









Nuclear receptor-SINE B1 network modulates expanded pluripotency in blastoids and blastocysts

Received: 8 October 2023

Accepted: 4 November 2024

Published online: 19 November 2024



Ka Wai Wong ^{1,14}, Yingying Zeng^{1,2,14}, Edison Tay¹, Jia Hao Jackie Teo¹, Nadia Omega Cipta ¹, Kiyofumi Hamashima ¹, Yao Yi¹, Haijun Liu³, Tushar Warrier ¹, Minh T. N. Le^{4,5,6}, Soon Chye Ng^{3,7,8}, Qi-Jing Li ^{6,9}, Hu Li ¹⁰ & Yui-Han Loh ^{1,3,11,12,13} 

Embryonic stem cells possess the remarkable ability to self-organize into blastocyst-like structures upon induction. These stem cell-based embryo models serve as invaluable platforms for studying embryogenesis and therapeutic developments. Nevertheless, the specific intrinsic regulators that govern this potential for blastoid formation remain unknown. Here we demonstrate an intrinsic program that plays a crucial role in both blastoids and blastocysts across multiple species. We first establish metrics for grading the resemblance of blastoids to mouse blastocysts, and identify the differential activation of gene regulons involved in lineage specification among various blastoid grades. Notably, abrogation of nuclear receptor subfamily 1, group H, member 2 (Nr1h2) drastically reduces blastoid formation. Nr1h2 activation alone is sufficient to rewire conventional ESC into a distinct pluripotency state, enabling them to form blastoids with enhanced implantation capacity in the uterus and contribute to both embryonic and extraembryonic lineages in vivo. Through integrative multi-omics analyses, we uncover the broad regulatory role of Nr1h2 in the transcriptome, chromatin accessibility and epigenome, targeting genes associated with embryonic lineage and the transposable element SINE-B1. The Nr1h2-centred intrinsic program governs and drives the development of both blastoids and early embryos.

Preimplantation embryonic development requires sophisticated control of lineage commitment via dynamic interactions between cell signaling and transcription factors. Trophectoderm (TE) specification is largely mediated by the cell contact-based Hippo signaling in controlling nuclear localization of Yap in the outer cells^{1,2}. Yap coactivates Tead4 to induce TE-specific gene expression, such as *Cdx2*, establishing trophoblast cell fate^{1,3,4}. Epiblast (EPI) and primitive endoderm (PE) segregate within inner cell mass (ICM). EPI cells could give rise to embryonic stem cells, which are tightly regulated by the core pluripotency network factors such as *Oct4* and *Nanog*⁵. *Nanog* initiates EPI cell fate by coordinating pluripotency gene expression in the ICM

cells⁶. It also regulates *Fgf4* in EPI cells^{7–10}, which activates sequential transcription factor expression, including *Gata6*, *Sox17* and *Gata4*, in neighboring cells to commit them towards the PE lineage^{11–13}. Recent studies have reported blastoid formation via three-dimensional (3D) differentiation of stem cells by administering cocktails of ligands and/or inhibitors that drive EPI-, TE- or PE-specific transcriptional programs and cell line derivation from each embryonic lineage^{14–24}. Despite the close resemblance of blastoids to blastocysts in terms of gross morphology and the spatial organization of cells from all embryonic lineages, notable differences exist in blastoid formation efficiency, presence of intermediate cell states, and the transcriptomic

A full list of affiliations appears at the end of the paper. ✉ e-mail: yhloh@imcb.a-star.edu.sg

discrepancies between blastoid cells and bona fide embryonic lineages^{19–25}. These differences suggest external cues alone cannot guarantee the faithful differentiation of stem cells into blastoids. Furthermore, the intrinsic programs regulating the differentiation and self-organizing capacity of blastoid formation, such as transcriptional and epigenetic regulations of embryonic stem cells^{26,27}, remain to be elucidated.

Aside from the prominent transcription factors, transposable elements also form essential regulatory networks during embryo development. Highly abundant transposable elements reside in the host genome where some exert cis-regulatory functions by acting as enhancers, promoters, or insulators^{28–31}. Some transposable elements are bound by pluripotency core factors (Oct4, Sox2 and Nanog) to regulate pluripotency²⁹. Endogenous retroviruses (ERVs) are involved in critical transcriptional programs during embryonic development. Murine Endogenous Retrovirus with Leucine tRNA Primer (MERVL) transcription becomes activated along with other 2-cell embryo-specific transcripts during zygotic genomic activation^{32,33}. MERVL is repressed by zinc-finger protein Trim28 and the H3K9 methyltransferase Setdb1 which mediates H3K9me3 epigenetic repressive mark^{34–37}. Consequentially, Trim28-depletion leads to accumulation of activating histone marks at MERVL³⁸. MERVL was recently reported to be regulated by ribosomal proteins³⁹. Nonetheless, the role of this sophisticated genome-wide regulation via coordination of transposable elements, transcription factors and epigenetic landscape remains largely unknown in expanded pluripotency of embryonic stem cells (ESC) and during blastoid formation.

Genomic approaches are useful in deconstructing the complex transcription regulatory networks in stem cells as well as their dynamic interaction with extrinsic signals^{40–42}. Here we categorize blastoids into three types based on their resemblance to blastocysts in morphology, lineage segregation, and single-cell transcriptome. Differential gene expressions and active gene regulons exist between the blastoid types. Our study identifies *Nr1h2* as a crucial transcription factor regulating blastoid formation through a loss-of-function screen. *Nr1h2* depletion drastically reduced blastoid formation, whereas *Nr1h2* activation significantly enhanced blastoid efficiency. Importantly, *Nr1h2* activation empowers conventional ESC to form blastoids, resulting in the generation of more faithful TE-like cells without intermediate cell populations. These blastoids exhibit a significantly higher efficiency of implantation into the mouse uterus compared to EPS-blastoids. Additionally, *Nr1h2* activation enables ESC to contribute to both embryonic and extraembryonic lineages in chimeric blastocysts. *Nr1h2* agonist enhances blastocyst rate and morphokinetics in both mouse and porcine. Our integrative multi-omic analyses uncover the broad regulatory mechanisms of *Nr1h2* activation in rewiring the transcriptomic and epigenetic landscape of ESC. *Nr1h2* corroborates with coactivators such as Kdm1a, as well as transposable element SINE-B1, to upregulate genes crucial for embryonic lineage specification, trophoblast maturation and blastocyst development. Downstream functional target genes, *Scd1* and *Grrn*, are essential for blastoid formation. This study illustrates the significance of a *Nr1h2*-centred intrinsic program conserved in stem cell-based and native embryos across species.

Results

Transcriptome-based classification of blastoids

We differentiated expanded potential stem cell (EPSC) into blastoids which consist of epiblast-like (EPI), primitive endoderm-like (PE), and trophoblast-like (TE) lineages based on established protocol²¹. “Blastoid” was previously defined as cavitated sphere-like structures with a diameter of 60–170 μm ⁴³. We further classified the blastoids into three grading types based on their resemblance to blastocysts in morphology and lineage marker expression (Fig. 1a and Supplementary Fig. 1a–g): (1) Type I blastoid has a comparable size, aspect ratio

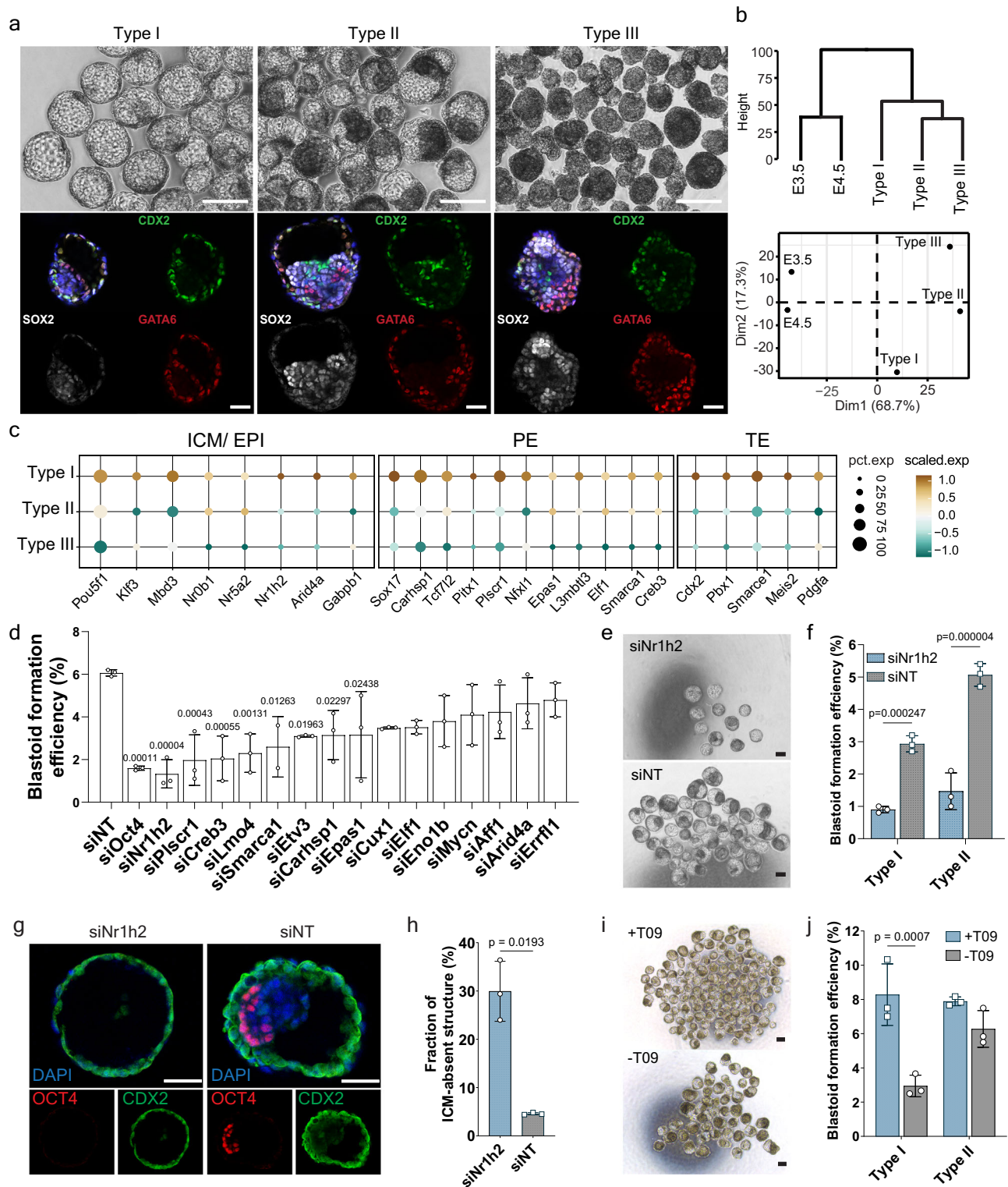
and cell ratio to the blastocyst. It consists of a large cavity space and a small acentric ICM. It displays proper spatial lineage segregation with an outer layer of Cdx2⁺ TE-like cells, and an ICM compartment with Sox2⁺ EPI-like cells enclosed by Gata6⁺ PE-like cells. (2) Type II blastoid has varied shapes and abnormal EPI-like/ TE-like cell ratios compared to type I blastoid and blastocyst. Unlike blastocyst, it contains a small cavity space and a large ICM. Cdx2⁺ cells are detected in the outer layer of type II blastoid, as well as in the ICM compartment which contains mixture of Sox2⁺ cells and Gata6⁺ cells. (3) Type III organoid is a compact cell mass without recognizable cavity space, in which Cdx2⁺, Sox2⁺, and Gata6⁺ cells localize randomly. This classification highlights the varying degrees of similarity between blastoids types and blastocysts.

We next performed single-cell RNA-sequencing (scRNA-seq⁴⁴) of type I and II blastoids and type III organoids (~10,000 single cells per sample type) using the 10 \times Chromium platform and compared their single-cell transcriptome to public dataset of mouse blastocysts. Integrated analysis using SEURAT showed that cells from type I and II blastoids and type III organoids largely overlapped with each other (Supplementary Fig. 1h). Clustering analysis categorized all cells into five clusters. Based on their expression patterns of lineage marker genes, we annotated the clusters as ICM/EPI-like (*Sox2*), PE-like (*Col4a1*), polar TE-like (*Cdx2*), mural TE-like (*Gata2*), and intermediate cells which showed mixed expression of different lineage marker genes (Supplementary Fig. 1h, i, Supplementary Data 1). Both unsupervised hierarchical clustering (Fig. 1b, top) and principal component analysis (PCA) (Fig. 1b, bottom) demonstrated that the transcriptome of type I blastoids is more similar to blastocysts, distinguishing it from type II blastoids and type III organoids. The alignment between morphological features and transcriptomic analysis provides further validation for the reliability of our blastoid classification metrics.

Critical regulons conserved across natural and stem cell-based embryos

Next, we examined the lineage genes detected in E3.5 and/or E4.5 blastocysts for their differentially expressed genes (DEGs) in the ICM/EPI-like, PE-like, and TE-like cells of blastoids (type I and II) and type III organoids (Fig. 1c). Fifty-nine ICM/EPI-specific DEGs (e.g., *Pou5f1*, *Dppa3*), 231 PE-specific DEGs (e.g., *Sox17*, *Klf4*) and 233 TE-specific DEGs (e.g., *Cdx2*) were upregulated in type I blastoid, which are enriched with functional GO terms such as DNA modification, blastocyst development, vesicle organization, apical junction assembly and anterior/posterior axis specification (Supplementary Fig. 1k–m, Supplementary Data 1). In particular, we detected a subset of DEGs which are highly expressed in type I blastoid with known roles in blastocyst development, anterior/posterior axis, apical junction assembly, and aquaporin proteins for blastocoel formation (Supplementary Fig. 1j). The higher resemblance of type I blastoid to blastocyst could be attributed to these critical transcriptome differences.

We then analyzed the single-cell gene regulatory networks—regulons, in both blastoids and blastocysts using SCENIC⁴⁵. Clustering analysis uncovered three regulon modules in type I blastoid, where module 1, 2 and 3 were enriched with lineage specification regulators for PE (*Gata6* and *Foxa2*), EPI (*Sox2* and *Pou5f1*), and TE (*Cdx2*, *Gata2*, *Zbta7a*⁴⁶) respectively (Supplementary Fig. 1n). Comparative analysis of regulons revealed that known master regulators of embryonic lineages are active across blastoids (type I and II), type III organoids, E3.5 and E4.5 blastocysts, such as *Nanog* and *Sox2* in ICM/EPI, *Gata6* in PE, *Cdx2* and *Gata3* in TE. Out of 96 top active regulons in blastocyst, 60 regulons were enriched in type I blastoids while 55 regulons were enriched in type III organoids (Supplementary Fig. 1o). Interestingly, a subset of active regulons in blastocysts was exclusively detected in type I blastoids but not in type II blastoids and type III organoids, including transcription factors (*Tcf7l1*), and nuclear receptor (*Rarg*)



which are essential regulators of blastocyst cell fate^{47–51}. These findings collectively highlight the importance of intrinsic program, which has functional implications in blastocyst development, in fine-tuning the transcriptional landscape in blastoids.

We next sought to uncover transcriptional regulatory factors facilitating stem cells in forming type I blastoids. To identify crucial regulons or genes required for blastoid formation, we investigated the top active regulons exclusively detected in type I blastoids and blastocysts, as well as the top-ranked upregulated lineage-specific DEGs in type I blastoids. To the best of our knowledge, these candidates were

not reported for their functions in blastocyst or blastoid development. We performed siRNA loss-of-function screening for 15 candidates to test their effects on blastoid formation using established protocol⁵². EPSC was seeded into Aggrewells in blastoid differentiation media at ~5 cells/ microwell and transfected with siRNA. At Day 5 post transfection, blastoids were harvested and assessed for changes in morphology and blastoid formation efficiency. *Oct4* knockdown, serving as a positive control for its critical role in stem cell pluripotency and blastocyst development, resulted in significant reduction of blastoid yield. Intriguingly, *Nr1h2* knockdown elicited the lowest blastoid formation

Fig. 1 | Classification of type I to III blastoids/ organoids and identification of key regulon in blastoid formation. **a** Phase-contrast images of EPS-blastoids (type I and II) and type III organoids (top, scale bar, 100 μ m) and immunofluorescence for markers of the TE (Cdx2), ICM/EPI (Sox2) and PE (Gata6) in type I to III (bottom, scale bar, 50 μ m). 5 independent experiments were repeated with similar results. **b** Hierarchical clustering dendrogram (top) and principle-component analysis (PCA) plot (bottom) of scRNA-seq data from cells of type I-III, public datasets of E3.5 and E4.5 blastocysts⁷⁰. **c** Dot plot indicating the expression of differentially expressed genes between Type I, II blastoids and Type III organoids for each lineage (EPI, PE, and TE). **d** Quantification of the blastoid efficiency after individual siRNA treatment. One-way ANOVA with Bonferroni correction for multiple comparisons. Data are represented as mean \pm s.d.; $n = 3$ independent assays. **e** Representative images of blastoids harvested at Day 5 of differentiation after transfection of

siNr1h2 and siNT respectively at Day 0 in a 24-well AggreWell. Scale bar, 150 μ m. **f** Quantification of type I and II blastoid formation efficiency upon *Nr1h2* knock-down. Two-way ANOVA with Bonferroni correction for multiple comparisons. Data are represented as mean \pm s.d.; $n = 3$ independent assays. **g** Immunofluorescence for markers of ICM/EPI (Oct4) and TE (Cdx2) in siNr1h2 and siNT blastoids. Scale bar, 50 μ m. **h** Quantification of the percentage of cystic structures without inner cell mass (ICM-absent) formed at Day 5 differentiation upon *Nr1h2* knockdown. Two-tailed Welch's *t*-test. Data are represented as mean \pm s.d.; $n = 3$ independent assays. **i** Representative images of blastoids harvested at Day 5 of differentiation with or without T09 treatment. Scale bar, 150 μ m. **j** Quantification of the type I and II blastoid formation efficiency with or without T09 treatment. Two-way ANOVA with Bonferroni correction for multiple comparisons. Data are represented as mean \pm s.d.; $n = 3$ independent assays. Source data are provided as a Source Data file.

efficiency corresponding to a drastic 75% reduction (Fig. 1d). Notably, *Nr1h2* abrogation significantly reduced the yield of both type I and II blastoids and the average blastoid diameter (Fig. 1e, f, Supplementary Fig. 1p, q). 30% of the cystic structures in *Nr1h2* knockdown sample displayed a single layer of Cdx2⁺ cells without any Oct4⁺ ICM (Fig. 1g, h).

