

A continuous totipotent-like cell-based embryo model recapitulates mouse embryogenesis from zygotic genome activation to gastrulation

Received: 28 March 2025

Accepted: 25 September 2025

Published online: 15 October 2025



Yixuan Ren^{1,2,8}, Xuyang Wang^{3,8}, Haiyin Liu^{3,8}, Yaxing Xu^{4,8}, Ruoqi Cheng^{5,8}, Shengnan Ren^{6,8}, Zining Li¹, Yunfei Huo¹, Bo Li¹, Jingyang Guan¹✉, Cheng Li¹✉, Hongkui Deng^{1,7}✉ & Jun Xu^{1,3}✉

The development of stem-cell-derived models of mammalian embryogenesis has provided invaluable tools for investigating embryo development. However, constructing embryo models that can continuously recapitulate the developmental trajectory, from zygotic genome activation to gastrulation, remains challenging. Here we report the development of a chemical cocktail to induce totipotent-like cells with robust proliferative ability and leverage these cells to establish a stepwise protocol for generating a continuous embryo model. This model sequentially mimics mouse embryogenesis from embryonic day 1.5 to 7.5. It recapitulates key developmental milestones, including zygotic genome activation in 2-cell embryos, the diversification of embryonic and extraembryonic lineages from 4-cell to 64-cell stages, the formation of blastocysts and the subsequent development into post-implantation egg cylinders. Notably, these structures undergo gastrulation, as indicated by the formation of a primitive streak-like structure and the subsequent emergence of several early organogenesis hallmarks. Our study opens avenues for modelling mammalian embryogenesis in vitro.

The use of early stem cell types to construct in vitro embryo models has opened avenues for exploring embryo development^{1–4}. Embryo-like structures can be generated through the self-organization of pluripotent stem cells^{5–12}, or by assembling them with extraembryonic stem cells^{13–18}. These approaches have led to the development of blastoids^{9–14,19}, gastruloids^{5–7} and other types of embryoid^{8,15–22}. Although these models provide powerful tools for studying mouse embryogenesis, the assembly of different stem cell types does not fully replicate the natural process of embryo formation. Furthermore, current embryo models are limited to recapitulating specific developmental stages, and generating models

that mimic the complete process of mouse embryogenesis remains challenging.

An alternative strategy for generating embryo models is to utilize totipotent-like stem cells^{23–26}, which correspond to an earlier developmental stage than blastocyst-derived stem cells. In principle, totipotent-like stem cells hold promise for modelling the entire developmental trajectory of embryogenesis. Notably, we and other groups have demonstrated that mouse totipotent-like stem cells can be induced into blastoids^{23,25,27,28}, bypassing the need for mixing different stem cell types and more closely mimicking natural blastocyst formation. However, it remains unknown whether mouse totipotent-like stem

A full list of affiliations appears at the end of the paper. ✉e-mail: guanjingyang@hsc.pku.edu.cn; cheng_li@pku.edu.cn; hongkui_deng@pku.edu.cn; jun_xu@bjmu.edu.cn

cells can be used to continuously model development from zygotic genome activation (ZGA) to gastrulation.

Recently, we reported that mouse totipotent potential stem (TPS) cells can be induced using a chemical cocktail²³. However, these cells proliferate slowly, hindering their applications in constructing embryo models. Here, we sought to utilize small molecules to capture totipotent-like cells with improved proliferative ability. Leveraging these cells, we developed a stepwise protocol to generate embryo models that sequentially recapitulate mouse embryogenesis from ZGA to the gastrulation.

Results

Identification of a chemical cocktail enabling the induction of totipotent-like cells with improved proliferative ability

Previously, we established TPS cells from mouse extended pluripotent stem (EPS) cells²⁹ and mouse two-cell embryos using the CPEC condition, which includes CD1530, VPA, EPZ004777 and CHIR-99021²³. While transient treatment of mouse EPS cells with the CPEC condition induced the expression of totipotent markers³⁰, these cells showed poor proliferative capacity after subsequent passaging, and continuous adaptation in the CPEC condition was required for generating stable totipotent-like cell lines²³. Therefore, we aimed to identify a chemical condition capable of rapidly inducing totipotent-like cells with robust proliferative ability, providing a more suitable resource for embryo modelling.

To achieve this, we screened a small chemical library using mouse EPS cells carrying a *MuERV-L* reporter (Extended Data Fig. 1a,b and Supplementary Table 1), and the Wnt signalling agonist CHIR-99021 was added to support cell survival. Consistent with our previous report²³, the retinoic acid agonist CD1530 emerged as one of the most effective totipotency inducers (Extended Data Fig. 1b,c). However, CD1530 treatment also activated marker genes of primitive endoderm (PrE) (Extended Data Fig. 1d). To enhance totipotency induction while minimizing PrE induction, we explored combinations of CD1530 and CHIR-99021 with additional candidates identified from our screen. Notably, the inclusion of PD0325901, a MEK inhibitor, further enhanced the expression of totipotency marker genes (Extended Data Fig. 1e). Moreover, this combination markedly suppressed PrE induction (Extended Data Fig. 1f), consistent with the reported inhibition of PrE lineage specification by PD0325901 treatment³¹.

Next, we evaluated the proliferation ability of cells upon the treatment of CD1530, CHIR-99021 and PD0325901. However, we noticed PD0325901 treatment increased the doubling time of cells compared with the treatment of CD1530 and CHIR-99021 (Extended Data Fig. 1g). Therefore, we tested whether further combination with other totipotency inducers identified in the screen can reverse the inhibitory effect of PD0325901 on cell proliferation. Notably, elvitegravir was found to decrease the doubling time of cells treated with CD1530, CHIR-99021 and PD0325901 (Extended Data Fig. 1g). Moreover, the doubling time

of cells treated with this chemical cocktail (12.75 h) is significantly shorter than that of cells treated with the CPEC condition (22.15 h) (Extended Data Fig. 1g), which approximates the cleavage dynamics of early mouse embryos (11–18 h)^{32,33}. We further confirmed that this combination activates totipotency marker genes at higher levels compared with treatment with each compound alone (Extended Data Fig. 1c). Collectively, these data indicate that we identified a chemical cocktail that supports rapid induction of totipotent-like cells with improved proliferative ability.

Chemically induced totipotent-like cells exhibit key features of totipotency

Using this cocktail, we attempted to induce totipotent-like cells from mouse EPS cells cultured in AggreWell plates under three-dimensional conditions during stage 1 (S1) (Fig. 1a,b). Treatment of EPS cells with this cocktail led to the rapid formation of cell aggregates (Extended Data Fig. 2a). However, the average diameters of these aggregates are shorter than that of E (embryonic day) 1.5 embryos (Extended Data Fig. 2b), which may be attributed to the relatively larger cellular volume of early blastomeres compared with that of cultured cells (Fig. 1c). A high percentage of these aggregates (114 out of 121 aggregates, 94.2%) expressed the totipotency marker ZSCAN4 (Extended Data Fig. 2c), known to be highly expressed in mouse two-cell embryos (Extended Data Fig. 2d). *MuERV-L-Gag* was also detected in these aggregates (Fig. 1d), which is specifically enriched in two-cell embryos (Fig. 1e).

To evaluate the developmental potentials of these totipotent-like cells, we performed chimeric experiments, which showed that these cells can integrate into embryonic and extraembryonic parts of E4.5 and E6.5 mouse embryos (Fig. 1f and Extended Data Fig. 2e–g). In addition, immunofluorescent analysis of chimeric mouse E4.5 embryos revealed the expression of the trophectoderm (TE) marker CDX2 and the PrE marker SOX17 in the derivatives from totipotent-like cells (Fig. 1f and Extended Data Fig. 2e).

We further performed single-cell RNA sequencing (RNA-seq) to characterize the totipotent-like population, revealing that these cells are transcriptionally distinct from the original EPS cells (Extended Data Fig. 2h). Notably, the pluripotency transcriptional programme was downregulated in the totipotent-like population (Extended Data Fig. 2h and Supplementary Table 2), suggesting exit from the pluripotent state. Consistent with bulk RNA-seq results, totipotency marker genes were found to be upregulated in the totipotent-like population (Extended Data Fig. 2i). In addition, we confirmed the absence of PrE and TE marker gene expression in these cells (Extended Data Fig. 2j), suggesting that extraembryonic lineages were not prematurely activated in the totipotent-like population.

Next, we performed Uniform Manifold Approximation and Projection (UMAP) analysis and identified three distinct subpopulations within the totipotent-like cell population (Extended Data Fig. 2k).

