Peer Review Information:

Nature Cell Biology thanks Laurent David and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Editor summary:

Carbognin, Carlini, Panariello et al. report that ESRRB drives activation of the transcriptional program regulating the formative transition of naïve embryonic stem cells.

1. Extended Data

Figure or Table # Please group Extended Data items by type, in sequential order. Total number of items (Figs. + Tables) must not exceed 10.	Figure/Table title One sentence only	Filename Whole original file name including extension. i.e.: Smith_ED_Fig1.jpg	Figure/Table Legend If you are citing a reference for the first time in these legends, please include all new references in the main text Methods References section, and carry on the numbering from the main References section of the paper. If your paper does not have a Methods section, include all new references at the end of the main Reference list.
Extended Data Fig. 1	Gene signatures of different pluripotent states	Carbognin_ED_Fig1.jpg	a: Line chart showing dynamics of mRNA expression based on qPCR of four pluripotency markers (*Tfcp2l1, Esrrb, Sall4, Oct4*) in E14 cells during monolayer differentiation (withdrawal of either 2iL or 2i for 96h) both in 2iL (purple) and 2i (green). White circles indicate the mean of n=4 independent experiments, shown as dots. P-values indicate two-sided unpaired t-test between the indicated time points. b: Heatmaps showing Z-score normalised expression of all genes of each group (defined in Fig. 1d) in E14 cells differentiating from 2iL (purple box) and 2i (green box). Integration of n=2 independent biological replicates for each time point. See also Supplementary Table 2 for the biological processes enriched in the 6 signatures.
Extended Data Fig. 2	Transcriptional response	Carbognin_ED_Fig2.jpg	a: Bar plot showing the number of AP positive colonies in the clonal assay of

	changes during commitment		cells cultured in 2iL and during differentiation (purple bars) and of cells in which 2iL was re-applied for 24h at the indicated time point (yellow bars, re-induction). Bars indicate mean +/-SD of n=8 independent experiments, shown as dots. Only the sample '24' was measured in n=4 independent experiments. Two-sided unpaired Student t-test. b: Heatmaps showing Z-score normalised expression of selected genes for each group (naive, formative, committed) during differentiation and re-induction. Integration of n=4 independent experiments for each time point. c: Barplots showing expression by RNAseq of Jak/Stat direct targets (Socs3 and Stat3, orange), WNT targets (Cdx2 and Axin2, green) and FGF targets (Dusp6 and Spry4, purple) in differentiating cells and after re-induction with 24h of LIF. Mean +/-SD of n=4 independent experiments. d: UCSC genome browser visualisation of normalised ATAC-seq profiles at the indicated loci. Rectangles indicate peaks found only in 2iL (green) or only at 48h (red). Integration of n=2 biological replicates. e: Volcano plot summarising published RNA-seq data[98] of ESCs cultured in Serum+LIF (S+L) or 2iL. Data were interpolated with the six groups of genes identified in Fig. 1 (naive early and late, formative early and late, committed early and late). f: Schematic representation of experimental strategy. Cells overexpressing pluripotency genes were mixed and differentiated for 96h. The clonal assay was then performed and cells were collected after 4 days. PCR on

			genomic DNA was used to identify factors enriched in pluripotent colonies. g: Bar plot showing quantification of AP positive colonies of cells overexpressing an empty vector or pluripotency factors, either maintained in 2iL or differentiated for 96h. Bars indicate mean n=2 independent experiments, shown as dots.
Extended Data Fig. 3	Characterisation of ESC differentiation and regulation of Esrrb expression	Carbognin_ED_Fig3.jpg	a: Representative images of immunostaining for EpiSCs markers (Oct4 and T) in WT cells maintained in 2iL or differentiated for 96h in N2B27 or in presence of CHIR and Activin A to induce T expression. Scale bar=25μm. Similar results were obtained in n=2 independent experiments. b: Barplots showing expression by RNA-seq of key EpiSCs markers in WT cells maintained in 2iL or differentiated for 96h upon 2iL withdrawal. Mean of n=2 independent biological replicates is shown. n.d. indicates samples in which expression was undetectable or below 5 CPM. c: Violin plot showing quantification of mean intensity (arbitrary units) for ESRRB in E14 cells cultured in 2iL or differentiated for 48h, 96h or 120h (48, 96 120) or after reinduction with 2iL for 24h (48+24 and 96+24). At least 3 randomly selected fields for each sample have been measured. N=3 independent experiments were analysed. Each violin indicates an independent experiment. d: Left: Representative images of clonal assay followed by Alkaline Phosphatase staining of cells either maintained in 2iL or differentiated for 96h with or without the Gsk3 inhibitor CHIR (96+CHIR). Centre: Bar plot showing quantification of AP positive colonies. Bars indicate mean of 2 biological replicates, shown as dots. Right: Bar plot showing relative

			mRNA expression, measured by qPCR, for *Esrrb*. Bars indicate mean of 2 biological replicates, shown as dots. e: Barplot showing expression by qPCR of Esrrb in E14 cells cultured in 2iL, N2B27, ActivinA (20ng/ml), FGF2 (12.5 ng/ml) and inhibitors of TGF-beta (A83-01, 1 μM) and FGF signalling pathways (PD173074, 0.5 μm) for 48h. Mean +/- SD of 3 independent biological replicates are shown as dots. f: ChIP-PCR analysis of E14 cells cultured in 2iL and differentiated for 24h, 48h, 72h and 96h in N2B27. Immunoprecipitation was performed using anti-ESRRB and anti-H3K27ac antibody followed by PCR with primers located on *Esrrb intron or Tfcp2l1, Utf1* and *Tcf15* promoter regions. Fold-enrichment over a negative region is plotted. Mean +/-SD of n=4 independent experiments, shown as dots.
Extended Data Fig. 4	Regulation of Esrrb expression	Carbognin_ED_Fig4.jpg	a: Genome browser snapshot of histone modifications on regulatory regions on Esrrb gene in naive (2iL, blue) and formative (48h, red) states. Integration of n=2 biological replicates. b: Top: Representative images of immunostaining for H3K27ac (green) and ESRRB (red) in E14 cells cultured in 2iL, differentiated for 84h (84) or after a pulse with 2iLIF for 24h at 84h (2iL pulse), with or without Sodium Butyrate (NaButy or H2O) treatment. Nuclei were identified by DAPI staining (blue). Scale bar: 25μm. Bottom: Barplot showing quantification of mean intensity for H3K27ac (blue) and ESRRB (red) immunostaining normalised to the 2iL H2O samples. Mean +/-SD of n=3 independent experiments, shown as dots.

			c: Plots showing abundance of the indicated histone modifications detected by CUT&RUN and DNA methylation in naive (2iL, left) and formative cells (48, right), on regions bound by ESRRB only in 2iL ('2iL'), only at 48h ('48'), or in 2iL and after 48h ('2iL - 48'), identified in Fig. 4a-b. Integration of n=2 biological replicates. For '2iL' and '2iL - 48' regions we observed enrichment for H3K4me3, H3K27ac, H3K27me3 and H3K9me3 in naive cells. In formative cells, H3K4me3 and H3K9me3 decreased by ~50% while H3K27me3 was lost, while DNA methylation substantially increased. Those regions, where Esrrb binding increases at 48h ('48'), are heavily DNA methylated and pre-decorated by H3K4me3 and H3K27ac in naive cells, while the repressive marks H3K9me3 and H3K27me3 are absent.
Extended Data Fig. 5	*Esrrb* KO clones characterisation	Carbognin_ED_Fig5.jpg	a: Left: Bar plot showing the number of AP positive colonies after clonal assay of cells with loxP sites flanking the second exon of both alleles of *Esrrb* (*Esrrb* fl/fl, dark blue) and *Esrrb* KO cells generated by Cre-mediated recombination (light blue), cultured in 2iL and differentiated for 24h, 48h and 72h in N2B27. Mean +/- SD of n=3 independent experiments, shown as dots. Right: Barplots showing expression measured by qPCR of Esrrb in *Esrrb* fl/fl (dark blue) and *Esrrb* KO (light blue) cells cultured in 2iL and differentiated for 24h, 48h and 72h. Mean of n=2 independent experiments. b: Schematic representation of edited alleles of 3 CRISPR-generated *Esrrb* KO clones. The edited genome is indicated in red. The blue sequence is an insertion. Black bars indicate deletions. c: Bright field images of 3 CRISPR-generated *Esrrb* KO clones, cultured in 2iL and after 48h of 2iL withdrawal. Scale

			bar: 300µm. d: Barplot showing expression by RNAseq of naive markers in WT E14 cells and in 3 CRISPR-generated *Esrrb* KO clones. WT values were set as 1. Mean of n=2 biological replicates.
Extended Data Fig. 6	Proliferation and viability analysis of *Esrrb* KO clones and FS differentiation	Carbognin_ED_Fig6.jpg	a: Left: Proliferation assay over 4 days of WT cells and *Esrrb* KO clones, cultured in 2iL. Mean +/-SD of n=3 independent experiments is shown. Right: Barplot showing percentage of dead cells measured by Propidium Iodide staining in two WT cell lines and 2 *Esrrb* KO clones. Boiled cells (95 degrees Celsius for 5 min) were used as positive control. Mean +/-SD of n=3 independent experiments is shown. P-value calculated with One-way ANOVA followed by Tukey multiple pairwise-comparisons. b: Gene Set Enrichment Analysis (GSEA) of key markers of Apoptosis and cell stress in WT and *Esrrb* KO cells cultured in 2iL (naive) and 48h (formative), which failed to detect any significant differences between WT and KO cells. P-values calculated by the GSEA software. c: Expression measured by qPCR of selected naive and formative genes in WT E14 cells (grey) and three *Esrrb* KO clones (blue) cultured in 2iL and after 24h, 48h and 72h of differentiation in N2B27. Mean of n=2 biological replicates is shown. d: Expression measured by qPCR of naive and lineage markers in Conditional *Esrrb cells* kept in 2iL+DOX. WT cells and *Esrrb* KO expressing a DOX-inducible empty vector (iEmpty) kept in 2iL+DOX are used as controls. Mean +/- SD n=3 independent experiments (dots) is shown. e: Gene expression of formative genes

			measured by qPCR in Conditional *Esrrb* cells cultured in 2iL+DOX (3rd bar) and withdrawn of 2iL and DOX for 48h (4th bar). *Esrrb* KO and WT cells expressing an inducible Empty vector (iEmpty) differentiated for 48h were used as controls (2nd and 6th bars). Bars indicate mean +/-SD of n=5 independent experiments, shown as dots. One-way ANOVA followed by Tukey multiple pairwise-comparisons. f: Left: Experimental strategy used for FS cells generation from ESCs. Right: Representative images of WT cells cultures in AloXR medium for 3 passages (P1, P2 and P3). Scale bar: 25µm. Similar results were obtained in n=3 independent experiments. g: Relative mRNA expression measured by qPCR of naive and formative genes in E14 cells cultured in 2iL or AloXR medium for up to 3 passages. Mean of n=3 technical replicates.
Extended Data Fig. 7	FS cell differentiation of *Esrrb* KO clones	Carbognin_ED_Fig7.jpg	a: Gene Set Enrichment Analysis (GSEA) of key markers of Apoptosis and cell stress in WT and *Esrrb* KO cells cultured in 2iL (naive) and P1/48h (formative) failed to detect any significant differences between WT and KO cells. P-values calculated by the GSEA software. b: PCA of RNA sequencing data of WT and *Esrrb* KO cells during FS differentiation. Genes contributing to Principal Components PC1 and PC3 are indicated. N=3 independent biological replicates, shown as dots, for 2iL samples. N=4 for P1-P3 samples. N=2 for KO3 at P2 and P3. c: Heatmaps showing mean normalised relative mRNA expression measured by qPCR of naive and formative genes in WT and *Esrrb* KO cells cultured in 2iL or AloXR medium for up to 3 passages (P1,

			P2, P3). Mean of n=3 technical replicates. d: Representative images of immunostaining for TFCP2L1 and OTX2 in WT cells (left panels) and for OTX2 in *Esrrb* KO cells (right panels) cultured in 2iL or AloXR medium for up to 3 passages. Nuclei were identified by DAPI staining (blue). Scale bar: 25µm. Similar results were obtained in n=2 independent experiments. e: Left: Representative images of WT and *Esrrb* KO cells cultured in FGF2+ActivinA+XAV for at least 6 passages, to induce EpiSCs differentiation. Scale bar: 25µm. Right: Barplots showing gene expression measured by qPCR of naive (Esrrb and Klf4), general pluripotency (Oct4) and EpiSCs (Fgf5, T) markers in WT and *Esrrb* KO cells cultured in FGF2+ActivinA+XAV for at least 6 passages, to induce EpiSCs differentiation. Embryo-derived EpiSCs (OEC2 and GOF18) and WT E14 ESCs cultured in 2iL are used as controls. Mean +/-SD of N=4 biological replicates, shown as dots.
Extended Data Fig. 8	PGCLC differentiation of *Esrrb* KO clones	Carbognin_ED_Fig8.jpg	a: Normalised frequency of individual gRNAs (indicative of KO) targeting *Esrrb* during induction of PGCLC (CRISPR screening results from[57]). Dots indicate the mean of n=2 independent CRISPR screens. b: Left: Frequency of individual gRNA targeting Esrrb in EpiLC that have acquired correct formative status (Stella-) and EpiLC blocked from formative transition (*Stella*+). Note *Esrrb* gRNA (KO) are enriched in EpiLC that fail to acquire formative status, indicating a functional role for Esrrb in promoting the formative program. Right:

			Normalised frequency of individual gRNAs targeting Olfr568 as a representative negative control gene that should not influence the induction of PGCLC upon KO. Dots indicate the mean of n=2 independent CRISPR screens. c: Immunoblot of clonal lines derived from SGET ESC transiently transfected with Cas9 and gRNAs binding *Esrrb* coding sequence. Out of 5 independent *Esrrb* KO clones, 3 (A1.2, B2.1, A2.5) were randomly chosen for further validations. Beta-TUBULIN was used as loading control. The experiment was repeated 3 times with similar results. *Esrrb* KO clones do not display Esrrb protein expression, but a shorter mRNA can still be detected (Fig. 7d). d: Schematic representation of SGET activation during in vitro cell fate transitions of ESC (Stella+/Esg1+) into EpiLC (Esg1+) and early and late PGCLCs (Stella+) (adapted from Hackett et al., 2018). e: Total number of cells in SGET WT and *Esrrb* KO clones obtained after 3 days of PGCLC induction from EpiLC differentiation. Mean +/-SD of n=3 independent experiments (dots) is shown. f: Gene expression of selected genes in WT (grey) and n=3 independent *Esrrb* KO SGET lines (blue) at EpiLC, d3 and d5 PGCLC stages. g: Expression of the PGC-early (left) and PGC-late (right) geneset in EpiLC, d3 and d5 PGCLC from WT and *Esrrb* KO lines. Bars indicate the median, box indicates the 25th and 75th percentiles, whiskers represent median plus/minus the interquartile (25-75%) range multiplied by 2. Two-sided paired Student t-test, n.s. not significant.

			Integration of n=3 biological replicates for each sample. h: Gene expression of BMP direct targets in WT (grey) and n=3 independent *Esrrb* KO SGET lines (blue) at EpiLC, d3 and d5 PGCLC stages.
Extended Data Fig. 9	Differentiation of *Esrrb* KO clones in 2D and 3D	Carbognin_ED_Fig9.jpg	a: Heatmap showing Z-scored, mean-scaled, normalised gene expression, measured by RNA-seq, of master regulator genes for each of the three primary germ layers and trophoblast in WT cells and three Esrrb KO clones cultured in 2iL and after 24h, 48h and 72h of differentiation in N2B27. Integration of n=2 biological replicates for each sample. b: Representative images of WT cells cultured in N2B27 medium in matrigel for 48h, 72h or 96h, to allow 3D organisation and lumenogenesis. F-actin was labelled by Phalloidin staining (green) and immunostaining for the apical protein PODXL was performed (red). Scale bar: 30 μm. Similar results were obtained in n=5 independent experiments. c: Top: Barplot showing quantification of number of structured/field in WT and *Esrrb* KO cells cultured in N2B27 medium in matrigel for 48h, 72h or 96h. Bars indicate mean of 2 independent experiments, shown as dots. Centre: Violin plot showing quantification of Area (expressed in pixels) of >14 structures in WT and *Esrrb* KO cells. P-values calculated by two-way repeated measures ANOVA. Similar results were obtained in 3 independent experiments. Bottom: Violin plot showing quantification of the ratio of the 2 main diameters (roundness) of >17 structures in WT and *Esrrb* KO cells, as shown in the WT panel. P-values indicate two-sided unpaired t-test. Similar results

			were obtained in 3 independent experiments. Box plots show 1st, 2nd and 3rd quartile, whiskers represent median plus/minus the interquartile (25-75%) range multiplied by 1.5. d: Line plots showing quantification of F-ACTIN intensity along the diameter of 3D structures obtained by culturing WT and *Esrrb* KO cells in N2B27 in matrigel for 48h, 72h or 96h. At least 8 structures were quantified from n=2 independent experiments. The shades indicate the SD. e: Violin plots showing quantification of OTX2 intensity in 3D structures obtained from WT and *Esrrb* KO cells cultured in N2B27 in matrigel for 48h. N>380 nuclei for each sample. Two independent experiments are shown (left and right). Box plots show 1st, 2nd and 3rd quartile, whiskers represent median plus/minus the interquartile (25-75%) range multiplied by 1.5.
Extended Data Fig. 10	Network analysis of formative gene regulation by Esrrb	Carbognin_ED_Fig10.jpg	a: Barplot showing expression of Otx2 measured by qPCR in ES cells treated for 48h with ActivinA (20ng/ml), FGF2 (12.5 ng/ml) and inhibitors of TGF-beta (A83-01, 1 μM) and FGF signalling pathways (PD173074, 0.5 μm). Cells cultured in 2iL or N2B27 for 48h were used as controls. Mean +/-SD of n=3 independent biological replicates (dots) are shown. b: Genome browser snapshot of histone modifications at Otx2 enhancer (E) bound by Esrrb and promoter (P), in naive and formative cells. Profiles are the integration of n=2 biological replicates. c: ABN derived from a Pearson correlation threshold of 0.56 (see Methods). Solid black lines indicate required and definite interaction, dashed

			lines indicate optional interactions, red lines indicate disallowed interactions. Positive regulations are indicated by a black arrow, negative regulations are indicated by a black circle-headed line. d: Summary of 4 experimental constraints, each with initial (left column) and final (right column) conditions. Gene expression is discretized as high (blue) or low (white).

11

12 ## 2. Supplementary Information:

13 ### A. PDF Files

Item	Present?	Filename Whole original file name including extension. i.e.: Smith_SI.pdf. The extension must be .pdf	A brief, numerical description of file contents. i.e.: *Supplementary Figures 1-4, Supplementary Discussion, and Supplementary Tables 1-4.*
Supplementary Information	Yes	Carbognin_SI.pdf	Supplementary Figures 1-3
Reporting Summary	Yes	RS-Martello.pdf	
Peer Review Information	Yes	*TPRFile_Martello.pdf*	

14

15 ### B. Additional Supplementary Files

Type	Number Each type of file (Table, Video, etc.) should be numbered from 1 onwards. Multiple files of the same type should be listed in sequence, i.e.: Supplementary Video 1, Supplementary Video 2, etc.	Filename Whole original file name including extension. i.e.: *Smith_ Supplementary_Video_1.mov*	Legend or Descriptive Caption Describe the contents of the file
Supplementary Table	1	Carbognin_Supplementary_Tables_1_2_3_4_5_6.xlsx	Supplementary Table 1: 6 genes signatures identified in this study.

			Supplementary Table 2: Biological processes enriched in the 6 signatures identified. Supplementary Table 3: Biological processes enriched in genes bound by Esrrb. Supplementary Table 4: Primer List. Supplementary Table 5: gRNAs sequences. Supplementary Table 6: Antibody List.

16

17 **3. Source Data**

Parent Figure or Table	Filename Whole original file name including extension. i.e.: *Smith_SourceData_Fig1.xls,* or *Smith_ Unmodified_Gels_Fig1.pdf*	Data description i.e.: Unprocessed western Blots and/or gels, Statistical Source Data, etc.
Fig. 1	Carbognin_SourceData_Fig1.xlsx	Statistical Source Data
Fig. 2	Carbognin_SourceData_Fig2.xlsx	Statistical Source Data
Fig. 3	Carbognin_SourceData_Fig3.xlsx	Statistical Source Data
Fig. 5	Carbognin_SourceData_Fig5.xlsx	Statistical Source Data
Fig. 5	Carbognin_SourceData_Fig5.pdf	Unprocessed western Blots
Fig. 6	Carbognin_SourceData_Fig6.xlsx	Statistical Source Data
Fig. 7	Carbognin_SourceData_Fig7.xlsx	Statistical Source Data
Fig. 8	Carbognin_SourceData_Fig8.xlsx	Statistical Source Data

Extended Data Fig. 1	Carbognin_SourceData_ED_Fig1.xlsx	Statistical Source Data
Extended Data Fig. 2	Carbognin_SourceData_ED_Fig2.xlsx	Statistical Source Data
Extended Data Fig. 3	Carbognin_SourceData_ED_Fig3.xlsx	Statistical Source Data
Extended Data Fig. 4	Carbognin_SourceData_ED_Fig4.xlsx	Statistical Source Data
Extended Data Fig. 5	Carbognin_SourceData_ED_Fig5.xlsx	Statistical Source Data
Extended Data Fig. 6	Carbognin_SourceData_ED_Fig6.xlsx	Statistical Source Data
Extended Data Fig. 7	Carbognin_SourceData_ED_Fig7.xlsx	Statistical Source Data
Extended Data Fig. 8	Carbognin_SourceData_ED_Fig8.xlsx	Statistical Source Data
Extended Data Fig. 8	Carbognin_SourceData_ED_Fig8.pdf	Unprocessed western Blots
Extended Data Fig. 9	Carbognin_SourceData_ED_Fig9.xlsx	Statistical Source Data
Extended Data Fig. 10	Carbognin_SourceData_ED_Fig10.xlsx	Statistical Source Data

Esrrb guides naive pluripotent cells through the formative transcriptional program

Author list

Elena Carbognin[1,12], Valentina Carlini[2,3,12], Francesco Panariello[4,12], Martina Chieregato[5], Elena Guerzoni[5], Davide Benvegnù[5], Valentina Perrera[1], Cristina Malucelli[1], Marcella Cesana[4,6], Antonio Grimaldi[4], Margherita Mutarelli[4,7], Annamaria Carissimo[4,8], Eitan Tannenbaum[9], Hillel Kugler[9], Jamie A. Hackett[2,13], Davide Cacchiarelli[4,10,11,13] and Graziano Martello[5,13]

Affiliations

1: Department of Molecular Medicine, Medical School, University of Padua, Padua, Italy.

2: Epigenetics & Neurobiology Unit, European Molecular Biology Laboratory (EMBL)-Rome, Adriano Buzzati-Traverso Campus, Rome, Italy.

3: Collaboration for joint PhD degree between EMBL and Heidelberg University, Faculty of Biosciences, Heidelberg, Germany

4: Telethon Institute of Genetics and Medicine (TIGEM), Armenise/Harvard Laboratory of Integrative Genomics, Pozzuoli, Italy

5: Department of Biology, University of Padua, Padua, Italy.,

6: Department of Advanced Biomedical Sciences, University of Naples “Federico II”, Naples, Italy

7: Istituto di Scienze Applicate e Sistemi Intelligenti "Eduardo Caianiello", Consiglio Nazionale delle Ricerche, Pozzuoli, Italy

8: Istituto per le Applicazioni del Calcolo "Mauro Picone," Consiglio Nazionale delle Ricerche, Naples, Italy

9: The Faculty of Engineering, Bar-Ilan University, Ramat Gan, Israel

10: Department of Translational Medicine, University of Naples “Federico II”, Naples, Italy

11: School for Advanced Studies, Genomics and Experimental Medicine Program, University of Naples “Federico II”, Naples, Italy

12: These authors contributed equally

13: These authors jointly supervised this work

Correspondence to: jamie.hackett@embl.it (J.A.H.), d.cacchiarelli@tigem.it (D.C.), graziano.martello@unipd.it (G.M.)