To understand the molecular consequences of *Nr1h2* knockdown in relation to embryo development, we performed *Nr1h2* siRNA microinjection into zygotes followed by bulk RNA-seq of siNr1h2 and siNT morula embryos. 734 DEGs were upregulated while 900 DEGs were downregulated in siNr1h2 morula (Supplementary Fig. 1r, Supplementary Data 1). The downregulated DEGs in siNr1h2 morula were enriched with 8-cell embryo gene expression signature, among which about 25% of downregulated DEGs (230 genes) overlapped with the genes highly expressed in 8-/16-/32-cell embryos (Supplementary Fig. 1s). These DEGs include EPI markers (*Tdgfr*^{53,54}, *Pramel7*⁵⁵), PE markers (*Lrp2*⁵⁶, *Dab2*^{57–59}), TE markers⁶⁰ (*Dppa1*, *Krt8*, *Ptges*), lineage specification regulator (*Amotl2*^{61,62}, *Sumo2*^{63–66}), and genes essential for blastocyst development (*Setd2*⁶⁷, *Ccnbl*⁶⁸, *Scd1*⁶⁹) (Supplementary Data 1). We utilized a public single-cell transcriptome dataset of mouse pre-implantation development⁷⁰ and found that *Nr1h2* expression was first detected in 4-cell embryo, upregulated in 8-cell embryo and peaked at 16-cell stage embryo (Supplementary Fig. 1t). These indicate that *Nr1h2* knockdown impedes blastoid formation by downregulating broad gene sets crucial for early embryonic development.

We next tested if Nr1h2 activation would elicit opposite effects. Indeed, supplementing Nr1h2 agonist T0901317 (T09) to blastoid differentiation significantly improved blastoid formation efficiency from 9.2% to 16.2%. Of note, type I blastoid formation increased by ~3-fold with T09 treatment. (Fig. 1i, j). Together, our results indicate the critical regulatory role of *Nr1h2* activity to facilitate blastoid formation, especially in type I blastoid.

Nr1h2 activation is sufficient to capacitate blastoid formation

We next effect Nr1h2 activation in conventional ESC over multiple passages by supplementing T09 in ES culture media, followed by comprehensive blastoid forming assay. Unexpectedly, Nr1h2 activation alone was sufficient to rewire ESC towards an expanded pluripotent state with blastoid-forming competency. Specifically, Nr1h2-activated ESC (NrESC) exhibited 2.1% blastoid formation efficiency by 3rd passage, which was further elevated to 7.5% by 20th passage (Fig. 2a, b).

NrESC displays compact 3D dome-shaped colony morphology resembling ESC (Fig. 2c). Since Nr1h2 was reported for its broad regulatory effects in hepatic epigenome and transcriptome⁷¹, we characterized the NrESC through integrative analyses of transcriptome (RNA-seq), chromatin accessibility (ATAC-seq), histone marks ChIP-Seq, Nr1h2 ChIP-Seq, DNA methylome (whole-genome bisulfite sequencing) and Nr1h2 transcriptional interactome (coimmunoprecipitation (Co-IP) mass spectrometry) (Fig. 2d). The bulk transcriptome of NrESC showed 1099 upregulated DEGs and 835 downregulated

DEGs compared to ESC (Supplementary Data 2). PCA revealed that NrESC transcriptome is distinct from both EPSC and ESC (Fig. 2e). The upregulated DEGs in NrESC significantly overlapped with 8- and 16-cell gene sets, suggesting NrESC transcriptome is closer to earlier embryonic stages compared to ESC (Supplementary Fig. 2a). Heatmap of the marker gene expression fold-change, with reference to ESC transcriptome, showed that NrESC marker genes are highly enriched in NrESC but not in EPSC nor 2-cell-like cell (2CLC) while the marker genes of 2CLC (*Zscan4a*, *Zscan4b*, *Zscan4c*, *Nelfa*, etc) and EPSC (*Hi9*, *Col4a1*, *Dnmt3a*, etc) are not enriched in NrESC (Supplementary Fig. 2b). Given that NrESC gains increasing blastoid forming capacity under prolonged T09 treatment, we have further investigated the NrESC bulk transcriptome changes in early and prolonged T09-treated samples. The gene expression fold-changes (NrESC over ESC) of the DEGs in 16th passage NrESC were visualized in a heatmap for early and later NrESC passages. Two gene clusters were observed, which refer to down-DEGs (cluster 1) and up-DEGs (cluster 2) in 16th passage NrESC. Interestingly, cluster 2, which includes many top-ranked NrESC marker genes, showed increasing expression fold change over NrESC passages (Supplementary Fig. 2c, Supplementary Data 2), suggesting a time-dependent T09 treatment effect on NrESC transcriptome. Single-cell transcriptome analysis of NrESC showed a homogenous population expressing canonical pluripotency markers (*Pou5f1*, *Sox2*, *Nanog*), with no pre-differentiated PE or TE populations. Selected NrESC markers (*Scd1*, *Gm*, *Krt8*), which are among top ten up-DEGs or essential for blastocyst formation, were detected throughout the NrESC population (Supplementary Fig. 2d, e). In addition, NrESC under prolonged T09 treatment showed a normal karyotype and three germ-layer contribution in teratoma assay (Supplementary Fig. 2f, g). Our data suggests that NrESC possess a unique pluripotent transcriptome.

Nr1h2 binding sites and H3K27Ac mark in NrESC were examined in relation to transcriptional changes using Nr1h2 and H3K27Ac ChIP-seq. A total of 11,215 H3K27Ac peaks were upregulated in NrESC. The majority of these upregulated peaks are located at gene regions, with around 20% peaks located at promoter regions, and around 38% peaks located at enhancer regions. Around 40% of peaks are located within 10 kb of genes (Supplementary Fig. 2h–j, Supplementary Data 2). 1175 genes with upregulated H3K27Ac peaks showed increased Nr1h2 binding in NrESC (Fig. 2f, g), among which 363 annotated genes correspond to ~33% of total upregulated DEGs in NrESC (Supplementary Fig. 2k, Supplementary Data 2). Functional GOTERM analysis showed enrichment in known roles of Nr1h2 in lipid and cholesterol metabolism, such as *Abca1*, *Abcg1*, and *Scd1*^{72–74}. Regulation of cell adhesion and desmosome organization is also enriched, including genes such as *Dsg2*, *Krt17*, and *Krt8*, which are key structural proteins of desmosome known to regulate the adhesiveness of blastocyst trophectoderm^{75–77} (Supplementary Fig. 2l, Supplementary Data 2). Importantly, Nr1h2 activation also led to upregulated expression of transcription factors *Tead4* and *Gata4*, accompanied by increased H3K27Ac marks, which specify TE and PE lineage respectively (Fig. 2h, i).

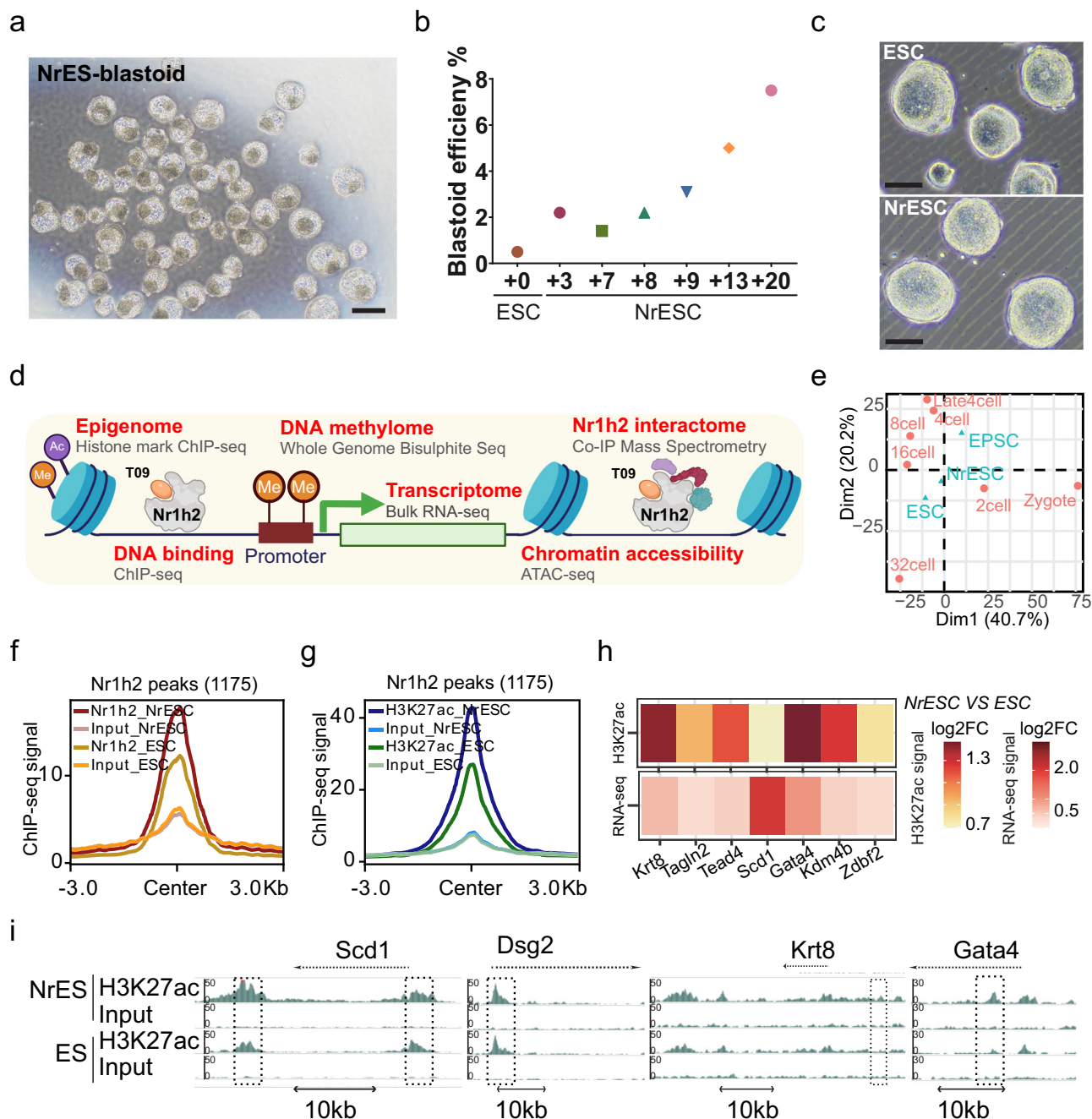


Fig. 2 | Multi-omic analyses of Nr1h2 activation in conferring expanded pluripotency. **a** Representative images of NrESC-derived blastoids harvested at Day 5 of differentiation. Scale bar, 150 μ m. 25 independent experiments were repeated with similar results. **b** Quantification of blastoid formation efficiency from different passages of NrESC. X-axis denotes number of passages cultured in presence of T09, where +0 referred to starting ESC. **c** Brightfield images of ESC and NrESC. Scale bar, 100 μ m. 100 independent cultures were repeated with similar results. **d** A schematic diagram showing the multi-omic approach of the molecular characterization of NrESC. Created in BioRender.com **e** Principle-component analysis (PCA) of bulk

RNA-seq data from ESC, NrESC (Passage 16), and EPSC along with single-cell RNA-seq data from development stage based average expression values. **f** Average Nr1h2 ChIP-seq signal comparison between NrESC (Passage 16) and ESC on Nr1h2 binding peaks with H3K27ac increases in NrESC (Passage 16). **g** Average H3K27ac ChIP-seq signal comparison between NrESC and ESC on Nr1h2 binding peaks with H3K27ac increases in NrESC. **h** Heatmap showing the representative genes upregulated in NrESC with differential H3K27ac peaks. **i** Representative genes upregulated in NrESC with increased H3K27ac signal in comparison to ESC. Source data are provided as a Source Data file.

Changes in genome-wide chromatin accessibility under Nr1h2 activation in NrESC were assessed by ATAC-seq. The number of detected ATAC-seq peaks for ESC, NrESC (Passage 1), NrESC (Passage 16) is 39,815, 18,942, and 23,016 respectively. The overall ATAC-seq signal was slightly decreased in NrESC (Supplementary Fig. 3a). NrESC gained accessibility in 751 genomic regions while losing accessibility in 799 open regions (Supplementary Fig. 3b). The upregulated ATAC

peaks are mainly located at intergenic regions, with around 30% peaks at gene regions (Supplementary Fig. 3c). We observed progressive upregulation of these peaks over prolonged T09 exposure in NrESC (Supplementary Fig. 3d). This is consistent with PCA showing that NrESC at 1st passage is closer to ESC and more distant by the 16th passage (Fig. 3a). Binding motifs of EPI and TE lineage specific transcription factors, such as *Pou5f1*, *Sox2* and *Tead4*, showed increased

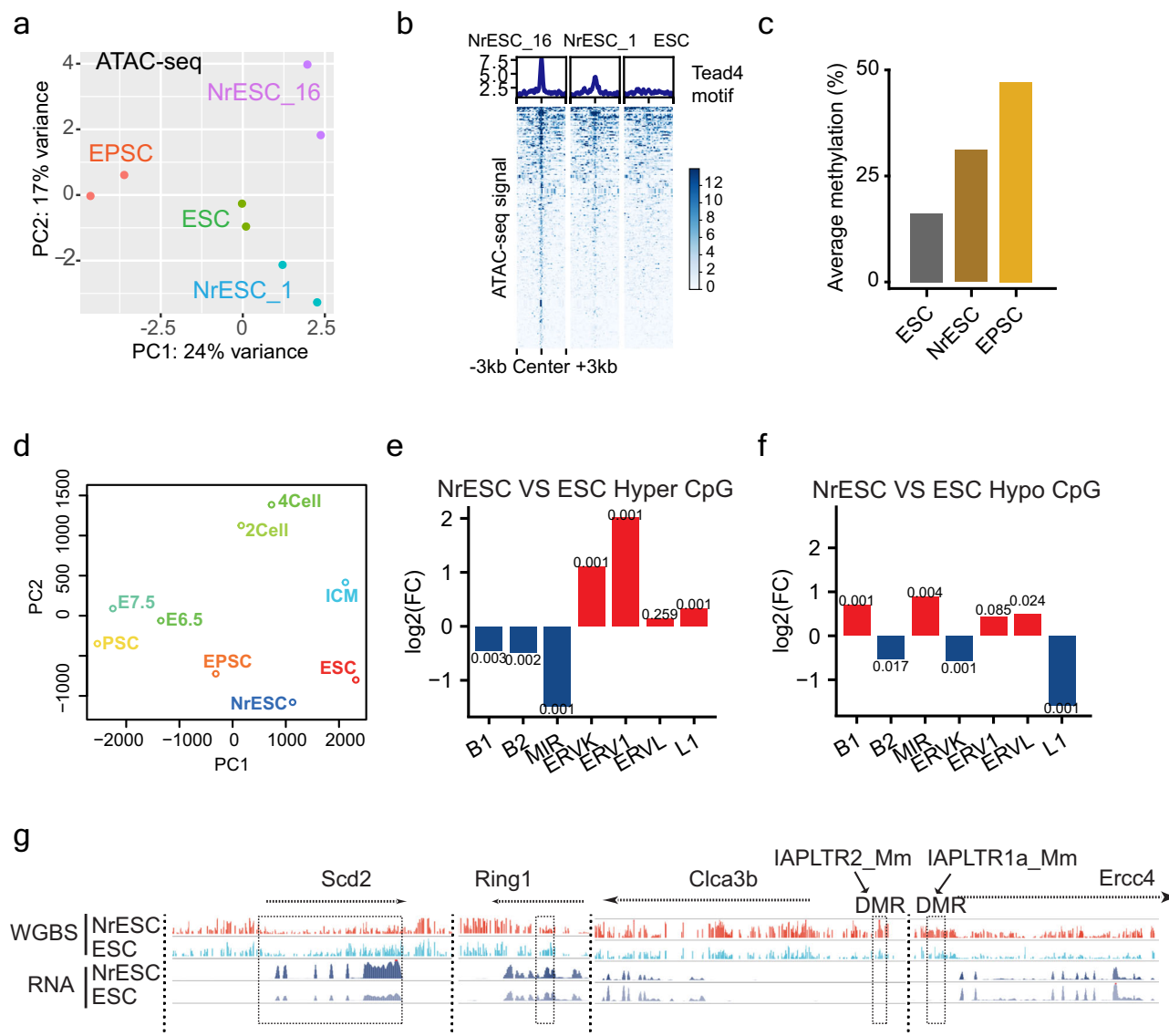


Fig. 3 | Global chromatin accessibility and DNA methylation changes in NrESC.

a PCA showing the chromatin accessibility divergence between ESC, EPSC, and NrESC at 1st and 16th passage. **b** Heatmap showing ATAC-seq signal intensity across upregulated accessible peaks with Tead4 motif in NrESC. **c** Histogram displaying the global CpG methylation levels of ESC, NrESC (Passage 16), and EPSC based on WGBS. **d** PCA analysis of ESC, NrESC (Passage 16), EPSC, mouse early embryos¹²⁹, and PSC in serum condition²³ based on the genome-wide methylomes. **e, f** Enrichment of genomic features at transposable element families on NrESC

differentially hypermethylated bases. Empirical *P*-values were calculated by a two-sided permutation test (**e**) and hypomethylated bases (**f**). The y-axis represents the log₂-enrichment for each genomic feature, and the value labelled on the bar represents the significance *p*-value, calculated with the Genomic Association Test (GAT) tool¹²⁸. **g** Genome browser view showing the WGBS and RNA-seq signal profiles for NrESC and ESC surrounding upregulated genes *Scd2* and *Ring1* with lower methylation signal in NrESC, and differentially methylated region (DMR)-enriched ERVKs annotated downregulated genes *Clca3b* and *Ercc4*.

accessibility in NrESC (Supplementary Fig. 3e). As a case in point, ATAC-seq signal of *Tead4* motif significantly increased over 1st to 16th passage of NrESC (Fig. 3b). Integrative analysis of ATAC-seq and H3K27ac ChIP-seq showed that 9242 H3K27ac peaks overlapped with ATAC peaks in NrESC (Passage 16), amongst which 206 peaks were upregulated ATAC peaks with higher H3K27ac signals (Supplementary Fig. 3f, g). Taken together, Nr1h2 activation alone is sufficient for ESC to attain a unique expanded pluripotency state capable of blastoid formation.