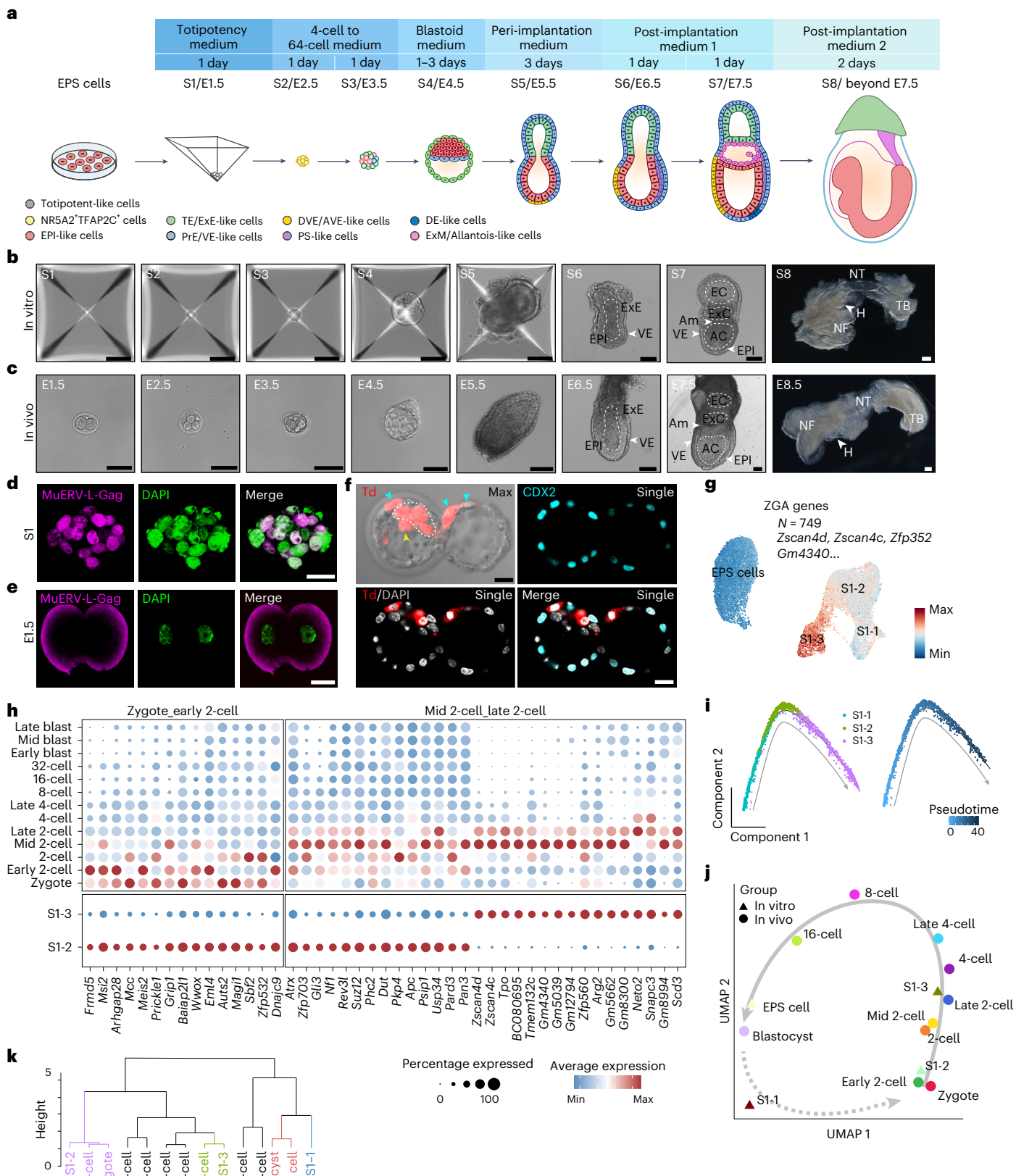
Fig. 1 | Chemically induced totipotent-like cells capture key features of mouse two-cell embryos.

a, Schematic of the stepwise protocol for using chemically induced totipotent-like stem cells to generate an embryo model mimicking different stages of mouse embryogenesis (E1.5 to E7.5 and beyond). PS, primitive streak; DE, definitive endoderm; ExM, extraembryonic mesoderm. **b,c**, Bright-field images of totipotent-like cell-derived embryo-like structures (**b**) at different stages and staged-matched natural embryos (**c**). Scale bars, 100 μ m. NF, neural fold; H, heart; NT, neural tube; TB, tail bud. **n** = 9 embryo-like structures from 9 experiments, **n** = 3 natural embryo from 3 experiments. **d,e**, Immunofluorescent image showing the expression of *MuERV-L-Gag* in induced totipotent-like cells at S1 (**d**) and E1.5 2-cell embryos (**e**). **n** = 5 S1 cell aggregates from 2 experiments, **n** = 3 E1.5 2-cell embryo from 2 experiments. Scale bars, 20 μ m. **f**, Bright-field and immunofluorescent image showing the presence of derivatives of induced totipotent-like cells in E4.5 mouse blastocysts. Td, tdTomato. Td⁺ cells were found to be present in the EPI (dashed white line), PrE (white arrowhead) and TE

(blue arrowhead) region. **n** = 5 from 3 experiments. Scale bars, 20 μ m. **g**, Feature plots showing the expression of ZGA genes in EPS cells and cell aggregates at S1. **h**, Dot plots showing the average levels and proportion of cells expressing specific marker genes for early, middle, and late two-cell embryos in subpopulations 2 and 3 from cell aggregates at S1. Specific marker genes for early, middle and late two-cell embryos were identified using published single-cell transcriptomic datasets (GSE45719 and GSE136714). Expressions of these marker genes in zygotes, early to late 2-cell, 4-cell, 8-cell and 16-cell embryos are shown as the control. **i**, Pseudotime trajectory reconstruction of single cells from the three subpopulations at S1. **j**, UMAP analysis of transcriptome of subpopulations from induced totipotent-like cells at S1. These samples were integrated with published single-cell transcriptomic datasets of early mouse embryo stages from zygotes to blastocysts (GSE45719 and GSE136714). EPS cells were also integrated. **k**, Hierarchical clustering of subpopulations 2 and 3 from induced totipotent-like cells at S1, EPS cells and pre-implantation embryos (GSE45719 and GSE136714).

The proportions of these subpopulations were 17.3% for subpopulation 1 (S1-1), 60.9% for subpopulation 2 (S1-2) and 21.8% for subpopulation 3 (S1-3) (Extended Data Fig. 2k). Further analysis of ZGA gene expression revealed that ZGA gene activation was most prominent in subpopulation 3 (Fig. 1g and Supplementary Table 3). Consistent

with this observation, genes upregulated in subpopulation 3 were predominantly expressed during the middle-to-late two-cell stage (Fig. 1h, Extended Data Fig. 2l and Supplementary Table 4), whereas genes downregulated in this subpopulation were more highly expressed at the blastocyst stage (Extended Data Fig. 2m and Supplementary Table 4).



Compared with subpopulation 3, ZGA gene activation in subpopulation 2 was relatively lower but still higher than that in the EPS cells (Fig. 1g). In addition, subpopulation 2 showed enrichment for marker genes associated with both early and middle-to-late two-cell embryos (Fig. 1h).

To explore the developmental relationship among the three subpopulations, we conducted pseudotime analysis, which revealed a sequential transition from subpopulation 1 to subpopulation 3, passing through subpopulation 2 (Fig. 1i). We further examined the transcriptional similarity between these subpopulations and totipotent blastomeres. Subpopulation 1 was distinct from EPS cells and positioned between blastocysts and totipotent blastomeres (Fig. 1j), suggesting it may represent an intermediate state during totipotency induction. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses revealed that genes enriched in subpopulation 1 were associated with phagocytosis, bacterial infection and placental development (Extended Data Fig. 2n and Supplementary Table 5). Meanwhile, subpopulation 2 mapped closely to early two-cell embryos, while subpopulation 3 aligned with late two-cell to early four-cell blastomeres (Fig. 1j). This developmental trajectory was further supported by hierarchical clustering analysis (Fig. 1k).

We further confirmed that the totipotent-like cells could be stably cultured for over 30 passages under the chemical cocktail comprising CD1530, PD0325901, CHIR-99021 and elvitegravir (Extended Data Fig. 2o,p). In addition, while the expression of pluripotency marker genes was reduced compared with EPS cells, totipotency-associated genes were upregulated (Extended Data Fig. 2p), suggesting that the totipotent-like state can be stably maintained in vitro.