Abstract

During embryonic development, naive pluripotent epiblast cells transit to a formative state. The formative epiblast cells form a polarised epithelium, exhibit distinct transcriptional and epigenetic profiles and acquire competence to differentiate into all somatic and germline lineages. However, we have limited understanding of how the transition to a formative state is molecularly controlled. Here we used murine ESC models to show that ESRRB is both required and sufficient to activate formative genes. Genetic inactivation of *Esrrb* leads to illegitimate expression of mesendoderm and extraembryonic markers, impaired formative expression and failure to self-organise in 3D. Functionally, this results in impaired ability to generate Formative Stem cells and primordial germ cells in the absence of *Esrrb*. Computational modelling and genomic analyses revealed that ESRRB occupies key formative genes in naive cells and throughout the formative state. In so doing, ESRRB kickstarts the formative transition, leading to timely and unbiased capacity for multi-lineage differentiation.

Main text

In mouse embryos, naive pluripotent cells exist from embryonic day (E) 3.5 to E4.75[1–3]. Upon implantation (E5.0 – E6.5) epiblast cells undergo a maturation phase, named “formative” pluripotency”, characterised by epithelial polarisation, lumenogenesis[4] and transcriptional changes in preparation for differentiation[2,3,5–7]. Cells in the formative phase downregulate naive pluripotency markers, upregulate epigenetic modifiers and become competent for both unbiased germ layer formation and primordial germ cell (PGC) specification (E6.5 - E7.5) [5,7–10]. What factors guide pluripotent cells through such a series of molecular changes?

Embryonic development entails transitions through pluripotent states, which can be captured in vitro. Mouse embryonic stem cell (ESCs) are obtained from naive cells of the preimplantation embryo, with whom they share similar transcriptional and epigenetic profiles, and metabolic activity[1,11–13]. Naive ESCs were originally cultured in the presence of the cytokine LIF together with Foetal Bovine Serum (Serum+LIF)[14,15]. More recently, a chemically defined culture condition, based on 2 inhibitors (2i, PD0325901 (PD)=Mek inhibitor and CHIR99021 (CHIR) =GSK3 inhibitor) allowed the expansion of a homogeneous population of naive ESCs[16]. Adding LIF to 2i (2iL) makes the culture more robust and delays the process of differentiation[17,18].
Formative Stem (FS) cells have been obtained from embryos or from ESCs using AloXR or other media[5,7,19]. They display inactivation of naive factors and induction of formative genes, including Otx2 that is crucial for their self-renewal.

We analysed gene expression changes during the irreversible exit from the naive state. After extinction of the naive pluripotency program a second program, compatible with a formative state, is transiently activated and cells irreversibly commit to differentiate. We found that Esrrb is both sufficient and required to induce formative genes. In the absence of Esrrb, differentiation is

91 skewed towards mesendoderm and trophectoderm, PGCs formation is severely impaired and FS
92 cells cannot be obtained, indicating a critical role of ESRRB in chaperoning pluripotent cells
93 through the initial phases of differentiation.

Transcriptional changes during exit from naive pluripotency

Upon withdrawal of 2iL, ESCs enter a reversible phase in which they self-renew if 2iL is reapplied. If 2iL withdrawal is further prolonged, all ESCs will commit to differentiation and lose self-renewal capacity[10,17,20] (Fig. 1a). We measured the kinetics of ESC commitment starting from 2i or 2iL. After signals withdrawal, cells were left in N2B27 medium, which induces neural differentiation[21]. Every 12 hours we quantified the percentage of self-renewing cells, forming undifferentiated colonies in a clonal assay. Cells differentiating from 2iL maintain full self-renewal capacity for up to 48 hours (h), despite the changes in morphology (Fig. 1b) and the strong reduction in the naive pluripotency markers[22–26] (Extended Data Fig. 1a). After 84h all cells committed and lost self-renewal capacity.

We performed a transcriptome analysis followed by Principal Component Analysis (PCA). Progression along Principal Component (PC) 1 indicated a reduction of naive genes *Tfcp2l1* and *Nanog* and induction of neuroectoderm genes *Sox1 and Sox11.* Along PC 2 we detected transient activation of the formative markers *Otx2*, *Pou3f1, Sox3, Fgf5, Lef1, Hes6 and Dnmt3a/b* [5,6,27–29] (Fig. 1c).

Unsupervised hierarchical clustering on samples differentiating from 2iL identified 6 groups of genes (‘gene signatures’, Fig. 1d, Extended Data Fig. 1b). Two signatures were highly expressed in 2iL and downregulated more or less rapidly (“naive early” and “naive late” gene signatures). Several formative markers (*Etv4, Etv5, Dnmt3a, Dnmt3b, Lef1, Pou3f1, Otx2, Sox3)* [6,27–30] were transiently upregulated during the reversible phase. Thus, we defined 2 signatures as “formative early” and “formative late”, reaching maximal expression at 24-36h and 60h. Genes lowly

expressed in 2iL, reaching their maximum in committed cells at 84h or 96h, formed the “early committed” and “late committed” signatures (Supplementary Table 1).

Full activation of the formative programs was associated with gradual loss of reversibility, as reported[10,31]. Cells differentiating from 2i show faster kinetics[17]. The 2iL and 2i time series displayed striking similarities in gene expression changes, with 2i cells showing anticipation by ~24h (Fig. 1b-c, Extended Data Fig. 1a-b and Supplementary Table 2).

Analysis of embryo data[32] revealed that the two naive gene signatures were enriched in E4.5 genes, the formative signatures were enriched in E5.5 genes and the committed signatures in E6.5 genes (Fig. 1d), indicating that our *in vitro* assay recapitulates transcriptional changes observed in the embryo.

We then asked whether commitment affected the response to 2iL. After 48h of differentiation cells responded to 2iL by upregulating naive markers, downregulating formative markers and were fully clonogenic (Extended Data Fig. 2a-b). In contrast, after 96h cells failed to form colonies after reinduction, failed to reactivate naive markers and further upregulated committed markers. However, analysis of direct targets of JAK/STAT, WNT and FGF revealed that the responsiveness of key signalling pathways was not significantly changed after commitment (Extended Data Fig. 2c). We conclude that commitment is associated with a change in the interpretation of external signals.

Reversibility is associated with Esrrb expression

Different transcriptional responses to signals could be due to changes in chromatin accessibility, thus we performed Assay for Transposase-Accessible Chromatin followed by sequencing (ATAC-seq). We identified regions accessible only in 2iL (*Tfcp2l1* locus), only at 48h, maintained

throughout the reversible phase (*Lef1* locus) or accessible only after commitment (Fig. 2a and Extended Data Fig. 2d). Transcription factors motif analysis (Fig. 2b) revealed an enrichment for SOXs, TCF3, ESRRB and KLFs in 2iL; in the reversible phase (2iL - 48h) we found motifs of ESRRB, KLFs and NFYA/B together with SMAD3 and ETS, downstream mediators of the TGF-beta and FGF pathways, which promotes formative transition[5,6,19,31]. Peaks found only at 48h were enriched for the formative transcription factors ZIC3 and OTX2, for ASCL2 and for ESRRB and MYB only on promoters. After commitment we observed enrichment for TEADs, SOXs, GLIS3, JUN/FOS, ZIC3 and OTX2. Thus, chromatin accessibility is dynamically regulated during commitment, with the persistence of accessible regions during the reversible phase, under the control of SMAD3, ETSs, KLFs and ESRRB.

Culture of ESCs under Serum+LIF conditions, in the absence of feeders, generates a heterogeneous population of pluripotent and partially differentiated cells[3,10,18,33–36]. Consistently, naive gene signatures were expressed at lower levels in Serum+LIF than in 2iL, while formative and committed genes were more abundant in Serum+LIF (Extended Data Fig. 2e).

We took advantage of such heterogeneity and analysed the genome-wide binding profiles of several pluripotency regulators and chromatin modifiers previously generated in Serum+LIF[37]. We calculated the relative enrichment of factors at the promoters of genes belonging to the 6 gene signatures, seeking potential regulatory mechanisms (Fig. 2c). The core pluripotency factors POU5F1, known as OCT4, and SOX2 were found significantly enriched at all signatures. Polycomb Repressive Comples 2 (PRC2) components were significantly bound to committed genes. Interestingly, both naive and formative signatures were significantly bound by naive pluripotency factors (KLF4, TFCP2L1, and ESRRB).

ESCs in the reversible phase express formative genes, which may be regulated by pluripotency factors. We asked whether the forced expression of pluripotency factors enriched at naive and

formative signatures could extend the reversible phase during ESC differentiation. We generated a pool of cells stably expressing individual factors, differentiated them for 96h and found that they were still able to form naive colonies (Extended Data Fig. 2f and Fig. 2d). We extracted genomic DNA from the pluripotent colonies, compared the frequency of genomic integration of each factor, and observed a strong enrichment for Esrrb integration. We generated lines stably expressing single factors, differentiated them individually and confirmed that Esrrb expression led to robust colony formation (Fig. 2e and Extended Data Fig. 2g).

Esrrb has been shown to efficiently reset primed Epiblast Stem cells (EpiSCs) to naive pluripotency[38,39]. Committed cells at 96h might be EpiSCs and Esrrb could be in fact resetting them. However, we failed to detect EpiSC-specific markers in committed cells (Extended Data Fig. 3a-b). We conclude that Esrrb forced expression prolongs the reversible phase during ESC differentiation.

We observed that *Esrrb* mRNA expression rapidly decreases upon 2iL withdrawal, with a ~90% reduction after 48h. ESRRB protein is still present in most cells after 48h of differentiation from 2iL, becoming barely detectable at 96h (Fig. 2f and Extended Fig. 3c). Moreover, we observed full reactivation after 2iL reinduction at 48h (48+24) and no response at 96h (96+24). These results indicate that cells in the reversible phase express ESRRB protein and are capable to reinduce *Esrrb* mRNA, while cells committed to differentiate have permanently lost *Esrrb* expression, as also reported under Serum+LIF conditions[40].

To dissect the mechanism controlling *Esrrb* expression during differentiation we analysed the role of both signalling pathways and epigenetics. *Esrrb* is a direct target of the repressor TCF7L1[24,41]. Treatment with CHIR, which causes derepression of TCF7L1 targets[42,43], elevated the expression of *Esrrb* for 96h, accompanied by extended reversibility (Extended Data Fig. 3d-e). FGF had a minor negative effect on *Esrrb* expression. We then analysed the epigenetic profile of the *Esrrb*

194 locus. In 2iL and at 48h ESRRB strongly binds to his intronic enhancer[44] (Extended Data Fig. 3f),
195 compatible with reactivation of *Esrrb* mRNA expression upon 2i treatment. In 2iL this enhancer is
196 enriched for H3K27ac and H3K4me3. Both activating marks are then lost and H3K27me3 is
197 gained, leading to a complete silencing (Extended data Fig. 4a-b). We treated fully committed
198 cells with Sodium Butyrate (NaButy), a histone deacetylase inhibitor, and with 2i. In committed
199 cells H3K27ac and ESRRB protein levels were undetectable (Extended Data Fig. 4b). NaButy
200 treatment maintained H3K27ac levels. Upon 2iL pulse, ESRRB protein was strongly detected.
201 We conclude that Esrrb expression is positively controlled by H3K27ac and 2i.
202

203 # Esrrb promotes the expression of formative genes

204 Next, we generated cells expressing an Esrrb-IRES-Venus (EIV) transgene under a Doxycycline
205 (DOX)-inducible promoter, to induce Esrrb expression *specifically* during differentiation and to
206 isolate by FACS pure populations of Esrrb-IRES-Venus positive cells upon DOX induction or pure
207 Esrrb-IRES-Venus negative cells in the absence of DOX. Upon replating in 2iL, EIV$^+$ cells robustly
208 formed naive colonies, confirming that *Esrrb* expression is sufficient to confer reversibility (Fig.
209 3a-b).
210 We performed transcriptional analysis and confirmed *Esrrb* induction in EIV$^+$ cells, while a panel
211 of naive markers were only mildly expressed. Commitment genes were strongly repressed by
212 *Esrrb* expression. Surprisingly, several genes of the formative signatures (n=57) were highly
213 induced (Fig. 3c-e). We conclude that *Esrrb* expression during differentiation results in activation
214 of both naive and formative genes.
215
216 Our results suggest a role of ESRRB as a positive activator of the formative gene program, beside
217 its role as a naive factor[22–24,45]. To further investigate this, we performed Chromatin
218 Immunoprecipation (ChIP)-sequencing for ESRRB. The large majority of ESRRB binding occurs

in 2iL and at 48h (Fig. 4a-b). Several formative genes promoters are bound by ESRRB in 2iL and at 48h (2iL – 48), in line with its capacity to induce their expression. ESRRB peaks on naive genes (e.g. *Tfcp2l1*) decrease after 48h, supporting its role in sustaining the naive transcriptional network (Fig. 4b-c and Extended Data Fig. 3f). On the contrary, peaks on formative genes are present in 2iL and the signal is maintained (or increased) at 48h. These observations endorse the concept of Esrrb as a direct activator of the formative gene program during the reversible phase. A search for biological processes enriched in genes bound by ESRRB identified Oxidative phosphorylation, Krebs cycle and glycolysis (Supplementary Table 3), as reported[46].

How is ESRRB binding dynamically regulated between the naive and formative states? Previous studies showed that ESRRB-bound regions are decorated by specific epigenetic marks[38,47]. Indeed, epigenetic profiling revealed differences in the levels of H3K27me3, H3K9me3 and DNA methylation in ESRRB-bound regions in naive or formative states (Extended Data Fig. 4c).

ESRRB activates both naive and formative programs

Next, we asked whether *Esrrb* inactivation would shorten the reversible phase leading to more rapid commitment. Transient knockdown of *Esrrb* (Fig. 5a) led to inability to form colonies after 48h and anticipated reduction of a naive marker. Similar results were obtained using *Esrrb* KO ESCs previously generated by gene-targeting[24] (Extended Data Fig. 5a). We generated 3 *Esrrb* KO clonal lines by CRISPR/Cas9 system in 2iL (Fig. 5b, Extended Data Fig. 5b), to exclude the possibility that multiple rounds of gene-targeting and long-term culture in Serum+LIF could have induced cell adaptation or selection. *Esrrb* KO clones showed no gRNAs off-target mutations (Supplementary Fig. 1) and no morphological differences (Extended Data Fig. 5c), they displayed long-term self-renewal in 2iL, although with a partial downregulation of some pluripotency markers (Extended Data Fig. 5d), as reported[24,48]. *Esrrb* KO lines showed reduced self-renewal capacity relative to parental cells, consistently with the anticipated commitment of *Esrrb* KO cells (Fig. 5b,

bottom). The reduction in colony number could be due to viability impairment, as reported upon acute *Esrrb* deletion[38,41,47]. The proliferation rate and viability of Esrrb KO clones did not differ from those of wild-type (WT) cells (Extended Data Fig. 6a). We obtained transcriptomes from *Esrrb* WT and KO clones. Analysis of Apoptosis and Cell Stress signatures revealed no differences (Extended Data Fig. 6b), further ruling out a viability impairment of *Esrrb* KO cells.

PCA indicated anticipated progression of *Esrrb* KO cells (Fig. 5c). Naive signatures were reduced (84 genes) and committed genes were upregulated in *Esrrb* KOs (278 genes), in line with the anticipated loss of self-renewal observed in clonal assays. Interestingly, formative genes were globally downregulated (148 genes) in multiple *Esrrb* KO clones (Fig. 5d-e and Extended Data Fig. 6c), confirming that Esrrb regulates both the naive and formative programs.

ESRRB is required for generation of FS cells

To uncouple the role of ESRRB in the maintenance of naive pluripotency from its role as activator of the formative program we generated *Esrrb* KO cells expressing a DOX-inducible *Esrrb* transgene ('Conditional Esrrb cells', Fig. 6a).

We gave a pulse of DOX between 24 and 48h, at the time of activation of the formative program and observed induction of formative genes (*Tcf15, Dnmt3a, Dnmt3b* and *Utf1,* Fig. 6b).

In a complementary strategy, we expanded Conditional Esrrb cells in 2iL+DOX. Expression of naive markers was comparable to the one observed in WT cells, with no spontaneous expression of lineage markers (Extended Data Fig. 6d). We withdrew 2iL and DOX and ESRRB protein was undetectable after 48h (Fig. 6c,d). We therefore asked whether acute loss of ESRRB would affect formative gene expression. Global transcriptome profiling revealed impaired induction of 128 formative genes, including *Otx2*, *Utf1*, *Dnmt3a* and *Dnmt3b*, compared to WT cells (Fig. 6e-f and Extended Data Fig. 6e). From these two experiments, whereby ESRRB is either specifically

added or acutely removed during formative gene activation, we conclude that ESRRB is an inducer of formative genes, independently from its role as a naive factor.

FS cells have been obtained from embryos or from ESCs using AloXR medium[19]. If Esrrb is a critical inducer of formative genes, it should not be possible to obtain *Esrrb* KO FS cells. After 3 passages in AloXR medium, WT cells rapidly downregulated naive genes and upregulated formative markers (Extended Data Fig. 6f-g). WT cells could be easily stabilised in AloXR for >9 passages. In contrast, *Esrrb* KO clones displayed an aberrant morphology and repeatedly collapsed soon after passage 3 (Fig. 6g).

We performed transcriptome analysis during FS conversion. Gene signatures of Apoptosis and Cell Stress revealed no differences between *Esrrb* WT and KO (Extended Data Fig. 7a), thus ruling out a general viability impairment of *Esrrb* KO cells. We looked for those early expression alterations distinguishing WT from *Esrrb* KO cells. The formative genes *Otx2*, *Dnmt3a* and *Dnmt3b* peaked after 48h (P1) and were then maintained at high levels in WT FS cells, but not in *Esrrb* KO cells (Fig. 6h and Extended data Fig. 7b-c). Lineage markers Nes, Hand1, Foxa2 and Cdkn1c were significantly upregulated in Esrrb KO cells from P1. We conclude that *Esrrb* genetic inactivation leads to inability to generate FS cells and impaired induction of formative genes.

Esrrb was detected only during the first 48h of FS cell generation (2iL and P1), and not in stably expanding WT FS cells, indicating that Esrrb plays an early role during establishment of FS cells and not for their maintenance. FS cells rely on the transcription factor Otx2 for their self-renewal[19]. In the absence of ESRBB most cells were devoid of OTX2 by passage 2 (Extended Data Fig. 7d), in agreement with their collapse. We conclude that during FS cell establishment, ESRRB is required for robust induction of formative genes.

Primed EpiSCs are an in vitro model of the peri-gastrulation epiblast [49–51] and are obtained from either post-implantation embryos or ESCs[39,52–54]. Otx2 is required for FS cell identity, but

dispensable in EpiSCs[19,30]. We asked whether EpiSCs could be obtained from *Esrrb* KO ESCs. We applied AFX conditions[55] and both *Esrrb* WT and KO cells robustly expanded for >10 passages. *Esrrb* and *Klf4* were downregulated, *Oct4* was maintained and EpiSC-specific markers *Fgf5* and *T* were partially induced (Extended Data Fig. 7e). We conclude that ESRRB is dispensable for EpiSCs generation.

PGCs specification is impaired by loss of ESRRB

Competence for PGC specification is acquired in the formative state[5–7,10]. To become responsive to inductive signals for germ cell specification, ESCs must extinguish their naive identity and transit to epiblast-like cells(EpiLCs). EpiLCs are molecularly similar to the formative epiblast [6,56]. ESRRB is crucial for the induction of the formative program, thus PGC specification should be impaired in the absence of Esrrb. Indeed, a CRISPR-Cas9 genome-wide screen showed that gRNAs targeting *Esrrb* were underrepresented both in formative EpiLCs and in PCGLCs[57] (Extended Data Fig. 8a-b).

Thus, we sought to confirm the role of ESRRB as a regulator of PGC specification by generating novel clonal *Esrrb KO* lines in the *Stella*-GFP:*Esg1*-tdTomato (SGET)[57] reporter ESC line (Extended Data Fig. 8c-d), which did not show any mutations in predicted off-targets (Supplementary Fig. 2).

Esrrb KO lines exhibited reduced capacity to specify *Stella*+ PGCLC at both day 3 and day 5, as compared to WT cells (Fig. 7a). This reduction was not due to impaired proliferation or survival of *Esrrb* KO cells, given that in our experiments we replated equal numbers or Esg1+ EpiLCs and observed no differences in the number of cells after PGCLC induction (Extended Data Fig. 8e).

We isolate those few *Stella*+ PGCLC that were derived from *Esrrb* KO lines by flow cytometry and analysed them by RNAseq. WT and KO cells follow a different trajectory during PGCs specification suggesting that the developmental programs are not appropriately activated in *Esrrb*

KO cells (Fig. 7b). Formative genes (*Fgf5, Dnmt3a, Etv5, Utf1, Otx2*) were significantly reduced in *Esrrb* KO EpiLCs (Fig. 7c-d and Extended Data Fig. 8f). Moreover, we observed significant downregulation of both early and late PGCs genesets in *Esrrb* KO PGCLC (Extended Data Fig. 8g), corroborating the idea that the full germline program was not appropriately activated. Several PGC markers, such as *Dazl, Kit, Nanog* and *Tfcp2l1* failed to activate fully in *Esrrb* KO cells, although we observed expected activation of the core PGC markers *Prdm14, Blimp1* and *Tfap2c/Ap2y*. We conclude that the absence of Esrrb leads to a loss of robustness in PGCs specification.

Previous studies have indicated that Esrrb promotes PGC specification via BMP4 production from the extraembryonic ectoderm[58] . In our system we provide excess of exogenous BMP4 and BMP8. We measured the expression of 6 BMP direct transcriptional targets[59] during PGCLC induction and found no differences upon *Esrrb* deletion (Extended Data Fig. 8h). We conclude that Esrrb promotes PGC specification both via BMP-independent and BMP-dependent mechanisms.

Further transcriptional analyses revealed that *Esrrb* KO cells robustly expressed somatic-mesodermal markers (*Pitx2, Pbx1, Lefty1, FoxP1*) at both EpiLC and PGCLC stages (Fig. 7d). Similarly, during neural differentiation in N2B27 all *Esrrb* KO cells display a robust expression of markers of mesoderm, endoderm and trophectoderm (Extended Data Fig. 9a).

Taken together these results endorse the role of ESRRB as a key coordinator of the formative gene program that is preparatory for unbiased germ-layer and germ-cell differentiation. In the absence of ESRRB, PGC specification is impaired and mesoderm, endoderm and trophectoderm markers are aberrantly activated.

Impaired lumenogenesis in *Esrrb* KO 3D structures

The formative epiblast in the embryo is a polarised epithelium with an apical domain facing a lumen, the proamniotic cavity[4,6]. Culture of ESCs in 3D allows for epithelialization and self-organisation similarly to the formative epiblast[4,60,61]. We asked whether the faulty activation of the formative program observed in *Esrrb* KO cells could affect morphogenesis.

WT ESCs plated in a hydrogel of extracellular matrix in N2B27 medium formed 3D structures with an apical domain marked by F-ACTIN accumulation and the formative gene PODXL (Supplementary Table 1), as observed in the peri-implantation formative epiblast[4,60] (Extended Data Fig. 9b, Fig. 8). *Esrrb* KO clones formed fewer structures with a reduced area, but comparable roundness (Extended Data Fig. 9c).