We next analyzed the global DNA methylation profiles of ESC, NrESC and EPSC using whole-genome bisulfite sequencing (WGBS). The global DNA methylation level in NrESC was higher than that in ESC but lower to that of EPSC (Fig. 3c, Supplementary Fig. 3h–k, Supplementary Data 3). The significant differentially hypermethylated (3923) and hypomethylated CpG sites (1878) were identified between NrESC

and ESC. The hypermethylated CpG sites were mainly detected in intergenic (62%) and intron (32%) regions of NrESC. Hypomethylated CpG sites were mostly detected in intron (47%) and intergenic (38%) regions (Supplementary Fig. 3l). Global DNA methylation level drops from 2-cell embryo stage to blastocyst ICM stage which has the lowest global methylation level before re-establishment of DNA methylation in post-implantation stages (E6.5 and E7.5)^{78,79}. ESC methylome under 2i condition in this study has low global methylation level, which is in proximity with ICM in the PCA plot. EPSC methylome is intermediate between ICM and post-implantation stage, which is consistent with the original report⁸⁰. The PCA and overall CpG methylation correlation heatmap plots showed that NrESC methylation status is in between ICM and E6.5/7.5. The DNA methylation level correlation value is higher between NrESC and E6.5/7.5 than 2-/4-cell stages, although NrESC is not as hypermethylated as E6.5/7.5 (Fig. 3d, Supplementary Fig. 3m).

NrESC showed intermediate average methylation profiles of promoters, exons and introns when compared to ICM (lowest) and E6.5/7.5 (highest) (Supplementary Fig. 3n).

Although we observed more downregulated genes with hypermethylation, there are moderate number of DEGs with significant changes in methylation level at the gene, promoter, exon and intron regions (Supplementary Fig. 3o). The methylation level of 2-cell markers *Zscan4d* and *Rarg* is higher in NrESC than ESC. The methylation level of pluripotency marker *Pou5f1* and NrESC marker *Scd1* is low in both NrESC and ESC compared to E6.5/7.5 (Supplementary Fig. 3p). These indicate that DNA methylation is not the direct effector of gene expression changes in NrESC.

Amongst the differentially hypermethylated CpG sites in NrESC compared to ESC, we have observed significant enrichment of multiple transposable elements such as ERVK, ERV1 and L1 (Fig. 3e). In contrast, B1 and MIR elements contained significantly higher levels of differentially hypomethylated CpG sites in NrESC (Fig. 3f). Hypomethylation at gene regions was detected in upregulated DEGs in NrESC, such as *Scd2* and *Ring1*. Hypermethylation was detected at ERVK families IAPLTR2_Mm and IAPLTR1a_Mm nearby *Clca3b* and *Ercc4* genes respectively, which are downregulated DEGs in NrESC (Fig. 3g). The differential gene methylation patterns observed support the notion that Nr1h2 activation regulates additional epigenetic feature in transposable elements through DNA methylation in NrESC. In summary, the integrative multi-omic analysis revealed the diverse regulatory role of Nr1h2 in rewiring the transcriptional and epigenetic landscape in ESC towards a distinct pluripotency state (NrESC) capable of blastoid formation.

Activation of Nr1h2 facilitates blastoid implantation and blastocyst development

We next characterized the transcriptome and assessed the in vivo development potential of NrESC-derived blastoids (Fig. 4a). NrESC-blastoid showed proper lineage segregation with a single outer layer of Cdx2⁺ TE-like cells, and Oct4⁺ ICM enclosed by a layer of Gata6⁺ PE-like cells (Fig. 4b). By integrating the scRNA transcriptome of published EPS-blastoid (Li et al. EPSblastoid), EPSC blastoid (current study), published E4.5 blastocyst and NrESC-blastoid, NrESC-blastoid displayed three major clusters (ICM/EPI-like, PE-like and TE-like) following the same annotation in EPS-blastoids, which largely overlapped with its blastocyst counterparts. Correlation coefficients and PCA showed that NrESC-blastoids transcriptome has overall higher correlation with type I EPS-blastoid and E4.5 blastocysts, compared to original Li et al. EPS-blastoids²¹ (Supplementary Fig. 4a, b). Of note, NrESC-blastoid cells do not contain intermediate cell clusters (Figs. 4c, S4c, Supplementary Data 4), unlike EPSC-derived blastoids (Supplementary Fig. 1h). NrESC-blastoid TE-like cells clustered more closely with blastocyst TE cells than EPSC-blastoid TE-like cells, which could be attributed to the expression of *Gata2*, a critical TE lineage specification factor, in the majority of TE-like cells in NrESC-blastoid but not EPSC-blastoid (Fig. 4d). A higher expression of *Gata2* (mural TE marker) at the mural region and *Cdx2* (polar TE marker) at the polar region was detected in NrESC-blastoids respectively (Supplementary Fig. 4d, e). This supports the presence of the polar-mural TE axis in NrESC-blastoids. Further analysis showed that most TE-like cells in NrESC-blastoid, but not EPSC-blastoid, express TE markers (*Tagln2*, *Krt8*, *Krt18*, *Gata2*) that are crucial in TE maturation and blastocyst implantation (Fig. 4e), suggesting that Nr1h2 activation induces more faithful TE differentiation. In addition, *Krt8* expression was exclusively detected in NrESC but not in ESC nor EPSC (Supplementary Fig. 4f).

In line with these observations, we have assessed the implantation potential of NrESC-blastoids by transferring them to surrogate females at 3.5 days post coitum (dpc). At 10.5 dpc, about 11% and 3% of transferred NrESC-blastoids and EPSC-blastoids, respectively, implanted and induced decidualization (Fig. 4f, g). The implantation rate of

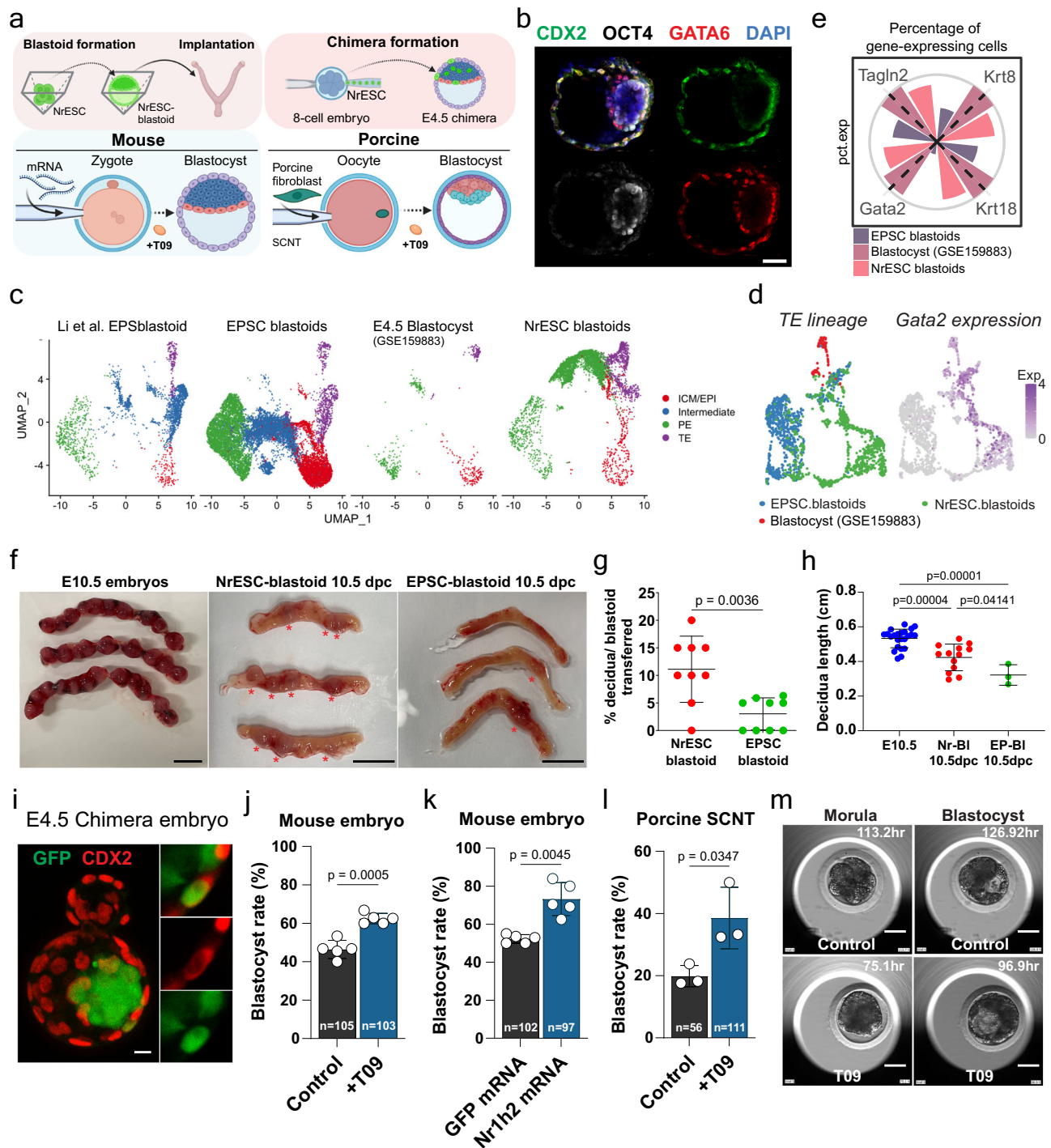
NrESC-blastoids is higher than the reported 7% implantation rate of EPSC-blastoids that were harvested at an earlier time point 7.5 dpc²¹. Moreover, the majority of NrESC-blastoid decidia were significantly larger than those induced by EPS-blastoid and were within the range of natural E10.5 decidia control (Fig. 4h). In addition, when GFP-labelled NrESCs were microinjected into 8-cell embryos, they contributed to both ICM and trophectoderm in about 10% of the chimeric blastocysts (Fig. 4i, Supplementary Fig. S4g, h). These showed that Nr1h2 activation alone is sufficient to capacitate expanded pluripotency in NrESC to form high quality blastoids with vastly improved implantation potential and to induce more faithful TE differentiation.

We have also tested the effect of Nr1h2 activation in blastocyst development by supplementing embryo culture media with T09. Mouse blastocyst rate was significantly improved from 48% to 62% when T09 treatment was maintained from zygote or 8-cell embryo to blastocyst stage (Fig. 4j, Supplementary Fig. 4i, j). This corroborates *Nr1h2* upregulation at 8-cell stage during blastocyst development. Similar result was achieved when *Nr1h2* mRNA was microinjected into 1-cell embryo, mouse blastocyst rate was significantly improved from 52% to 73% (Fig. 4k, Supplementary Fig. 4k). Importantly, porcine somatic cell nuclear transfer (SCNT) blastocyst rate nearly doubled from 19.9% to 38.6% upon T09 treatment post-SCNT (Fig. 4l). Morphokinetics of T09-treated SCNT embryos showed that T09 accelerated SCNT blastocyst maturation by about 30 h earlier than control (Fig. 4m, Supplementary Fig. 4l, m). These collectively demonstrated the importance of Nr1h2 activation in pre-implantation embryo development conserved across species.

Nr1h2-specific transcriptional regulation via coactivator complex and SINE-B1 element

We sought to understand the molecular mechanism of Nr1h2 activation regarding its multifaceted roles in blastoid and blastocyst formation. We first confirmed the specificity of ligand activated Nr1h2 in conventional ESC by transient Nr1h2-FLAG overexpression for 72 h, along with T09 activation. This approach allowed ESC to acquire blastoid formation ability with similar efficiency as NrESC (~7.5%) without requiring prolonged culture or extensive passages. Conversely, neither EGFP-FLAG control nor C-terminus truncated Nr1h2 mutant overexpressing ESC (Δ LBD-Nr1h2-FLAG), which lacks the ligand binding domain^{81,82}, yielded any increase in blastoid formation efficiency under T09 treatment (Fig. 5a, b, Supplementary Fig. 5a, b).

We next performed immunoprecipitation pulldown of Nr1h2-FLAG using an anti-FLAG antibody and subsequent mass spectrometry to identify Nr1h2 protein partners with or without T09. EGFP-FLAG was used as control. Proteins detected from triplicate runs of MS/MS were filtered by inclusion criteria of detection in at least one run, fold change ≥ 2 and p -value ≤ 0.05 , which resulted in total 801 and 66 proteins enriched with Nr1h2-FLAG immunoprecipitation with and without T09 respectively (Supplementary Fig. 5c). Of note, Nr1h2 and its known heterodimer partners Rxra and Rxrb were identified among the top 15 ranked hits (Supplementary Data 5). Interestingly, T09-mediated activation of Nr1h2 increased its binding with many coactivator (Ncoa1, Ncoa2, Ncoa3, Rbbp7, Kdm1a, etc) and corepressor (Nrip1, Hdac1, Dnmt1, Trim28, Setdb1, etc) complex proteins (Fig. 5c, Supplementary Data 5). Protein-protein interaction network showed that active Nr1h2 transcriptional interactome proteins are involved in negative regulation of gene expression/ epigenetic/ negative regulation of chromatin organization (Hdac1, Trim28, Dnmt1, Setdb1, Smarcd1), DNA-methylation dependent heterochromatin formation (Kdm1a, Ssrp1), *in utero* development, etc (Supplementary Fig. 5d). Knockdown of several coactivator complex proteins (Ncoa3, Rxra, Rxrb) led to exclusive reduction of blastoid formation efficiency in NrESC but not in EPSC (Fig. 5d). These suggest a dual role of Nr1h2 in partnering with both coactivator and corepressor complex for broad regulation of the pluripotent transcriptome and epigenome landscape.



Next, we performed FLAG ChIP-seq to identify binding regions of Nr1h2-FLAG with or without T09 treatment. 23,826 and 28,317 peaks were detected in Nr1h2-FLAG and Nr1h2-FLAG with T09 sample respectively, indicating an increase in overall Nr1h2 binding sites upon T09 treatment. In contrast, only 1345 and 2716 peaks were detected in control EGFP-FLAG and EGFP-FLAG with T09 samples, respectively, indicating a high signal-to-noise ratio of the FLAG ChIP-seq dataset. Consistently, the detected peaks of Nr1h2 ChIP-seq were largely captured in the Nr1h2-FLAG ChIP-seq dataset (Supplementary Fig. 5e), indicating high specificity of the FLAG-tagged Nr1h2. Differential Nr1h2-FLAG ChIP peaks with T09 treatment were analyzed to understand the changes in Nr1h2 binding upon activation (Supplementary Data 6).

In line with the differential DNA methylation pattern observed in transposable elements, Nr1h2 binding on B1 regions were significantly enriched with T09 treatment. (Supplementary Fig. 6a). H3K27Ac ChIP-seq showed that 1,283 and 4,663 H3K27Ac peaks at B1 elements were upregulated in T09-treated Nr1h2-FLAG ESC and NrESC respectively (Supplementary Fig. 6b). Among 1491 annotated genes nearby B1 elements with upregulated H3K27Ac signals, 158 genes were upregulated DEGs (Supplementary Fig. 6c). Consistently, genes within 20 kb distance to Nr1h2-bound SINE-B1 elements showed increased Nr1h2 binding, H3K27Ac marks, and upregulated gene expression than those with >20 kb distance (Supplementary Fig. 6d). 634 peaks in B1 regions overlapped between upregulated Nr1h2-FLAG ChIP peaks and H3K27Ac ChIP peaks in NrESC (Supplementary Fig. 6e).