Recapitulation of the diversification of embryonic and extraembryonic lineages from 4-cell to 64-cell stages using chemically induced totipotent-like cells

Next, we tested whether the cell aggregates at S1 could be prompted to exit the totipotent state. Therefore, we tested several small molecules and cytokines known to target signalling pathways and epigenetic regulators involved in early embryo development. Notably, treatment with CD1530, CHIR-99021, birabresib and 8-Br-cAMP for one day promoted further growth of cell aggregates derived from S1 (Fig. 1b and Extended Data Fig. 2a). However, their average diameters remained smaller than those of E2.5 embryos (Fig. 1c and Extended Data Fig. 2b). Importantly, the majority of these aggregates (107/110 aggregates, 97.3%) coexpressed NR5A2 and TFAP2C at stage 2 (S2) (Fig. 2a), resembling the expression pattern observed in E2.5 embryos (Fig. 2b). These factors are key bipotency activators that promote the diversification of embryonic and extraembryonic lineage during 2-cell to 16-cell embryo development^{34,35}. Consistent with this, we observed a transcriptional upregulation of genes regulated by *Nr5a2* and *Tfap2c* in S2 cells compared with subpopulation 3 at S1 (Fig. 2c and Supplementary Table 6). Transcriptome analysis further revealed the expression of marker genes for 4-cell, 8-cell and 16-cell embryos in these S2 cells (Fig. 2d), which

clustered with 4-cell to 16-cell blastomeres (Fig. 2e). These findings were further supported by hierarchical clustering analysis (Fig. 2f).

To investigate the relationship between cells from stages S1 and S2, we performed pseudotime analysis. Subpopulation 1 from S1, characterized by intermediate features, transitioned into subpopulation 2 from S1 (Extended Data Fig. 3a), which exhibited transcriptional signatures of early and middle-to-late 2-cell embryos (Fig. 1h). From this state, a subset of cells progressed to subpopulation 3 from S1, which expressed high levels of *Zscan4*, while others committed to the four-cell-like population at S2 (Extended Data Fig. 3a). The majority of 4-cell-like cells subsequently transitioned into the 16-cell-like population (Extended Data Fig. 3a). Importantly, subpopulation 3 from S1 appeared to be disconnected from the 4-cell- to 16-cell-like developmental trajectory (Extended Data Fig. 3a), consistent with a recent study reporting that only *Zscan4*-negative ZGA-like cells proceed to subsequent stages of development²⁸. In addition, we confirmed that genes enriched in subpopulation 1 from S1, which represents an intermediate state, were significantly downregulated in the subpopulations at S2 (Extended Data Fig. 3b). This observation, together with the pseudotime analysis, suggest that the intermediate cells from S1 are unlikely to have a direct influence on the formation of S2 cells.

Following an additional day of induction with CD1530, CHIR-99021, birabresib and 8-Br-cAMP, the cell aggregates continued to proliferate at stage 3 (S3). However, their average diameters remained smaller than those of E3.5 embryos (Fig. 1c and Extended Data Fig. 2b). Notably, immunofluorescence analysis showed that a high percentage of S3 aggregates (111 out of 127, 87.4%) contained cells positive for OCT4, CDX2 and SOX17 (Fig. 2g), exhibiting an expression pattern similar to that of E3.5 embryos (Fig. 2h). In support of this, these cells also expressed marker genes for 32-cell and 64-cell embryos (Fig. 2i). Analyses of UMAP and hierarchical clustering further showed that these cells were mapped to 32-cell to 64-cell embryos (Fig. 2j,k).

We further investigated the roles of CD1530, CHIR-99021, birabresib and 8-Br-cAMP during induction at S2 and S3. To this end, we systematically removed one compound at a time from the base combination and collected cells treated under each modified condition for bulk RNA-seq. Omission of birabresib resulted in a failure to downregulate totipotency marker genes (Extended Data Fig. 3c). In addition, removal of either 8-Br-cAMP or CD1530 led to reduced expression of *Tfap2c* and *Nr5a2* target genes (Extended Data Fig. 3d). We also assessed the impact of each compound on the expression of lineage-specific marker genes associated with the epiblast (EPI), TE and PrE at S3. While exclusion of 8-Br-cAMP led to a broad reduction in marker gene expression across all three lineages, omission of the other three compounds generally had the opposite effect (Extended Data Fig. 3e). In addition, CHIR-99021 showed a strong inductive effect on *Cdx2* expression (Extended Data Fig. 3e).

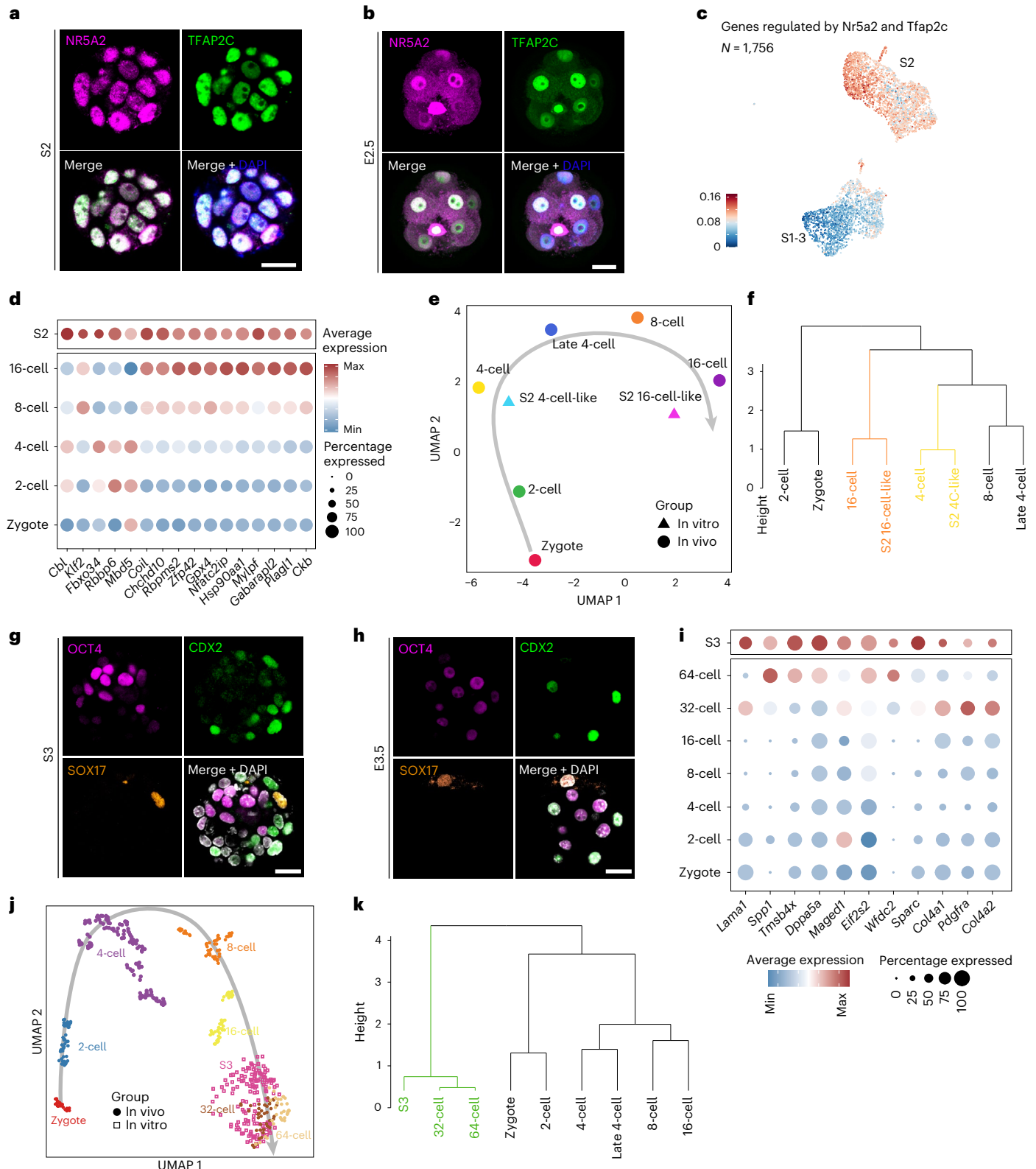
Fig. 2 | Totipotent-like cell-derived aggregates resemble mouse early embryos from the 4-cell to 64-cell embryo stages. a,b, Immunofluorescent analysis showing the expression of NR5A2 and TFAP2C in cell aggregates at S2 (a) and E2.5 embryos (b). *n* = 6 S2 cell aggregates from 2 experiments, *n* = 3 E2.5 embryos from 2 experiments. Scale bars, 20 μ m. **c**, Feature plots showing the expression of genes regulated by *Nr5a2* and *Tfap2c* in subpopulation 2 from S1 and cell aggregates at S2. Genes regulated by *Nr5a2* and *Tfap2c* were identified using published bulk transcriptomic datasets (GSE229740 and GSE216256). **d**, Dot plots showing the average levels and proportion of cells expressing stage-specific marker genes in cell aggregates at S2. Developmental stage-specific marker genes for zygotes, 2-cell, 4-cell, 8-cell and 16-cell embryos were identified using published single-cell transcriptomic datasets (GSE136714). Expressions of these marker genes in zygotes, 2-cell, 4-cell, 8-cell and 16-cell embryos are shown as the control. **e**, UMAP analysis of transcriptome of totipotent-like cell-derived cell aggregates at S2. These were integrated with published single-cell transcriptomic datasets from early mouse embryo stages from zygotes to

