LSR II and data was processed using (FlowJo, Version 9).
Cell population abundance	Sorted cells were visually inspected for GFP fluorescence and the sorting purity was typically greater than 95%.
Gating strategy	Gating strategy is provided in Extended data Fig 7c. E14 parental cell line (non-EGFP) was used to establish baseline fluorescence. A preliminary gating was done to determine and assess the degree of separation between the GFP- positive and -negative population, and to bias for singlets . To enrich for GFP positive population, a stringent sort where GFP positive cells that have fluorescence intensity that are well separated from baseline are collected for downstream processing.

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