Fig. 4 | Nr1h2 activation enhances blastoid implantation and multi-species blastocyst development. **a** In vivo approaches depicted to examine Nr1h2 activation in blastoid implantation and blastocyst development. Created in BioRender.com. **b** Immunofluorescence for markers of the TE (Cdx2), ICM/EPI (Oct4) and PE (Gata6) in NrESC blastoid derived from 16th passage NrESC. Scale bar, 50 μ m. 5 independent experiments were repeated with similar results. **c** A Umap plot showing the clustering of cells from Li et al. EPSCblastoid, EPSC blastoids (current study), E4.5 blastocyst (GSE159883), and NrESC blastoids single-cell transcriptome. **d** Umap plot showing *Gata2* expression in clustered TE lineage cells from (c). **e** Polar bar chart displaying the percentage of Krt8/Krt18/*Gata2*/*Tagln2* expressing cells in TE lineage for different blastoids or blastocysts. **f** Representative brightfield images showing the decidual formation in three mouse uteri at E10.5, as well as 7 days after NrESC-blastoid or EPSC-blastoid transfer at 3.5dpc. Red asterisks indicate deciduae. Scale bar, 1 cm. **g** Quantification of the decidual formation efficiency from the NrESC- or EPSC-derived blastoids at 10.5dpc. Two-tailed Welch's *t*-test. Data are represented as mean \pm s.d.; *n* = 9 independent assays. **h** Quantification of the decidual length induced by E10.5 natural embryo, NrESC- or EPSC-derived

blastoid 10.5dpc. One-way ANOVA with Bonferroni correction for multiple comparisons. Data are represented as mean \pm s.d.; *n* = 23 E10.5 natural embryo, *n* = 13 NrESC-derived blastoid 10.5dpc, *n* = 3 EPSC-derived blastoid 10.5dpc. **i** Left, maximum projection image of immunofluorescence for GFP and TE marker (Cdx2) in chimeric E4.5 blastocysts microinjected with psinGFP NrESC. Right, single confocal plane insets showing Cdx2-GFP double positive cell and GFP single positive cells. Scale bar, 20 μ m. Quantification of the developmental rate of mouse blastocysts after supplementing T09 in culture media from one-cell stage (**j**) or microinjection of *GFP* or *Nr1h2* mRNA into one-cell embryo (**k**). Two-tailed Welch's *t*-test. Data are represented as mean \pm s.d.; *n* = 5 independent assays. **l** Quantification of the developmental rate of porcine SCNT blastocysts with or without T09 in culture media from one-cell stage. Two-tailed Welch's *t*-test. Data are represented as mean \pm s.d.; *n* = 3 independent assays. **m** Representative images of porcine SCNT morula and blastocyst in T09-treated and control group post-SCNT. Scale bar, 50 μ m. 3 independent experiments were repeated with similar results. Source data are provided as a Source Data file.

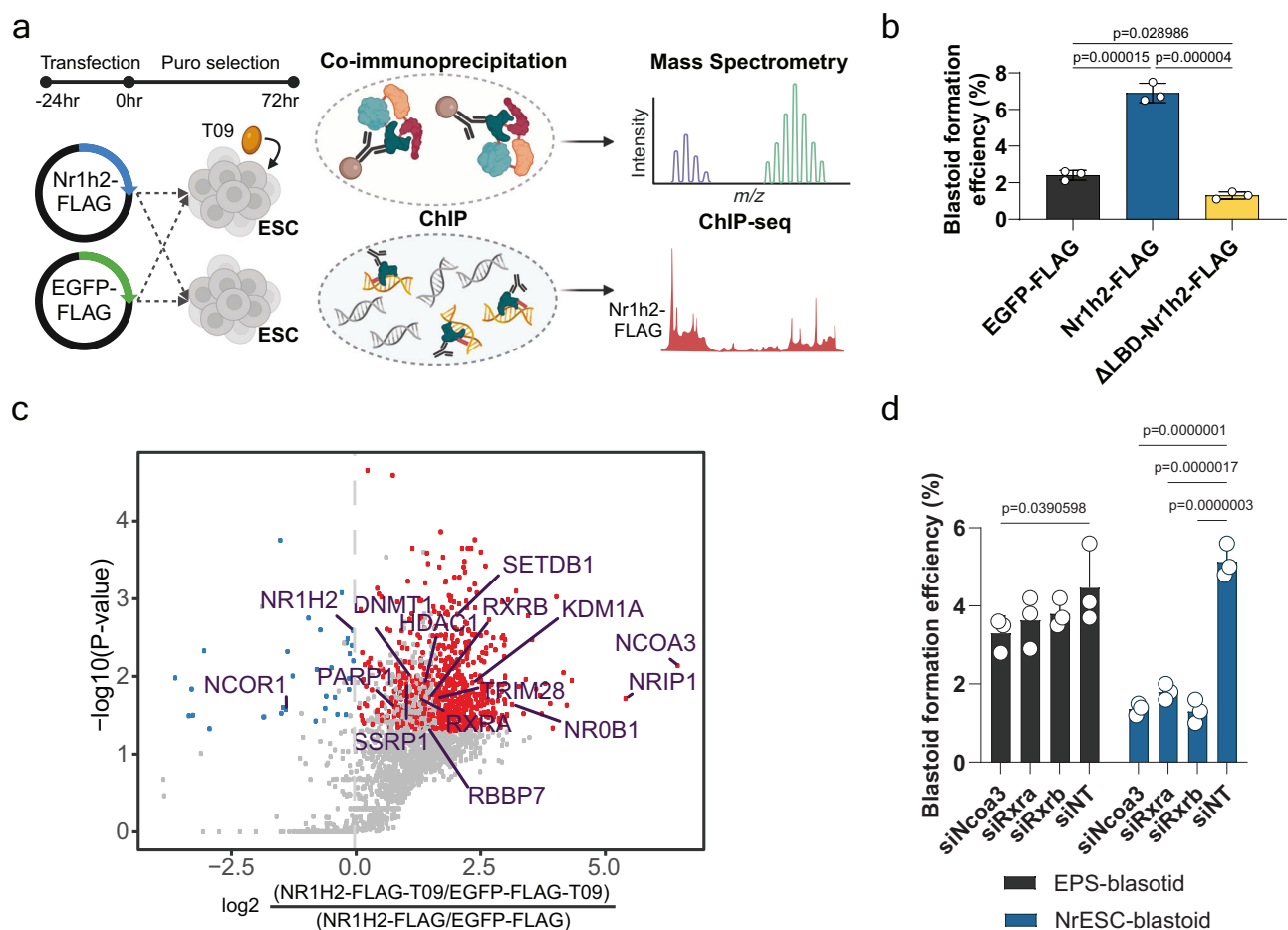
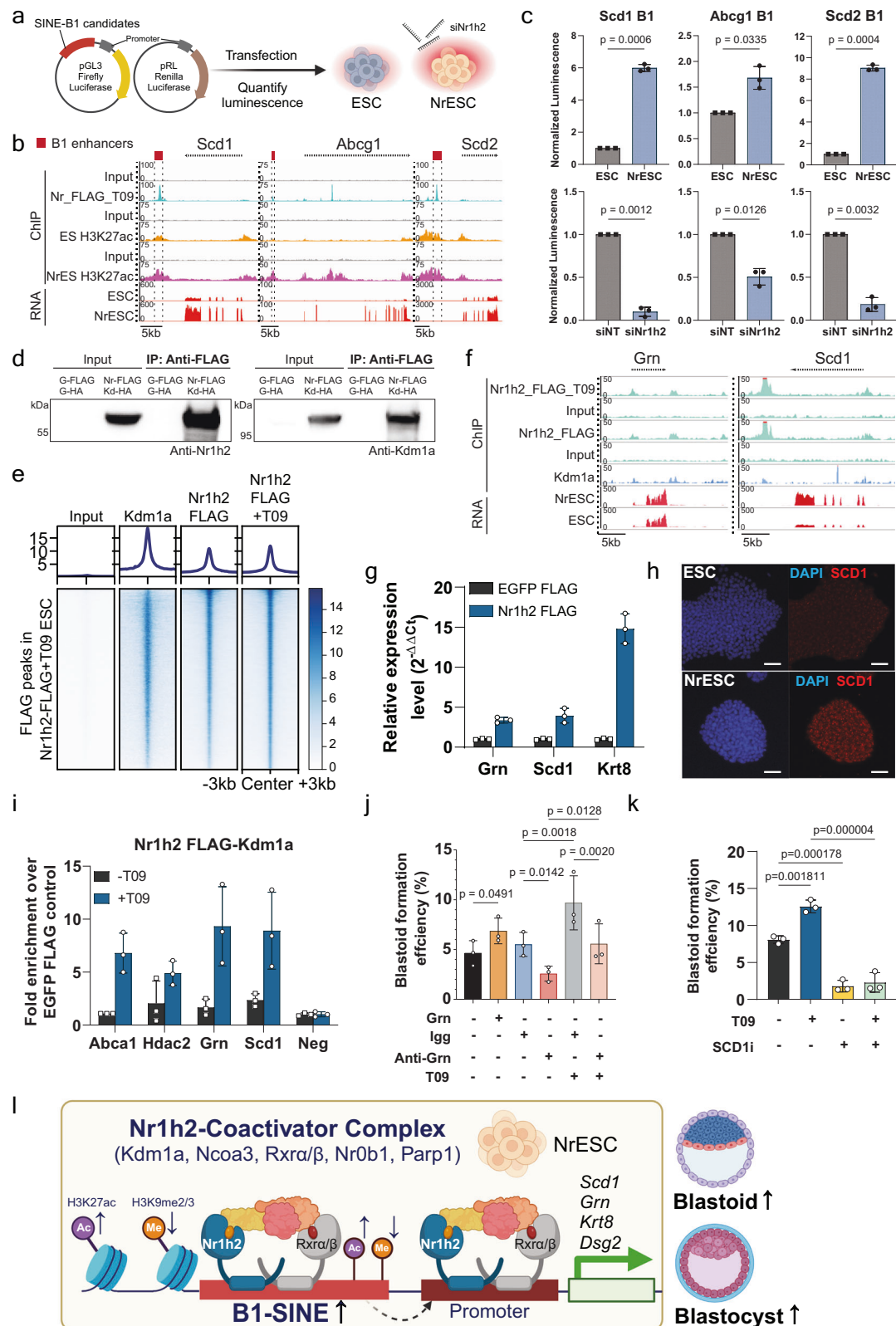


Fig. 5 | Active Nr1h2 partners with coactivator proteins to regulate blastoid formation. **a** A schematic diagram showing transient overexpression of Nr1h2-FLAG in ESC with T09 treatment, followed by pull-down mass spectrometry and ChIP-seq of FLAG-tagged Nr1h2. Created in BioRender.com. **b** Quantification of blastoid formation efficiency harvested at Day 5 of differentiation from ESC over-expressing EGFP-FLAG, Nr1h2-FLAG or ΔLBD-Nr1h2-FLAG under T09 treatment. One-way ANOVA with Bonferroni correction for multiple comparisons. Data are represented as mean \pm s.d.; *n* = 3 independent assays. **c** Volcano plot summarizing

the quantitative results of the proteins enriched significantly in Nr1h2-FLAG mass spectrometry data in ESC with T09 from three replicates. The specific proteins interacting with Nr1h2 in ESC with T09 were colored red (Two-sided *t*-test; *p*-value < 0.05 & FC > 2). **d** Quantification of the blastoid efficiency after individual siRNA treatment in EPSC-blastoid or NrESC-blastoid formation. Two-way ANOVA with Bonferroni correction for multiple comparisons. Data are represented as mean \pm s.d.; *n* = 3 independent assays. Source data are provided as a Source Data file.

We hypothesize that Nr1h2 activation regulates downstream gene expression via B1 elements as enhancers. We conducted a luciferase assay to validate whether T09-activated Nr1h2 regulates B1 enhancer activity which in turn upregulate downstream gene expression. The pGL3 luciferase reporter vector with SCP1 promoter (pGL3-SCP1) was

utilized to evaluate enhancer activities of various inserts during Nr1h2 activation (NrESC vs ESC) and Nr1h2 knockdown in NrESC (Fig. 6a). Candidate Nr1h2-B1 enhancer regions were selected based on three criteria: increased Nr1h2 binding, increased H3K27ac histone activation mark, and proximity (~20 kb) to upregulated genes in NrESC.



Approximately 2kbp candidate regions nearby target genes, including *Scd1*, *Abcg1*, and *Scd2*, were cloned into the pGL3-SCPI vectors (Fig. 6b). The results showed that these three top-ranked B1 enhancer candidates exhibited a significant increase in luciferase activity in NrESC compared to ESC (Fig. 6c, top). The luciferase activity under these candidates was reduced significantly upon Nr1h2 knockdown in NrESC, indicating the

specificity of Nr1h2 interaction with target B1 enhancers (Fig. 6c, bottom). These findings indicate that increased Nr1h2 binding to specific B1 sites boost their enhancer activities, thereby upregulating nearby Nr1h2 target gene expression in NrESC. Taken together, our data support the hypothesis that Nr1h2-B1 regulatory axis regulates downstream target genes by modulating B1 enhancer activity.

Fig. 6 | Molecular mechanism of the activation of Nr1h2 intrinsic program. **a** A dual-Glo luciferase assay depicted to test SINE-B1 enhancer activity in ESC and NrESC, as well as siNr1h2 and siNT NrESC samples. Created in BioRender.com. **b** Genome browser view of Nr1h2-FLAG ChIP-seq (turquoise), H3K27ac ChIP-seq (orange, ESC; purple, NrESC) and RNA-seq signal profile (red) surrounding NrESC up-DEGs (*Scd1*, *Abcg1* and *Scd2*). Red-colored boxes and dashed lines indicate the candidate SINE-B1 enhancer regions being tested. **c** Quantification of the normalized luminescence (normalized to Renilla signal and ESC signal) in three constructs testing for SINE-B1 candidates nearby *Scd1*, *Abcg1* and *Scd2* in ESC and NrESC (top), or siNT and siNr1h2 NrESC (bottom). Two-tailed Welch's *t*-test. Data are represented as mean \pm s.d.; *n* = 3 independent assays. **d** Western blot analysis showing that Kdm1a is only associated with Nr1h2-FLAG in T09-treated Nr1h2-FLAG and Kdm1a-HA overexpressing sample (Nr-FLAG and Kd-HA), compared to GFP control sample (G-FLAG and G-HA). 3 independent experiments were repeated with similar results. **e** Heatmap of ChIP-seq signal on Nr1h2-FLAG ChIP-seq peaks. The ChIP-seq of Kdm1a¹³³ was obtained from the GEO database. **f** Genome browser view of Nr1h2-FLAG ChIP-seq (turquoise), Kdm1a ChIP-seq (blue) and RNA-seq signal profile (red) surrounding NrESC markers (*Grn* and *Scd1*). **g** Expression of *Grn*, *Scd1* and *Krt8* in

ESC overexpressing EGFP-FLAG or Nr1h2-FLAG under T09 treatment. Expression is relative to *Gapdh* and normalized to EGFP-FLAG ESC expression. Data are represented as mean \pm s.d.; *n* = 3 technical replicates. **h** Immunofluorescence of *Scd1* in ESC and NrESC. Scale bar, 20 μ m. 3 independent experiments were repeated with similar results. **i** Sequential ChIP-qPCR indicating the increased co-binding of Kdm1a and Nr1h2-FLAG under T09 treatment at *Abca1*, *Hdac2*, *Grn* and *Scd1*, normalized to EGFP-FLAG control. Regions not bound by Kdm1a and Nr1h2-FLAG were used as negative control. Data are represented as mean \pm s.d.; *n* = 3 technical replicates. **j** Quantification of blastoid formation efficiency from NrESC treated with recombinant Grn, Igg, Anti-Grn, Igg + T09, or Anti-Grn + T09. One-way ANOVA with Bonferroni correction for multiple comparisons. Data are represented as mean \pm s.d.; *n* = 3 independent assays. **k** Quantification of blastoid formation efficiency from NrESC treated with T09 and/or *Scd1* inhibitor. One-way ANOVA with Bonferroni correction for multiple comparisons. Data are represented as mean \pm s.d.; *n* = 3 independent assays. **l** Model schematic showing multifaceted regulatory roles of Nr1h2-specific intrinsic program in NrESC. Created in BioRender.com. Source data are provided as a Source Data file.

On the other hand, RT-qPCR showed that B1 transcript expression level increased by ~2.5-fold in NrESC (Supplementary Fig. 6f). The regulation of B1 transcription by Nr1h2 was further validated in vivo by microinjecting *Nr1h2* mRNA into fertilized one-cell embryos, which showed ~2.5-fold increase of B1 transcript expression level at 8-cell embryo stage, compared to *EGFP* mRNA as control (Supplementary Fig. 6g). Unlike the signature peak ERVL transcript level at two-cell stage, B1 transcript level showed an increasing trend from zygote to blastocyst stage while another transposable element, ERV1, did not show major upregulation during preimplantation development (Supplementary Fig. 6h). These suggest Nr1h2 also induces B1 transcript upregulation.

To explore the gene regulatory relationship between Nr1h2 and partner proteins, we first validated the Nr1h2-Kdm1a protein interaction by FLAG co-immunoprecipitation, where Kdm1a was detected only in T09-treated Nr1h2-FLAG overexpressing sample but not in GFP control (Fig. 6d). Consistently, 10,211 peaks in Nr1h2-FLAG and 11,348 peaks in T09-treated Nr1h2-FLAG overlapped with Kdm1a ChIP-seq public dataset (Fig. 6e, Supplementary Data 6). Around 70% of Nr1h2 and Kdm1a overlapping regions are located in the gene promoter regions, and 16% of peaks are in distal intergenic regions (Supplementary Fig. 6i). We assessed whether Nr1h2 interacts with histone demethylase Kdm1a^{83,84} to demethylate target gene regions. Average H3K9me3 level was decreased at Nr1h2-bound sites in T09-treated Nr1h2-FLAG ESC (Supplementary Fig. 6j). Specifically, 3,523 Nr1h2-Kdm1a co-binding sites showed lower H3K9me3 level in T09-treated Nr1h2-FLAG ESC compared to GFP-FLAG control (Supplementary Fig. 6k). For example, *Dsg2* (NrESC marker) gene region is co-bound by Nr1h2 and Kdm1a, accompanied by decreased H3K9me3 level in T09-treated Nr1h2-FLAG ESC (Supplementary Fig. 6l). We also assessed the level of H3K9me2, which is specifically demethylated by Kdm1a, at sites with increased Nr1h2-Kdm1a co-binding in NrESC by performing H3K9me2 ChIP-qPCR. Regions near upregulated DEGs in NrESC, such as *Klhl24*, *Pcmt1d*, *Casp8*, *Pcsk9* and *Spats2*, displayed lower H3K9me2 level in NrESC compared to ESC control (Supplementary Fig. 6m).