We then analysed the polarisation and lumenogenesis of 3D structures. Most WT structures showed a strong and apically localised PODXL signal at 48h, that after 72h and 96h defined a central cavity (Fig. 8a-b). In *Esrrb* KO, the majority of structures failed to form a lumen. We then quantify F-ACTIN intensity along a diameter. In WT cells we found a prominent central peak at 48h, separating into two peaks by 96h, indicating the formation of apical domains facing a central lumen (Extended Data Fig. 9d). *Esrrb* KO clones failed to do so. Molecularly, *Esrrb* KO failed to fully activate formative genes, including *Otx2* (Fig. 8c-d and Extended Data Fig. 9e), which has been demonstrated to be both sufficient and required for lumen formation in 3D[62]. Our results reveal that *Esrrb* inactivation causes impaired activation of formative genes and lumenogenesis in 3D structures.

Esrrb inactivation results in inefficient activation of the key formative gene *Otx2*, under several different experimental conditions (Fig. 5e, 6f, 6e, 7c-d, 8c-d). How does ESRRB regulate *Otx2*? ESRRB might regulate the FGF and TGF-beta signals, which promote transition towards formative state[6,19,31]. However, *Otx2* levels were unchanged after stimulation or inhibition of both pathways (Extended Data Fig. 10a). *Otx2* expression might be controlled epigenetically. *Otx2* promoter was found bivalent in naive cells (Extended Data Fig. 10b), while an *Otx2* downstream

enhancer was bound by ESRRB. ESRRB binding was consolidated in formative cells (Fig. 4c), with a concomitant increase in H3K27ac (Extended Data Fig. 10b), indicating a potential direct regulation of *Otx2* expression by ESRRB.

To gain a more comprehensive understanding of how ESRRB regulates *Otx2* and the transition from naive to formative state we turned to computational modelling. We extended a gene regulatory network of naive pluripotency[17,39,63] by adding formative genes and we inferred interactions between components from RNAseq and ChIP-seq (Fig. 8e and Extended Data Fig. 10c). We then defined the naive, formative and committed states (Extended Data Fig. 10d) and constrained the model asking whether it could orderly proceed through them. The model started from the naive state (step 0), it gradually activated formative genes while naive genes were inactivated (step 9, Fig. 8f-g). Finally, all formative genes were inactivated (step 15).

The model showed that *Esrrb* KO cells fail to activate some formative genes, including *Otx2*, while naive genes were also inactivated more rapidly (Fig. 8f-g and Extended Data Fig. 10d), in agreement with the faster exit kinetics observed in *Esrrb* KO cells (Fig. 5).

We then focussed on how ESRRB regulates the expression of formative genes. ESRRB engages in positive interactions with formative genes already in the naive state, which are maintained until the formative state is reached (Fig. 8g). Thus, Esrrb is pre-wired to formative genes in the naive state, in line with ATAC-seq and ESRRB ChIP-seq results. In the absence of *Esrrb* these positive interactions do not take place and the activation of formative genes is impaired (Fig. 8g, bottom panels). Concerning *Otx2* regulation, we detected activating interactions from Esrrb and the formative factors Lef1, Utf1 and Dnmt3b. In *Esrrb* KO all those formative genes fail to activate. Thus, *Otx2* expression appears to be controlled by ESRRB, both directly and indirectly.

Discussion

ESC differentiation entails inactivation of naive genes followed by a phase of renovation, named formative pluripotency, in preparation for unbiased germ layer specification [6,10]. We have identified gene signatures of the formative state that are transiently activated during pluripotency progression (Supplementary Table 1). Subsequently, pluripotent cells express lineage-specific transcription factors and segregation of definitive embryonic lineages occurs. Characterisation of transcriptional and epigenetic profiles of early embryos revealed that in the formative state at E5.5 the enhancer landscape of epiblast cells is already set for the specification of neuroectoderm lineage[64]. Not surprisingly, some markers of the formative state (e.g. *Otx2, Pou3f1, Zic3*) are retained in the neuroectoderm lineage.

We also found that activation of formative genes leads to irreversible commitment to differentiate, as reported[10]. Of note, “reversibility’ indicates the capacity of early differentiated cells to revert to naive pluripotency[10,40]. This transition could be considered as reprogramming from formative to naive pluripotency. For instance, after 48h of 2iL withdrawal, cells can fully revert, or reprogram, back to the naive state when exposed to 2iL.

What orchestrates this ordered progression? What activates formative genes? Our unexpected findings are that ESRRB covers this role. ESRRB was first characterised as a pivotal transcription factor for the maintenance of the naive pluripotency network[17,22–24,45], acting downstream of both NANOG and the WNT pathway and its inactivation results in partial loss of naive markers, which is compensated by other naive factors, such as *Nr5a2*[48]. However *in vivo* studies showed that *Esrrb* expression arises in cleavage stage embryos[65,66] and is maintained in pluripotent cells until the peri-implantation stage[62], in line with the persistence of ESRRB for up to 48h of in vitro ESC differentiation (Fig. 2f). *Esrrb* expression pattern is consistent with a dual role both in the maintenance of naive pluripotency and in activation of formative pluripotency. Our ChIP-

sequencing experiments showed that ESRRB binds naive and formative genes in 2iL (Fig. 4). Upon 2iL withdrawal, binding on naive genes decreases while peaks on formative genes are maintained throughout the formative phase.

When overexpressed in ESCs, ESRRB ectopically induces both naive and formative gene expression (Fig. 3c-e). Also during resetting of EpiSCs by *Esrrb* overexpression, we noticed robust induction of formative markers *Otx2*, *Dnmt3b* and *Utf1*[39]. *Esrrb* inactivation leads to impaired induction of formative genes, less robust PGC specification and spontaneous activation of mesendoderm and trophectoderm markers. Furthermore, *Esrrb* KO ESCs failed to form FS cells. Such results clearly endorse the concept of ESRRB as a direct activator of the formative gene program and show that correct activation of the formative program is required for timely and unbiased multilineage differentiation of murine naive pluripotent cells.

We investigated the molecular mechanisms associated with dynamic binding of ESRRB in naive and formative cells. ESRRB-bound loci in naive cells were enriched for H3K4me3, H3K27ac and depleted of DNA methylation (Extended Data Fig. 4c), in agreement with a work from Atlasi and collaborators[47], reporting that 2iL specific enhancers are accessible regions, decorated by H3K27ac and enriched for ESRRB binding. Loci bound by ESRRB only in formative cells are heavily DNA methylated. ESRRB binds DNA-methylated regions, leading to gene activation[38] also during reprogramming. Of note, ESRRB binds key chromatin regions during cell divisions to preserve the transcriptional identity of ESCs[67]. This bookmarking activity might also explain the persistent binding of ESRRB on formative genes.

Embryos deficient for *Esrrb*, or deficient for its upstream regulator *Nanog*, showed reduced numbers of PGCs[25,58,68], which has been imputed to lower production of BMP4 by the extraembryonic ectoderm[58]. However, tetraploid complementation experiments, whereby extraembryonic tissues are provided by wild-type embryos, revealed a reduction in PGCs[68],

indicating that ESRRB must control PGCs number by additional, cell-autonomous mechanisms. We conducted *in vitro* assays in which BMP4 was exogenously provided and saw a reduction in PGCLCs upon *Esrrb* inactivation. Thus, we would propose that the reduction of PGCs in *Esrrb* mutant is due to both reduced BMP signalling and to faulty formative program activation.

Esrrb null embryos display placental defects at E8.5 and die by E10.5. Epiblast-specific deletion of *Esrrb* or tetraploid complementation allowed to rescue those defects, indicating that ESRRB is a critical regulator of the trophectoderm lineage[69–71]. Nonetheless, *Esrrb* null embryos were under-represented in a large tetraploid complementation study[68], suggesting additional Esrrb developmental functions in embryonic cells. ESRRB is detected in the epiblast up to E5.0 and *Esrrb* null embryos show growth defects at E6.0[58]. Diapause is a state of metabolic dormancy whereby epiblast cells are held between naive and formative state, and self-organise into polarised rosette-like structures[41,72]. *Esrrb* null embryos in diapause display a dramatic reduction in the number of epiblast cells, which fail to self-organise[41]. Consistently with these *in vivo* observation, *Esrrb* inactivation in a 3D *in vitro* model led to reduced number and size of structures, and impaired self-organisation.

We propose that ESRRB confers robustness to epiblast cells transitioning from naive to formative pluripotency. The absence of ESRRB causes alterations (e.g. embryo size and number, PGC number) with partial penetrance. However, when the progression of epiblast cells is delayed or blocked, as in diapause or in FS cells, *Esrrb* becomes strictly required for self-renewal of pluripotent cells. *Esrrb* and *Nr5a2* are two orphan nuclear receptors with overlapping functions, both in totipotent and naive pluripotent cells[48,66]. Interestingly, *Nr5a2* is expressed both in the naive and formative epiblast and *Nr5a2* null embryos are under-represented and display severe defects at E6.5[73] indicating a role for Nr5a2 in the control of pluripotency progression. It would be therefore interesting to investigate the effect of combined inactivation of *Esrrb* and *Nr5a2*.

Reprogramming studies further support a role for ESRRB in multiple embryonic lineages, showing that Esrrb localises preferentially near genes expressed in the epiblast, extraembryonic lineages and PGCs[38,74]. ESRRB plays also a role, in combination with NR5A2, in the zygotic genome activation of totipotent blastomeres[66]. These observations indicate a multifaceted function of ESRRB in the control of several early embryonic lineages, raising the question of how ESRRB activity can be interpreted in different contexts, possibly thanks to a combination of cell signalling, epigenetic and metabolic regulation.

Acknowledgements

We thank Austin Smith for critical reading of the manuscript. This work was supported by Fondazione Telethon Core Grant, Armenise-Harvard Foundation Career Development Award, European Research Council (grant agreement 759154, CellKarma), and the Rita-Levi Montalcini program from MIUR to D.C and M.C. We thank TIGEM NGS core, NEGEDIA and Anna Manfredi for genomic library preparation and sequencing run. Work in J.A.H's laboratory is supported by programme grants from the European Molecular Biology Laboratory (EMBL). Work in H.K.'s group is supported by the ISRAEL SCIENCE FOUNDATION (Grant no. 190/19). G.M.'s laboratory is supported by grants from the Giovanni Armenise–Harvard Foundation, the Telethon Foundation (GJC21157), Microsoft Research and an ERC Starting Grant (MetEpiStem).

Author Contributions Statement

G.M. and E.C. conceived the project. E.C., G.M., D.C. and J.A.H. designed the experiments and interpreted the results. E.C. and E.G. performed ESC experiments and analyses. C.M. and D.B.

performed molecular analyses of *Esrrb* KO cells. V.P. performed ChIP-PCR experiments. M.Chieregato performed 3D experiments. V.C. performed PGCLC assays. F.P. and A.C. performed bioinformatic analysis of transcriptomic data. M.Cesana performed ChIP-seq experiments. A.G. performed ATAC-seq experiments. M.M. performed ChIP-seq and ATAC-seq analysis. G.M., H.K and E.T. performed computational modelling. E.C., F.P. and J.A.H. prepared figures. G.M. and E.C. wrote the manuscript with help from all authors. G.M., D.C. and J.A.H. secured fundings and supervised the project.

Competing interests Statement

The authors declare no competing interests.

Figure Legends

Fig. 1: Transcriptional changes associated with irreversible exit from naive pluripotency.

a: Schematic representation of the first stages of exit from the naive state. Upon 2i or 2iL withdrawal, cells transit through a reversible phase before being irreversibly committed to differentiate.

b: Top: Morphology and AP staining images after clonal assay of E14 cells cultured in 2iL and after 2iL withdrawal. Bottom: Barplot showing the relative number of AP positive pluripotent colonies after clonal assay of E14 cells cultured both in 2iL (purple) and 2i (green) and after the withdrawal of either 2iL or 2i every 12h for 96h. Mean +/-SD of n=3 independent experiments. Unpaired two-sided t-test '2iL 0' vs '48' p=0.97, '2iL 0' vs '96' p=0.0096. Scale bars= 30 μm.

c: PCA of RNA sequencing data of cells differentiating from 2iL (purple) and from 2i (green). Genes contributing to the first two Principal Components are indicated. N=2 independent biological replicates for each time point, shown as dots.

d: Top: Line plot showing expression dynamics of differentially expressed genes during differentiation, grouped by hierarchical clustering based on Pearson Correlation. Grey shades represent a 95% bootstrap confidence interval around mean values. Integration of n=2 independent biological replicates for each time point.

Pie charts represent the intersection of the gene signatures with published gene sets of mouse embryo development at E4.5, E5.5 and E6.5. Bottom: Heatmaps show the sum of the log2-scaled normalised expression values of the intersection lists shown in the pie charts, averaged by different time points.

Fig. 2: Differentiation reversibility is associated with Esrrb expression

a: UCSC genome browser visualisation of normalised ATAC-seq profiles at the indicated loci. Rectangles indicate peaks found only in 2iL (green), only at 48h, both in 2iL and 48h (purple). Integration of n=2 independent biological replicates for each time point.

b: Ballon plot summarising the percentage of ATAC peaks containing a given motif, and the associated p-value, at the indicated time points. Peaks on promoters, peaks at 10Kb from TSS and all peaks were analysed and are represented in blue, green and orange respectively. Integration of n=2 independent biological replicates for each time point.

c: Balloon plot summarising published ChIP-seq data of ESCs cultured in Serum+LIF from the Codex compendium[37]. The size of each balloon indicates the fold enrichment, the colour indicates the statistical significance.

d: left: Bar plot showing the number of AP positive colonies after clonal assay performed on cells overexpressing a pool of pluripotency genes and maintained in 2iL (grey bars) or differentiated for 96h (blue bars). Cells overexpressing an empty vector were used as control (empty). Bars indicate mean +/-SD of n=3 independent experiments, shown as dots. Right: Bar plot showing enrichment of genomic integrations of 8 naive genes in cells differentiated for 96h and plated for clonal assay compared to cells in 2iL. Bars indicate mean +/-SD of n=3 independent experiments, shown as dots.

e: Representative images of clonal assay followed by Alkaline Phosphatase staining of cells overexpressing an empty vector or pluripotency factors, either maintained in 2iL or differentiated for 96h. N=2 independent experiments, quantified in Extended Data Fig. 2g.

f: Immunostaining for ESRRB (green) in E14 cells cultured in 2iL or differentiated for 48, 96 or 120h (48, 96 120) or after reinduction with 2iL for 24h (48+24 and 96+24). Nuclei were identified by DAPI staining (blue). Scale bar: 25μm. N=3 independent experiments, quantified in Extended Data Fig. 3c.

Fig. 3: Esrrb promotes the expression of formative genes

a: Schematic representation of E14 cells transfected with an inducible Esrrb-Ires-Venus vector (EIV) and differentiated for 96h with or without Doxycycline (DOX) treatment. Cells were sorted for presence or absence of Venus expression (EIV+ and EIV- respectively) and further characterised.

b: Left: Representative images of Alkaline phosphatase staining after clonal assay of cells expressing EIV or an inducible Empty vector (iEmpty), cultured in 2iL or without 2iL for 96h (96), in the presence of DOX. Right: Barplot showing number of AP positive colonies in cells expressing EIV or iEmpty, cultured in 2iL or without 2iL for 96h, in the presence or absence of DOX (+D or -D). Mean +/-SD of 4 biological replicates. Two-sided unpaired Student t-test.

c: Heatmaps showing mean-scaled normalised expression levels, measured by RNA-seq, of selected naive, formative and committed genes in E14 cells expressing EIV cultured in 2iL or differentiated for 96h in the presence or absence of DOX (96-DOX or 96+DOX respectively). Integration of n=2 independent experiments.

d: Expression levels of selected naive and formative genes measured by qPCR in cells treated as described in Fig. 3b. Mean of n=2 independent experiments. Expression of naive genes is normalised to iEmpty cells kept in 2iL -D. Formative genes are normalised to E14 cells differentiated for 48h.

e: Scatter plot showing transcriptome analysis of E14 cells expressing EIV cells differentiated for 96h in N2B27 with or without DOX. Down-regulated (Log2FC < -1 and p-value < 0.01) and Up-regulated (Log2FC > 1 and p-value < 0.01) genes are plotted on the left or right part of the panel respectively. The Y-axis indicates the mean expression on a log scale. Genes belonging to the 6 genes signatures described are represented by coloured dots. Selected genes are highlighted. Integration of n=2 independent experiments.

Fig. 4: Esrrb promotes the expression of formative genes

a,b: ChIP-seq analysis of E14 cells cultured in 2iL and differentiated for 48h and 96h in N2B27. Time points are colour-coded in blue (2iL), cyan (48h) and yellow (96h). N=1 biological replicate.

a: Venn diagram showing the intersection of significant ESSRB peaks for each given time point obtained by ChIP-seq analysis of E14 cells cultured in 2iL and differentiated for 48h and 96h in N2B27.

b: Binding heatmaps displaying the read coverage density of ESSRB peaks along with average intensity. Peaks are grouped by the presence in one or multiple time points. For example, the “2iL - 48” group contains peaks found both in 2iL and after 48h of differentiation.

c: Representative genome browser snapshots of selected gene loci bound by ESSRB in each given time point. Both reads distributions as line plots and peak intervals are displayed. N=1 biological replicate.

Fig. 5: Esrrb coordinates the activation of naive and formative programs.

a: Left: Bar plot showing number of AP positive colonies after clonal assay of E14 cells transfected with a non-targeting control siRNA (siCo, dark grey), siGFP (light grey) or si*Esrrb* (blue) and cultured in 2iL or differentiated for 24h and 48h. Centre and Right: Expression analysis by qPCR of *Esrrb* and *Tfcp2l1* genes in E14 cells transfected with siCo, siGFP and si*Esrrb* and differentiated for 24h or 48h. Bars indicate mean +/-SD of n=4 independent experiments, shown as dots. The 24h sample was analysed in n=3 experiments.

b: Top: Immunoblot of clonal lines derived from E14 cell population stably expressing Cas9 and transfected with 2 gRNAs flanking *Esrrb* DNA-binding region. Three *Esrrb* KO clones were chosen (KO1, KO2, KO3). GAPDH was used as a loading control. Bottom: Barplot showing number of AP positive colonies after clonal assay of E14 cells (WT) and 3 *Esrrb* KO clonal lines cultured in 2iL and after 24h, 48h and 72h of differentiation. Mean +/-SD from n=3 independent experiments, shown as dots.

c: PCA of RNA sequencing data obtained from E14 (WT) and 3 *Esrrb* KO clones cultured in 2iL and after 24h, 48h and 72h of differentiation. N=2 biological replicates for each data point, shown as dots.

d: Transcriptome analysis of *Esrrb* KO cells cultured in 2iL and after 24h, 48h and 72h of differentiation in N2B27. Down-regulated and Up-regulated genes (P-value <0.05, FC >1 or <-1, compared to WT) are plotted on the left or right part of each panel. Genes belonging to the 6 genes signatures are highlighted with coloured dots. Integration of n=2 biological replicates for each cell line. Mean of the 3 independent *Esrrb* KO clones.

e: Heatmaps showing mean-scaled normalised expression levels, measured by RNA-seq, of selected naive, formative and committed genes in WT and *Esrrb* KO clones. Integration of n=2 biological replicates for each cell line. Mean of the 3 independent *Esrrb* KO clones.

Fig. 6: Esrrb is required for generation of FS cells

a: Experimental strategy used to induce Esrrb specifically at the time of activation of the formative program.

b: Relative expression of *Esrrb* and selected formative genes measured by qPCR in Conditional Esrrb cells differentiated for 72h and treated with a pulse of DOX between 24 and 48h. *Esrrb* KO clones expressing an inducible Empty vector (iEmpty) served as controls. Bars indicate mean +/- SD of n=5 independent experiments for *Tcf15* and n=3 for all other markers, shown as dots.

c: Experimental strategy used to remove Esrrb at the time of activation of the formative program.

d: Immunostaining for ESRRB in WT cells in 2iL and in Conditional Esrrb cells cultured either in the presence or absence of DOX for 48h. *Esrrb* KO cells expressing an Empty vector served as negative controls. Scale bar= 25 μm. Representative images from 3 independent experiments.

e: Volcano plot depicting DEGs (adjusted P-value<0.05) in *Esrrb* conditional cells kept without DOX vs WT cells after 48h of differentiation. Genes belonging to the 6 gene signatures are highlighted with coloured dots. Total number of Formative early and Formative late genes down and up-regulated in Conditional Esrrb cells are coloured in green and purple in the bottom corners. Integration of n=5 independent experiments.

f: RNAseq analysis in *Esrrb conditional cells* withdrawn of 2iL and DOX for 48h (3rd bar). WT cells expressing an inducible Empty vector (iEmpty) cultured either in 2iL or differentiated for 48h were used as controls (1st and 2nd bar). Bars indicate mean +/-SD of n=5 independent experiments, shown as dots.

g: Representative images of WT and *Esrrb* KO cells cultured in AloXR for 3 or 9 passages. Scale bar: 25μm. Esrrb KO cells collapsed between passage 4 and 6 in n=3 independent attempts.

h: Line plots showing gene expression during FS differentiation. Mean +/- SD of n=4 independent biological replicates, shown as dots.

Fig. 7: Differentiation towards PGCs is impaired by loss of Esrrb

a: Left: Quantification of the percentage of *Stella*-GFP positive PGCLC at day3 (early) and day5 (late) of independent WT (grey) and *Esrrb* KO SGET lines (blue) (n=3 independent KO clones and matching WT controls, shown as dots). Bar indicates mean value. p-values indicate two-sided unpaired t-test. Right: Flow cytometry plots showing impaired induction of PGCLC in *Esrrb*-knockout (KO) cells. Percentages of *Stella*+ PGCLC are shown in representative plots of three independent WT and KO lines.

b: PCA showing the developmental trajectory of independent *Esrrb* KO (blue) and matched-WT (grey) SGET lines during induction of PGCLC, based on the global transcriptome of n=3 independent KO clones and matching WT controls, shown as dots.

c: Volcano plots depicting DEGs in *Esrrb* KO EpiLC, d3 and d5 PGCLC. Down-regulated and Up-regulated genes (adjusted p-value <0.05) are plotted on the left or right part of each panel respectively. Formative -early and -late and PGC -early and -late signatures (genesets) are highlighted with coloured dots indicating general shifts in activity of specific programs. Integration of n=3 independent KO clones and matching WT controls.

d: Barplots showing gene expression of selected genes in independent WT (grey) and *Esrrb* KO lines (blue) at EpiLC, d3 and d5 PGCLC stages. Integration of n=3 independent KO clones and matching WT controls. See also Extended Data Fig. 8.

Fig. 8: Impaired lumenogenesis in Esrrb KO 3D structures

a: Representative images of WT and *Esrrb* KO cells cultured in N2B27 medium in matrigel for 96h. Nuclei were identified by DAPI staining (blue). Left: Filamentous (F) ACTIN was labelled by Phalloidin staining. Right: Immunostaining for PODXL. Scale bars: 30μm. Similar results were obtained with 2 Esrrb KO clones in n=2 independent experiments.

b: Barplot showing quantification of number of structures with an apically localised PODXL (polarised), with a defined central cavity (lumen) or negative for PODXL (negative) in WT and *Esrrb* KO cells cultured in N2B27 medium in matrigel for 48h, 72h or 96h. Bars indicate mean of n=2 independent experiments, shown as dots.

c: Heatmaps showing mean normalised expression measured by qPCR of the naive gene *Tfcp2l1* and formative genes in WT and *Esrrb* KO cells in 2iL or cultured in N2B27 in matrigel for 24h, 48h, 72h or 96h. Mean of n=3 independent experiments. Stars indicate p-value<0.05 calculated by two-sided unpaired Student t-test.

d: Representative images of immunostaining for OTX2 in WT and *Esrrb* KO cells cultured in N2B27 in matrigel for 48h. Scale bar: 30μm. Similar results were obtained with 2 Esrrb KO clones in n=2 independent experiments.

e: Heatmap showing Pearson's correlation of naive, core and formative genes, obtained from RNAseq data of 2iL withdrawal (Fig. 1c-d)

f: Left, trajectory followed by WT ESCs from the naive (step 0) to formative (step 9) to committed state (step 14) in a representative model. Right, trajectory followed by Esrrb KO cells.

g: Network representation of the model used to calculate the trajectories shown in f. Black solid lines indicate interactions from active components. Grey lines indicate interactions not present, as they emanate from inactive components. Positive regulations are indicated by a black arrow, negative regulations are indicated by a black circle-headed line.