16-cell embryos (GSE136714). **f**, Hierarchical clustering of cell aggregates at S2 and early mouse embryo stages from zygotes to 16-cell embryos (GSE136714). **g,h**, Immunofluorescent analysis showing the expression of OCT4, CDX2 and SOX17 in cell aggregates at S3 (g) and E3.5 embryos (h). *n* = 6 S3 cell aggregates from 2 experiments, *n* = 3 E3.5 embryos from 2 experiments. Scale bars, 20 μ m. **i**, Dot plots showing the average levels and proportion of cells expressing stage-specific marker genes in cell aggregates at S3. Developmental stage-specific marker genes for zygotes, 2-cell, 4-cell, 8-cell, 16-cell, 32-cell and 64-cell embryos were identified using published single-cell transcriptomic datasets (GSE136714 and GSE84892). Expressions of these marker genes in embryos from zygotes to 64-cell embryos are shown as the control. **j**, UMAP analysis of transcriptome of totipotent-like cell-derived cell aggregates at S3. These samples were integrated with published single-cell transcriptomic datasets of early mouse embryo stages from zygotes to 64-cell embryos (GSE136714 and GSE84892). **k**, Hierarchical clustering of cell aggregates at S3 and early mouse embryo stages from zygotes to 64-cell embryos (GSE136714 and GSE84892).

Further induction of blastoids from totipotent-like cell-derived 64-cell-like aggregates

To further promote the formation of blastocyst-like structures (blastoids) from 64-cell-like aggregates, we tested different combinations of compounds targeting signalling pathways involved in blastocyst development. Notably, we identified a combination of signalling agonists for WNT (CHIR-99021), YAP (GA-017), FGF (bFGF) and BMP

(BMP-4) that efficiently induced blastoid formation, with a success rate of 75.9% following 3 days of treatment at stage 4 (S4) (Figs. 1b and 3a and Extended Data Fig. 2a). Immunostaining analysis revealed that these blastoids expressed key marker genes for EPI (OCT4), TE (CDX2 and CK8) and PrE (SOX17 and GATA6) (Fig. 3b,d and Extended Data Fig. 4a), mimicking the expression patterns observed in E4.5 embryos (Fig. 3c,e and Extended Data Fig. 4b). Moreover, both



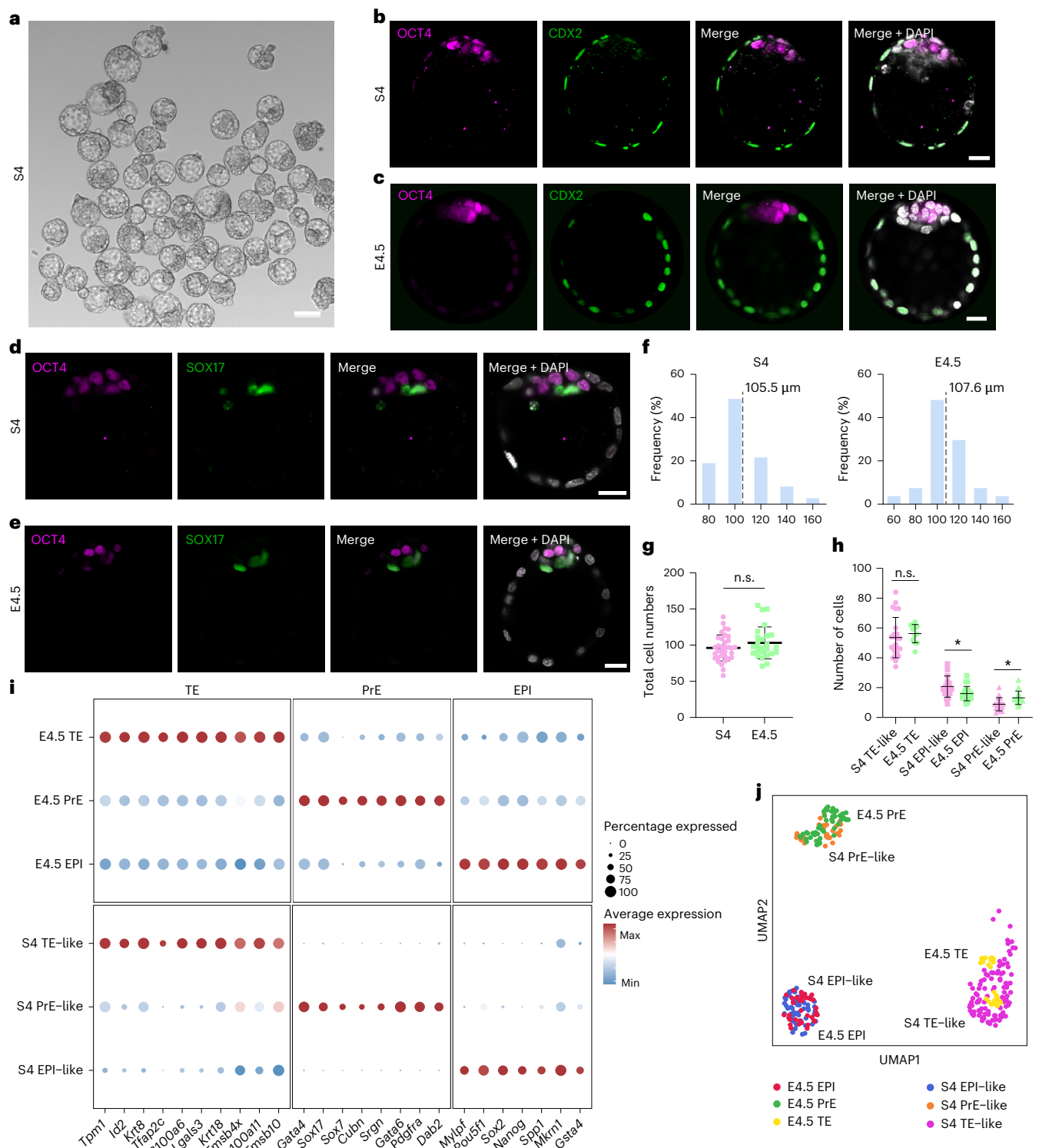


Fig. 3 | Induction of blastocyst-like structures from chemically induced totipotent-like cells. a, Bright-field image of blastoids at S4. Scale bar, 100 μm . Similar images were obtained in at least three independent experiments. **b, c**, Immunofluorescent analysis showing the expression of OCT4 and CDX2 in blastoids at S4 (**b**) and E4.5 blastoids (**c**). $n = 9$ S4 blastoids from 6 experiments, $n = 3$ E4.5 blastoids from 2 experiments. Scale bars, 20 μm . **d, e**, Immunofluorescent analysis showing the expression of OCT4 and SOX17 in blastoids at S4 (**d**) and E4.5 blastoids (**e**). $n = 7$ S4 blastoids from 5 experiments, $n = 3$ E4.5 blastoids from 2 experiments. Scale bars, 20 μm . **f**, Histograms showing the distribution of the diameters of S4 blastoids and E4.5 blastoids. $n = 37$ S4 blastoids, $n = 27$ E4.5 blastoids. The vertical dashed line denotes the mean of the group. **g**, Histograms showing the total cell number of S4 blastoids

and E4.5 blastoids. Data are presented as mean $\pm$ s.d. $n = 34$ S4 blastoids, $n = 26$ E4.5 blastoids. n.s., non-significant (two-sided Student's t -test). **h**, Histograms showing the indicated cell number of S4 blastoids and E4.5 blastoids. Data are presented as mean $\pm$ s.d. $n = 62$ S4 blastoids, $n = 47$ E4.5 blastoids. * $P < 0.05$; n.s., non-significant (two-sided Student's t -test). **i**, Dot plots showing the average levels and proportion of cells expressing lineage-specific marker genes in blastoids at S4. Lineage-specific marker genes for EPI, TE and PrE in E4.5 mouse blastocysts were identified using published single-cell transcriptomic datasets (GSE123046). Expressions of these marker genes in EPI, TE and PrE from E4.5 mouse blastocysts are shown as the control. **j**, UMAP analysis of transcriptome of cells from blastoids at S4. These samples were integrated with published single-cell transcriptomic datasets from E4.5 mouse blastocysts (GSE123046).

the average diameters and total cell numbers of blastoids were comparable to those of E4.5 embryos (Fig. 3f,g and Extended Data Fig. 2b). However, we also observed a higher proportion of EPI-like cells and a lower proportion of PrE-like cells in blastoids compared with E4.5 blastocysts (Fig. 3h), a phenomenon that has also been reported in previous blastoid studies^{13,14,28}.