We analyzed the Nr1h2-Kdm1a co-binding peaks in NrESC DEGs and showed that T09 increased Nr1h2 binding to the same promoter/enhancer regions where Kdm1a binds at NrESC markers such as *Grn* and *Scd1* (Fig. 6f). Both genes were upregulated under Nr1h2-FLAG overexpression with T09 (Fig. 6g), and downregulated upon Nr1h2 knockdown (Supplementary Fig. 6n). *Scd1* expression is upregulated in NrESC (Fig. 6h), mouse embryos microinjected with *Nr1h2* mRNA, and porcine SCNT embryos treated with T09 (Supplementary Fig. 6o, p). Importantly, the role of Kdm1a as a coactivator was illustrated by sequential ChIP of Nr1h2-FLAG ChIP followed by Kdm1a-ChIP. Upon T09 treatment, Kdm1a showed increased binding to the Nr1h2-Kdm1a

co-bound regions at *Abca1*, *Hdac2*, *Grn* and *Scd1* (Fig. 6i). Furthermore, exogenous addition of recombinant Grn protein significantly improved NrESC blastoid formation efficiency. Conversely, anti-Grn antibody treatment significantly reduced the blastoid formed. The beneficial effect of T09 treatment boosting blastoid formation efficiency was offset by anti-Grn antibody treatment (Fig. 6j, Supplementary Fig. 6q). On the other hand, *Scd1* inhibitor significantly abolished blastoid formation even with T09 treatment (Fig. 6k, Supplementary Fig. 6r). These indicate that Nr1h2 recruits Kdm1a and possibly other coactivator proteins to remove repressive H3K9me2/3 histone marks, which leads to upregulation of functional downstream targets, such as *Grn* and *Scd1*, facilitating blastoid formation.

Discussion

Intertwining intrinsic and extrinsic pathways facilitate cells to respond and feedback to extrinsic inductions during cell fate specification. For example, EPI/PE segregation involves non-cell autonomous Fgf4 signaling, stimulated by *Oct4*, *Sox2*, and *Nanog* in EPI cells, to induce PE specification (*Gata6* expression) in neighboring cells and repress *Nanog* expression in EPI cells^{7,85–88}. This segregation is further reinforced by the cell-autonomous reciprocal transcriptional repression and self-activation of *Nanog* and *Gata6* in EPI and PE cells^{10,85,89–91}.

Various recipes of exogenous cues in mouse blastoid media combined signaling ligands (such as Fgf4, Tgf β 1, Bmp4) and activators/inhibitors of signaling pathways (such as Wnt activator CHIR99201 and TGF β inhibitor A83-01) that were presumed to induce pre-implantation embryonic lineage specification in blastoids^{19–24}. Nonetheless, blastoids contain cells expressing post-implantation markers; for example, mesoderm marker Brachyury is detected in TE-like cluster. This correlates well with the use of Bmp4 and Tgf β in 2D differentiation studies to model post-implantation embryonic lineages such as mesoderm and parietal endoderm^{92–95}.

We classified different types of blastoids formed from the same pool of EPSC based on morphology, lineage segregation and single-cell transcriptome. ICM/EPI-like, PE-like and TE-like lineages are present in blastoids (type I and II) and type III organoids, which manifest the expanded potential of EPSC to differentiate into all three embryonic lineages regardless of blastoid grade. Hence, exogenous cues in media do not guarantee proper self-organization of stem cells during blastoid development. Single-cell transcriptome analysis reveals lineage-specific DEGs and differential active regulons enriched in functional terms related to lineage specification and blastocyst development. The blastoid classification presented in our current study thus provides more clarity, enabling exploration of the heterogeneous response during blastoid induction. We hypothesize that intrinsic regulations during the dynamic process of differentiation and self-organization are

crucial for blastoid formation. Regulon analysis has revealed many active master transcriptional regulators in all three embryonic lineages, albeit with significant morphological and transcriptional differences among blastoids (type I and II) and type III organoids. This indicates that additional active regulons are essential to guide stem cells through the dynamic differentiation process and facilitate their self-organization into type I blastoids.

We then sought out exclusive active regulons detected in type I blastoids and blastocysts. To this end, we uncovered the essential role of *Nr1h2* in blastoid formation. *Nr1h2*, or LXR β (Liver X receptor β), is a transcription factor belonging to nuclear receptors superfamily. Its natural ligands include oxysterols, oxidized cholesterol. Synthetic agonist T09 activates *Nr1h2* by dimerizing it with retinoid X receptors (RXR), which then bind on to target gene locus on LXR-responsive elements (LXREs)⁹⁶. *Nr1h2* regulates genes involved in cholesterol transport and metabolism in liver and other tissues, and controls transcriptional landscape by regulating chromatin accessibility. It is also a lineage-determining factor for Kupffer cells⁷¹. In our study, ~25% downregulated genes upon *Nr1h2* knockdown during mouse blastocyst development are enriched in early embryo signature markers. The improved blastocyst rate with T09 treatment corroborates the functional significance in embryo development, from the molecular changes induced by *Nr1h2* activation.

In addition to the characterization of the *Nr1h2* regulon in blastoids, we discovered that *Nr1h2* activation rewires ESC into a unique pluripotency state. NrESC requires only three passages in culture to acquire expanded pluripotency, compared to other blastoid-competent cell types such as EPSC, totipotent blastomere-like cell (TBLC) or totipotent potential stem (TPS) cell that require maintenance for at least five passages in their respective culture media^{23,24}. In particular, NrESC is established through treatment with a single gene-specific agonist, contrasting with the use of multiple non-gene specific signaling activators and inhibitors in EPSC and TPS cell media, or pladienolide B-induced widespread splicing inhibition in TBLC²³. NrESC shows no alteration of pluripotent genes *Oct4*, *Sox2* and *Nanog*, which are downregulated in TBLC and TPS cell resembling 2CLC state^{23,24}. Furthermore, NrESC retains similar proliferation rate as ESC, which was markedly reduced in TBLC²³ and EPSC⁸⁰. No observable morphological differences were found between NrESC and ESC, in contrast to the flattened EPSC colonies. NrESC displays a unique transcriptional landscape that is different from 2CLC or EPSC signatures, which mimics 2-cell or 4- to 8-cell embryo signature, respectively^{74,75}. Importantly, this distinct pluripotency state in NrESC is functionally competent in self-organization to form blastoids in vitro and contribution to both embryonic and extraembryonic lineage in vivo.

Integrative analysis of transcriptome, DNA methylation, chromatin accessibility and epigenome^{97,98} revealed genome-wide molecular consequences upon *Nr1h2* activation. *Nr1h2* is well-known to activate target gene transcription by binding to their DR4 motifs. Our data showed that *Nr1h2* in ESC binds to and upregulates classical target genes in lipid metabolism (*Abca1*, *Fasn*, *Srebf1*, and *Scd1*)^{99–101}, which is in line with a recent study reporting the importance of lipid homeostasis during peri-implantation development¹⁰². *SCD1* inhibition was reported to robustly eliminate human pluripotent stem cells as well as impeding mouse blastocyst development, while supplementing oleic acid, a monounsaturated fatty acid synthesized by *Scd1*, improves blastocyst development⁶⁹. These findings align with our observation that *Scd1* inhibition drastically reduces blastoid formation even with T09 treatment. Unexpectedly, active *Nr1h2* upregulates multiple genes important for lineage segregation (*Bmp4*), trophectoderm architecture (*Dsg2* and *Krt8*), and blastocyst development (*Grn*). *Nr1h2* activation also alters the epigenetic landscape by adding H3K27Ac enhancer marks and increasing chromatin accessibility at genes specific to EPI (*Pou5f1*, *Sox2*), PE (*Gata4*) and TE (*Krt8*, *Tagln2*, *Tead4*) lineages. These findings indicate that *Nr1h2* regulates genes vital for

stem cell pluripotency and the development of both blastoids and blastocysts.

While the above integrative analyses provide a valuable overview of native NrESC expanded pluripotency, we adopted an exogenous overexpression system (*Nr1h2*-FLAG) with robust performance in capturing the *Nr1h2* protein and its partners, thereby facilitating subsequent mechanistic studies of the *Nr1h2* protein. *Nr1h2*, upon ligand binding, interacts with multiple coactivator proteins and epigenetic regulators such as *Ncoa3* and *Kdm1a* to upregulate target genes expression. *Ncoa3* is well known for its essential coactivator role in mediating the functions of another nuclear receptor, *Esrrb*, in the pluripotency gene regulatory network^{103,104}. Our findings suggest additional roles of *Ncoa3* in the *Nr1h2* regulatory network in mouse-expanded pluripotency and blastoid formation. On the other hand, *Nr1h2* activation recruits *Kdm1a* to target gene loci, especially in functional target genes *Grn* and *Scd1*. The *Nr1h2*-*Kdm1a* partnership also modulates di-/tri-methylated histone H3K9 landscape in NrESC, where *Kdm1a* functions as a co-activator by demethylating target gene regions.

We have discovered a regulatory mode of *Nr1h2* in expanded pluripotency via transposable elements. *Nr1h2* activation leads to increased *Nr1h2* binding, DNA demethylation and H3K27Ac enhancer marks at specific SINE-B1 sites. These result in increased SINE-B1 enhancer activities which in turn upregulate nearby gene expression. Interestingly, active *Nr1h2* upregulates SINE-B1 transcript level in both NrESC and mouse 8-cell embryos. This *Nr1h2*-SINE-B1 cis-regulatory axis aligns with a recent study reporting another nuclear receptor, *Nr5a2*, as an essential pioneering factor broadly regulating ZGA (zygotic genome activation) genes through binding onto SINE-B1 elements¹⁰⁵. These collectively elucidate the multifaceted regulatory roles of *Nr1h2* in capacitating ESC to self-organize into blastoids (Fig. 6). Further studies on *Nr1h2* potential regulations in other DNA and histone modifications^{106,107}, genome topology^{108–110} as well as epitranscriptomic landscape¹¹¹ would provide additional insights in NrESC and its developmental potential.

In species such as porcine, which lack similar elements to mice's SINE-B1, we speculate that *Nr1h2* regulates the transcriptome mainly through direct binding to target gene promoter and enhancer regions. Nonetheless, a recent study discovered a full-length PREO-SS, a member of SINE families, in the 3'-UTR of the pig *PDK1* gene, which aligns with previous reports that Alu and B1 regulate *PDK1* orthologs¹¹². Future mechanistic studies of *Nr1h2* relationship with PREO-SS in transcriptional regulation are thus required.

We hypothesize that the *Nr1h2*-specific intrinsic program improves blastoid quality. Single-cell transcriptome analysis showed absence of intermediate cell types in NrESC-blastoids. This could be attributed to the gene-specific activation in NrESC compared to multiple broad-range signaling manipulation in EPSC. NrESC-blastoid TE-like cells clustered more closely with blastocyst TE compared to EPS-blastoid TE-like cells. The majority of NrESC-blastoid TE-like cells express classical TE markers such as *Gata2* at a level similar in blastocyst TE. This aligns with a recent review study showing low percentage of TE-like cells in EPS-blastoid without proper expression of TE markers²⁵. Of note, most NrESC-blastoid TE-like cells express *Tagln2* which is essential for blastocyst implantation¹¹³. This correlates well with the higher implantation efficiency of NrESC-blastoids into uterus, suggesting NrESC gives rise to more faithfully differentiated TE-like cells. *Nr1h2* activation also benefits both mouse and porcine SCNT blastocyst development with improved blastocyst rate and morphokinetics. These demonstrate the pivotal role of *Nr1h2* regulatory program conserved in both stem cell-based embryos and embryos across species. Further exploration of additional intrinsic programs will benefit the current rapid development and application of stem cell-based embryo models.

Methods

Cell culture

Experiments were done using the following cell lines; mouse ESC lines: E14 ESC (ES-E14TG2a, ATCC, CRL-1821), E14 ESC transduced with pSIN-GFP virus. ESCs were cultured on gelatin-coated plates in ESC medium with DMEM basal medium (Thermo Fisher Scientific, 11330-032) supplemented with 15% ESC-screened FBS (Thermo Fisher Scientific, 16141-079), 1X NEAA (Thermo Fisher Scientific, 11140-050), 1X L-glutamine (Thermo Fisher Scientific, 25030-081), 0.1 mM 2-mercaptoethanol (Thermo Fisher Scientific, 21985-023), 10^3 units/mL recombinant mouse LIF (Sigma-Aldrich, ESG1107), 3 μ M CHIR99021 (Tocris, 4953) and 1 μ M PD0325901 (Tocris, 4192). The cells were incubated at 37 °C with 5% CO₂ and replaced with fresh media every 24 h. Mouse ESC were converted to EPSC through culturing in EPSC medium for 5 passages in gelatin-coated 6-well plates as previously reported⁸⁰. EPSC medium is made up of DMEM/F-12 basal media supplemented with 20% knockout serum replacement (Thermo Fisher Scientific, 10828-033), 1X L-glutamine, 1X NEAA, 0.1 mM 2-mercaptoethanol, 10^3 units/mL recombinant mouse LIF, 3 μ M CHIR99021, 1 μ M PD0325901, 4 μ M JNK Inhibitor VIII (Tocris, 3222), 10 μ M SB203580 (Tocris, 1402), 0.3 μ M A-419259 (Santa Cruz, SC-361094) and 5 μ M XAV939 (Tocris, 3748). Mouse ESC were converted into NrESC through culturing in NrESC medium for 3 passages in gelatin-coated 6-well plates. NrESC medium is composed of ESC medium supplemented with 10 μ M T0901317 (MedChemExpress, HY-10626). Cells were routinely passaged every two to three days. All cell cultures have been routinely tested negative for mycoplasma.

Blastoid formation

Blastoid formation was based on established protocol²¹, which was performed on 6-well AggreWell 400 plates (STEMCELL Technologies, 34425). On Day 5, blastoids were manually picked with a Cook Flexipet pipette under a stereomicroscope for analysis. For blastoid experiments involving Nr1h2 agonist treatment, 10 μ M T0901317 (MedChem Express, HY-10626) was added to the complete blastoid medium. For blastoid experiments involving recombinant Grn protein or Grn neutralizing antibody, 5 μ g/mL of Recombinant Mouse Progranulin Protein (R&D Systems, 2557-PG-050), Mouse Progranulin Antibody (R&D Systems, MAB2557) or Rat IgG2A Isotype Control (R&D Systems, MAB006) was added to the complete blastoid medium. For blastoid experiments involving SCD1 inhibitor, 5 μ M SCD1 inhibitor (Abcam, ab142089) was added to the complete blastoid medium.

Gene knockdown during blastoid formation

For blastoid generation experiments involving gene knockdown by siRNA transfection, a transfection mix containing DharmaFECT 1 transfection reagent (Horizon Discovery, T-2001-03) and 50 nM siRNA diluted in Opti-MEM was added to the blastoid seeding medium on Day 0 following manufacturer's protocol. Gene knockdown efficiency was verified at 48 h post transfection.

Animal Work

All procedures related to animals were performed following the ethical guidelines of the A*STAR. Animal protocols were reviewed and approved by A*STAR Institutional Animal Care and Use Committee (IACUC) for mouse work (IACUC #231771), and A*STAR Institutional Biosafety Committee (IBC) for porcine work (IBC #HSE-IBC-ESCAR-05-01). All mice were housed in a specific pathogen-free facility in the Biological Resource Centre, A*STAR, with temperature maintained at 21 ± 1 °C, 45–70% humidity and a 12/12 h dark/light cycle.

Microinjection of ESC and NrESC

8-cell embryos were flushed out of the uterus of pregnant female C57BL/6J. Injections were performed with a Nikon Eclipse TE2000-U inverted microscope unit. The micromanipulator was an Eppendorf InjectMan NI2. The outside diameter of the Holding pipette, with

95–105 μ m outer diameter (O.D.) and 20–25 μ m inner diameter (I.D.), and injection pipette with 18–20 μ m O.D. and 16–18 μ m I.D. were used. psinGFP-ESC/ NrESC microinjection was performed at room temperature in an injection drop of 1.0 mL M2 medium (Sigma, M7167) with 7.5 μ g cytochalasin B (Sigma, C6762). Each 8-cell embryo was injected with five to eight ESC or NrESC. All mouse embryos are cultured in M16 media, covered with mineral oil, in humidified incubator at 37 °C with 5% CO₂.

mRNA and siRNA microinjection and drug treatment in mouse embryo

EGFP and *Nr1h2* mRNA were synthesized in-house using mMESSAGE mMACHINE™ T7 ULTRA Transcription Kit (Invitrogen, AM1345). SMARTPool siNT and siNr1h2 siRNA (Horizon Discovery, D-001206-13, M-042839-01) were used for Nr1h2 knockdown in mouse embryos. Capillary glass tubes (outer diameter 1 mm, inner diameter 0.8 mm, and length 10 cm) with cut diameter 30–70 μ m were used. 4–5 μ L of 100 ng/ μ L mRNA or 100 μ M siRNA was injected into fertilized one-cell embryos. Injected embryos were then transferred into a droplet of M16 medium covered with mineral oil in a 30 mm dish, and cultured until 8-cell, morula or blastocyst embryo stage before harvest. For T09 treatment, 1-cell or 8-cell embryos were cultured in M16 media supplemented with 10 μ M T0901317.