References

1. Boroviak, T., Loos, R., Bertone, P., Smith, A. & Nichols, J. The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat. Cell Biol.* **16**, 516–528 (2014).
2. Kinoshita, M. & Smith, A. Pluripotency Deconstructed. *Dev. Growth Differ.* **60**, 44–52 (2018).
3. Morgani, S., Nichols, J. & Hadjantonakis, A.-K. The many faces of Pluripotency: in vitro adaptations of a continuum of in vivo states. *BMC Dev. Biol.* **17**, 7 (2017).
4. Bedzhov, I. & Zernicka-Goetz, M. Self-organizing properties of mouse pluripotent cells initiate morphogenesis upon implantation. *Cell* **156**, 1032–1044 (2014).
5. Endoh, M. & Niwa, H. Stepwise pluripotency transitions in mouse stem cells. *EMBO Rep.* **23**, e55010 (2022).
6. Smith, A. Formative pluripotency: the executive phase in a developmental continuum. *Development* **144**, 365–373 (2017).
7. Pera, M. F. & Rossant, J. The exploration of pluripotency space: Charting cell state transitions in peri-implantation development. *Cell Stem Cell* **28**, 1896–1906 (2021).
8. Buecker, C. *et al.* Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 838–853 (2014).
9. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532 (2011).
10. Kalkan, T. *et al.* Tracking the embryonic stem cell transition from ground state pluripotency. *Dev. Camb. Engl.* **144**, 1221–1234 (2017).
11. Meissner, A. Epigenetic modifications in pluripotent and differentiated cells. *Nat. Biotechnol.* **28**, 1079–1088 (2010).

12. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
13. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 7634–7638 (1981).
14. Smith, A. G. *et al.* Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688–690 (1988).
15. Williams, R. L. *et al.* Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684–687 (1988).
16. Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).
17. Dunn, S.-J., Martello, G., Yordanov, B., Emmott, S. & Smith, A. G. Defining an essential transcription factor program for naïve pluripotency. *Science* **344**, 1156–1160 (2014).
18. Wray, J., Kalkan, T. & Smith, A. G. The ground state of pluripotency. *Biochem. Soc. Trans.* **38**, 1027–1032 (2010).
19. Kinoshita, M. *et al.* Capture of Mouse and Human Stem Cells with Features of Formative Pluripotency. *Cell Stem Cell* **28**, 453-471.e8 (2021).
20. Betschinger, J. *et al.* Exit from Pluripotency Is Gated by Intracellular Redistribution of the bHLH Transcription Factor Tfe3. *Cell* **153**, 335–347 (2013).
21. Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol.* **21**, 183–186 (2003).
22. Festuccia, N. *et al.* Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. *Cell Stem Cell* **11**, 477–490 (2012).
23. Ivanova, N. *et al.* Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533–538 (2006).

24. Martello, G. *et al.* Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell* **11**, 491–504 (2012).
25. Zhang, M. *et al.* Esrrb Complementation Rescues Development of Nanog-Null Germ Cells. *Cell Rep.* **22**, 332–339 (2018).
26. Qiu, D. *et al.* Klf2 and Tfcp2l1, Two Wnt/β-Catenin Targets, Act Synergistically to Induce and Maintain Naive Pluripotency. *Stem Cell Rep.* **5**, 314–322 (2015).
27. Yang, P. *et al.* Multi-omic Profiling Reveals Dynamics of the Phased Progression of Pluripotency. *Cell Syst.* **8**, 427-445.e10 (2019).
28. Kalkan, T. *et al.* Complementary Activity of ETV5, RBPJ, and TCF3 Drives Formative Transition from Naive Pluripotency. *Cell Stem Cell* **24**, 785-801.e7 (2019).
29. Lackner, A. *et al.* Cooperative genetic networks drive embryonic stem cell transition from naïve to formative pluripotency. *EMBO J.* **40**, e105776 (2021).
30. Acampora, D. *et al.* Loss of the Otx2-Binding Site in the Nanog Promoter Affects the Integrity of Embryonic Stem Cell Subtypes and Specification of Inner Cell Mass-Derived Epiblast. *Cell Rep.* **15**, 2651–2664 (2016).
31. Mulas, C., Kalkan, T. & Smith, A. NODAL Secures Pluripotency upon Embryonic Stem Cell Progression from the Ground State. *Stem Cell Rep.* **9**, 77–91 (2017).
32. Mohammed, H. *et al.* Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. *Cell Rep.* **20**, 1215–1228 (2017).
33. Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230–1234 (2007).
34. Hayashi, K., de Sousa Lopes, S. M. C., Tang, F., Lao, K. & Surani, M. A. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* **3**, 391–401 (2008).
35. Kolodziejczyk, A. A. *et al.* Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell Stem Cell* **17**, 471–485 (2015).

36. Papatsenko, D. *et al.* Single-Cell Analyses of ESCs Reveal Alternative Pluripotent Cell States and Molecular Mechanisms that Control Self-Renewal. *Stem Cell Rep.* **5**, 207–220 (2015).
37. Sánchez-Castillo, M. *et al.* CODEX: a next-generation sequencing experiment database for the haematopoietic and embryonic stem cell communities. *Nucleic Acids Res.* **43**, D1117-1123 (2015).
38. Adachi, K. *et al.* Esrrb Unlocks Silenced Enhancers for Reprogramming to Naive Pluripotency. *Cell Stem Cell* **23**, 266-275.e6 (2018).
39. Dunn, S.-J., Li, M. A., Carbognin, E., Smith, A. & Martello, G. A common molecular logic determines embryonic stem cell self-renewal and reprogramming. *EMBO J.* **38**, (2019).
40. Festuccia, N. *et al.* Esrrb extinction triggers dismantling of naïve pluripotency and marks commitment to differentiation. *EMBO J.* **37**, e95476 (2018).
41. Fan, R. *et al.* Wnt/Beta-catenin/Esrrb signalling controls the tissue-scale reorganization and maintenance of the pluripotent lineage during murine embryonic diapause. *Nat. Commun.* **11**, 5499 (2020).
42. Shy, B. R. *et al.* Regulation of Tcf7l1 DNA binding and protein stability as principal mechanisms of Wnt/β-catenin signaling. *Cell Rep.* **4**, 1–9 (2013).
43. Wray, J. *et al.* Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nat. Cell Biol.* **13**, 838–845 (2011).
44. Martello, G., Bertone, P. & Smith, A. Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. *EMBO J.* **32**, 2561–2574 (2013).
45. Zhang, X., Zhang, J., Wang, T., Esteban, M. A. & Pei, D. Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells. *J. Biol. Chem.* **283**, 35825–35833 (2008).

773 46. Sone, M. *et al.* Hybrid Cellular Metabolism Coordinated by Zic3 and Esrrb Synergistically
774 Enhances Induction of Naive Pluripotency. *Cell Metab.* **25**, 1103-1117.e6 (2017).
775 47. Atlasi, Y. *et al.* Epigenetic modulation of a hardwired 3D chromatin landscape in two naive
776 states of pluripotency. *Nat. Cell Biol.* **21**, 568–578 (2019).
777 48. Festuccia, N., Owens, N., Chervova, A., Dubois, A. & Navarro, P. The combined action of
778 Esrrb and Nr5a2 is essential for murine naïve pluripotency. *Dev. Camb. Engl.* **148**,
779 dev199604 (2021).
780 49. Kojima, Y. *et al.* The transcriptional and functional properties of mouse epiblast stem cells
781 resemble the anterior primitive streak. *Cell Stem Cell* **14**, 107–120 (2014).
782 50. Tsakiridis, A. *et al.* Distinct Wnt-driven primitive streak-like populations reflect in vivo lineage
783 precursors. *Dev. Camb. Engl.* **141**, 1209–1221 (2014).
784 51. Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4**, 487–492
785 (2009).
786 52. Brons, I. G. M. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos.
787 *Nature* **448**, 191–195 (2007).
788 53. Guo, G. *et al.* Klf4 reverts developmentally programmed restriction of ground state
789 pluripotency. *Dev. Camb. Engl.* **136**, 1063–1069 (2009).
790 54. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human
791 embryonic stem cells. *Nature* **448**, 196–199 (2007).
792 55. Sumi, T., Oki, S., Kitajima, K. & Meno, C. Epiblast ground state is controlled by canonical
793 Wnt/β-catenin signaling in the postimplantation mouse embryo and epiblast stem cells. *PloS*
794 *One* **8**, e63378 (2013).
795 56. Magnúsdóttir, E. & Surani, M. A. How to make a primordial germ cell. *Dev. Camb. Engl.*
796 **141**, 245–252 (2014).
797 57. Hackett, J. A. *et al.* Tracing the transitions from pluripotency to germ cell fate with CRISPR
798 screening. *Nat. Commun.* **9**, 4292 (2018).

58. Okamura, E. *et al.* Esrrb function is required for proper primordial germ cell development in presomite stage mouse embryos. *Dev. Biol.* **455**, 382–392 (2019).
59. Miyazono, K., Maeda, S. & Imamura, T. BMP receptor signaling: Transcriptional targets, regulation of signals, and signaling cross-talk. *Cytokine Growth Factor Rev.* **16**, 251–263 (2005).
60. Shahbazi, M. N. *et al.* Pluripotent state transitions coordinate morphogenesis in mouse and human embryos. *Nature* **552**, 239–243 (2017).
61. Wang, X. *et al.* Formative pluripotent stem cells show features of epiblast cells poised for gastrulation. *Cell Res.* **31**, 526–541 (2021).
62. Neagu, A. *et al.* In vitro capture and characterization of embryonic rosette-stage pluripotency between naive and primed states. *Nat. Cell Biol.* **22**, 534–545 (2020).
63. Yordanov, B. *et al.* A Method to Identify and Analyze Biological Programs through Automated Reasoning. *NPJ Syst. Biol. Appl.* **2**, 16010- (2016).
64. Argelaguet, R. *et al.* Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* **576**, 487–491 (2019).
65. Boroviak, T. *et al.* Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. *Dev. Cell* **35**, 366–382 (2015).
66. Gassler, J. *et al.* Zygotic genome activation by the totipotency pioneer factor Nr5a2. *Science* **378**, 1305–1315 (2022).
67. Festuccia, N. *et al.* Mitotic binding of Esrrb marks key regulatory regions of the pluripotency network. *Nat. Cell Biol.* **18**, 1139–1148 (2016).
68. Mitsunaga, K. *et al.* Loss of PGC-specific expression of the orphan nuclear receptor ERR-beta results in reduction of germ cell number in mouse embryos. *Mech. Dev.* **121**, 237–246 (2004).

69. Byerly, M. S. *et al.* Estrogen-related receptor β deletion modulates whole-body energy balance via estrogen-related receptor γ and attenuates neuropeptide Y gene expression. *Eur. J. Neurosci.* **37**, 1033–1047 (2013).
70. Latos, P. A. *et al.* Fgf and Esrrb integrate epigenetic and transcriptional networks that regulate self-renewal of trophoblast stem cells. *Nat. Commun.* **6**, 7776 (2015).
71. Luo, J. *et al.* Placental abnormalities in mouse embryos lacking the orphan nuclear receptor ERR-beta. *Nature* **388**, 778–782 (1997).
72. Bulut-Karslioglu, A. *et al.* Inhibition of mTOR induces a paused pluripotent state. *Nature* **540**, 119–123 (2016).
73. Gu, P. *et al.* Orphan nuclear receptor LRH-1 is required to maintain Oct4 expression at the epiblast stage of embryonic development. *Mol. Cell. Biol.* **25**, 3492–3505 (2005).
74. Benchetrit, H. *et al.* Direct Induction of the Three Pre-implantation Blastocyst Cell Types from Fibroblasts. *Cell Stem Cell* **24**, 983-994.e7 (2019).

Methods

ESCs culture

All cell lines were routinely cultured on plates coated with 0.2% Gelatin (Sigma-Aldrich, G1890) in N2B27 medium (DMEM-F12 and Neurobasal at 1:1 ratio (Life Technologies), 1X N2 Supplement (Life Technologies), 1X B27 Supplement (Life Technologies), 2mM L-Glutamine (Life Technologies), 0.1mM 2-mercaptoethanol) with the addition of 2iL (3uM CHIR99021 (Axon), 1uM PD0325901 (Axon), 1uM LIF (Qkine)). Media was replaced every other day and cells were passaged every 3 days at $1x10^4$ cells/cm^2 density, following dissociation with Accutase (Life Technologies, cat. A1110501). E14IVc mouse ESCs were kindly provided by Austin Smith's laboratory[75]. Esrrb fl/fl and Esrrb-/- cells (Extended Data Fig.5a) were provided by Hitoshi Niwa's laboratory[24].

Monolayer differentiation, clonal assay, ESCs to EpiSCs differentiation and FS cells generation

For monolayer differentiation experiments, 5000 cells/12 well were plated at single cell density on 0.2% Gelatin coated plates in N2B27 medium with 2iL.

For clonal assays, cells were dissociated at indicated time points and 300 cells/12 well were plated at single cell density on 0.2% Gelatin coated plates in KSR medium (GMEM (Sigma Aldrich, G5154) supplemented with 10% KnockOut Serum Replacement (Life Technologies), 2% FBS (Sigma Aldrich), MEM non-essential amino acids (Life Technologies), 1mM Sodium Pyruvate (Life Technologies), 2mM L-Glutamine, 0.1mM 2-mercaptoethanol) with 2iL. After 4 days, cells were fixed and stained for Alkaline Phosphatase (AP) (Sigma, 86R 1KT) according to manufacturer instructions. Plates were scanned with Epson Scanner and AP positive colonies were scored manually.

For ESCs to EpiSCs differentiation, 35000 cells/12 well were plated on fibronectin (Sigma Aldrich FC010)-coated plates in N2B27 medium supplemented with 20 ng/ml ActivinA (Qkine), 12.5 ng/ml FGF2 (Qkine) and 1uM XAV939 (Axon Medchem). After three days cells were passaged in clumps using Accutase and were subsequently passaged every two days in a ratio of 1:5 or 1:10 depending on the cell density. ROCK inhibitor (Y27632 dihydrochloride, Axon Medchem) was added one hour before detaching the cells and for 6-12 hours after plating.

For Formative Stem (FS) cells generation, mouse ESCs were plated at standard density in fibronectin-coated well in N2B27 medium. The next day, medium was changed to AloXR (3 ng/ml of activin A, 2 μM XAV, 1.0 μM BMS439 in N2B27 medium). The next day cells were dissociated into clumps with Accutase and plated at higher density than established cultures (1/5 ratio). Medium was changed every day and cells split every other day.

3D structures generation

10000 mouse ESCs were resuspended in a 20uL drop of Matrigel (Corning, 356231) in a 8-well chamber slide (Life Technologies, 154534PK) and placed at 37° for 3 minutes to allow polymerisation. Each well was then filled with 300uL of N2B27 to allow 3D structures formation. Medium was changed after 2 days.

RNA extraction, cDNA synthesis and quantitative PCR

Total RNA was isolated using Total RNA Purification kit (Norgen Bioteck, cat.37500) and complementary DNA was synthesised using M-MLV Reverse Transcriptase (Life Technologies, cat.28025-013) and random hexamers. qPCR was performed with SYBR Green Master mix (Bioline BIO-94020). Expression levels were normalised to Gapdh. See also Supplementary Table 4 for primer details.

RNA-sequencing library preparation and sequencing for ESCs experiments

Total RNA was quantified using the Qubit 2.0 fluorimetric Assay (Thermo Fisher Scientific). Libraries were prepared from 250 ng of total RNA using the 3'DGE mRNA-seq sequencing service (TIGEM NGS Core) which included library preparation, quality assessment, and sequencing on a NovaSeq 6000 sequencing system using a single-end, 100-cycle strategy (Illumina Inc.).

RNA-sequencing data pre-processing and analysis for ESCs experiments

Illumina NovaSeq base call (BCL) files were converted in fastq file through bcl2fastq (http://emea.support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/bcl2fastq/bcl2fastq2-v2-20-software-guide-15051736-03.pdf - (v2.20.0.422). The raw data were analyzed by Next Generation Diagnostic srl proprietary 3'DGE mRNA-seq pipeline (v2.0) which involves a cleaning step by quality filtering and trimming (https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/usage-guide/ - bbmap suite 37.31), alignment to mm10 reference genome assembly[76], and counting by gene[77] using mm10 Ensembl assembly (release 93). Differential expression analyses were performed using edgeR[78] on genes having more than 1 CPM in more than the minimum number of samples belonging to one condition minus 1 and less than 20% of multi-mapping reads, simultaneously.

PCA was performed on $\log_2$(CPM), after filtering out genes with average raw counts across the dataset less than 5, using *prcomp* function from R (v. 4.2).

Time course differential expression analysis was performed through Gaussian process regression [79]after voom[80] transformation of $\log_2$(CPM). Genes with log-ratio of marginal likelihood greater than 5 were considered differentially expressed and used for clustering analysis. Gene clusters were first defined by dividing genes in three groups, based on their maximum expression (maximum at time 0, between 24 and 72 hours and between 84 and 96 hours respectively). A hierarchical clustering based on Pearson correlation, performed on the aforementioned groups,

allowed the identification of the 6 gene signatures (naive early, naive late, formative early, formative late, committed early, committed late) used in this study.

Pathway and gene sets enrichment analysis was conducted using the R package Enrichr(v. 3.0) on KEGG and Wikipathways (wikipathways.org) databases.

To perform the enrichment of cell-stress (GO:0033554) and apoptosis (KEGG) gene signatures in KO experiments, we used pre-ranked Gene Set Enrichment Analysis (GSEA) from *fgsea* (v 1.14.0) R package. Pre-ranked lists for each time-point or passage were generated by multiplying $\log_2$ fold change (L2FC) and false discovery rate (FDR) values as obtained by the differential expression analysis between KO and WT samples[81–85].

Generation of Overexpression (O/E) ESCs

For DNA transfection, Lipofectamine 2000 (Life Technologies, cat. 11668019) was used and reverse transfection was performed. Briefly, cells were dissociated with Accutase and resuspended in N2B27 +2iL. $1.5x10^5$ cells were mixed with 750uL 2iL, 3uL Lipofectamine 2000 (Life Technologies), 125uL of Optimem, 500ng of transposon and 500ng of transposase and plated in a Gelatin coated well of a 12 well plate. Media was changed to 2iL after overnight incubation.

siRNA transfection

For siRNA transfection, E14 cells were dissociated and resuspended in N2B27 + 2iL. $25x10^3$ cells were mixed with 400uL 2iL, 1.5 uL Lipofectamine 2000 (Life Technologies), 100uL of Optimem (Life Technologies), siRNA at a final concentration of 40nM and plated in a Gelatin coated well of a 12 well plate. Media was changed to 2iL after overnight incubation. 2iL was withdrawn for differentiation assay after 7h and replaced with N2B27 medium.

Generation of Knock-out ESCs using CRISPR-Cas9-mediated mutagenesis

gRNAs were cloned into U6 vector. Two gRNAs targeting different exons of a gene were co-transfected into E14 ESCs stably expessing Cas9, using Lipofectamine 2000 (Life Technologies). Transfected cells were selected using G418/Neomicin (50 ug/ml). Clones were picked and expanded in 2iL. gRNA sequences are listed in Supplementary Table 5.

Immunofluorescence staining and Image Analysis

Immunofluorescence was performed on Fibronectin (Merck, cat. FC010) coated glass coverslips. Cells were fixed for 10 min with 4% Formaldehyde at RT, followed by permeabilization in PBS + 0.5% Triton X-100 for 20 min at RT and blocking with PBS+0.5% Triton X-100 + 3% FBS for 1 hour at RT. Cells were incubated with primary antibodies in blocking solution overnight at 4°C. After washing with PBS, cells were incubated for 30 min at RT with Alexa Fluor-conjugated secondary antibodies (Life Technologies) used at 1:500 dilution. After washing with PBS, cells were mounted with Fluoroshield with DAPI (Sigma Aldrich, F6057). Images were acquired with Leica SP5 confocal microscope equipped with a charge-couples device camera, using the LAS AF software. Antibodies are listed in Supplementary Table 6. For image analysis, Cell Profiler was used to quantify mean fluorescence intensity for each nucleus.

Western blot

Total cell extracts were obtained by lysing cell pellets in Sonication buffer (10mM HEPES pH7.8, 150mM NaCl, 5mM EDTA, 5% glycerol, 0.5% NP40) with the addition of protease inhibitors (Sigma Aldrich, P8340) and 1mM DTT, and sonicated briefly with Bioruptor. PVDF membranes were blocked with 5% milk in TBST 0.5% for 1h at RT. Primary antibodies were incubated overnight at 4°C. HRP-conjugated secondary antibodies were incubated for 1h at RT. Images were digitally acquired using a ImageQuant LAS 4000 (GE Healthcare) and its proprietary software. Antibodies are listed in Supplementary Table 6.

Chromatin immunoprecipitation for ChIP-PCR

For ChIP-PCR, cell suspension was crosslinked with 1/10 volume of fresh Formaldehyde solution (11% methanol free Formaldehyde, 100mM Nacl, 1mM EDTA, 0.5mM EGTA, 50mM HEPES) in culture medium and incubated for 8 min at RT on an orbital shaker. Formaldehyde was quenched by adding 1/20 volume of 2.5M Glycine. After 2 min incubation at RT, cells were centrifuged at 500g for 3 min and cell pellet was washed twice with ice cold PBS+PIC before storage at -80°C. To obtain nuclear lysates, pellets from fixed cells were resuspended (1ml/10^7 cells) in ice cold LB1 buffer (10mM NaCl, 1mM EDTA, 50mM HEPES pH 7.5, 10% glycerol, 0.5% NP40, 0.25% Triton X-100) +PIC, rotated for 20min at 4°C and spun for 5min at 1500g in a table top centrifuge. Cell pellet was resuspended (1mL/10^7 cells) in ice cold LB2 buffer (10mM Tris Hcl pH8, 200mL Nacl, 1mM EDTA, 0.5mM EGTA) +PIC, rotated for 10 min, spun at 1500g for 5 min and resuspended in LB3 buffer (10mM Tris-Hcl pH8, 1mM EDTA, 100mM NaCl, 0.5mM EGTA, 0.1% Na-deoxycholate, 0.5% N-lauroylsarcosine) +PIC. Nuclear lysates were sonicated with Branson Sonifier 450A for 8 cycles (1minON/ 2min OFF) to obtain DNA fragments with an average size of 400bp.