Next, we performed single-cell RNA-seq analysis on S4 blastoids. Consistent with the immunostaining result, we identified three distinct subpopulations that expressed marker genes for the EPI, TE and PrE lineages (Fig. 3i). Furthermore, UMAP analysis showed that these subpopulations resembled the EPI, TE and PrE lineages found in E4.5 blastocysts (Fig. 3j).

We further examined the roles of CHIR-99021, GA-017, bFGF and BMP-4 in blastoid induction at S4. Removal of any single compound from the S4 cocktail resulted in a significant reduction in cavity formation (Extended Data Fig. 4c,d). Consistent with the established role of the Hippo–YAP signalling pathway in TE specification^{36,37}, omission of the LATS1/2 inhibitor GA-017 produced the most pronounced decrease in blastoid formation efficiency (Extended Data Fig. 4c,d). Similarly, removal of the WNT signalling agonist CHIR-99021 led to a dramatic reduction in blastoid formation (Extended Data Fig. 4c,d), in agreement with previous studies highlighting the role of WNT signalling in regulating blastoid induction^{12,14,23,27}.

To assess the developmental relevance of S4 blastoids, we first examined whether the process of blastoid formation could respond to developmental perturbations in a manner similar to natural embryos. Given that endogenous retinoic acid activity is critical for ZGA and developmental progression in early mouse embryos³⁸, we treated S1 cultures with the RAR γ antagonist LY2955303. Notably, this pharmacological inhibition impaired the expression of totipotency marker genes at S1 (Extended Data Fig. 5a). Moreover, consistent with the effect of LY2955303 observed in natural early mouse embryos (Extended Data Fig. 5b), treatment with LY2955303 also significantly reduced blastoid formation efficiency at stage S4 (Extended Data Fig. 5c,d). We next evaluated the role of RHO–ROCK signalling, which is reported to suppress TE formation in mouse blastocysts through activation of the Hippo signalling pathway³⁹. Consistent with observations in mouse early embryos (Extended Data Fig. 5e), treatment with the ROCK inhibitor Y27632 reduced the efficiency of blastoid formation (Extended Data Fig. 5f,g). To further explore the developmental competence of S4 blastoids, we transplanted them into pseudo-pregnant mice at 2.5 days post coitum (dpc). These blastoids induced decidualization in vivo (Extended Data Fig. 5h). However, only degenerated structures were detected within the decidua.

Early-stage blastoids derived from totipotent-like cells can form post-implantation embryoids

Next, we investigated whether the blastoids could further develop into post-implantation embryo-like structures. Because the removal of mural TE from the E4.5 blastocysts facilitates their development into post-implantation embryos in vitro⁴⁰, we reasoned that treatment with the blastoid medium for 3 days could induce numerous mural TE-like cells in the blastoids, potentially hindering the transition of blastoids into post-implantation stages. Therefore, we focused on using early-stage blastoids induced by a 1-day treatment with the blastoid medium.

Given that the proliferation of trophoblast stem cells is critical for the formation of peri-implantation embryos⁴¹, we included FGF-4 and Activin A in the peri-implantation medium, which are beneficial for proliferation of trophoblast stem cells^{42,43}. In addition, BMP-4 was added because BMP signalling plays a crucial role in both embryonic and extraembryonic development during the peri-implantation stage⁴⁴. We also incorporated XAV939, a WNT signalling inhibitor, because it has been reported to promote the transition of EPI from naive pluripotency to primed pluripotency in peri-implantation embryos⁴⁵. Notably, after combined treatment with these four compounds, 50.2% of the early

blastoids developed into embryoids that morphologically resembled E5.0–E5.5 embryos at stage 5 (S5), with the presence of egg cylinder-like structures surrounded by a layer of visceral endoderm (VE)-like cells (Fig. 1b and Extended Data Fig. 2a). While the average width of these egg cylinder-like structures was comparable to those of E5.5 embryos, their average length was smaller (Extended Data Fig. 6a).

To further analyse the early egg cylinder-like embryoids, we performed immunostaining. Similar to E5.5 embryos, OCT4⁺ and CDX2⁺ cells formed two separate rosette-like structures (Fig. 4a,b). Single-cell transcriptomic analysis showed the presence of EPI-, extraembryonic ectoderm (ExE)- and VE-like cells within these structures (Fig. 4c and Extended Data Fig. 6b). GO analysis also showed that genes enriched in the EPI-, ExE- and VE-like cells are involved in the specification and development of these lineages in vivo (Extended Data Fig. 6c and Supplementary Table 7). Moreover, UMAP analysis revealed that these EPI-, ExE- and VE-like cells clustered with their in vivo counterparts (Fig. 4d). Notably, we also detected the expression of *Cer1* in the VE-like cells (Fig. 4e), suggesting the initiation of distal VE (DVE)/anterior VE (AVE) formation.

Next, we evaluated the individual contributions of Activin A, BMP-4, FGF-4 and XAV939 in inducing E5.5 embryo-like structures. Removal of any single component from the S5 cocktail resulted in a marked downregulation of ExE marker genes (Extended Data Fig. 6d). In terms of VE development, omission of BMP-4, FGF-4 or XAV939 impaired the expression of VE markers, whereas removing Activin A led to an increase in their expression (Extended Data Fig. 6d). Regarding EPI development, exclusion of Activin A reduced the expression of EPI-associated genes, while omission of XAV939 resulted in their upregulation (Extended Data Fig. 6d).

Post-implantation embryoids derived from totipotent-like cells can progress to the gastrulation stage

The successful generation of E5.5-like embryoids at S5 raised the question of whether gastrulation could occur in these structures. To investigate this, we transferred the E5.5-like embryoids from the AggreWell plates to six-well non-adherent suspension culturing plates at stage 6 (S6) (Fig. 1a), and switched the culture medium into the modified IVC1 medium⁴⁰. After 1 day of culturing, we observed 30.4% of the E5.5 embryoids can develop into embryo-like structures that morphologically resemble E6.5 embryos (Fig. 1b and Extended Data Fig. 2a), with the emergence of an expanded pro-AC surrounded by a VE-like layer. However, the average length of the E6.5-like embryoids was smaller than those of E6.5 embryos (Extended Data Fig. 6a).

Consistent with the observed morphological changes, immunostaining analysis revealed expanded OCT4⁺ and CDX2⁺ lumens within the E6.5-like embryoids (Fig. 4f), mimicking the expression pattern of OCT4 and CDX2 in E6.5 embryos (Fig. 4g). In addition, 70.5% of the E6.5-like embryoids contained both OCT4⁺ and SOX17⁺ cells, with 35.0% of these structures exhibiting an expanded OCT4-positive lumen surrounded by a SOX17-positive layer (Fig. 4h and Extended Data Fig. 6e), mimicking the spatial organization in E6.5 embryos (Fig. 4i). To assess whether the anterior–posterior axis had been successfully established in the E6.5-like embryoids, we examined the expression pattern of CER1, a marker of the AVE. 65.2% of S6 embryoids contained CER1-positive cells (Extended Data Fig. 6f). Among these, 45.6% displayed localization of CER1-positive cells to the distal region, while 54.4% showed an asymmetric pattern of CER1 expression on one side of the structure (Fig. 4j and Extended Data Fig. 6f), resembling the expression pattern observed in E6.5 embryos (Fig. 4k).

After 2 days of culturing in the modified IVC1 medium, 42.8% of the embryoids became more complex at stage 7 (S7), displaying different compartments resembling the amnion (Am), amniotic cavity (AC), exocoelomic cavity (ExC) and ectoplacental cavity (EC) (Fig. 1b and Extended Data Figs. 2a and 7a). However, both the average width and length of S7 embryoids were smaller compared with those of E7.5 embryos (Extended Data Fig. 6a).