Blastoid transfer

Blastoids (NrESC-blastoid or EPSC-blastoid) were manually picked up under a stereomicroscope and transferred into M16 media droplets using a Cook Flexipet pipette. The surrogate female at 3.5 dpc was anesthetized with avertin. Uterine horn was exposed surgically and punctured with 27G surgical needle. Blastoids, which were washed three times in M16, were loaded to the pipette with air bubble and transferred to the uterine horn. 15–20 blastoids were transferred into each uterine horn. 9 female mouse replicates for each group. A C-section was performed at 10.5 dpc, and the uterus was dissected out. Blastoid-induced decidua has a minimum length of 24 mm.

Teratoma assay

Around 2.5 million ESCs or NrESCs in 200 μ L Matrigel suspension were subcutaneously injected into one side of the dorsal flanks of NSG mice (5-week old). 3 mouse replicates for each group. Approximately 4 weeks post injection, primary teratomas (diameter ~20 mm) were harvested and fixed in 4% paraformaldehyde solution overnight at 4 °C. Fixed teratoma samples were embedded in paraffin and sectioned for hematoxylin-eosin (HE) staining and histology analysis.

In vitro maturation of porcine oocytes

Immature pig ovaries were collected from a local abattoir and transported to the laboratory in PBS around 30 °C in a thermal box. Cumulus-oocyte complexes (COCs) were released by scarification of the ovarian follicles into TCM199-Hepes medium (Gibco, Grand Island, NY, USA) with 4 mg/mL BSA. Only oocytes with homogeneous-dark ooplasm surrounded by at least two layers compact cumulus cells were selected for in vitro maturation (IVM). Subsequently, 30–40 COCs were allocated to one well of four-well plates (Nunc, Roskilde, Denmark) containing 1 mL of TCM-199 supplemented with 4 mg/mL BSA, 1 mM Sodium Pyruvate, 1.2 mM Cysteine, 0.5 IU/mL FSH, 0.5 IU/mL LH, 20 ng/mL EGF, 100 IU/mL Penicillin, 100 μ g/mL Streptomycin, and cultured at 38.5 °C in a humidified atmosphere of 5% CO₂, 5% O₂, 90% N₂. After culture for 42–44 h, cumulus cells surrounding the oocytes were removed with the help of 0.5% hyaluronidase. Oocytes excluded first polar bodies with normal morphology were selected for SCNT.

Somatic cell nuclear transfer for porcine

Porcine skin fibroblasts with passage 2 or 3 were used as donor cells, which were purchased from Cell Biologics (Chicago, IL, USA). Before

enucleation, metaphase II oocytes were incubated in TCM-199 medium (Grand Island, NY, USA) supplemented with 10 µg/ml Hoechst 33342 and 5 µg/ml cytochalasin B for 15 min. Metaphase II spindle was removed with the aid by exposure to UV light for location. Fibroblasts were introduced into the perivitelline space of enucleated oocytes. Fusion for the oocyte–fibroblast couplets was induced with two pulses of 1.2 KV/cm for 30 µs each with a 1-s interval in fusion medium (0.3 M mannitol, 0.1 mM CaCl₂ and 0.1 mM MgSO₄). The fused couplets were activated at the same time.

In vitro culture of porcine SCNT embryos

The reconstructed embryos were cultured individually in 25 µl droplets with PZM-3 medium in a EmbryoScope™ time-lapse incubator (Vitrolife, Gothenburg, Sweden) at 38.5 °C with 5% CO₂, 5% O₂, 90% N₂ for 160 h. Images of embryos were captured every 5 min in five different focal planes.

Karyotypic analysis

Karyotyping preparation was carried out following previous protocols¹¹⁴. Briefly, logarithmic growth phase cells were treated with 0.03 µg/ml Colcemid, trypsinization, hypotonic 0.075 M KCl for 15 min at 37° followed by fixation (3:1 methanol: acetic acid). We follow public guidelines (<http://www.pathology.washington.edu/research/cytopages/idiograms/mouse/>) used for karyotypic designations and chromosome breakpoint determination of murine metaphase chromosomes.

Bulk RNA-seq library preparation

Total RNA was extracted from each sample using Monarch® Total RNA Miniprep Kit (New England Biolabs, T2010S). Isolated RNA was then used for RNA-seq library preparation using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (New England Biolabs, E7775L), following manufacturer's instructions. For mouse morula bulk RNA-seq libraries, morulae samples were processed for RNA-seq library preparation using SMART-Seq® v4 Ultra® Low Input RNA Kit for Sequencing and Nextera XT DNA Library Preparation Kit, following manufacturers' instructions.

Single-cell RNA-sequencing library generation

Blastoids were manually picked by Flexipet pipette and washed three times in PBS containing 0.04% BSA. All blastoids were divided into type I-II blastoids and type III organoids based on the metrics presented in current study, with around 150 blastoids or organoids per sample group. All samples were dissociated into single cells with enzyme mix composed of 0.5X versene (Lonza, 17711E), 0.5X Accumax (Innovative Cell Tech, AM105), and 0.05X Dnase (STEMCELL Technologies, 07900) at 37 °C for 30 min with agitation. NrESC was dissociated into single cells with Accumax at 37 °C for 3–5 min. Single cell suspensions were centrifuged at 400 × g, and the resultant cell pellets were washed three times and resuspended in PBS + 0.04% BSA. Cell count was performed by an automated cell counter. Approximately 10,000 cells for each sample were loaded into the Chromium Next GEM Chip G (10X Genomics, 2000177) and processed with the Chromium controller (10X Genomics, 1000204) to generate gel beads in-emulsions according to manufacturer's instructions. Libraries were generated using the Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit v3.1 (10X Genomics, PN-1000121) according to manufacturer's protocol. The libraries were pooled and sequenced using HiSeq3000.

Transfection

Transfection was carried out using Lipofectamine 3000 (Invitrogen), following the manufacturer's instructions. Briefly, 50,000 cells were seeded 24 h prior to transfection. A total of 3000 ng of plasmid DNA (pCAG-EGFP-FLAG or pCAG-Nr1h2-FLAG) was mixed with 125 µL of Opti-MEM medium and 6 µL of P3000 reagent. In a separate tube, 6 µL

of Lipofectamine 3000 was combined with 125 µL of Opti-MEM medium. Subsequently, the Lipofectamine 3000 mixture was added to the plasmid DNA mixture. Following a 20-min incubation period at room temperature, the transfection mixture was added dropwise to the wells containing the seeded cells. After 24 h, the culture medium was replaced with fresh medium with antibiotic selection. Puromycin (0.6 µg/mL) was applied to the cells for 48 h before they were harvested for further analysis. For Nr1h2-FLAG Co-IP experiment, pCAG-EGFP-HA and pCAG-EGFP-FLAG or pCAG-Kdm1a-HA and pCAG-Nr1h2-FLAG plasmids were transfected into HEK293T cells with same transfection protocol.

Immunofluorescence staining

Blastoid samples were fixed with 4% PFA in PBS for 20 min at room temperature, washed and permeabilized with 0.2% Triton X-100 in PBS for 15 min. Samples were then blocked with blocking buffer (PBS containing 3% BSA and 0.1% Tween 20) at room temperature for 2 h or overnight at 4 °C. Primary antibodies diluted in blocking buffer were added to the samples and incubated overnight at 4 °C. Samples were washed for three times with PBS containing 0.1% Tween 20 followed by the incubation with fluorescence conjugated secondary antibodies diluted in blocking buffer for 2 h at room temperature. Samples were washed for three times with PBS containing 0.1% Tween 20. Nuclei were counterstained with DAPI at 1 µg/mL. Image acquisition was performed using a Zeiss LSM 700 confocal microscope. Images were processed using Zen (Zeiss). The primary antibodies and dilutions used were: Mouse anti-CDX2 (1:100; BioGeneX, MU392A-5UC), Rabbit anti-CDX2 (1:100; Abcam, ab76541), Rabbit anti-SOX2 (1:200; Abcam, ab97959), Mouse anti-GATA2 (1:50; Santa Cruz Biotechnology, sc-267), Goat anti-GATA6 (1:200; R and D Systems, af1700), Mouse anti-OCT4 (1:200; Santa Cruz Biotechnology, sc-5279) and Rabbit anti-GFP Alexa Fluor® 488 (1:100; Abcam, ab225314). The secondary antibodies and dilutions used were: Alexa Fluor 488 Donkey anti-Mouse IgG (H+L) (1:1000; Invitrogen, A32766), Alexa Fluor 488 Donkey anti-Rabbit IgG (H+L) (1:1000; Invitrogen, A32790), Alexa Fluor 647 Donkey anti-Rabbit IgG (H+L) (1:1000; Invitrogen, A32795), Alexa Fluor 555 Donkey anti-Goat IgG (H+L) (1:1000; Invitrogen, A32816) and Alexa Fluor 647 Donkey anti-Mouse IgG (H+L) (1:1000; Invitrogen, A32787).

ATAC-seq library preparation

ATAC-seq was performed with Illumina tagment DNA TDE1 Enzyme and Buffer kit (Cat No: 20034197) and Nextera DNA library preparation kit (Cat No.:FC-121-1030) following manufacturer's instruction. ES, EPSC, and NrES cells were lysed with 50 µl lysis buffer (10 mM Tris-Cl (pH 7.4), 10 mM NaCl, 3 mM MgCl₂, 0.1% (v/v) Igepal CA-630), and spun down at 500 × g for 10 min at 4 °C. The supernatant was removed. The cell pellet was resuspended with 50 µl transposition reaction mix (25 µl TD, 2.5 µl TDE1, 22.5 µl Nuclease-free water) and incubated at 37 °C for 30 min. Subsequent DNA purification was performed with Qiagen Minielute PCR purification kits. The purified DNA was then prepared into sequencing libraries following the protocol of the Nextera DNA library preparation kit.

Whole genome bisulfite sequencing library preparation

Genomic DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen, 69556), followed by sonication to generate ~200 bp fragments using Bioruptor® Pico (Diagenode, B01080010). Libraries were prepared using Premium WGBS Kit (Whole Genome Bisulfite Sequencing) (Diagenode, C02030034). All procedures follow manufacturer's instructions.

Co-immunoprecipitation

Fifty million transfected embryonic stem (ES) cells or 130 million 293T cells were lysed using IPH lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, protease inhibitor cocktail

(Roche)) and incubated on ice for 30 minutes. Subsequently, the lysed cells were subjected to resuspension using a 21G needle, repeating the process 50 times. The lysate was then centrifuged at $16,000 \times g$ at 4°C for 30 min, and the resulting supernatant was collected for immunoprecipitation, with some kept as input. Dynabeads Protein G (Invitrogen; 50 μL per sample) were washed three times with IPH lysis buffer and incubated with anti-Flag (Merck, F3165) and IgG (Santa Cruz #sc2025) antibodies at room temperature for three hours. After incubation, the beads were isolated using a magnetic stand and washed three times with IPH lysis buffer. During each wash, the beads were resuspended in the buffer, rotated for 5 min, and placed on the magnetic stand for 30 s to allow for bead separation. The washed beads were then incubated with the collected lysate at 4°C overnight. Following incubation, the beads were washed three more times with IPH lysis buffer using the same procedure as described earlier. After the final wash, the beads were resuspended in 50 μL elution buffer (100 mM Tris pH7.5, 4% SDS and 40 mM DTT) and incubated at 99°C for 10 min for elution. The eluted proteins were subsequently sent for mass spectrometry (MS) analysis or subjected to Western blot analysis to determine their composition and relative abundance.

Western blot

Co-IP samples were denatured by boiling at 99°C in Laemmli buffer (Biorad) for 10 min before electrophoresis. 2.5 μL of input and 20 μL of IP samples were resolved using 4–12% NuPAGE™ Bis-Tris Protein gels (Invitrogen) before transfer to a nitrocellulose membrane. Membranes were blocked with 5% blotting-grade blocker (Biorad) in TBST (Thermo Scientific) for 2 h at room temperature with shaking, followed by overnight incubation at 4°C with the following antibodies and dilutions: anti-Kdm1 1:10,000 (Abcam, ab129195) or anti-Nr1h2 1:3000 (Novus Biologicals, NB100-1464). After three five-minute washes in TBST, the membranes were then incubated with the corresponding HRP-conjugated secondary antibodies at 1:10,000 dilution for 2 h at room temperature. Three more washes in TBST were performed before visualization using Clarity Western ECL substrate (Biorad) and imaged using iBright (Invitrogen).

Chromatin Immunoprecipitation (ChIP) and ChIP-seq library preparation

For preparation of chromatin the protocol previously described was utilized as the template⁵. Cells were trypsinized, harvested and the cell number was estimated. Cross-linking was performed for 10×10^6 cells using 1% formaldehyde for 10 min at room temperature followed by quenching with 0.125 M Glycine. Crosslinked cells were resuspended in 1 ml of FA lysis buffer (10 mM Tris–Cl (pH 8), 100 mM NaCl, 10 mM EDTA, 0.25% Triton X-100 and protease inhibitor cocktail (Roche)), nutated at 4°C for 15 min and spun down at $1300 \times g$ for 5 min at 4°C . Supernatant was removed and the pellet was re-suspended in 1% SDS lysis buffer (50 mM HEPES–KOH (pH 7.5), 150 mM NaCl, 1% SDS, 2 mM EDTA, 1% Triton X-100, 0.1% NaDOC and protease inhibitor cocktail). After 15 min of incubation, the pellet was spun down and washed two times with 0.1% lysis buffer (50 mM HEPES–KOH (pH 7.5), 150 mM NaCl, 0.1% SDS, 2 mM EDTA, 1% Triton X-100, 0.1% NaDOC and protease inhibitor cocktail). Cell pellet was sonicated by Biorupter (Diagenode) for 14 cycles (30 s on and 60 s off for each cycle). Cell debris was removed from the sheared chromatin by centrifugation at $21,000 \times g$ at 4°C for 30 min. Protein G Dynabeads (Life Technologies) was incubated with antibodies (Nr1h2 ChIP antibody (NB100-1464, Novus Biologicals), FLAG ChIP antibody (F1804, Sigma-Aldrich), H3K9me2 ChIP antibody (Abcam, ab1220), H3K9me3 ChIP antibody (Abcam, ab8898) and H3K27Ac ChIP antibody (ab4729, Abcam)) for 3 h at room temperature and washed three times by PBS with 0.1% TritonX-100. 5% of the sheared chromatin was separated as input while the remaining was subject to overnight incubation with antibody-bound Dynabeads. In total 7 washes were performed to remove

unspecific bindings, including three washes with 0.1% SDS lysis buffer, one with 0.1% SDS lysis buffer/0.35 M NaCl, one with 10 mM Tris–Cl (pH 8.0), 1 mM EDTA, 0.5% NP40, 0.25 LiCl, 0.5% NaDOC, and two with TE buffer (pH 8.0). Dynabeads were resuspended in elution buffer (50 mM Tris–HCl (pH 7.5), 10 mM EDTA, 1% SDS) and incubated at 68°C for 1 h with 1400 rpm shaking. Eluted samples were de-crosslinked at 42°C for 2 h and 67°C for 6 h with Pronase (Sigma). DNA purification was further performed with QIAGEN PCR Purification kit (for inputs) and the QIAGEN MinELute PCR Purification kit (for ChIP samples). Quantitative PCR with CFX384 Real-time System (Bio-Rad) and Kapa SYBR Fast qPCR kit (Kapa Biosystems). For ChIP-seq library preparation, ACCEL-NGS®2S Plus DNA library kit (IDT) and 2S™ Adapter Set A (IDT) was used as per manufacturer's instruction starting with 5–10 ng DNA for both input and ChIP sample.

Sequential chromatin immunoprecipitation (Sequential ChIP)

Equimolar ratio of Protein A and Protein G Dynabeads (Invitrogen) were combined in 50 μL per sample. The beads were washed three times with PBS with 0.1% Triton X-100 and then resuspended in 600 μL of Pre-Adsorption buffer (equal amounts of ChIP lysis buffer and ChIP dilution buffer) (1% Triton X-100, 2 mM EDTA, 20 mM Tris–HCl, 150 mM NaCl, 1X Protease inhibitor cocktail). BSA was added to the resuspended beads for blocking non-specific binding at a final concentration of 200 $\mu\text{g}/\text{ml}$, followed by overnight incubation at 4°C . The beads were subsequently washed with ChIP dilution buffer. The first antibody (FLAG) was added to 100 μL of the bead mixture, which was incubated at room temperature for 3 h. Sonicated E14 DNA (see Chromatin Immunoprecipitation method) was pre-cleared with 100 μL of the beads for 3 h at 4°C . After pre-clearing, 50 μL of the chromatin was isolated as input while the rest was added to the antibody-bound beads for overnight incubation at 4°C . For the first elution, the chromatin-antibody-bound beads were washed three times with 0.1% SDS lysis buffer, once with 0.1% SDS lysis buffer/0.35 M NaCl, once with 10 mM Tris–Cl (pH 8), 1 mM EDTA, 0.5% NP40, 0.25 LiCl, 0.5% NaDOC, and once with TE buffer (pH 8.0). The beads were collected by centrifugation at $800 \times g$ for 1 min, and were resuspended in 75 μL of elution buffer (50 mM Tris–HCl (pH 7.5), 10 mM EDTA, 1% SDS) at 37°C for 30 min. The eluted DNA was separated from the beads by a brief centrifugation at $1000 \times g$ for 2 min. 15 μL of the eluted DNA was used for validation of the first ChIP, whereas the remaining 60 μL was diluted by 20 times with ChIP dilution buffer to 1200 μL total volume. The second antibody (Kdm1a) was added to incubate with 100 μL freshly pre-adsorbed Protein A and G Dynabeads at room temperature for 3 h. To maintain a 1:19 Input:IP ratio, 63 μL of the first eluted sample was separated out. The remaining 12 μL sample was added to incubate with the antibody-bound beads overnight at 4°C . DNA elution for the second ChIP follows the same procedure as the first ChIP, which was then incubated at 68°C for 60 min while shaking at 1400 rpm. Subsequently, the samples were de-crosslinked, purified and used for ChIP qPCR.