60ug of DNA were diluted in 600uL LB3+PIC and 10 uL of Dynabeads protein G (Life Technologies) were added. Samples were rotated for 3h at 4°C. Supernatants were then incubated overnight at 4°C with 2ug of appropriate primary antibody or IgG in a rotating wheel, followed by incubation with 10uL of Protein G Dynabeads for 2h at 4°C. Beads were collected with a magnetic stand and were then washed 3 times (5min each on a rotating wheel) with 1mL of low salt wash buffer (0.1% SDS, 2mM EDTA, 1% Triton X-100, 20mM Tris pH8, 150mM NaCl) at 4°C, followed by one wash in high salt wash buffer (0.1% SDS, 2mM EDTA, 1% Triton X-100, 20mM Tris ph8, 500mM Nacl), one wash with LiCl buffer (250mM LiCl, 1% NP40, 1mM EDTA, 10mM Tris pH8) and one wash with 1mL TE buffer +50mM NaCl. Chromatin-antibody complexes were eluted by incubating in 210uL of Elution buffer (50mM Tris-HCl pH8, 10mM EDTA, 1% SDS) on a shaker block at 65°C for 20min followed by centrifugation at 16000g for 1min. Reverse

crosslinking was performed by incubating at 65°C overnight. DNA was purified with Qiagen QIAquick PCR Purification kit.

The following antibodies and cell number were used for each ChIP replicate: ESRRB ChIP: $3x10^7$ cells, 2ug mouse ESRRB antibody (Perseus cat.PP-H6705-00); H3K27ac ChIP: $3x10^6$ cells, 2.5ug rabbit H3K27ac antibody (Abcam cat.ab4729).

Chromatin immunoprecipitation library preparation and sequencing

10^7 cells were fixed with 1% formaldehyde for 15 minutes at room temperature. Samples for ChIP-seq were prepped as previously described[86].

Libraries were prepared from 10 ng of DNA using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (New England Biolabs). Quality of libraries was assessed by using Bioanalyzer DNA Analysis (Agilent Technologies), and quantified by using Qubit 4 Fluorometer (Thermo Fisher Scientific).

Libraries were sequenced on a NovaSeq 6000 sequencing system using a paired-end (PE) 100 cycles flow cell (Illumina Inc.).

ChIP-sequencing bioinformatic analyses

Paired sequencing reads were aligned on mouse mm10 reference genome using BWA (v0.7.17)[87] and filtered with samtools v(1.9) [88] to remove unmapped read pairs, not primary alignment, reads failing platform quality, with mapping quality score below 30 and duplicate reads were then removed using picard MarkDuplicates ("Picard Toolkit." 2019. Broad Institute, GitHub Repository. https://broadinstitute.github.io/picard/) (v.2.18.27).

Each sample was equally split in two pseudoreplicates, peaks were called with MACS2 (v2.2.5)[89] with $p < 0.1$ on both samples and pseudoreplicates and filtered after Irreproducible Discovery Rate analysis with a threshold of 0.05. Coverage signal profile was generated with deeptools[90] using CPM normalisation.

ATAC-seq and bioinformatic analyses

ATAC-seq libraries were prepared starting from cryopreserved cells aliquots containing approximately 0.4 million cells. Each aliquot was thawed by brief warming at 37°C in a water bath, dropwise transferred into 10 mL warm 1X PBS (Euroclone, #ECB4004) and then centrifuged at 200 rcf for 5 minutes. Cell pellet was resuspended in 1 mL of warm 1X PBS and 10 uL of suspension were used for cell counting. ATAC-seq libraries (two for each condition) were generated starting from 100,000 live cells for each sample[91] . Ten PCR cycles were performed for each library using a SimplyAmp (Applied Biosystems, #A24811) thermal cycler. Finally, the libraries were run on a 2% agarose gel using the E-gel electrophoresis system (Thermofisher, #G6400EU) for size selection. Fragments ranging from 200 bp to 700 bp were then purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research, #D4007). Sequencing was performed using 2x50 bp reads on an Illumina Novaseq-6000. Libraries pool was load on a S1 flowcell at the final concentration of 250 pM using 1% PhiX.

Paired sequencing reads were aligned on mouse mm10 reference genome using BWA (v0.7.17)[87] and filtered with samtools (v1.9)[88] to remove unmapped read pairs, not primary alignment, reads failing platform quality, with mapping quality score below 30 and duplicate reads were then removed using picard MarkDuplicates (Picard Toolkit." 2019. Broad Institute, GitHub Repository. https://broadinstitute.github.io/picard/) (v2.18.27). Peaks were called using the ENCODE ATAC-seq pipeline[92] (https://github.com/ENCODE-DCC/atac-seq-pipeline) (v1.10.0) using an IDR threshold of 0.05.

TF motifs enrichment analysis on promoter-annotated peaks (+-2kb from the TSS) was carried out using Homer's findMotifsGenome[93] (-size given) for de novo motif discovery. Only transcription factors with a percentage of targets > 8 were considered.

CUT&RUN-seq analysis

Cleavage Under Targets and Release Using Nuclease sequencing (CUT&RUN-seq)[94] data, generated in Gretarsson and colleagues[95], were analysed as follows: raw Fastq-sequences were trimmed to remove adaptors with TrimGalore (v0.4.3.1, -phred33 --quality 20 --stringency 1 -e 0.1 --length 20), quality checked and aligned to the mouse mm10 genome using Bowtie2 (v2.3.4.2, -I 50 -X 800 --fr -N 0 -L 22 -i 'S,1,1.15' --n-ceil 'L,0,0.15' --dpad 15 --gbar 4 --end-to-end --score-min 'L,-0.6,-0.6'). Analysis of the mapped sequences was performed using Seqmonk (v1.46.0) and R statistical software (v4.0.4)

Generation of *Esrrb*-knockout in SGET ESC for PGCLC specification

For analysis of PGCLC induction, *Esrrb* KO cells were generated in the *Stella*-GFP:*Esg1*-tdTomato (SGET) compound-reporter mESC line[57]. Briefly, two sp*Cas9* plasmids (pX459 Addgene #62988) carrying gRNAs that induce deletions of exons 2 and 3 (gRNAs sequences are listed in Supplementary Table 5) were transiently transfected with lipofectamine 3000, following the manufacturers guidelines. Transfected cells were selected with puromycin (1.2 µg/ml) for 60 hours and subsequently seeded at low density (1000 cells per 9.6cm^2) for single colony picking. After clonal expansion, homozygous knock-out clones were identified by PCR genotyping and confirmed by western blot. Three independent clonal WT and *Esrrb* KO cell lines were selected for further analysis.

Cell culture and PGCLC induction

Esrrb KO SGET ESCs were routinely maintained and regularly passaged on gelatin in 2i/L culture media (NDIFF 227 supplemented with PD0325901 (1 µM), CHIR99021 (3µM), LIF (1000 U/ml), FBS (1%) and penicillin/streptomycin (1%)); filtered through 0.22 µM filter) in a humified CO_2 incubator at 37°C. Epiblast-like cells (EpiLC) were induced by seeding 3x10^4 naive ESC per cm^2 on fibronectin-coated plates and maintained in EpiLC media (NDIFF 227 supplemented with knockout serum replacement (KSR) (1%), ActivinA (20 ng/ml), bFGF (12.5 ng/ml) and

penicillin/streptomycin (1%)) for 46 hours. For subsequent induction of PGCLC, $1.5x10^6$ EpiLC were seeded per well of an ultra-low attachment microwell 6-well plate (Iwaki 4810-900) using PGCLC culture media (GMEM supplemented with KSR (15%), NEAA (0.1 mM), Sodium Pyruvate (1 mM), Penicillin/Streptomycin(1%), B-mercaptoethanol (0.1mM), L-glutamine (1mM), BMP4 (500 ng/ml), LIF (1000 U/ml), SCF (100 ng/ml), BMP8a (500 ng/ml), EGF (50 ng/ml). A half-media change was performed every day. See Supplementary Fig. 3 for the gating strategy used.

Flow-cytometry of *Esrrb* KO SGET lines

After dissociation in single cell suspension using TrypLE, *Esrrb* KO or WT SGET PGCLC at 3- or 5- days induction were resuspended in PBS/1%FBS and filtered. PGCLC were isolated using fluorescent activated cell sorting with FACS Aria III (Becton Dickinson) and FACS diva software (v9.0), according to *Stella*-GFP (SG) and *Esg*-tdTomato (ET) expression that is indicative of authentic PGCLC (SG^+ET^{low}). Data analysis was performed with FlowJo software (v10.7.1).

RNA sequencing of *Esrrb* KO EpiLC and PGCLC

Total RNA was collected from *Esrrb* KO or WT bulk ESC, EpiLC and from sorted SG^+ET^{low} PGCLC of three independent biological replicates lines using the PicoPure RNA isolation kit (Applied Biosystems KIT0204), following the manufacturer instructions. After quantification of total RNA with Qubit III and quality check with high sensitivity RNA Screen Tape (Agilent 5067-5579) to ensure RIN > 8.5, 100ng of RNA was used for library preparation for NGS sequencing with NEB next Ultra II Directional RNA protocol for Poly(A) mRNA magnetic Isolation Module (NEB #E7490) following manufacturer guidelines. Multiplexed amplified libraries were sequenced on an Illumina NextSeq (SE50).

Bioinformatics analysis of RNA sequencing data for EpiLC and PGCLC

After removal of the adaptors with TrimGalore (0.4.3.1) raw reads were mapped to the mm10 (GRCm38) reference genome with RNAstar (2.6.0b-2) using default settings. Reads associated with a MAPQ score <20 were discarded. The data was quantified using the RNA-seq quantification pipeline for directional libraries in seqmonk software (v1.46.0) to generate $\log_2$ reads per million (RPM) or gene-length-adjusted (RPKM) gene expression values. We determined differentially expressed genes (DEG) using the DESeq2 package (version, 1.24.0) and applying a multiple-testing adjusted *p-value* (FDR) <0.05 significance threshold.

Computational modelling

The computational modelling was performed using the reasoning engine for interaction networks (RE:IN)[17,63,96] . This approach supports the modelling of gene networks via Abstract Boolean Networks (ABN), allowing to specify partially known networks by specifying certain interactions as definite while other interactions are designated as possible. In this logical modelling setting, an ABN contains a set of components which can be active or inactive (represented by a Boolean value). The networks are constrained by experimental observations obtained from experimental measurements, and formal verification methods are utilised to handle the large state space of candidate solutions and identify consistent models when they exist or prove inconsistency which requires a revision of the model. The methodology has proven to be applicable to study stem cell systems and the implementation has been extended to support integration of new analysis procedures and support the use of computational notebooks. The source code is publicly available on GitHub (https://github.com/fsprojects/ReasoningEngine), together with the files used to build the ABN (https://github.com/kuglerh/Esrrb).

The initial ABN (Extended Data Fig. 8g) was constructed from the experimentally validated ABN (0.717 cABN) described by Dunn and collaborators[39]. We added the formative genes Etv5, Tcf15, Dnmt3a/b, Otx2, Utf1, Lef1, Pou3f1[19,28,29]. We kept the 0.717 cABN definite interactions, connecting the signals LIF, CH and FGF to the network, and added a positive interaction between

MEK/ERK and ETV5. All these interactions are based on previous experimental studies that identified the direct targets of these signals[24,26,28,44,97,98]. We also kept the additional 0.717 cABN interactions as possible. We then derived a set of possible interactions from time course gene expression data (Fig. 1c-d), by calculating Pearsons' correlation coefficients (shown in Fig. 7e). A positive interaction between two components was defined as possible when the coefficient between the two components was above a threshold value. A negative possible interaction was set when the Pearson's coefficient was below the negative of the threshold value. A threshold value of 0.55 was determined by constructing a set of experimental constraints (Extended Data Fig. 8h),and looking for the maximum Pearson coefficient threshold that could satisfy those constraints, as described before in[17,63]. Additional possible interactions were obtained from ChIP-seq data (Fig. 3f-g).

Experimental constraints were obtained by discretising expression measurements. We started from the 2iL time course data, calculated the maximum expression value of each component and for each time point we assigned to that component a Boolean value of 1 if its expression was above 0.5 of the maximum value. The naive, formative and committed states correspond to 2iL, 48h and 96h time points, and we also included protein expression levels for Esrrb (Fig. 2h). Esrrb KO constraints were derived from time course RNAseq data of Fig. 5c-e, following the procedure described above, while Esrrb conditional KO data are based on Extended Data Fig. 5b. After verifying that all experimental constraints were satisfiable, we identified required and disallowed interactions (shown in Extended Data Fig 8g). We then generated a set of model solutions and picked a representative one, which was used to calculate trajectories and generate the network diagrams shown in Fig. 7f-g.

Statistics and reproducibility

No statistical method was used to predetermine sample size, but our sample sizes are similar to those reported in previous publications[28,48,60]. No data were excluded from the analyses. Data

distribution was assumed to be normal but this was not formally tested. For experiments with cells lines, we randomly allocated a fraction of each cell population to different biological replicates. For the analysis of immunostaining and flow cytometry data, we analysed random fields or random fraction of cells. Other kind of experiments were not randomized. Data collection and analysis were not performed blind to the conditions of the experiments, but data analyses have been performed with identical parameters and software. Data representation and statistical analyses were performed using R software, unless stated otherwise. The statistical tests used are indicated in figure legends. The number of biological replicates and independent experiments, both >2, is indicated in figures legends. Key experimental results have been obtained by 2 independent operators.

Data availability

Sequencing data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE184137. Previously published RNAseq and CUT&RUN data that were re-analysed here are available under accession codes GSE23943 and GSE146863. Primers, oligonucleotides sequences and antibodies are present in Supplementary Tables 4, 5 and 6. Source data are provided with this study. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

Code availability

Data files and models used to build the ABN are available at https://github.com/kuglerh/Esrrb. The code used to build the ABN has been described in[17,39,63] and made publicly available on GitHub at: https://github.com/fsprojects/ReasoningEngine.

Methods-only references

75. Handyside, A. H., O'Neill, G. T., Jones, M. & Hooper, M. L. Use of BRL-conditioned medium in combination with feeder layers to isolate a diploid embryonal stem cell line. *Rouxs Arch. Dev. Biol. Off. Organ EDBO* **198**, 48–56 (1989).

76. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–21 (2013).

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

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

79. Kalaitzis, A. A. & Lawrence, N. D. A simple approach to ranking differentially expressed gene expression time courses through Gaussian process regression. *BMC Bioinformatics* **12**, 180 (2011).

80. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).

81. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

82. Fast gene set enrichment analysis | bioRxiv. https://www.biorxiv.org/content/10.1101/060012v3.

83. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

84. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90-97 (2016).
85. Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Curr. Protoc.* **1**, e90 (2021).
86. Cacchiarelli, D. *et al.* Integrative Analyses of Human Reprogramming Reveal Dynamic Nature of Induced Pluripotency. *Cell* **162**, 412–424 (2015).
87. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **25**, 1754–1760 (2009).
88. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
89. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
90. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160-165 (2016).
91. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
92. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
93. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38**, 576–589 (2010).
94. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).
95. Gretarsson, K. H. & Hackett, J. A. Dppa2 and Dppa4 counteract de novo methylation to establish a permissive epigenome for development. *Nat. Struct. Mol. Biol.* **27**, 706–716 (2020).
96. Yordanov, B., Dunn, S.-J., Gravill, C., Kugler, H. & Wintersteiger, C. M. An SMT-Based Framework for Reasoning About Discrete Biological Models. in *Bioinformatics Research and*

Applications (eds. Bansal, M. S., Cai, Z. & Mangul, S.) 114–125 (Springer Nature Switzerland, 2022). doi:10.1007/978-3-031-23198-8_11.

97. H, N., K, O., D, S. & K, A. A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature* **460**, (2009).

98. Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722–737 (2009).

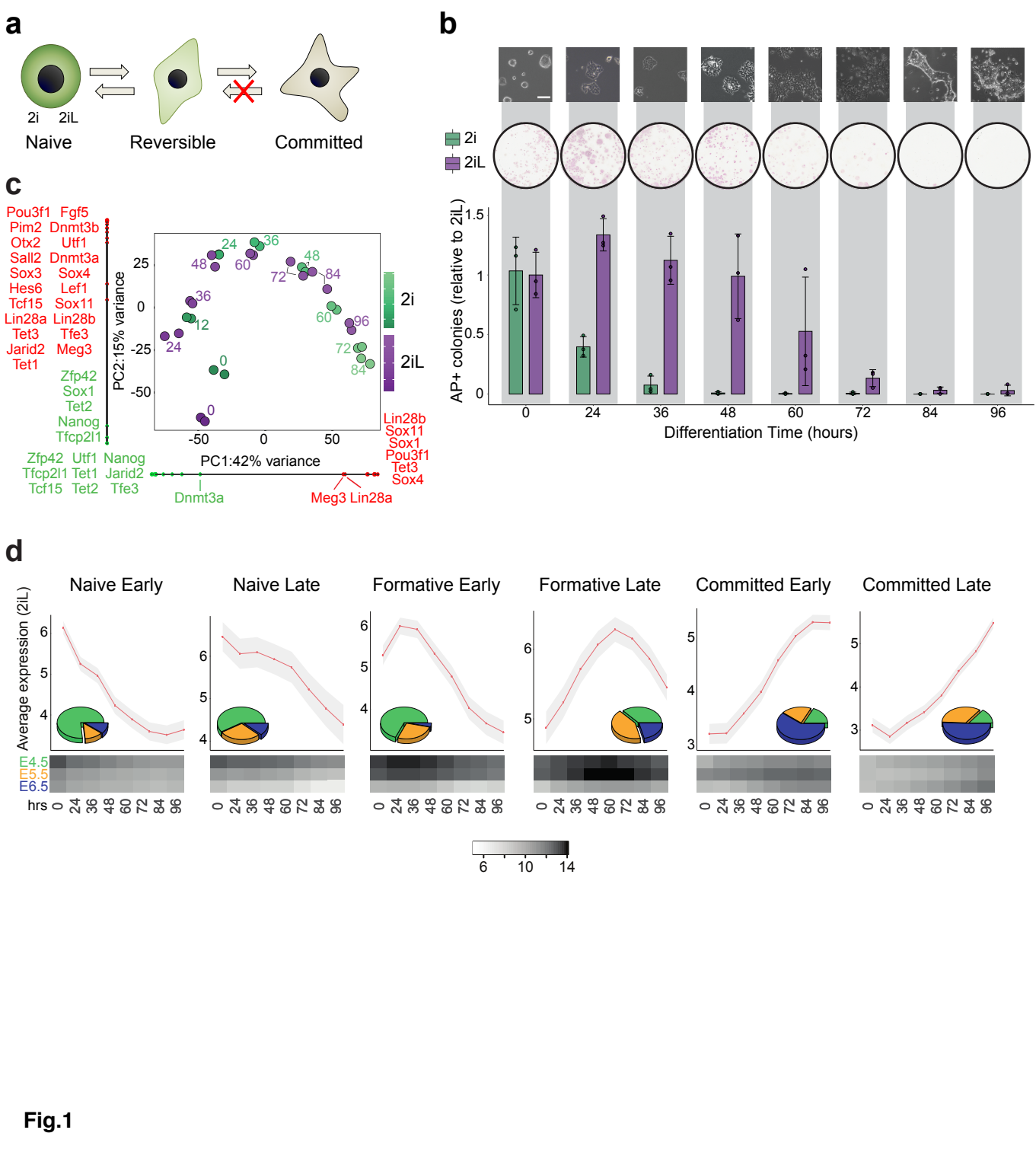

a
2i
2iL
Naive
Reversible
Committed
b
2i
2iL
AP+ colonies (relative to 2iL)
Differentiation Time (hours)
c
Pou3f1 Fgf5
Pim2 Dnmt3b
Otx2 Utf1
Sall2 Dnmt3a
Sox3 Sox4
Hes6 Lef1
Tcf15 Sox11
Lin28a Lin28b
Tet3 Tfe3
Jarid2 Meg3
Tet1
Zfp42
Sox1
Tet2
Nanog
Tfcp2l1
PC2:15% variance
PC1:42% variance
Zfp42 Utf1 Nanog
Tfcp2l1 Tet1 Jarid2
Tcf15 Tet2 Tfe3
Dnmt3a
Meg3 Lin28a
Lin28b
Sox11
Sox1
Pou3f1
Tet3
Sox4
d
Naive Early
Naive Late
Formative Early
Formative Late
Committed Early
Committed Late
Average expression (2iL)
E4.5
E5.5
E6.5
hrs

Fig.1

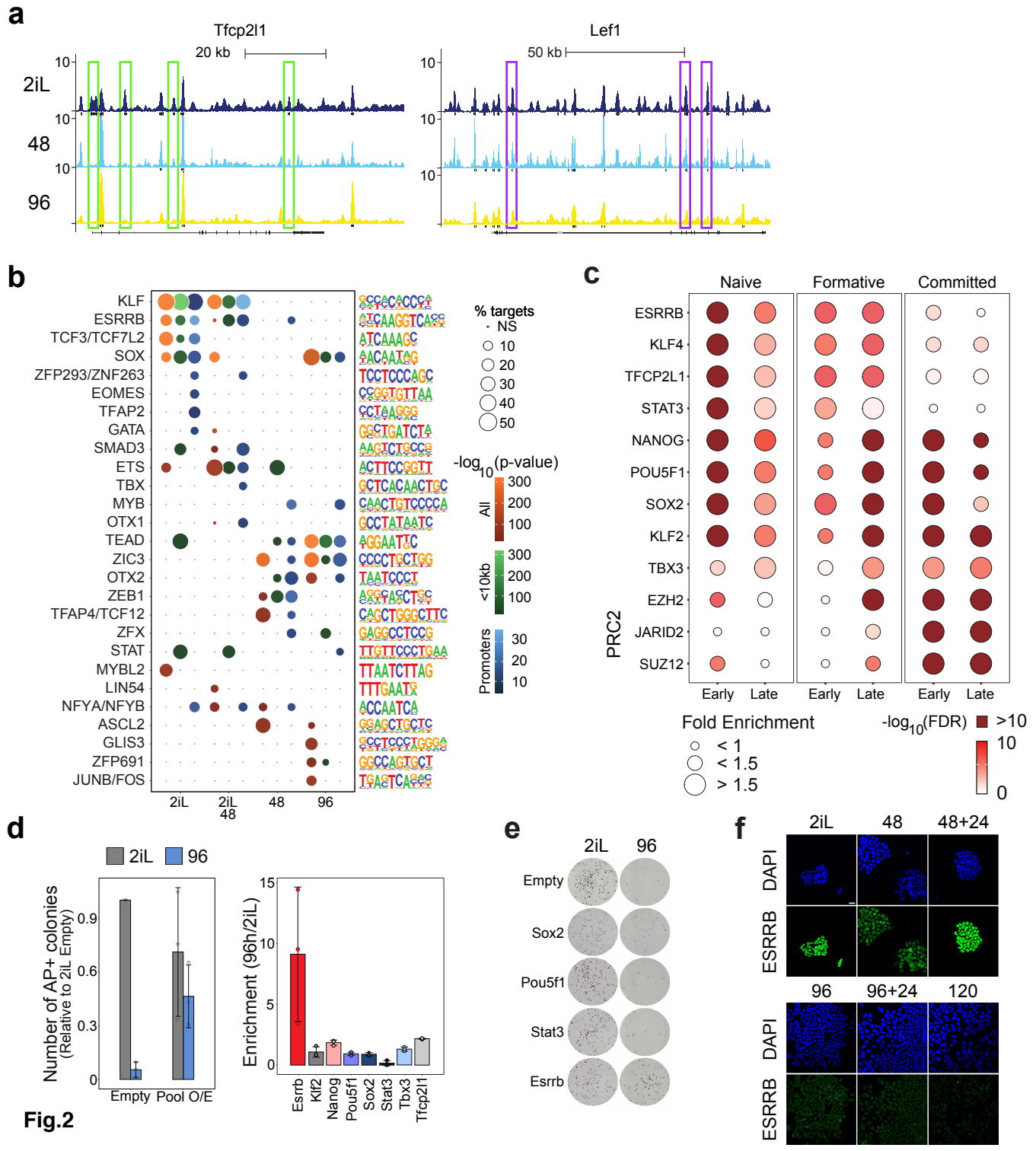

a
Tfcp2l1
20 kb
Lef1
50 kb
2iL
48
96
b
KLF
ESRRB
TCF3/TCF7L2
SOX
ZFP293/ZNF263
EOMES
TFAP2
GATA
SMAD3
ETS
TBX
MYB
OTX1
TEAD
ZIC3
OTX2
ZEB1
TFAP4/TCF12
ZFX
STAT
MYBL2
LIN54
NFYA/NFYB
ASCL2
GLIS3
ZFP691
JUNB/FOS
2iL
2iL 48
48
96
% targets
NS
10
20
30
40
50
-log10(p-value)
All
<10kb
Promoters
c
Naive
Formative
Committed
ESRRB
KLF4
TFCP2L1
STAT3
NANOG
POU5F1
SOX2
KLF2
TBX3
EZH2
JARID2
SUZ12
PRC2
Early
Late
Fold Enrichment
< 1
< 1.5
> 1.5
-log10(FDR)
>10
10
0
d
2iL
96
Number of AP+ colonies
(Relative to 2iL Empty)
Empty
Pool O/E
Enrichment (96h/2iL)
Esrrb
Klf2
Nanog
Pou5f1
Sox2
Stat3
Tbx3
Tfcp2l1
e
2iL
96
Empty
Sox2
Pou5f1
Stat3
Esrrb
f
2iL
48
48+24
DAPI
ESRRB
96
96+24
120
DAPI
ESRRB