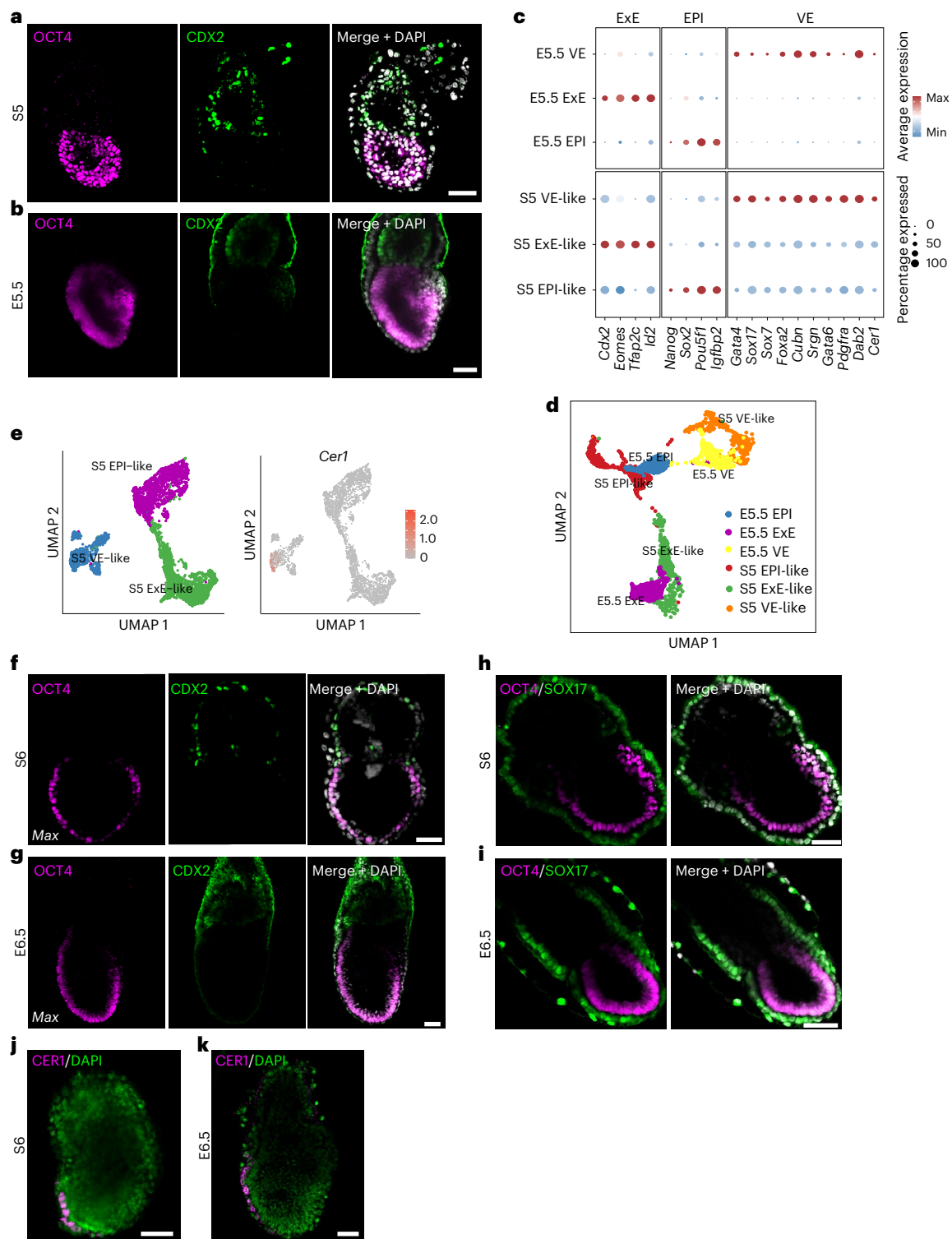


Fig. 4 | Totipotent-like cell-derived early blastoids develop into peri- and post-implantation embryoids. a, b, Immunofluorescent analysis showing the expression of OCT4 and CDX2 in peri-implantation embryo-like structures at S5 (a) and E5.5 embryos (b). $n = 9$ S5 embryos from 5 experiments, $n = 3$ E5.5 embryos from 2 experiments. Scale bars, 50 μ m. **c,** Dot plots showing the average levels and proportion of cells expressing lineage-specific marker genes in embryo-like structures at S5. Lineage-specific marker genes for EPI, ExE and VE in E5.5 mouse embryos were identified using published single-cell transcriptomic datasets (GSE123046). Expressions of these marker genes in EPI, ExE and VE from E5.5 mouse embryos are shown as the control. **d,** UMAP analysis of transcriptome of cells from peri-implantation embryo-like structures at S5. These were integrated with published single-cell transcriptomic datasets from E5.5 mouse

embryos (GSE123046). **e,** UMAP analysis showing the expression of *Cer1* in VE-like cells from embryo-like structures at S5. **f, g,** Immunofluorescent analysis showing the expression of OCT4 and CDX2 in post-implantation embryo-like structures at S6 (f) and E6.5 embryos (g). $n = 7$ S6 embryos from 6 experiments, $n = 3$ E6.5 embryos from 2 experiments. Scale bars, 50 μ m. **h, i,** Immunofluorescent analysis showing the expression of OCT4 and SOX17 in embryo-like structures at S6 (h) and E6.5 embryos (i). $n = 6$ S6 embryos from 4 experiments, $n = 2$ E6.5 embryos from 2 experiments. Scale bars, 50 μ m. **j, k,** Immunofluorescent analysis showing the expression of CER1 in post-implantation embryo-like structures at S6 (j) and E6.5 embryos (k). $n = 6$ S6 embryos from 2 experiments, $n = 2$ E6.5 embryos from 2 experiments. Scale bars, 50 μ m.