SINE-B1 enhancer luciferase reporter assay

Genomic DNA was extracted using QIAGEN DNeasy Blood & Tissue Kit. Candidate B1-enhancers were amplified from genomic DNA by PCR using KAPA HiFi HotStart ReadyMix (Roche). Using NEBuilder® HiFi DNA Assembly (NEB), the PCR products were inserted into a modified pGL3-basic plasmid, which contains SCP1 minimal promoter upstream of the luciferase gene. 12 h before transfection, cells were seeded into a 96 well-plate, at 7000 cells per well density. The cells were then co-transfected with 150 ng pGL3-B1-enhancer luciferase reporter plasmid, 1 ng pRL renilla luciferase plasmid (Promega), and 4.5 pmol siRNA (Dharmacon) according to Lipofectamine 2000 (Invitrogen) user's protocol. Luciferase signal was measured 48 h after transfection, using Dual-Glo Luciferase Assay System kit (Promega) and GloMax Explorer (Promega) plate reader.

RNA extraction, cDNA synthesis and quantitative RT-PCR

RNA was extracted using the Monarch® Total RNA Miniprep Kit (New England Biolabs, T2010S), followed by cDNA synthesis using iScript™ Reverse Transcription Supermix (Bio-Rad Laboratories, 1708841). qPCR reactions were performed using SYBR Green qPCR Master Mix on QuantStudio 5 Real-Time PCR System. Gene expression level was quantified using the relative cycle threshold (Ct) method. Relative expression levels were normalized to GAPDH. The primers used for the qPCR analysis are listed in Supplementary DataSupplementary Data 7.

Blastoids scRNA-seq analysis

For 10X analyses, alignment and UMI counting were performed by using Cellranger program with default parameters and mouse mm10 reference genome. Feature-barcode count matrix obtained from Cellranger were used for the downstream analyses using Seurat¹¹⁵ in R. For the quality control, cells whose number of expressed genes <500, percentage of mitochondrial genes >10 were treated as abnormal ones and filtered out. Genes expressed in less than three cells were removed. UMAP was used to do the visualization with 20 dimensions after dimension reduction using PCA. Differentially expressed genes in each cluster or cell type were identified using FindAllMarkers. Genes that were differentially expressed between samples were identified by FindMarkers using Wilcoxon's rank sum tests with a minimum upregulation of 0.1 log-fold and a *p*-value of 0.01. DoHeatmap was used to visualize the top markers for each lineage. To integrate single cell datasets between blastoids and blastocysts, Seurat's FindIntegrationAnchors and IntegrateData functions were used. UMAP were used for the dimensionality reduction. Transcriptome correlation was calculated using the integrated gene expression values and average expression values were calculated for each cell type in each sample. Based on the average expression values of cell types, PCA and hierarchical clustering were performed between samples. The AddModuleScore function was used to calculate the average overexpression of a set of gene signatures.

SCENIC analysis

Regulons were inferred using pySCENIC^{45,116} in Python. First, the gene regulatory network for each sample was constructed using “grn” function in pyscenic with the loom file produced from Seurat and mouse transcription factor list. The regulons were then identified using the “ctx” function. Lastly, the “aucell” function was used with standard settings to compute the area under each curve, which represents each regulon's activity. To obtain a cell type-specific ranking of regulons, regulon specificity scores were computed per regulon and cell type. The correlation of the regulon activity is visualized using pheatmap package in R.

RNA-seq analyses

For bulk RNA-seq analysis, reads were mapped to the mm9 genome using STAR¹¹⁷ with the default parameters. GTF annotation file of genes (mm9) was downloaded from the UCSC genome browser. Cufflinks (v2.2.1)¹¹⁸ were used for gene assembly and gene quantification to generate fragments per kilobase per million mapped reads (FPKM) table based on mapped bam files and GTF annotation file. Count numbers of transcripts were computed by Htseq¹¹⁹ software (v0.12.4). DESeq2¹²⁰ was used to perform differential gene analysis for samples with replicates. The cutoff *p*-value was set as 0.05, and the cutoff fold change was set as 1.5. EdgeR¹²¹ was used to perform differential gene analysis for samples without replicates. The cutoff *p*-value was set as 0.05. For siNr1h2 embryo and prolonged cultured ESC/NrESC analysis, RSEM¹²² was used for the quantification of count numbers of transcripts. K-means gene clusters were identified by R function pheatmap.

ATAC-Seq analysis

The sequencing reads were mapped to the mouse mm9 genome using STAR with the following parameters: --alignIntronMax 1 --alignEndsType EndToEnd. Peak calling was performed by MACS2¹²³. The common ATAC-seq peaks for replicates were kept for further analysis. ATAC-seq read counts for each sample were then calculated using featureCounts¹²⁴. The differential peaks were identified using DESeq2. Significant differential ATAC-seq peaks were determined using cutoff *p*-value < 0.05 and |log2FC| > 1. Replicates were merged into one sample, and DeepTools¹²⁵ was used to create bigwig files, metaplots and heatmaps.

Whole genome bisulfite sequencing (WGBS) data analysis

Raw paired-end reads were aligned to the mouse mm10 genome reference with Bismark¹²⁶/Bowtie2 mode with default parameters. Duplicate reads were removed by the command deduplicate_bismark from Bismark. The methylation calls were extracted and bedGraph files were produced by bismark_methylation_extractor. The differential hypermethylated/hypomethylated CpGs were identified by methylKit¹²⁷ with parameters: minimum read coverage of 10, minimum methylation difference of 20%, and *q*-value cutoff of 0.05. PCA based on WGBS data was done by methylKit. The coordinates of genomic features (promoters, exons, introns, CpG islands, and transposable elements) were downloaded from the UCSC browser. GAT¹²⁸ was used to perform the nucleotide overlap between genomic features and differentially hyper/hypomethylated bases and calculated the log2-fold change and significance *p*-value by comparing with 10,000 random samplings of background genome regions. The bar plot of the enrichment result was plotted using R. The DNA methylation data of preimplantation embryo cells were downloaded from the GEO database (GSE56697)¹²⁹. WGBS data of PSC cultured in serum condition was downloaded from the GEO database (GSE168728)²³.

Mass spectrometry analysis

The average MS/MS counts between replicates for proteins detected in IP MS samples were calculated. And log2-fold change between Nr1h2-FLAG and EGFP-FLAG was computed. The statistical significance was assessed using *t* test. Protein-protein interaction (PPI) network of Nr1h2 and its interacting proteins were constructed based on STRINGdb¹³⁰. The network was visualized using Cytoscape.

ChIP-seq data analysis

FastQC was used to check the sequence quality. Reads were mapped to the mm9 genome using STAR software with the parameters ‘--alignIntronMax 1 --alignEndsType EndToEnd’. MACS2 were used to call H3K27ac peaks (*q*-value < 0.05) and Nr1h2 peaks (*p*-value < 0.01). H3K27ac peaks from samples were merged with bedtools, and read counts for each sample on merged peaks were then calculated using featureCounts. The differential peaks were identified by EdgeR. Significant differential H3K27ac peaks were determined using cutoff *p*-value < 0.05 and |log2FC| > 1.5. Bigwig files were produced by DeepTools and visualized in WashU Epigenome Browser¹³¹. Motif enrichment analysis was performed using Homer¹³² with function findMotifsGenome.pl. The *p*-value of known motifs was used for the visualization.

The enrichment analysis of genomic features on differentially upregulated/downregulated Nr1h2 peaks was performed using Fisher's exact test. The log2-odds ratio and significance *p*-value was shown in the bar plot. The public ChIP-seq data of Kdm1a¹³³ (GSE27841) used in Fig. 6 and S6 was downloaded from GEO database. For the analysis of Nr1h2-FLAG and EGFP-FLAG cells, significant differential H3K27ac peaks were determined using cutoff *p*-value < 0.05 and |log2FC| > 1. Genome-wide H3K9me3 peaks were first determined by MACS2 and featureCounts followed by edgeR workflow similarly using cutoff *p*-value < 0.05 and |log2FC| > 1. The overlapping peaks with Nr1h2 bound sites were then selected using bedtools intersect function.

Principle component analysis

PCA was performed on RNA-seq counts with at least five mapped reads across at least 10% of samples. Batch effects were adjusted with the ComBat function in the sva Bioconductor package (v.3.34.0). FactoMineR R package was used to perform PCA analysis. The cluster dendrogram was plotted after hierarchical clustering was performed with the hclust function.

Gene ontology analysis

Duplicate genes were removed from the gene list of interest and entered into Metascape for gene ontology analysis¹³⁴ (<http://metascape.org>). Gene ontology terms were sorted according to $[-\log(p\text{-value})]$, and visualized in R using ggplot2.

Interaction networks

In order to obtain protein-protein interactions, the gene symbols were entered into the STRING database¹³⁵. From the interaction networks, a.tsv file was created, and Cytoscape¹³⁶ was used to visualize the networks.

GSEA analysis

Gene set enrichment analysis (GSEA)¹³⁷ was used to determine if gene marker sets are significantly enriched in the transcriptome. Python package gseapy¹³⁸ was used to do the analysis with parameters: 'gseapy prerank -f pdf -max-size 100000 -r a.mnk -g b.gmt -o result'.

Reporting summary

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

Data availability

The RNA-seq, scRNA-seq, ATAC-seq, ChIP-seq and WGBS data generated during this study were deposited into the Gene Expression Omnibus database are available at GEO.: RNA-seq: [GSE210788](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE210788); Single-cell RNA-seq: [GSE217826](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE217826); ATAC-seq: [GSE217825](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE217825); ChIP-seq: [GSE217824](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE217824); WGBS: [GSE244019](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE244019). The Co-IP mass spectrometry dataset was deposited to the ProteomeXchange Consortium via the PRIDE¹³⁹ partner repository with the dataset identifier [PXD057257](https://www.ebi.ac.uk/pride/archive/projects/PXD057257). Published E3.5 and E4.5 blastocyst single cell RNA-seq data is obtained from [GSE123046](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE123046)⁷⁰ [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE123046>]. Raw datasets of early development stage single cell RNA-seq are obtained from [GSE136714](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE136714)¹⁴⁰ [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE136714>]. Early development stage bulk RNA-seq datasets are downloaded from [GSE98150](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98150)¹⁴¹ [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98150>]. 2C-like cells RNA-seq raw data is downloaded from [GSE85627](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE85627)¹⁴² [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE85627>]. Published EPS-blastoids and blastocysts single cell RNA-seq data from [GSE135701](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE135701)²¹ [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE135701>]. Source data are provided with this paper.

References

- Nishioka, N. et al. The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. *Dev. Cell* **16**, 398–410 (2009).
- Zhao, B. et al. Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control. *Genes Dev.* **21**, 2747–2761 (2007).
- Yagi, R. et al. Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. *Development* **134**, 3827–3836 (2007).
- Nishioka, N. et al. Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. *Mech. Dev.* **125**, 270–283 (2008).
- Loh, Y.-H. et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **38**, 431–440 (2006).
- Allègre, N. et al. NANOG initiates epiblast fate through the coordination of pluripotency genes expression. *Nat. Commun.* **13**, 3550 (2022).
- Frankenberg, S. et al. Primitive endoderm differentiates via a three-step mechanism involving Nanog and RTK signaling. *Dev. Cell* **21**, 1005–1013 (2011).
- Albertini, P. et al. Effect of acetyl-L-carnitine treatment on brain ribonucleic acid levels in methylazoxymethanol-microencephalic rats. *Drugs Exp. Clin. Res.* **15**, 429–434 (1989).
- Krawchuk, D., Honma-Yamanaka, N., Anani, S. & Yamanaka, Y. FGF4 is a limiting factor controlling the proportions of primitive endoderm and epiblast in the ICM of the mouse blastocyst. *Dev. Biol.* **384**, 65–71 (2013).
- Yamanaka, Y., Lanner, F. & Rossant, J. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* **137**, 715–724 (2010).
- Plusa, B., Piliszek, A., Frankenberg, S., Artus, J. & Hadjantonakis, A.-K. Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. *Development* **135**, 3081–3091 (2008).
- Niakan, K. K. et al. Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes Dev.* **24**, 312–326 (2010).
- Artus, J., Piliszek, A. & Hadjantonakis, A.-K. The primitive endoderm lineage of the mouse blastocyst: sequential transcription factor activation and regulation of differentiation by Sox17. *Dev. Biol.* **350**, 393–404 (2011).
- Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
- 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. USA* **78**, 7634–7638 (1981).
- Tanaka, S., Kunath, T., Hadjantonakis, A. K., Nagy, A. & Rossant, J. Promotion of trophoblast stem cell proliferation by FGF4. *Science* **282**, 2072–2075 (1998).
- Niakan, K. K., Schrode, N., Cho, L. T. Y. & Hadjantonakis, A.-K. Derivation of extraembryonic endoderm stem (XEN) cells from mouse embryos and embryonic stem cells. *Nat. Protoc.* **8**, 1028–1041 (2013).
- Cho, L. T. Y. et al. Conversion from mouse embryonic to extraembryonic endoderm stem cells reveals distinct differentiation capacities of pluripotent stem cell states. *Development* **139**, 2866–2877 (2012).
- Rivron, N. C. et al. Blastocyst-like structures generated solely from stem cells. *Nature* **557**, 106–111 (2018).
- Sozen, B. et al. Self-organization of mouse stem cells into an extended potential blastoid. *Dev. Cell* **51**, 698–712.e8 (2019).
- Li, R. et al. Generation of blastocyst-like structures from mouse embryonic and adult. *Cell Cult. Cell* **179**, 687–702.e18 (2019).
- Kime, C. et al. Induced 2C expression and implantation-competent blastocyst-like cysts from primed pluripotent stem cells. *Stem Cell Rep.* **13**, 485–498 (2019).
- Shen, H. et al. Mouse totipotent stem cells captured and maintained through spliceosomal repression. *Cell* **184**, 2843–2859.e20 (2021).
- Xu, Y. et al. Derivation of totipotent-like stem cells with blastocyst-like structure forming potential. *Cell Res.* **32**, 513–529 (2022).
- Posfai, E. et al. Evaluating totipotency using criteria of increasing stringency. *Nat. Cell Biol.* **23**, 49–60 (2021).