Fig.2

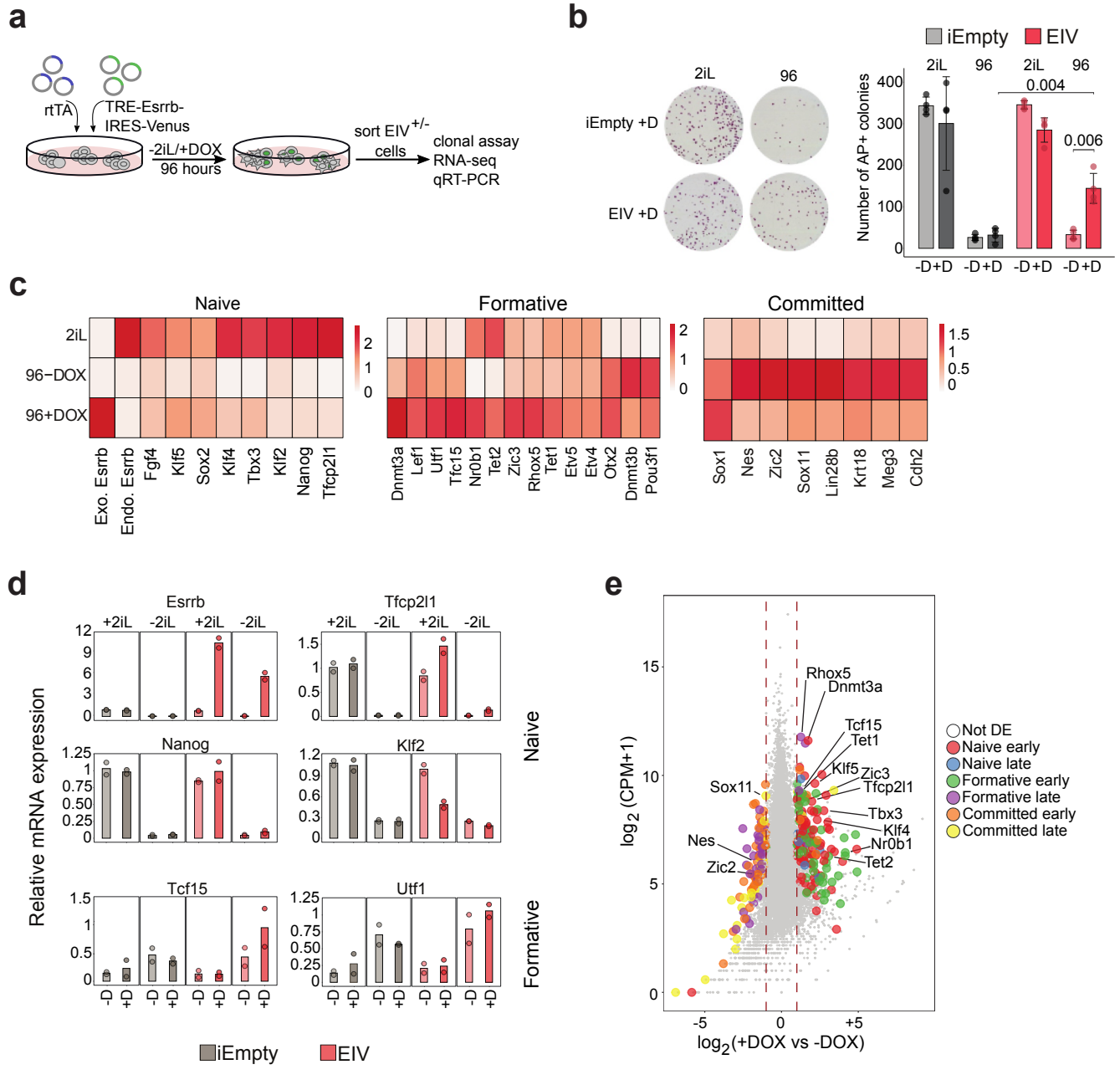

a
rtTA
TRE-Esrrb-IRES-Venus
-2iL/+DOX
96 hours
sort EIV+/- cells
clonal assay
RNA-seq
qRT-PCR
b
2iL
96
iEmpty +D
EIV +D
iEmpty
EIV
Number of AP+ colonies
0.004
0.006
-D+D
c
Naive
Formative
Committed
2iL
96−DOX
96+DOX
Exo. Esrrb
Endo. Esrrb
Fgf4
Klf5
Sox2
Klf4
Tbx3
Klf2
Nanog
Tfcp2l1
Dnmt3a
Lef1
Utf1
Tfc15
Nr0b1
Tet2
Zic3
Rhox5
Tet1
Etv5
Etv4
Otx2
Dnmt3b
Pou3f1
Sox1
Nes
Zic2
Sox11
Lin28b
Krt18
Meg3
Cdh2
d
Esrrb
Tfcp2l1
Nanog
Klf2
Tcf15
Utf1
+2iL
-2iL
Relative mRNA expression
Naive
Formative
iEmpty
EIV
e
log2 (CPM+1)
log2(+DOX vs -DOX)
Rhox5
Dnmt3a
Tcf15
Tet1
Klf5
Zic3
Tfcp2l1
Sox11
Tbx3
Klf4
Nr0b1
Nes
Zic2
Tet2
Not DE
Naive early
Naive late
Formative early
Formative late
Committed early
Committed late

Fig.3

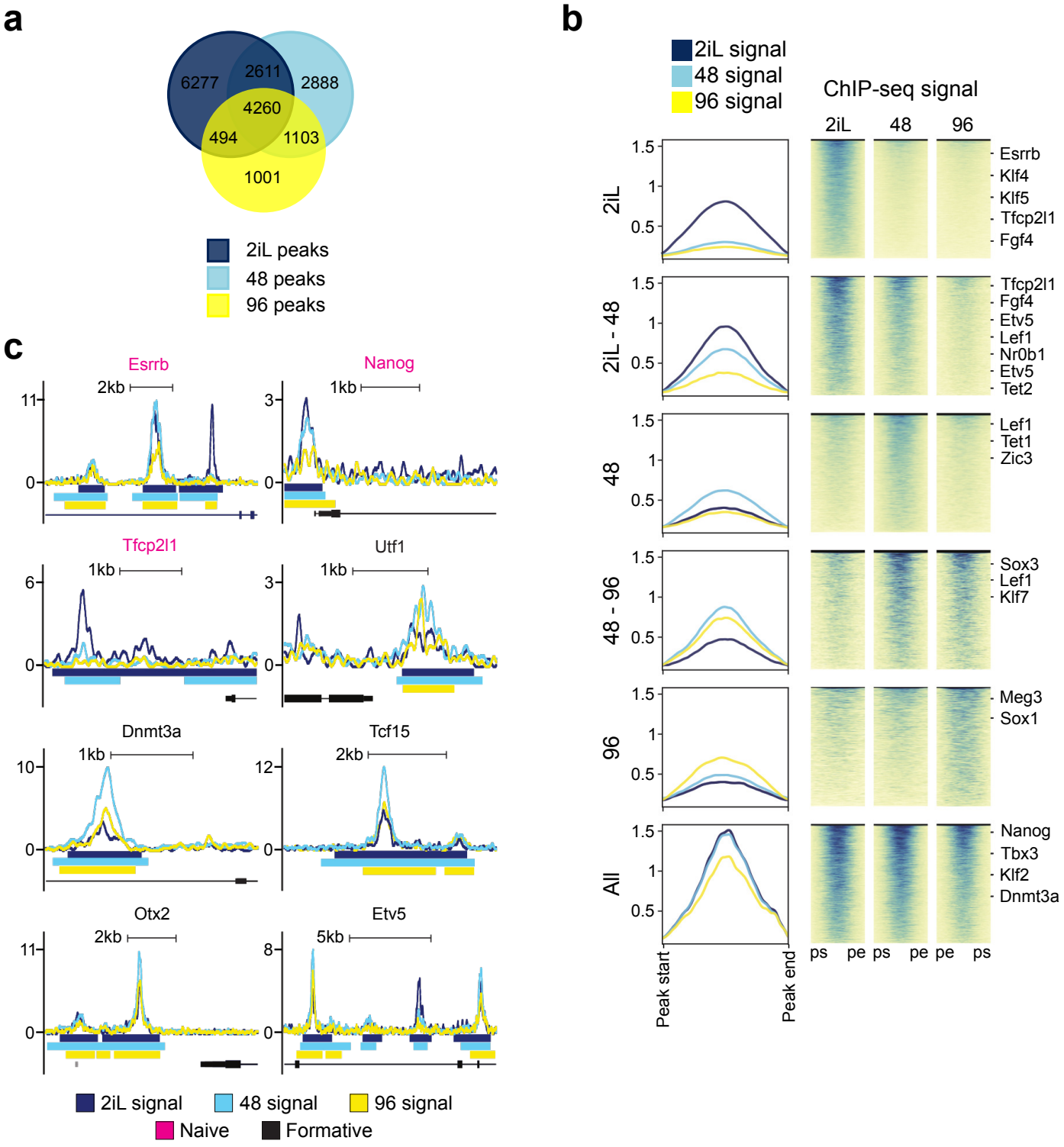

a
6277
2611
2888
4260
494
1103
1001
2iL peaks
48 peaks
96 peaks
b
2iL signal
48 signal
96 signal
ChIP-seq signal
2iL
48
96
2iL
2iL - 48
48
48 - 96
96
All
Esrrb
Klf4
Klf5
Tfcp2l1
Fgf4
Tfcp2l1
Fgf4
Etv5
Lef1
Nr0b1
Etv5
Tet2
Lef1
Tet1
Zic3
Sox3
Lef1
Klf7
Meg3
Sox1
Nanog
Tbx3
Klf2
Dnmt3a
Peak start
Peak end
ps
pe
ps
pe
pe
ps
c
Esrrb
Nanog
Tfcp2l1
Utf1
Dnmt3a
Tcf15
Otx2
Etv5
2iL signal
48 signal
96 signal
Naive
Formative

Fig.4

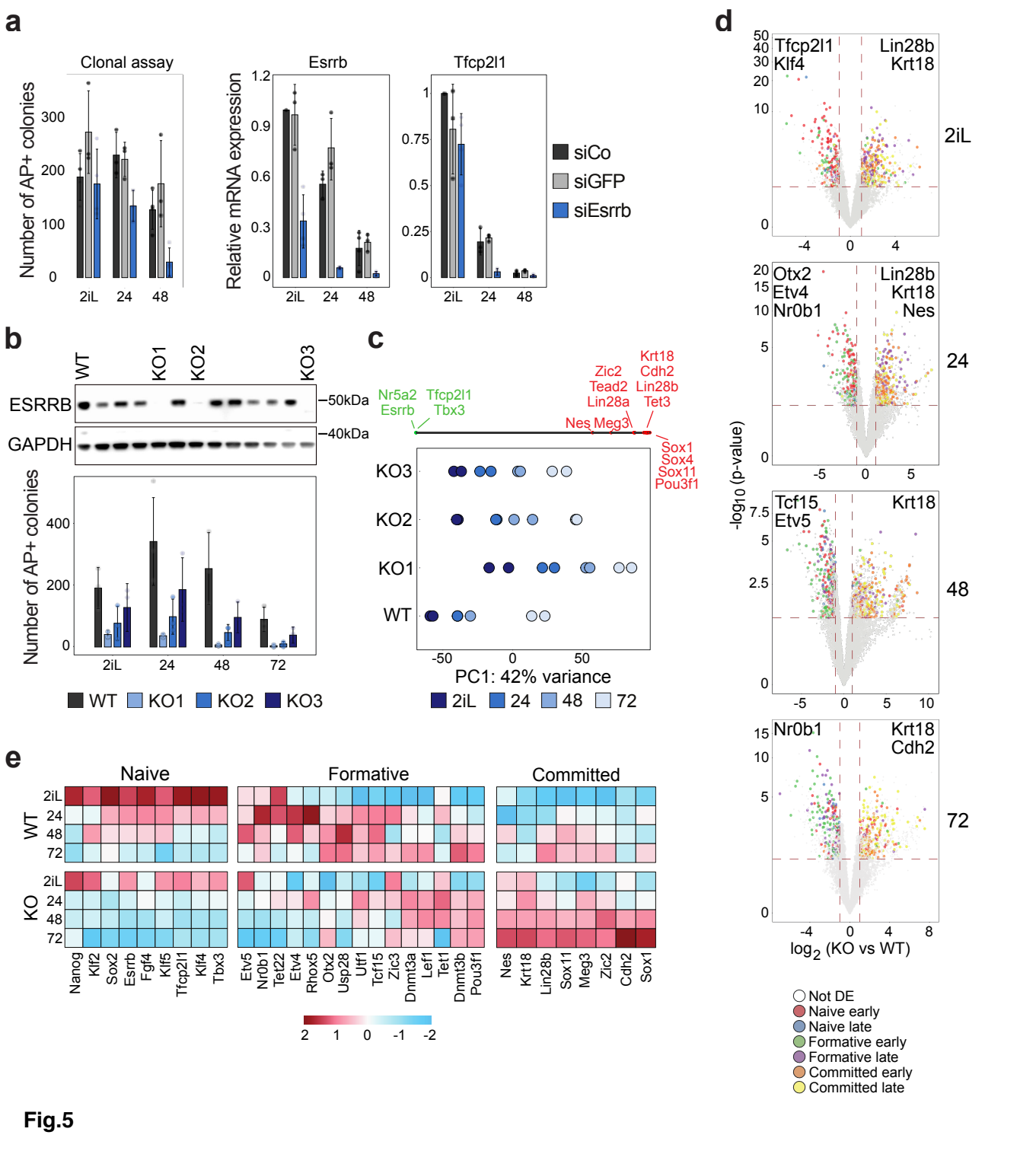

a
Clonal assay
Number of AP+ colonies
2iL
24
48
Esrrb
Tfcp2l1
Relative mRNA expression
siCo
siGFP
siEsrrb
b
WT
KO1
KO2
KO3
ESRRB
GAPDH
50kDa
40kDa
Number of AP+ colonies
72
WT
KO1
KO2
KO3
c
Nr5a2
Esrrb
Tfcp2l1
Tbx3
Zic2
Tead2
Lin28a
Krt18
Cdh2
Lin28b
Tet3
Nes
Meg3
Sox1
Sox4
Sox11
Pou3f1
PC1: 42% variance
2iL
24
48
72
d
Tfcp2l1
Klf4
Lin28b
Krt18
2iL
Otx2
Etv4
Nr0b1
Nes
24
Tcf15
Etv5
48
Nr0b1
Cdh2
72
-log10 (p-value)
log2 (KO vs WT)
Not DE
Naive early
Naive late
Formative early
Formative late
Committed early
Committed late
e
Naive
Formative
Committed
Nanog
Klf2
Sox2
Esrrb
Fgf4
Klf5
Tfcp2l1
Klf4
Tbx3
Etv5
Nr0b1
Tet2
Etv4
Rhox5
Otx2
Usp28
Utf1
Tcf15
Zic3
Dnmt3a
Lef1
Tet1
Dnmt3b
Pou3f1
Nes
Krt18
Lin28b
Sox11
Meg3
Zic2
Cdh2
Sox1

Fig.5

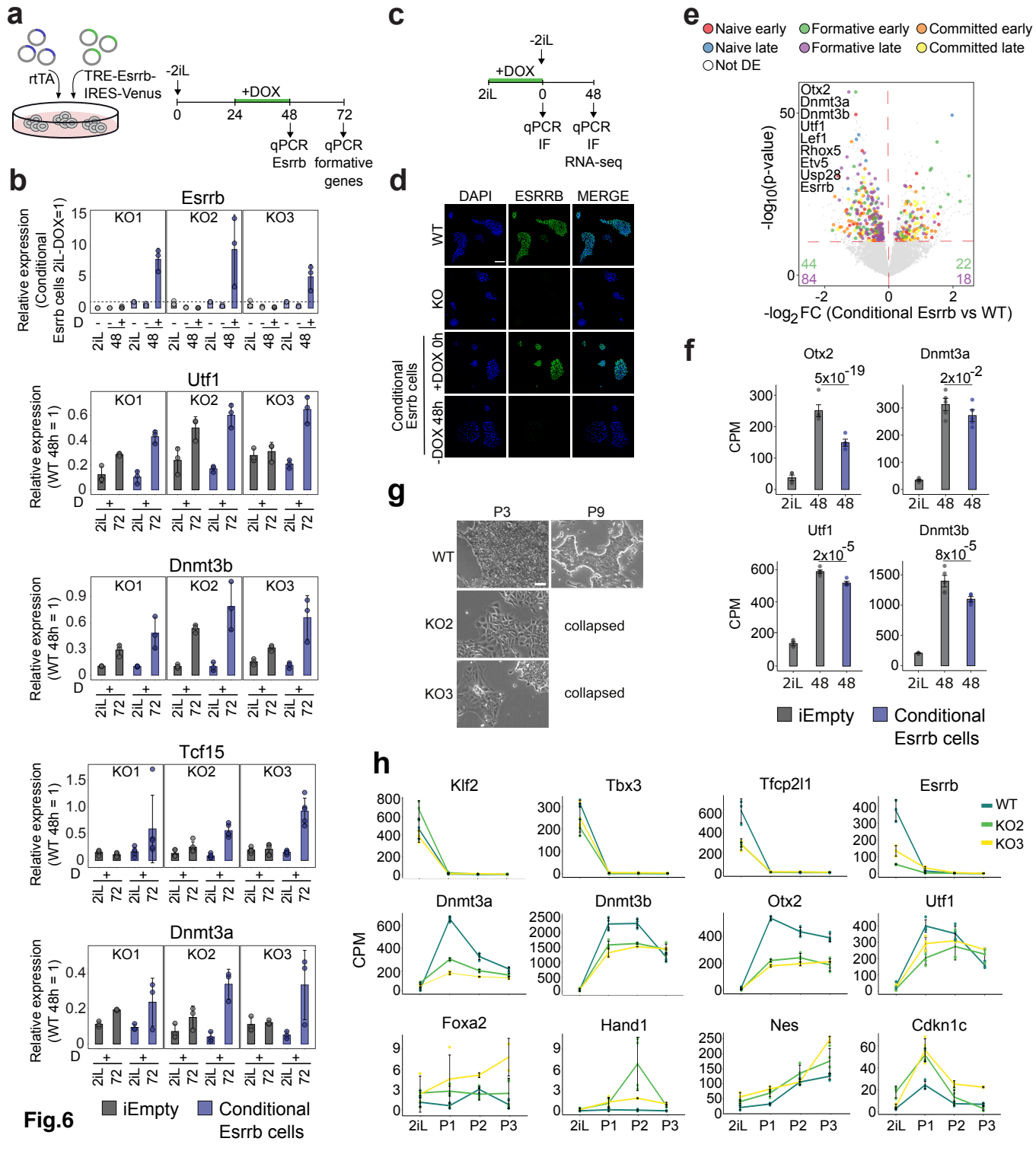

a
rtTA
TRE-Esrrb-IRES-Venus
-2iL
+DOX
0
24
48
72
qPCR Esrrb
qPCR formative genes
b
Esrrb
Utf1
Dnmt3b
Tcf15
Dnmt3a
KO1
KO2
KO3
Relative expression (Conditional Esrrb cells 2iL-DOX=1)
Relative expression (WT 48h = 1)
D
2iL
48
72
iEmpty
Conditional Esrrb cells
c
-2iL
+DOX
2iL
0
48
qPCR IF
qPCR IF
RNA-seq
d
DAPI
ESRRB
MERGE
WT
KO
Conditional Esrrb cells
+DOX 0h
-DOX 48h
e
Naive early
Naive late
Not DE
Formative early
Formative late
Committed early
Committed late
Otx2
Dnmt3a
Dnmt3b
Utf1
Lef1
Rhox5
Etv5
Usp28
Esrrb
-log10(p-value)
-log2FC (Conditional Esrrb vs WT)
44
22
84
18
f
Otx2
Dnmt3a
Utf1
Dnmt3b
5x10-19
2x10-2
2x10-5
8x10-5
CPM
iEmpty
Conditional Esrrb cells
g
P3
P9
WT
KO2
KO3
collapsed
h
Klf2
Tbx3
Tfcp2l1
Esrrb
Dnmt3a
Dnmt3b
Otx2
Utf1
Foxa2
Hand1
Nes
Cdkn1c
CPM
WT
KO2
KO3
2iL
P1
P2
P3