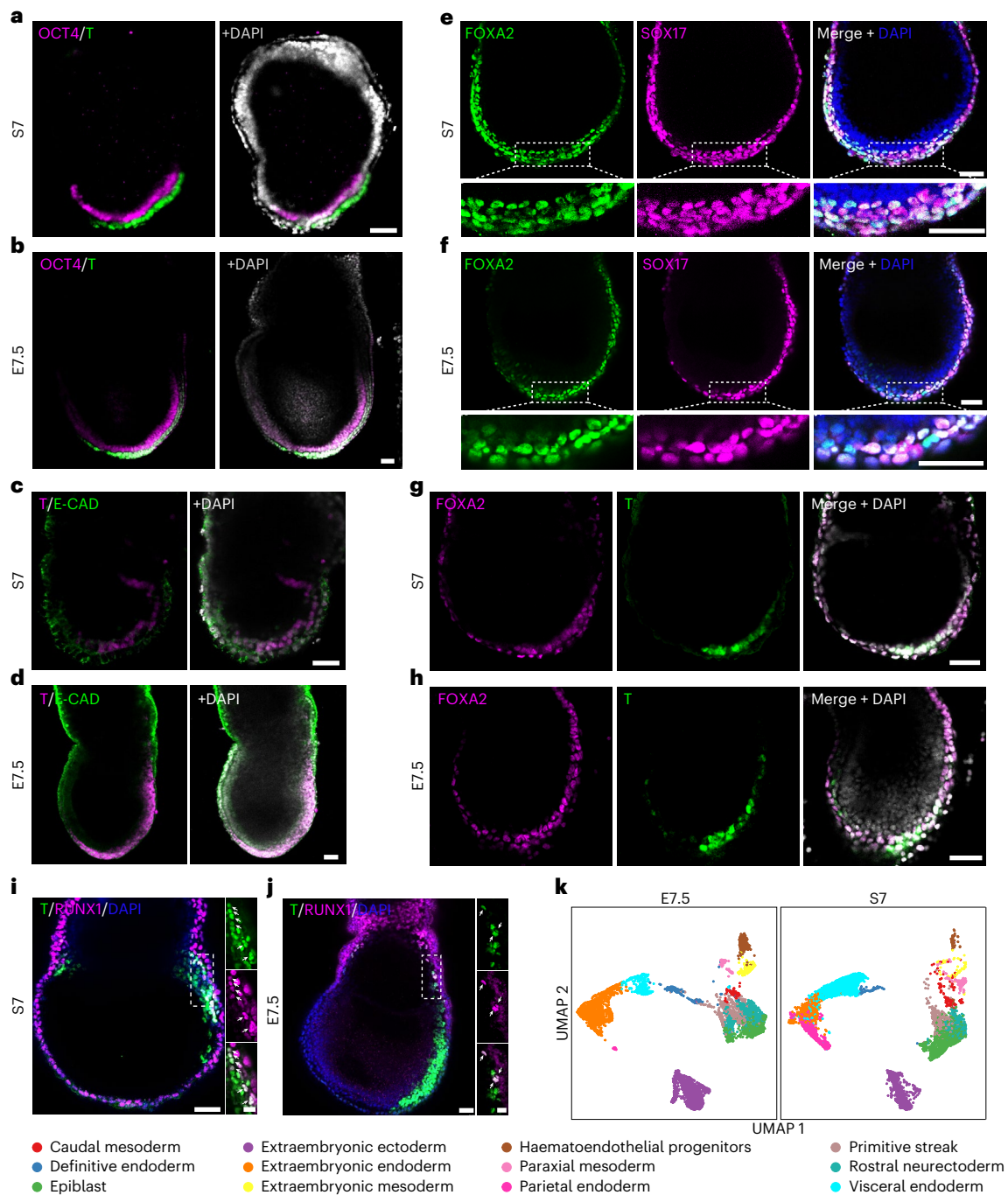


Fig. 5 | Totipotent-like cell-derived post-implantation embryo-like structures capture key developmental events during gastrulation. a,b, Immunofluorescent analysis showing the expression of OCT4 and T in post-implantation embryo-like structures at S7 (**a**) and E7.5 embryos (**b**). $n = 9$ S7 embryo-like structures from 4 experiments, $n = 4$ E7.5 embryos from 2 experiments. Scale bars, 50 μm . **c,d**, Immunofluorescent analysis showing the expression of T and E-CADHERIN in post-implantation embryo-like structures at S7 (**c**) and E7.5 embryos (**d**). Down-regulation of E-CADHERIN in T⁺ cells indicate epithelial-to-mesenchymal transition in the primitive streak (PS) area. $n = 2$ S7 embryo-like structures from 2 experiments, $n = 2$ E7.5 embryos from 2 experiments. Scale bars, 50 μm . **e,f**, Immunofluorescent analysis showing the expression of SOX17 and FOXA2 in post-implantation embryo-like structures at S7 (**e**) and E7.5 embryos (**f**). Bottom images, magnified field showing the presence of DE-like cells in the anterior parts of the PS area. $n = 6$ S7 embryo-like structures from 2 experiments, $n = 4$ E7.5 embryos from

2 experiments. Scale bars, 50 μm . **g,h**, Immunofluorescent analysis showing the expression of FOXA2 and T in post-implantation embryo-like structures at S7 (**g**) and E7.5 embryos (**h**). FOXA2⁺T⁺ cells indicate the presence of axial mesoderm-like cells. $n = 4$ S7 embryo-like structures from 3 experiments, $n = 3$ E7.5 embryos from 2 experiments. Scale bars, 50 μm . **i,j**, Immunofluorescent analysis showing the expression of T and RUNX1 in post-implantation embryo-like structures at S7 (**i**) and E7.5 embryos (**j**). Insets are enlargements of the dashed boxes, indicating extraembryonic mesoderm. Arrow, T- and RUNX1-double-positive cells. $n = 6$ S7 embryo-like structures from 2 experiments, $n = 2$ E7.5 embryos from 2 experiments. Scale bars, 50 μm (main image) and 20 μm (magnified view). **k**, UMAP analysis of transcriptome of cells from embryo-like structures at S7. These samples were integrated with published single-cell transcriptomic datasets from E7.5 mouse embryos (E-MTAB-6967).

To further investigate the progression of gastrulation in post-implantation embryoids, we examined the expression of the primitive streak marker *T* (Brachyury, *Bra*). In total, 53.6% of the analysed S7 embryoids contained *T*-positive cells (Extended Data Fig. 7b). Among these *T*-positive structures, 41.6% exhibited asymmetric *T* expression extending from the posterior side of the EPI to its distal region (Fig. 5a and Extended Data Fig. 7b). Furthermore, co-staining for *T* and E-CADHERIN revealed the downregulation of E-CADHERIN in *T*-positive cells (Fig. 5c), implying epithelial-to-mesenchymal transition in the primitive streak region. In addition, 53.5% of S7 embryoids displayed coexpression of *SOX17* and *FOXA2* in the distal region of the primitive streak (Fig. 5e and Extended Data Fig. 7c), suggesting the emergence of definitive endoderm-like cells. We also detected the presence of *T* and *FOXA2* double-positive cells in the distal region of the primitive streak-like area (Fig. 5g), suggesting the formation of axial mesoderm-like cells. Moreover, coexpression of *T* and *RUNX1* was also observed (Fig. 5i), marking extraembryonic mesoderm-like cells committing to the hematopoietic lineage. The expression patterns of these analysed lineage markers in S7 embryoids mimic those observed in E7.5 embryos (Fig. 5b,d,f,h,j).

To assess the lineage composition of S7 embryoids, we performed single-cell RNA-seq analysis. UMAP analysis revealed the presence of different embryonic and extraembryonic lineages (Fig. 5k), which contained cell types from all three germ layers, as well as extraembryonic lineages corresponding to the ExE and VE (Fig. 5k). We further confirmed that the embryonic and extraembryonic lineages expressed corresponding marker genes of E7.5 embryos (Extended Data Fig. 7d,e). Consistent with these findings, most of the lineages displayed a high correlation ($R > 0.9$) with the corresponding lineages in E7.5 embryos (Extended Data Fig. 7f). Within the ExE cluster, we also detected diverse subpopulations expressing marker genes characteristic of chorion progenitors, chorion, ectoplacental cone and progenitors of trophoblast giant cells (Extended Data Fig. 7g).

Chemically induced totipotent-like cell-based embryo model recapitulates the developmental trajectories from ZGA to gastrulation

To determine whether our embryo model recapitulates the developmental trajectories of mouse embryogenesis, we first identified gene sets specific to the 2-cell, 4-cell to 16-cell, 32/64-cell, E4.5, E5.5 and E7.5 stages respectively (Extended Data Fig. 8a,b and Supplementary Table 8). Notably, cell aggregates and embryo-like structures, from S1 to S7, sequentially exhibited stage-specific enrichment of gene sets corresponding to their *in vivo* counterparts (Fig. 6a). Consistent with this, GO analysis also showed stage-specific shared GO terms between *in vitro* cell aggregates or embryo-like structures and their *in vivo* counterparts (Fig. 6a and Extended Data Fig. 8a,b).

To further investigate the developmental trajectories of *in vitro* cell aggregates and embryo-like structures, we performed pseudotime analysis. This analysis revealed a sequential progression from S1 to S3 via S2 (Fig. 6b), suggesting recapitulation of the developmental transition from the 2-cell to the 64-cell stage. Moreover, we observed a continuous trajectory from S3 to S7 (Fig. 6c), suggesting further progression from the pre-implantation stage to the onset of gastrulation. Notably, pseudotime analysis also captured the progression of extraembryonic lineages from S5 to S7 (Extended Data Fig. 8c,d). Consistent with this, a Sankey diagram depicting lineage transitions from S5 to S7 showed that both embryonic and extraembryonic cells at S5 gave rise to their respective lineages at S7 (Fig. 6d).

Next, we integrated single-cell transcriptomic data from stages S1 to S7 and performed UMAP analysis to visualize the transcriptional landscape. Cells from S1, S2 and S3 exhibited closer clustering compared with cells from later stages (Fig. 6e). Blastoid cells from S4 segregated into three distinct clusters corresponding to EPI-, TE- and PrE-like lineages (Fig. 6e). Furthermore, the post-implantation derivatives of

these lineages from S5 to S7 were found to cluster more closely with their respective progenitor populations at S4 than with unrelated lineages (Fig. 6e), supporting the notion of lineage continuity throughout the modelled developmental progression.

Totipotent-like cell-derived post-implantation embryoids can develop beyond the gastrulation stage

Finally, we attempted to extend the *in vitro* development of E7.5-like embryoids using a three-dimensional roller culture system⁴⁶. After two additional days of culturing, 59.6% of the S7 embryoids progressed into more complex embryo-like structures surrounded by an expanded yolk sac-like structure, reaching stage 8 (S8) (Fig. 7a and Extended Data Fig. 2a). Morphological abnormalities were also observed, such as empty yolk sacs or embryo-like structures emerging outside the yolk sac (Extended Data Fig. 9a). The average width of these S8 embryoids was comparable to those of E8.5 embryos, although their average length remained smaller (Extended Data Fig. 6a). Within the properly developed embryonic regions, structures displaying distinct morphological features of headfolds and tail buds were observed (Fig. 1b).