26. Ng, J.-H., Heng, J.-C. D., Loh, Y.-H. & Ng, H.-H. Transcriptional and epigenetic regulations of embryonic stem cells. *Mutat. Res./Fundam. Mol. Mech. Mutagen.* **647**, 52–58 (2008).
27. Maury, J. J. P. et al. RING1B O-GlcNAcylation regulates gene targeting of polycomb repressive complex 1 in human embryonic stem cells. *Stem Cell Res.* **15**, 182–189 (2015).
28. Simonti, C. N., Pavlicev, M. & Capra, J. A. Transposable element exaptation into regulatory regions is rare, influenced by evolutionary age, and subject to pleiotropic constraints. *Mol. Biol. Evol.* **34**, 2856–2869 (2017).
29. Gerdes, P., Richardson, S. R., Mager, D. L. & Faulkner, G. J. Transposable elements in the mammalian embryo: pioneers surviving through stealth and service. *Genome Biol.* **17**, 100 (2016).
30. Castro-Diaz, N. et al. Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev.* **28**, 1397–1409 (2014).
31. Thompson, P. J., Macfarlan, T. S. & Lorincz, M. C. Long terminal repeats: from parasitic elements to building blocks of the transcriptional regulatory repertoire. *Mol. Cell* **62**, 766–776 (2016).
32. Schulz, K. N. & Harrison, M. M. Mechanisms regulating zygotic genome activation. *Nat. Rev. Genet.* **20**, 221–234 (2019).
33. Svoboda, P. et al. RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev. Biol.* **269**, 276–285 (2004).
34. Rowe, H. M. & Trono, D. Dynamic control of endogenous retroviruses during development. *Virology* **411**, 273–287 (2011).
35. Schoorlemmer, J., Pérez-Palacios, R., Climent, M., Guallar, D. & Muniesa, P. Regulation of mouse retroelement MuERV-L/MERVL expression by REX1 and epigenetic control of stem cell potency. *Front Oncol.* **4**, 14 (2014).
36. Maksakova, I. A. et al. Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. *Epigenetics Chromatin* **6**, 15 (2013).
37. Yang, B. X. et al. Systematic identification of factors for provirus silencing in embryonic stem cells. *Cell* **163**, 230–245 (2015).
38. Rowe, H. M. et al. TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res.* **23**, 452–461 (2013).
39. Yi, Y. et al. Ribosomal proteins regulate 2-cell-stage transcriptome in mouse embryonic stem cells. *Stem Cell Rep.* **18**, 463–474 (2023).
40. Loh, Y.-H. et al. Genomic approaches to deconstruct pluripotency. *Annu Rev. Genom. Hum. Genet.* **12**, 165–185 (2011).
41. Chen, X. et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
42. Mzoughi, S. et al. PRDM15 safeguards naive pluripotency by transcriptionally regulating WNT and MAPK–ERK signaling. *Nat. Genet.* **49**, 1354–1363 (2017).
43. Li, R. & Izpisua Belmonte, J. C. Protocol for the generation of blastocyst-like structures from mouse extended pluripotent stem cells. *STAR Protoc.* **2**, 100745 (2021).
44. Gautam, P. et al. Multi-species single-cell transcriptomic analysis of ocular compartment regulons. *Nat. Commun.* **12**, 5675 (2021).
45. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
46. Yu, S. et al. BMP4 resets mouse epiblast stem cells to naive pluripotency through ZBTB7A/B-mediated chromatin remodelling. *Nat. Cell Biol.* **22**, 651–662 (2020).
47. Goolam, M. et al. Heterogeneity in Oct4 and Sox2 targets biases cell fate 4-cell mouse embryos. *Cell* **165**, 61–74 (2016).
48. Lim, L. S. et al. Zic3 is required for maintenance of pluripotency in embryonic stem cells. *Mol. Biol. Cell* **18**, 1348–1358 (2007).
49. Hoffman, J. A., Wu, C.-I. & Merrill, B. J. Tcf7l1 prepares epiblast cells in the gastrulating mouse embryo for lineage specification. *Development* **140**, 1665–1675 (2013).
50. Grover, A. & Adamson, E. D. Evidence for the existence of an early common biochemical pathway in the differentiation of F9 cells into visceral or parietal endoderm: modulation by cyclic AMP. *Dev. Biol.* **114**, 492–503 (1986).
51. Athanasouli, P. et al. The Wnt/TCF7L1 transcriptional repressor axis drives primitive endoderm formation by antagonizing naive and formative pluripotency. *Nat. Commun.* **14**, 1210 (2023).
52. Toh, C.-X. D. et al. RNAi reveals phase-specific global regulators of human somatic cell reprogramming. *Cell Rep.* **15**, 2597–2607 (2016).
53. Ding, J. et al. Cripto is required for correct orientation of the anterior-posterior axis in the mouse embryo. *Nature* **395**, 702–707 (1998).
54. Fiorenzano, A. et al. Cripto is essential to capture mouse epiblast stem cell and human embryonic stem cell pluripotency. *Nat. Commun.* **7**, 12589 (2016).
55. Graf, U. et al. Prame17 mediates ground-state pluripotency through proteasomal-epigenetic combined pathways. *Nat. Cell Biol.* **19**, 763–773 (2017).
56. Gerbe, F., Cox, B., Rossant, J. & Chazaud, C. Dynamic expression of Lrp2 pathway members reveals progressive epithelial differentiation of primitive endoderm in mouse blastocyst. *Dev. Biol.* **313**, 594–602 (2008).
57. Moore, R., Cai, K. Q., Tao, W., Smith, E. R. & Xu, X.-X. Differential requirement for Dab2 in the development of embryonic and extra-embryonic tissues. *BMC Dev. Biol.* **13**, 39 (2013).
58. Yang, D.-H. et al. Disabled-2 is essential for endodermal cell positioning and structure formation during mouse embryogenesis. *Dev. Biol.* **251**, 27–44 (2002).
59. Morris, S. M., Tallquist, M. D., Rock, C. O. & Cooper, J. A. Dual roles for the Dab2 adaptor protein in embryonic development and kidney transport. *EMBO J.* **21**, 1555–1564 (2002).
60. Posfai, E. et al. Position- and Hippo signaling-dependent plasticity during lineage segregation in the early mouse embryo. *eLife* **6**, 1–24 (2017).
61. Hirate, Y. et al. Polarity-dependent distribution of angiominin localizes Hippo signaling in preimplantation embryos. *Curr. Biol.* **23**, 1181–1194 (2013).
62. Leung, C. Y. & Zernicka-Goetz, M. Angiominin prevents pluripotent lineage differentiation in mouse embryos via Hippo pathway-dependent and -independent mechanisms. *Nat. Commun.* **4**, 2251 (2013).
63. Theurillat, I. et al. Extensive SUMO modification of repressive chromatin factors distinguishes pluripotent from somatic cells. *Cell Rep.* **33**, 108251 (2020).
64. Cossec, J.-C. et al. SUMO safeguards somatic and pluripotent cell identities by enforcing distinct chromatin states. *Cell Stem Cell* **23**, 742–757.e8 (2018).
65. Antonio Urrutia, G. et al. ZFP451-mediated SUMOylation of SATB2 drives embryonic stem cell differentiation. *Genes Dev.* **35**, 1142–1160 (2021).
66. Borkent, M. et al. A Serial shRNA Screen for Roadblocks to Reprogramming Identifies the Protein Modifier SUMO2. *Stem Cell Rep.* **6**, 704–716 (2016).
67. Li, C., Huang, Z. & Gu, L. SETD2 reduction adversely affects the development of mouse early embryos. *J. Cell Biochem.* **121**, 797–803 (2020).
68. Strauss, B. et al. Cyclin B1 is essential for mitosis in mouse embryos, and its nuclear export sets the time for mitosis. *J. Cell Biol.* **217**, 179–193 (2018).
69. Ben-David, U. et al. Selective elimination of human pluripotent stem cells by an oleate synthesis inhibitor discovered in a high-throughput screen. *Cell Stem Cell* **12**, 167–179 (2013).
70. Nowotschin, S. et al. The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* **569**, 361–367 (2019).

71. Bideyan, L. et al. Integrative analysis reveals multiple modes of LXR transcriptional regulation in liver. *Proc. Natl Acad. Sci. USA* **119**, 1–11 (2022).
72. Kennedy, M. A. et al. ABCG1 has a critical role in mediating cholesterol efflux to HDL and preventing cellular lipid accumulation. *Cell Metab.* **1**, 121–131 (2005).
73. Lee, D.-K. et al. Stearoyl-coenzyme A desaturase 1 is required for lipid droplet formation in pig embryo. *Reproduction* **157**, 235–243 (2019).
74. Zhao, C. & Dahlman-Wright, K. Liver X receptor in cholesterol metabolism. *J. Endocrinol.* **204**, 233–240 (2010).
75. Eshkind, L. et al. Loss of desmoglein 2 suggests essential functions for early embryonic development and proliferation of embryonal stem cells. *Eur. J. Cell Biol.* **81**, 592–598 (2002).
76. Fleming, Tom & Cell, P. adhesion in the preimplantation mammalian embryo and its role in trophoblast differentiation and blastocyst morphogenesis. *Front. Biosci.* **6**, d1000 (2001).
77. Wu, G. et al. Initiation of trophoblast lineage specification in mouse embryos is independent of Cdx2. *Development* **137**, 4159–4169 (2010).
78. Morgan, H. D., Santos, F., Green, K., Dean, W. & Reik, W. Epigenetic reprogramming in mammals. *Hum. Mol. Genet.* **14**, R47–R58 (2005).
79. Reik, W., Dean, W. & Walter, J. Epigenetic reprogramming in mammalian development. *Science* **293**, 1089–1093 (2001).
80. Yang, J. et al. Establishment of mouse expanded potential stem cells. *Nature* **550**, 393–397 (2017).
81. Calkin, A. C. & Tontonoz, P. Transcriptional integration of metabolism by the nuclear sterol-activated receptors LXR and FXR. *Nat. Rev. Mol. Cell Biol.* **13**, 213–224 (2012).
82. Williams, S. et al. X-ray crystal structure of the liver X receptor beta ligand binding domain: regulation by a histidine-tryptophan switch. *J. Biol. Chem.* **278**, 27138–27143 (2003).
83. Ancelin, K. et al. Maternal LSD1/KDM1A is an essential regulator of chromatin and transcription landscapes during zygotic genome activation. *eLife* **5**, 1–24 (2016).
84. Ciccone, D. N. et al. KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature* **461**, 415–418 (2009).
85. Wicklow, E. et al. HIPPO pathway members restrict SOX2 to the inner cell mass where it promotes ICM fates in the mouse blastocyst. *PLoS Genet.* **10**, e1004618 (2014).
86. Messerschmidt, D. M. & Kemler, R. Nanog is required for primitive endoderm formation through a non-cell autonomous mechanism. *Dev. Biol.* **344**, 129–137 (2010).
87. le Bin, G. C. et al. Oct4 is required for lineage priming in the developing inner cell mass of the mouse blastocyst. *Development* **141**, 1001–1010 (2014).
88. Frum, T. et al. Oct4 cell-autonomously promotes primitive endoderm development in the mouse blastocyst. *Dev. Cell* **25**, 610–622 (2013).
89. Chazaud, C., Yamanaka, Y., Pawson, T. & Rossant, J. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Dev. Cell* **10**, 615–624 (2006).
90. Kang, M., Piliszek, A., Artus, J. & Hadjantonakis, A.-K. FGF4 is required for lineage restriction and salt-and-pepper distribution of primitive endoderm factors but not their initial expression in the mouse. *Development* **140**, 267–279 (2013).
91. Nichols, J., Silva, J., Roode, M. & Smith, A. Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. *Development* **136**, 3215–3222 (2009).
92. Liu, C., Peng, G. & Jing, N. TGF- β signaling pathway in early mouse development and embryonic stem cells. *Acta Biochim Biophys. Sin.* **50**, 68–73 (2018).
93. Roelen, B. A. J., Goumans, M.-J., Zwijsen, A. & Mummery, C. L. Identification of two distinct functions for TGF- β in early mouse development. *Differentiation* **64**, 19–31 (1998).
94. Watabe, T. & Miyazono, K. Roles of TGF- β family signaling in stem cell renewal and differentiation. *Cell Res.* **19**, 103–115 (2009).
95. Bernardo, A. S. et al. BRACHYURY and CDX2 mediate BMP-induced differentiation of human and mouse pluripotent stem cells into embryonic and extraembryonic lineages. *Cell Stem Cell* **9**, 144–155 (2011).
96. Wang, B. & Tontonoz, P. Liver X receptors in lipid signalling and membrane homeostasis. *Nat. Rev. Endocrinol.* **14**, 452–463 (2018).
97. Xing, Q. R. et al. Diversification of reprogramming trajectories revealed by parallel single-cell transcriptome and chromatin accessibility sequencing. *Sci. Adv.* **6**, 1–18 (2020).
98. Xing, Q. R. et al. Parallel bimodal single-cell sequencing of transcriptome and chromatin accessibility. *Genome Res.* **30**, 1027–1039 (2020).
99. Joseph, S. B. et al. Direct and indirect mechanisms for regulation of fatty acid synthase gene expression by liver X receptors. *J. Biol. Chem.* **277**, 11019–11025 (2002).
100. Liang, G. et al. Diminished hepatic response to fasting/refeeding and liver X receptor agonists in mice with selective deficiency of sterol regulatory element-binding protein-1c. *J. Biol. Chem.* **277**, 9520–9528 (2002).
101. Repa, J. J. et al. Regulation of mouse sterol regulatory element-binding protein-1c gene (SREBP-1c) by oxysterol receptors, LXRalpha and LXRbeta. *Genes Dev.* **14**, 2819–2830 (2000).
102. Mau, K. H. T. et al. Dynamic enlargement and mobilization of lipid droplets in pluripotent cells coordinate morphogenesis during mouse peri-implantation development. *Nat. Commun.* **13**, 3861 (2022).
103. Wu, Z. et al. Role of nuclear receptor coactivator 3 (Nco3) in pluripotency maintenance. *J. Biol. Chem.* **287**, 38295–38304 (2012).
104. Percharde, M. et al. Nco3 functions as an essential Esrrb coactivator to sustain embryonic stem cell self-renewal and reprogramming. *Genes Dev.* **26**, 2286–2298 (2012).
105. Gassler, J. et al. Zygotic genome activation by the totipotency pioneer factor Nr5a2. *Science* **378**, 1305–1315 (2022).
106. Zhou, X., Sam, T. W., Lee, A. Y. & Leung, D. Mouse strain-specific polymorphic provirus functions as cis-regulatory element leading to epigenomic and transcriptomic variations. *Nat. Commun.* **12**, 6462 (2021).
107. Viswanathan, R. et al. DARE SOME enables concurrent profiling of multiple DNA modifications with restriction enzymes in single cells and cell-free DNA. *Sci. Adv.* **9**, eadi0197 (2023).
108. Warwick, T., Schulz, M. H., Gilsbach, R., Brandes, R. P. & Seuter, S. Nuclear receptor activation shapes spatial genome organization essential for gene expression control: lessons learned from the vitamin D receptor. *Nucleic Acids Res.* **50**, 3745–3763 (2022).
109. Warriar, T. et al. SETDB1 acts as a topological accessory to Cohesin via an H3K9me3-independent, genomic shunt for regulating cell fates. *Nucleic Acids Res.* **50**, 7326–7349 (2022).
110. Feng, J. et al. PHF2 regulates genome topology and DNA replication in neural stem cells via cohesin. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkae457> (2024).
111. Hamashima, K. et al. Single-nucleus multiomic mapping of m6A methylomes and transcriptomes in native populations of cells with sn-m6A-CT. *Mol. Cell* <https://doi.org/10.1016/j.molcel.2023.08.010> (2023).
112. Zhao, P. et al. Young SINEs in pig genomes impact gene regulation, genetic diversity, and complex traits. *Commun. Biol.* **6**, 894 (2023).
113. Liang, X. et al. Transgelin 2 is required for embryo implantation by promoting actin polymerization. *FASEB J.* **33**, 5667–5675 (2019).

114. Testa, J. R. et al. Spontaneous transformation of rat ovarian surface epithelial cells results in well to poorly differentiated tumors with a parallel range of cytogenetic complexity. *Cancer Res.* **54**, 2778–2784 (1994).
115. Stuart, T. et al. Comprehensive Integration of Single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
116. van de Sande, B. et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247–2276 (2020).
117. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
118. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
119. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
120. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
121. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
122. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **12**, 323 (2011).
123. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
124. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
125. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
126. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
127. Akalin, A. et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012).
128. Heger, A., Webber, C., Goodson, M., Ponting, C. P. & Lunter, G. GAT: a simulation framework for testing the association of genomic intervals. *Bioinformatics* **29**, 2046–2048 (2013).
129. Wang, L. et al. Programming and inheritance of parental DNA methylomes in mammals. *Cell* **157**, 979–991 (2014).
130. Szklarczyk, D. et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).
131. Li, D. et al. WashU epigenome browser update 2022. *Nucleic Acids Res.* **50**, W774–W781 (2022).
132. 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).
133. Whyte, W. A. et al. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* **482**, 221–225 (2012).
134. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).
135. Szklarczyk, D. et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368 (2017).
136. Cline, M. S. et al. Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.* **2**, 2366–2382 (2007).
137. Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci.* **102**, 15545–15550 (2005).
138. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
139. Perez-Riverol, Y. et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552 (2022).
140. Wang, Y. et al. Single-cell multiomics sequencing reveals the functional regulatory landscape of early embryos. *Nat. Commun.* **12**, 1247 (2021).
141. Wang, C. et al. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat. Cell Biol.* **20**, 620–631 (2018).
142. Hendrickson, P. G. et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934 (2017).

Acknowledgements

We thank Arfayunmy Abu Bakar and the Biological Resource Centre, A*STAR for their technical support in the mouse embryo work. K.H. was supported by the NMRC grant MOH-000937-00 and A*STAR grant C210812003. M.T.N.L. was supported by the Industry Alignment Fund - Prepositioning (IAF-PP) [H23J2a0097]. H.L. was supported by grants from the Mayo Clinic Center for Biomedical Discovery, Center for Individualized Medicine, the Mayo Clinic Comprehensive Cancer Center (NIH; P30CA015083), the Mayo Clinic Center for Cell Signaling in Gastroenterology (NIH: P30DK084567), the Mayo Clinic Nutrition Obesity Research Program, the Glenn Foundation for Medical Research, the Eric & Wendy Schmidt Fund for AI Research & Innovation and the National Institutes of Health (NIH; U19AG74879, P50CA136393, R03OD038392); Y-H.L. was supported by the National Research Foundation, Singapore (NRF) Investigatorship award [NRFI2018- 02]; National Medical Research Council [NMRC/OFIRG21nov-0088]; Singapore Food Story (SFS) R&D Programme [W22W3D0007]; A*STAR Biomedical Research Council, Central Research Fund, Use-Inspired Basic Research (CRF UIBR); Competitive Research Programme (CRP) [NRF-CRP29-2022-0005]; Industry Alignment Fund - Prepositioning (IAF-PP) [H23J2a0095, H23J2a0097].

Author contributions

K.W.W. and Y.Y.Z. designed and performed the research, analysed the data and wrote the paper. E.T., J.H.J.T., N.O.C., K.H., Y.Y., H.J.L., and T.W. designed and conducted research. M.T.N.L., S.C.N., Q-J.L., and H.L. analysed data. Y-H.L. designed research analysed data and wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-54381-0>.

Correspondence and requests for materials should be addressed to Yuin-Han Loh.

Peer review information *Nature Communications* thanks the anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024

¹Cell Fate Engineering and Therapeutics Lab, Institute of Molecular and Cell Biology (IMCB), Agency for Science, Technology and Research (A*STAR), 61 Biopolis Drive, Proteos, Singapore 138673, Republic of Singapore. ²School of Biological Sciences, Nanyang Technological University, Singapore, Republic of Singapore. ³Endangered Species Conservation via Assisted Reproduction (ESCAR) Lab, Institute of Molecular and Cell Biology (IMCB), Agency for Science, Technology and Research (A*STAR), 61 Biopolis Drive, Proteos, Singapore 138673, Republic of Singapore. ⁴Department of Pharmacology and Institute for Digital Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Republic of Singapore. ⁵Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Republic of Singapore. ⁶Institute of Molecular and Cell Biology (IMCB), Agency for Science, Technology and Research (A*STAR), 61 Biopolis Drive, Proteos, Singapore 138673, Republic of Singapore. ⁷Department of Obstetrics & Gynaecology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Republic of Singapore. ⁸Sincere Healthcare Group, Singapore, Republic of Singapore. ⁹Singapore Immunology Network (SigN), Agency for Science, Technology and Research (A*STAR), 8A Biomedical Grove, Immunos, Singapore 138648, Republic of Singapore. ¹⁰Department of Molecular Pharmacology & Experimental Therapeutics, Mayo Clinic, Rochester, MN, USA. ¹¹Department of Biological Sciences, National University of Singapore, Singapore, Republic of Singapore. ¹²NUS Graduate School's Integrative Sciences and Engineering Programme, National University of Singapore, Singapore, Republic of Singapore. ¹³Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Republic of Singapore. ¹⁴These authors contributed equally: Ka Wai Wong, Yingying Zeng. ✉e-mail: yhloh@imcb.a-star.edu.sg