Fig.6

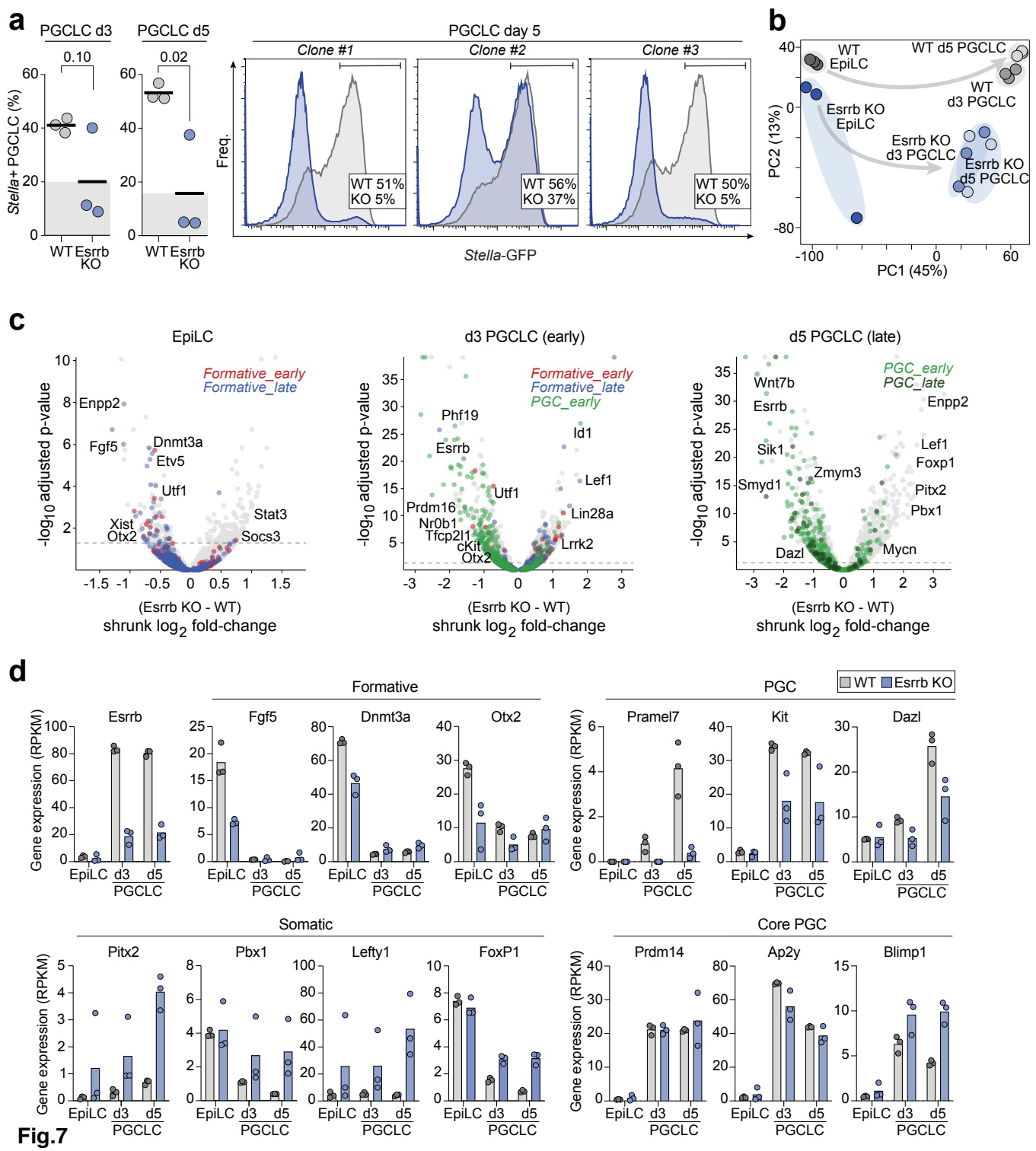
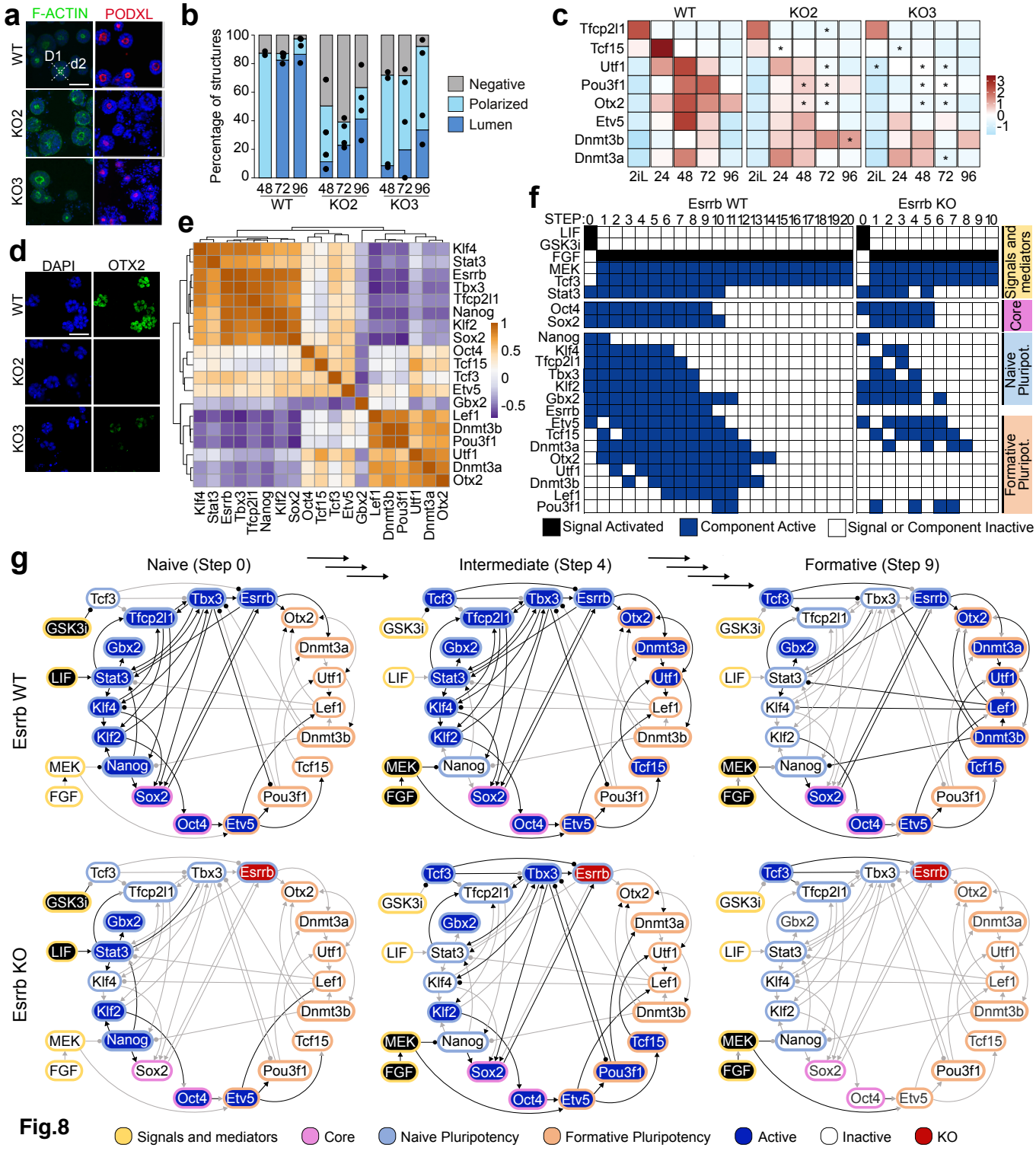
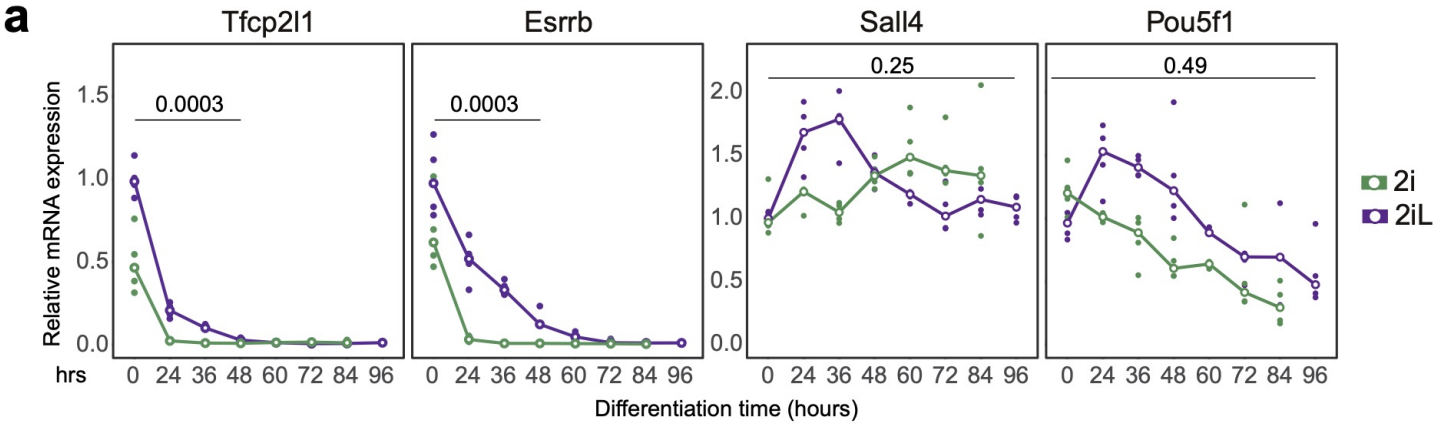
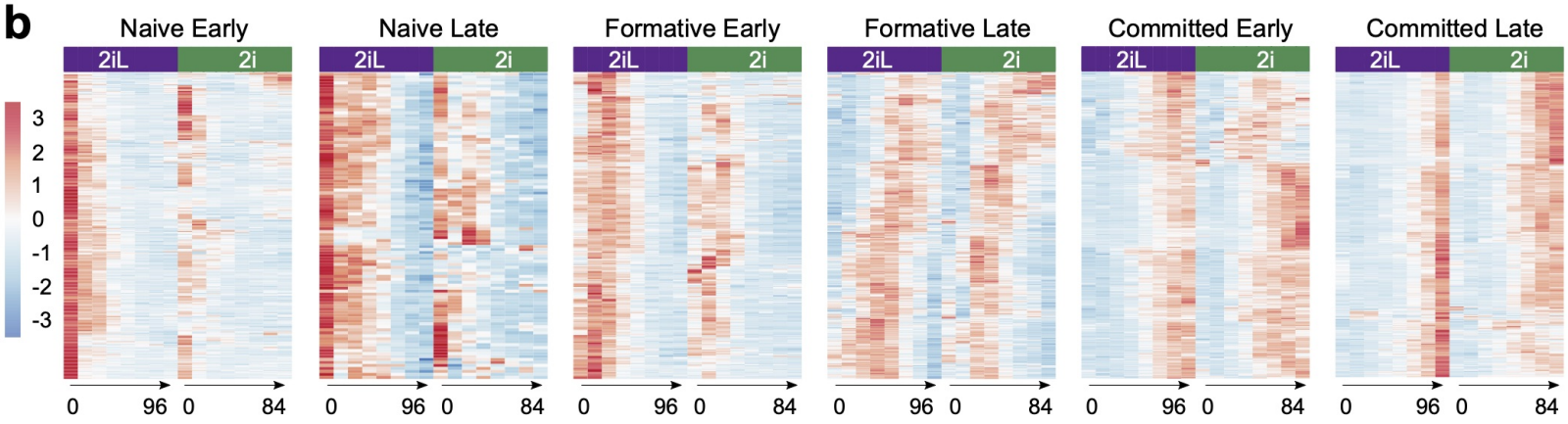

a
PGCLC d3
PGCLC d5
0.10
0.02
Stella+ PGCLC (%)
WT
Esrrb KO
PGCLC day 5
Clone #1
Clone #2
Clone #3
Freq.
WT 51%
KO 5%
WT 56%
KO 37%
WT 50%
KO 5%
Stella-GFP
b
WT EpiLC
WT d5 PGCLC
WT d3 PGCLC
Esrrb KO EpiLC
Esrrb KO d3 PGCLC
Esrrb KO d5 PGCLC
PC2 (13%)
PC1 (45%)
c
EpiLC
d3 PGCLC (early)
d5 PGCLC (late)
Formative_early
Formative_late
PGC_early
PGC_late
-log10 adjusted p-value
(Esrrb KO - WT)
shrunk log2 fold-change
Enpp2
Fgf5
Dnmt3a
Etv5
Utf1
Stat3
Xist
Otx2
Socs3
Phf19
Esrrb
Id1
Lef1
Prdm16
Nr0b1
Tfcp2l1
cKit
Lin28a
Lrrk2
Wnt7b
Sik1
Zmym3
Smyd1
Foxp1
Pitx2
Pbx1
Dazl
Mycn
d
WT
Esrrb KO
Formative
PGC
Somatic
Core PGC
Gene expression (RPKM)
Esrrb
Fgf5
Dnmt3a
Otx2
Pramel7
Kit
Dazl
Pitx2
Pbx1
Lefty1
FoxP1
Prdm14
Ap2y
Blimp1
EpiLC
d3
d5
PGCLC

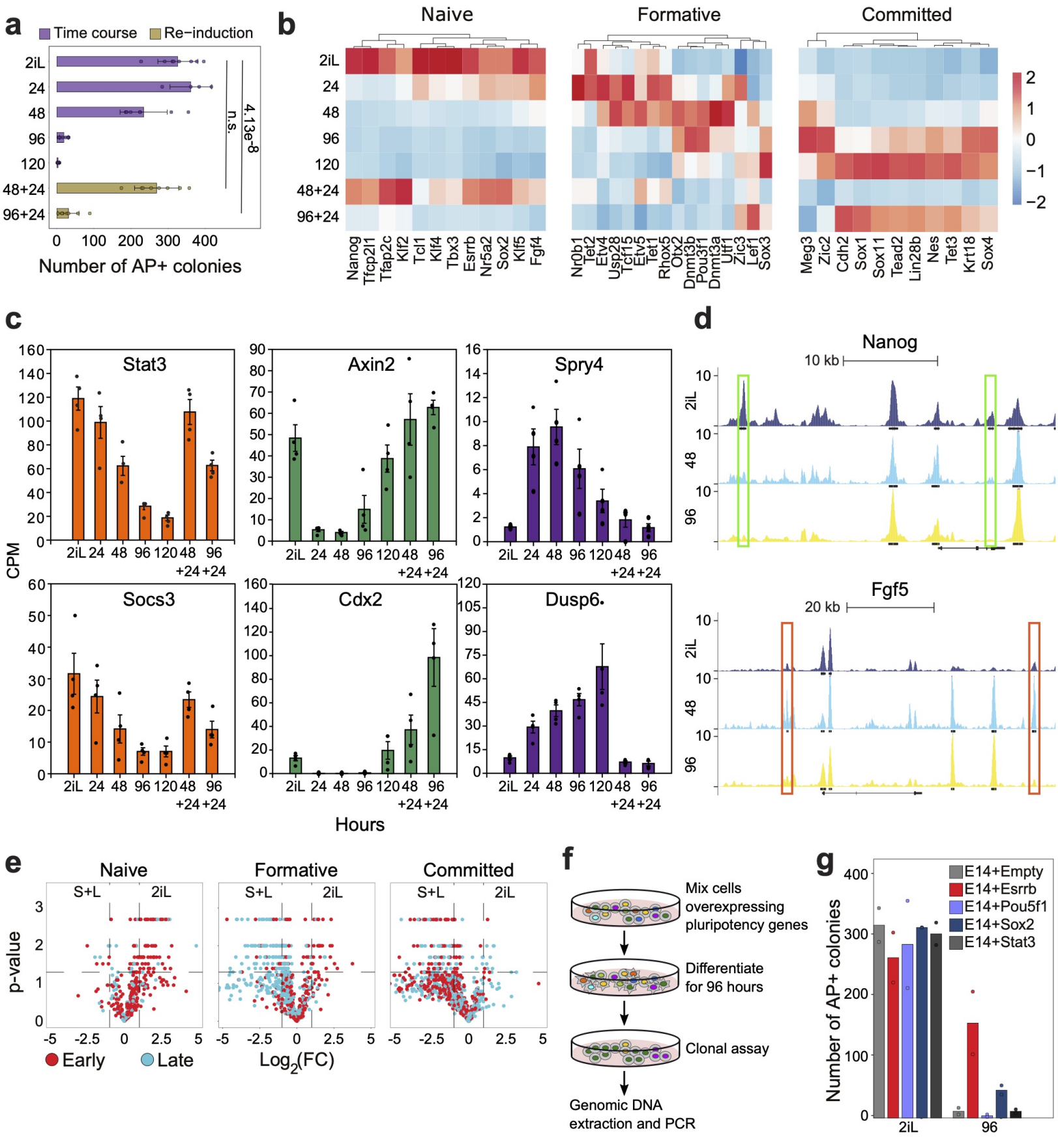
Fig.7


a
F-ACTIN
PODXL
WT
KO2
KO3
D1
d2
b
Percentage of structures
Negative
Polarized
Lumen
48 72 96
WT
KO2
KO3
c
WT
KO2
KO3
Tfcp2l1
Tcf15
Utf1
Pou3f1
Otx2
Etv5
Dnmt3b
Dnmt3a
2iL 24 48 72 96
d
DAPI
OTX2
e
Klf4
Stat3
Esrrb
Tbx3
Tfcp2l1
Nanog
Klf2
Sox2
Oct4
Tcf15
Tcf3
Etv5
Gbx2
Lef1
Dnmt3b
Pou3f1
Utf1
Dnmt3a
Otx2
f
Esrrb WT
Esrrb KO
STEP:
LIF
GSK3i
FGF
MEK
Tcf3
Stat3
Oct4
Sox2
Nanog
Klf4
Tfcp2l1
Tbx3
Klf2
Gbx2
Esrrb
Etv5
Tcf15
Dnmt3a
Otx2
Utf1
Dnmt3b
Lef1
Pou3f1
Signals and mediators
Core
Naive Pluripot.
Formative Pluripot.
Signal Activated
Component Active
Signal or Component Inactive
g
Naive (Step 0)
Intermediate (Step 4)
Formative (Step 9)
Esrrb WT
Esrrb KO
Fig.8
Signals and mediators
Core
Naive Pluripotency
Formative Pluripotency
Active
Inactive
KO

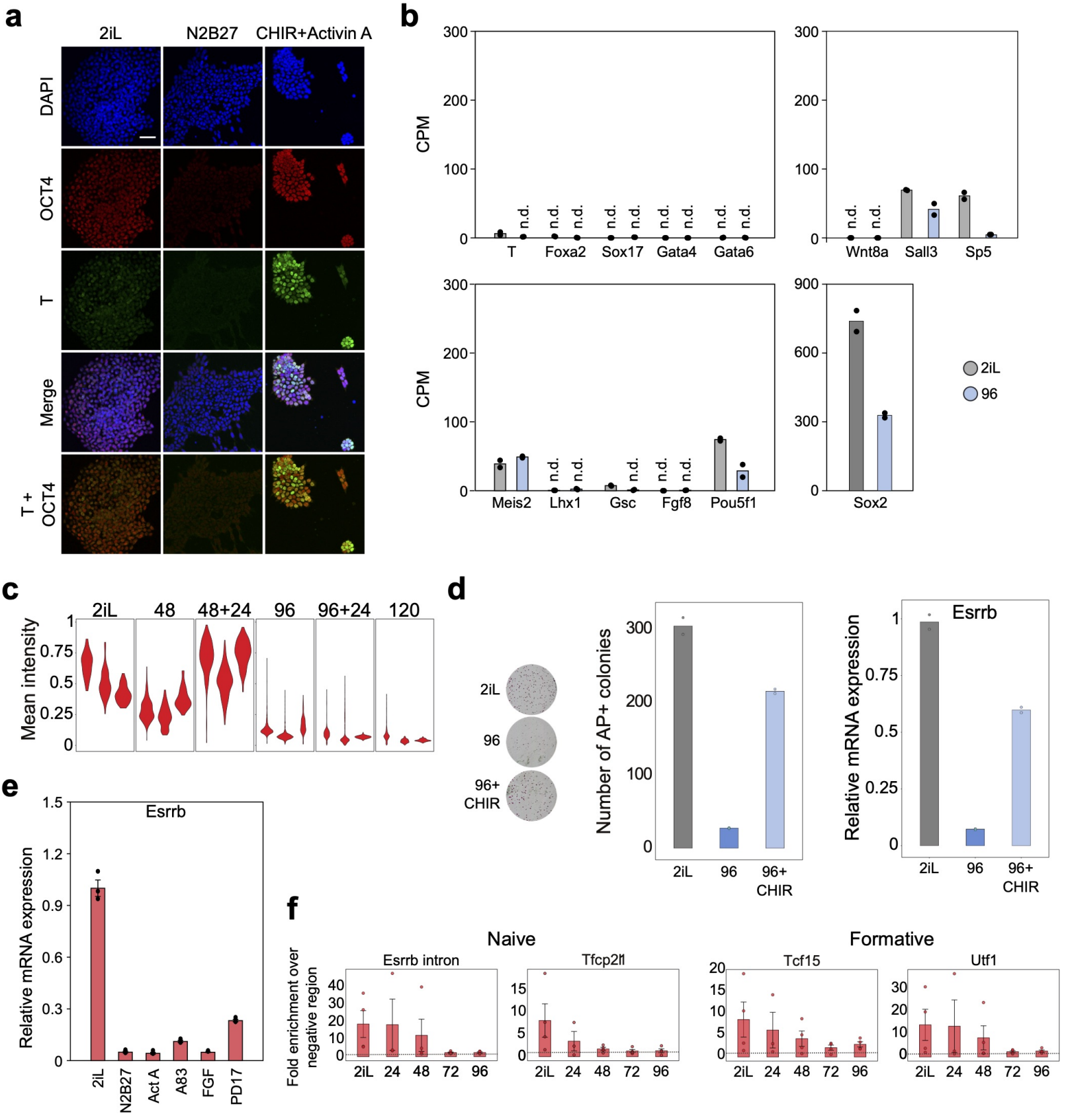




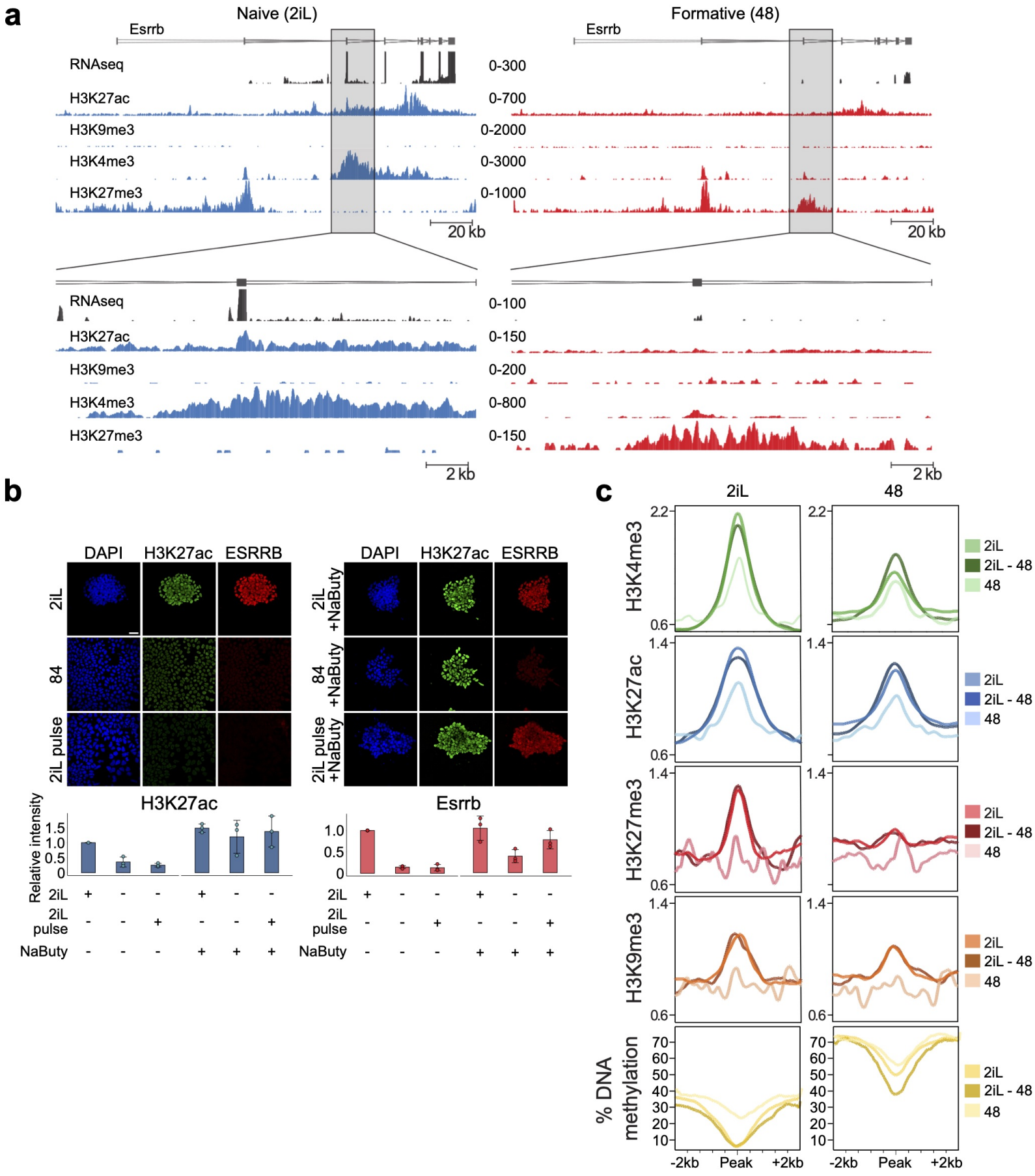
Extended Data Fig.1



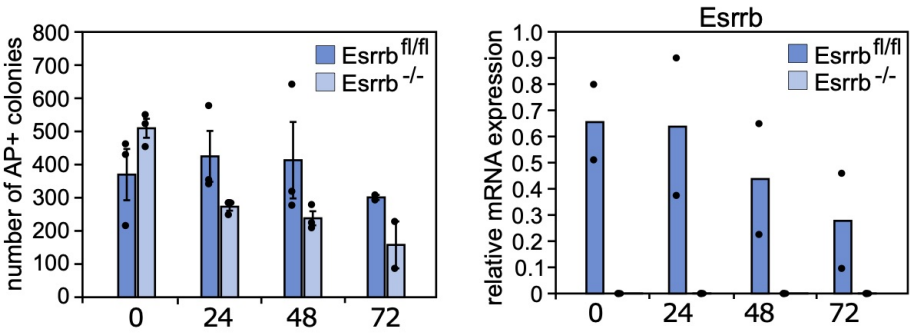
Extended Data Fig.2



Extended Data Fig.3



Extended Data Fig.4



b

gRNA1 gRNA2

Wild-type GTTCTTCAAGAGAACCATTCAAG GCAACATCGAGTACAACTGCCC....ATTCATGAAATGCCTCAAAGTGGGGATGCTGAAGGAAG

Edited KO1 alleles GTTCTTCAAGAGAACCATTCAAG —— GTGTG

gRNA1 gRNA2

Wild-type GTTCTTCAAGAGAACCATTCAAG ATGAATGTGAGATCACCAAACGGAGGCGCAAGTCCTGTCAGGCCTGCCGATTCATGAAATGC GTGTG

Edited #1 KO2 allele GTTCTTCAAGAGAACCATTCAAG ATGAATGTGAGATCACATGAATCGGCAGGCCTGACAGGACTTGCGCC—TCATGAAATGC GTGTG

Wild-type GTTCTTCAAGAGAACCATTCAAG GCAACATCGAGTACAACTGCCC....ATTCATGAAATGCCTCAAAGTGGGGATGCTGAAGGAAG

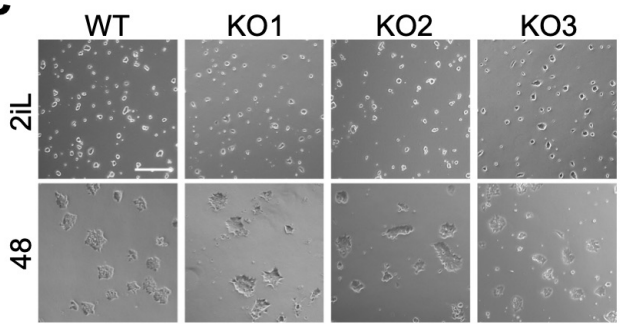
Edited #2 KO2 allele GTTCTTCAAGAGAACCATTCAAG —— GTGTG

gRNA1 gRNA2

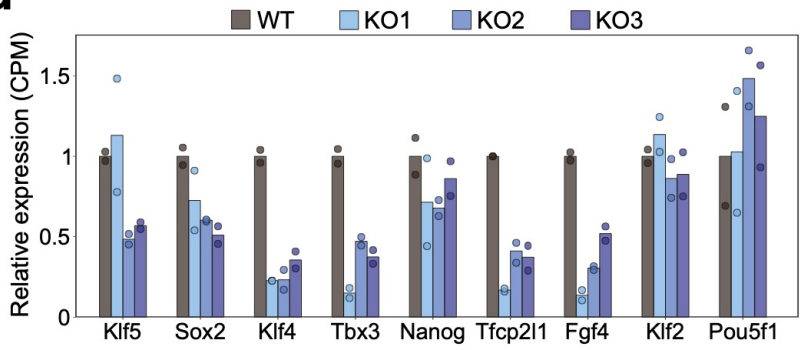
Wild-type CTGGCTACCACTACGGAGTGGCCTCCTGCGAGGCTTGCAA GGCCACCAATGAATGTGAGATCACCAAACGGAGGCGCAAGTCCTGT

Edited KO3 alleles CTGGCTACCACTA—GAGTGGCCTCCTGCGAGGCTTGCAA GGCCACCAATGAATGTGAGATCA—AACGGAGGCGCAAGTCCTGT

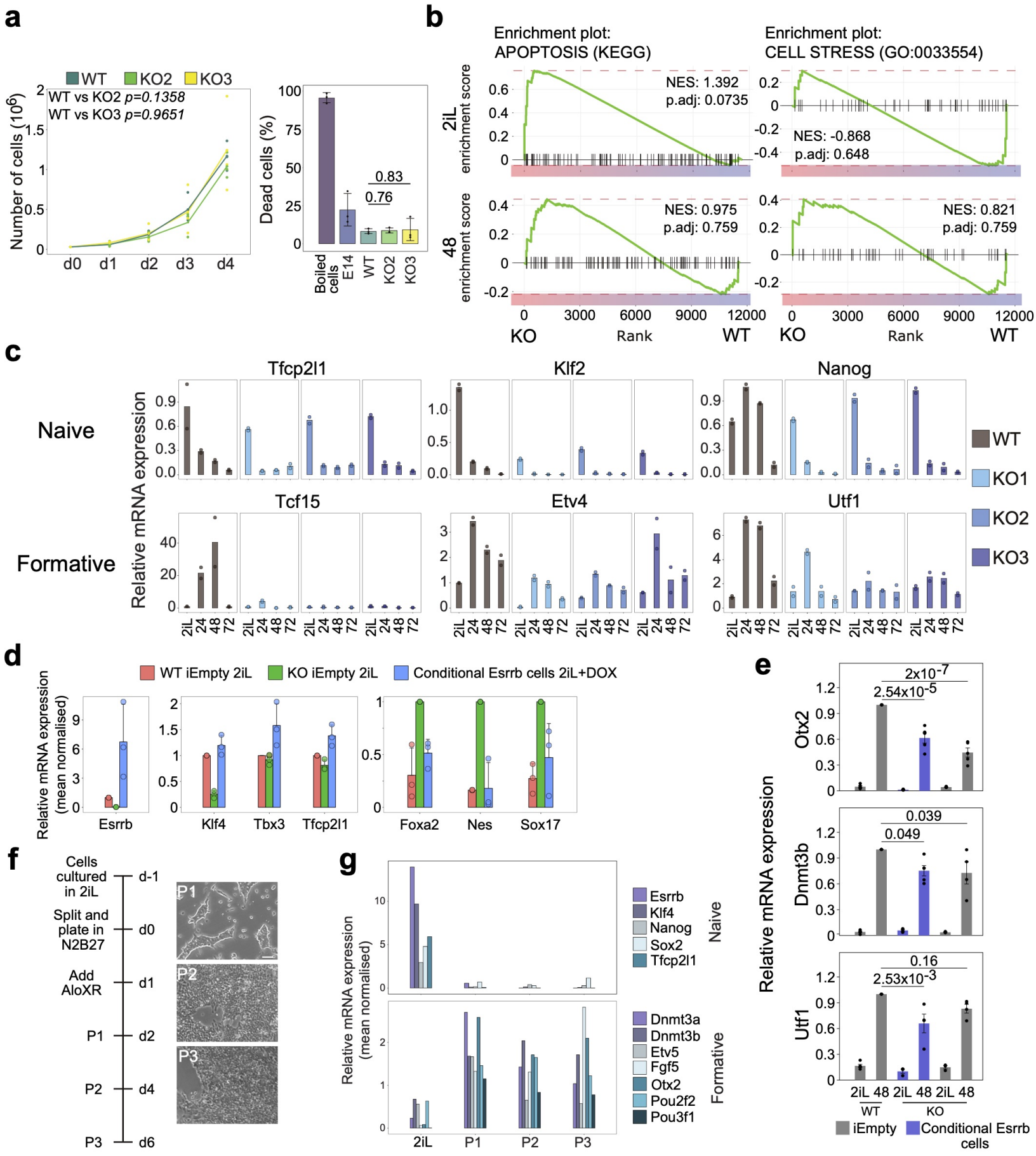
c



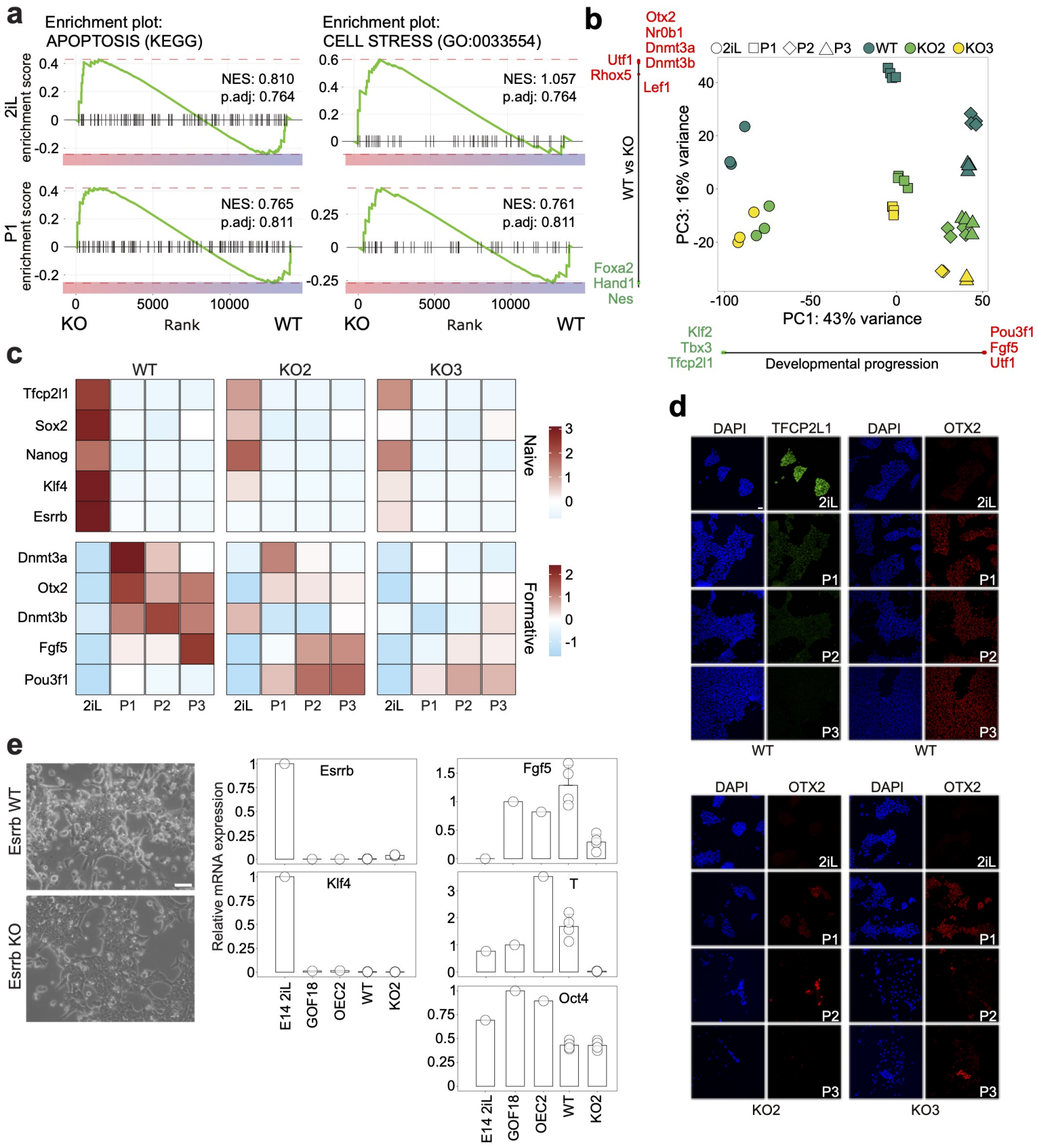
d



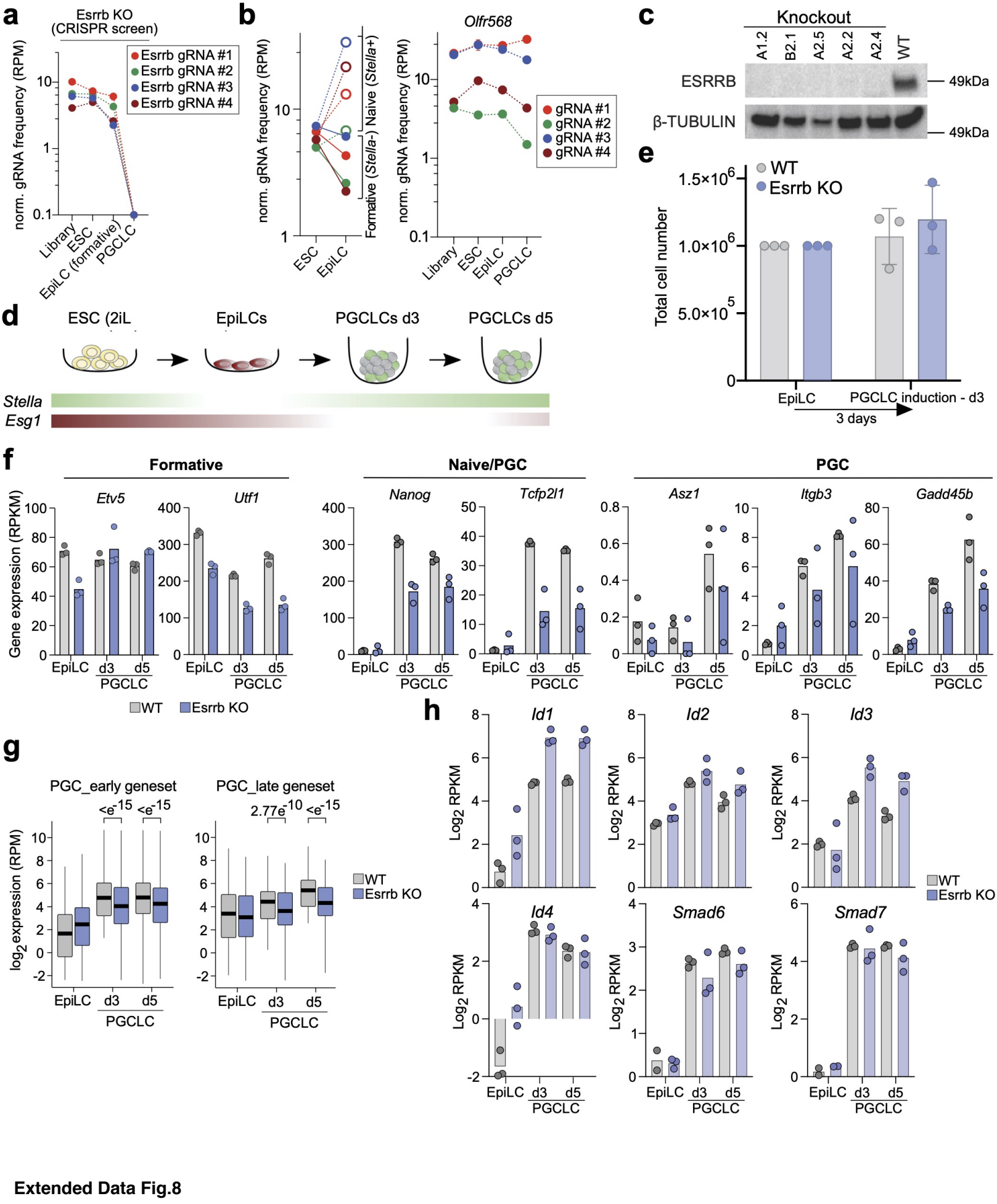
Extended Data Fig.5



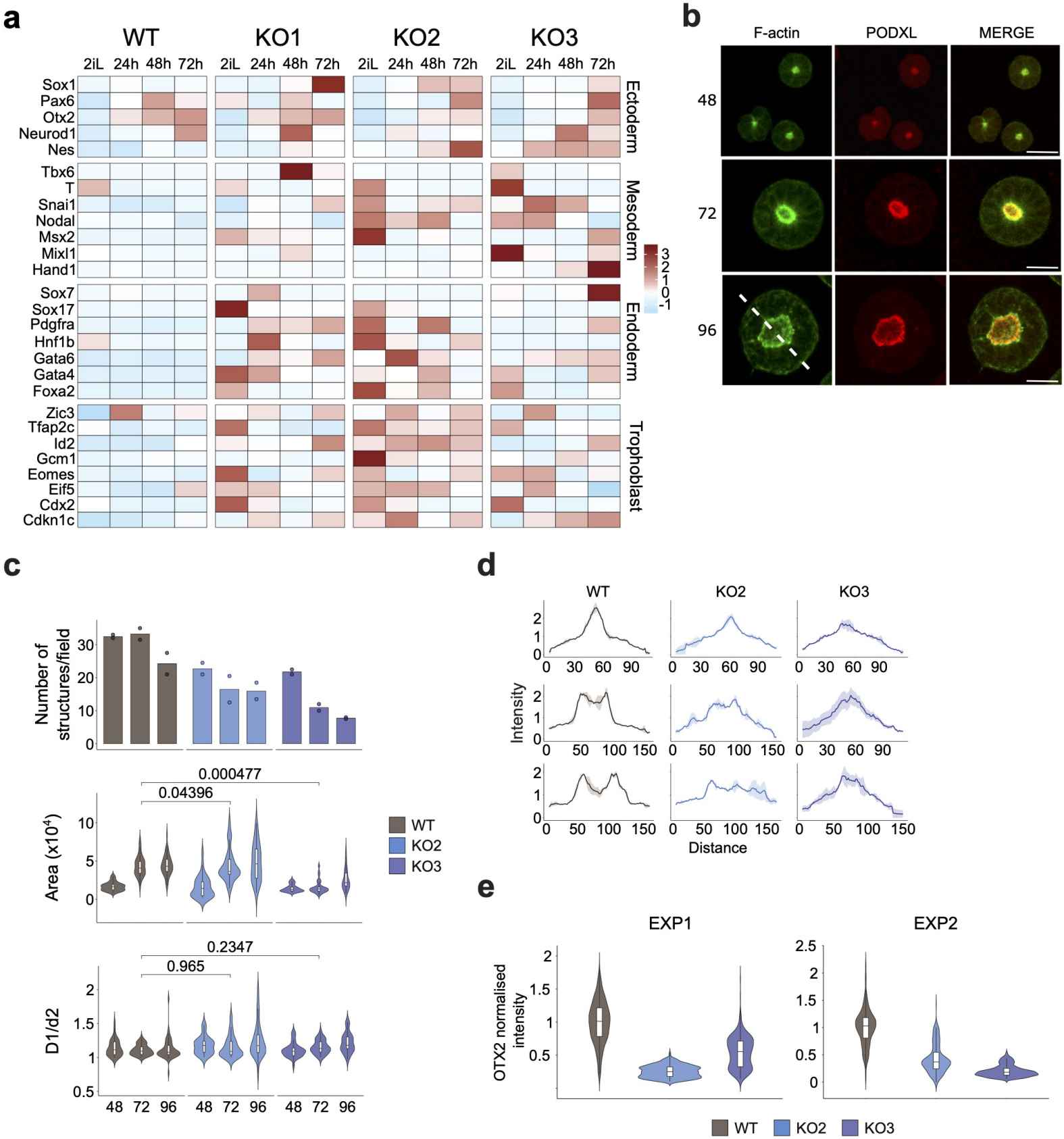
Extended Data Fig.6



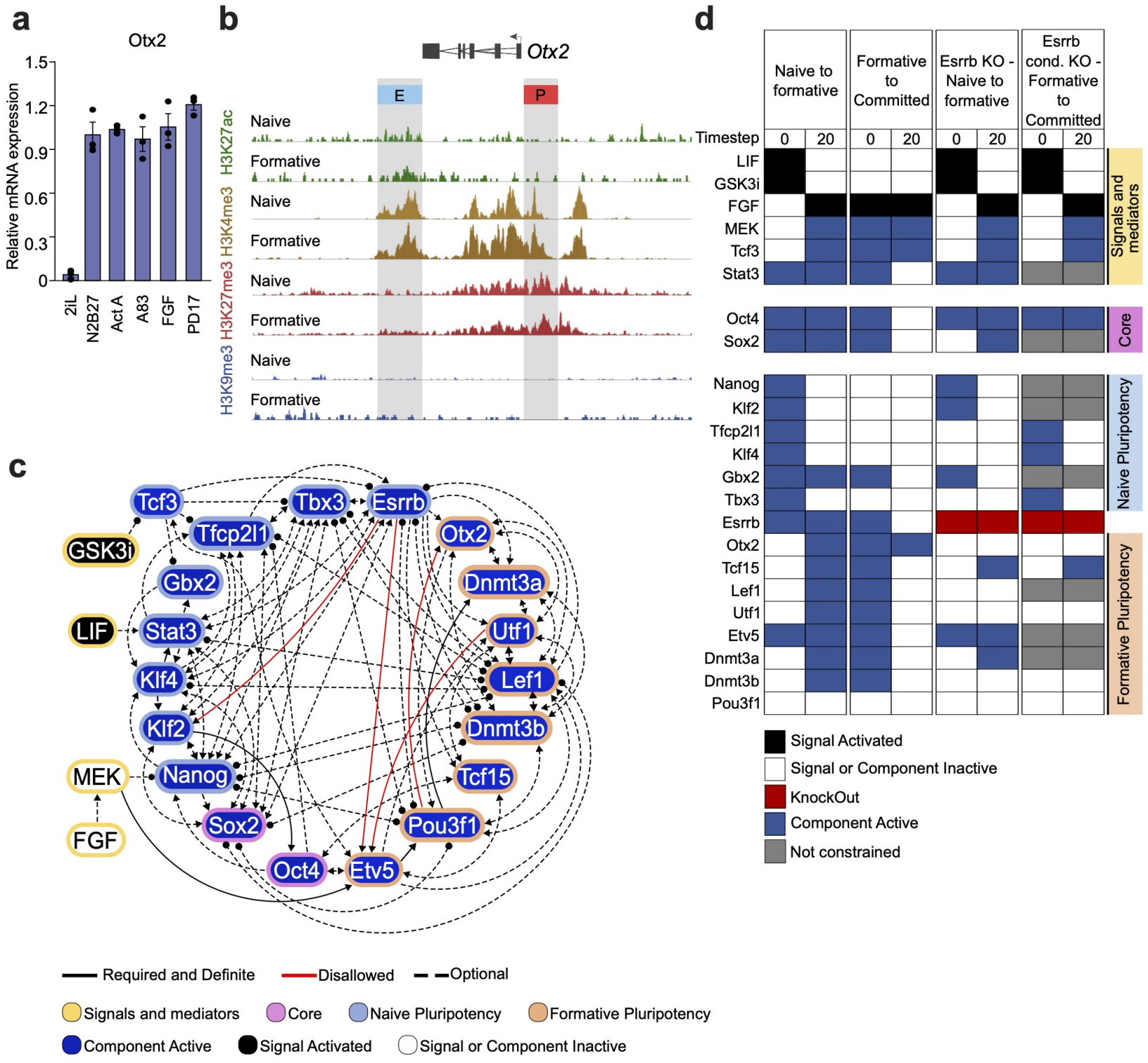
Extended Data Fig.7



Extended Data Fig.8



Extended Data Fig.9



Extended Data Fig.10