To determine whether lineage specification associated with early organogenesis occurs in S8 embryoids, we performed immunostaining to assess the spatiotemporal expression of lineage-specific markers. In the headfold-like region, *OTX2* was expressed in the anterior part (Fig. 7b), consistent with its role in marking the forebrain and midbrain in E8.5 embryos (Fig. 7c). In addition, the neuroepithelium marker *SOX1* was detected along the anteroposterior axis and within rostral neural fold-like structures (Fig. 7d and Extended Data Fig. 9b), mimicking the expression pattern observed in E8.5 embryos (Fig. 7e and Extended Data Fig. 9c). In the anterior ventral region, coexpression of the cardiac markers Myosin Heavy Chain II (MHC-II) and *GATA6* was detected (Extended Data Fig. 9d), suggesting initiation of heart tube-like structures, as observed in E8.5 embryos (Extended Data Fig. 9e). At the posterior end, coexpression of *SOX1* and *T* was observed in the tail bud-like region (Fig. 7f), implying a neuro-mesodermal progenitor population that is observed in E8.5 embryos (Fig. 7g). In addition to the embryonic compartment, we also detected *RUNX1*-positive hematopoietic progenitors within the yolk sac (Fig. 7h), similar to those observed in E8.5 embryos (Fig. 7i).

To characterize the cellular diversity within the embryonic compartment of S8 embryoids, we performed single-cell RNA-seq analysis using the embryo part from S8 embryoids. Clustering based on differentially expressed genes (DEGs) revealed distinct cell states, which were annotated using known marker genes for major lineages present in E8.5 mouse embryos (Fig. 7j). This analysis identified 25 lineage clusters present in both S8 embryoids and E8.5 embryos (Fig. 7j), encompassing derivatives of all three germ layers (Fig. 7j). Each lineage expressed marker genes consistent with their *in vivo* counterparts (Extended Data Fig. 9f). Furthermore, Pearson correlation analysis showed that most lineages in S8 embryoids exhibited a high transcriptional similarity to those in E8.5 embryos (Extended Data Fig. 9g).

Collectively, these findings indicate that post-implantation embryoids derived from totipotent-like cells can develop beyond the gastrulation stage and initiate several hallmarks of early organogenesis. Notably, the stepwise induction protocol established in this study was independently reproduced by another laboratory, supporting its robustness and reproducibility (Extended Data Fig. 10).

Discussion

In this study, we established a chemical cocktail to induce totipotent-like cells with improved proliferation ability, and utilized these cells to construct an embryo model that recapitulates key developmental events during mouse embryogenesis, including ZGA, morula and blastocyst formation, and subsequent generation of post-implantation egg cylinders. Furthermore, these post-implantation embryo-like structures could progress to E7.5-like embryos and model gastrulation.

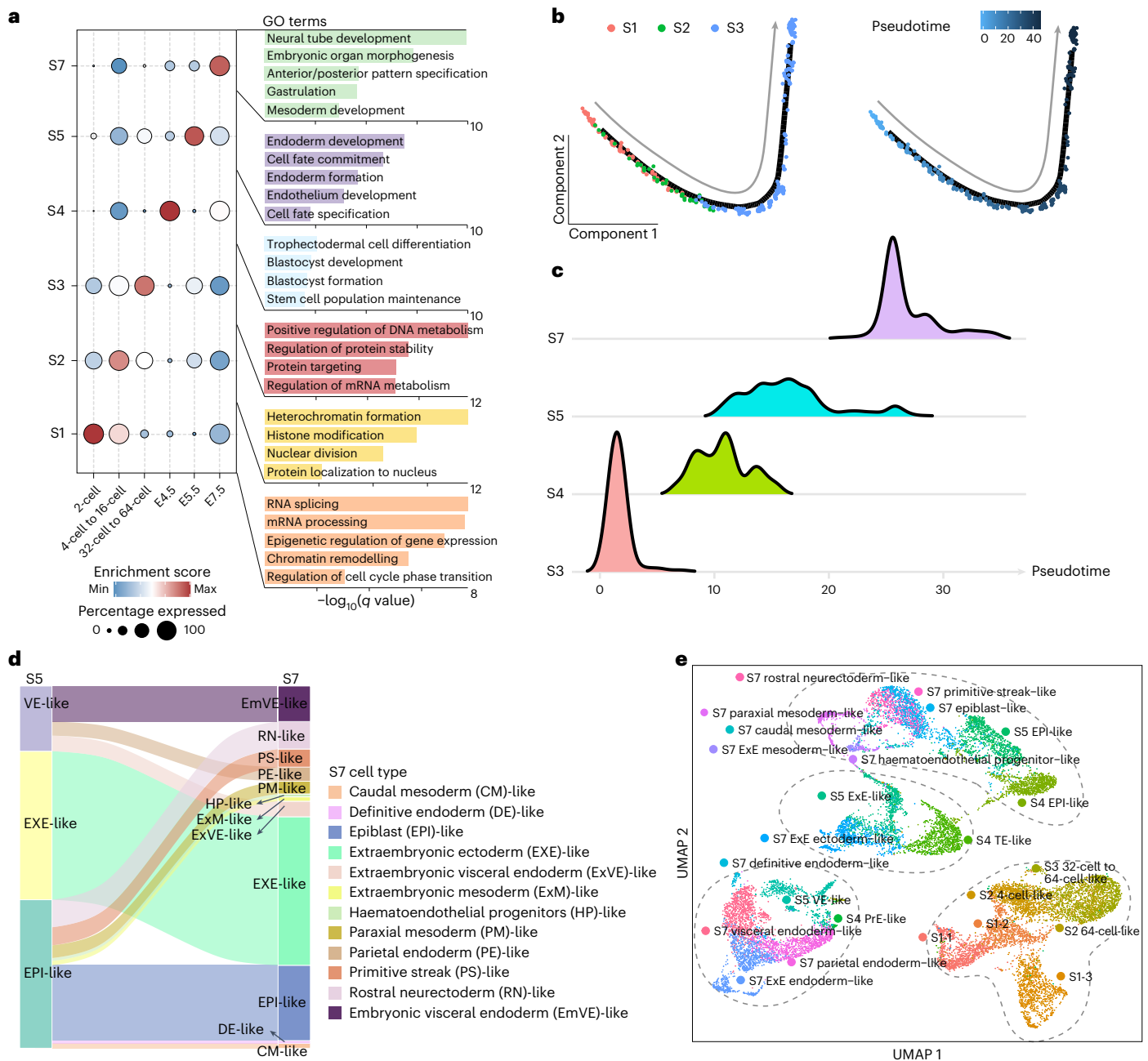


Fig. 6 | Totipotent-like cell-based embryo model recapitulates the developmental trajectories from ZGA to gastrulation. **a**, Dot plots showing the average levels and proportion of cells expressing stage-specific gene sets in cell aggregates or embryo-like structures from S1 to S7 (left) and natural embryos from 2-cell embryo stage to E7.5 (right). Stage-specific gene sets for 2-cell, 4-cell to 16-cell, 32-cell to 64-cell, E4.5, E5.5 and E7.5 mouse embryos were identified using published single-cell transcriptomic datasets (GSE136714, GSE45719, GSE84892, GSE100597, GSE123046 and E-MTAB-6967). GO terms that

are specifically enriched in cell aggregates or embryo-like structures from each in vitro stage are shown, which are also shared by their corresponding in vivo counter parts. **b**, Pseudotime trajectory reconstruction of single cells from S1 to S3. **c**, Joy plots showing the pseudotime distribution in samples of S3, S4, S5 and S7. **d**, Application of a Sankey diagram to visualize developmental trajectory with single-cell transcriptomic data of samples from S5 and S7. Different developmental branches are simultaneously displayed on the Sankey diagram. **e**, UMAP analysis of transcriptome of cells from S1 to S7.

One major advantage of our embryo model is its ability to continuously recapitulate mouse embryogenesis in vitro. Activation of ZGA genes was observed at S1 (Fig. 1g), followed by the diversification of embryonic and extraembryonic lineage at S2 and S3 (Fig. 2a,g). Furthermore, 32-cell-to-64-cell-like aggregates were induced into blastocyst-like structures at S4 (Fig. 3a,b), which further progressed to form peri-implantation embryo-like structures at S5 (Fig. 4a). More importantly, the S5 embryoids could initiate gastrulation from S6 to S7 (Figs. 4f,h,j and 5a,c,e,g,i). Notably, pseudotime analysis further supports the continuous reconstruction of developmental

trajectories across mouse embryogenesis in our model (Fig. 6b,c). Moreover, embryoids can further develop beyond the E7.5 stage in a roller culture system (Fig. 7 and Extended Data Fig. 9). While these results demonstrate the feasibility of using totipotent-like cells to continuously model mouse embryogenesis, the induction conditions for generating these cells require further optimization, such as extending the induction duration or refining the chemical cocktail, to eliminate the intermediate cell populations observed at the S1 stage, which may interfere with the subsequent modelling. Furthermore, more rigorous analyses, such as lineage tracing and comprehensive profiling of

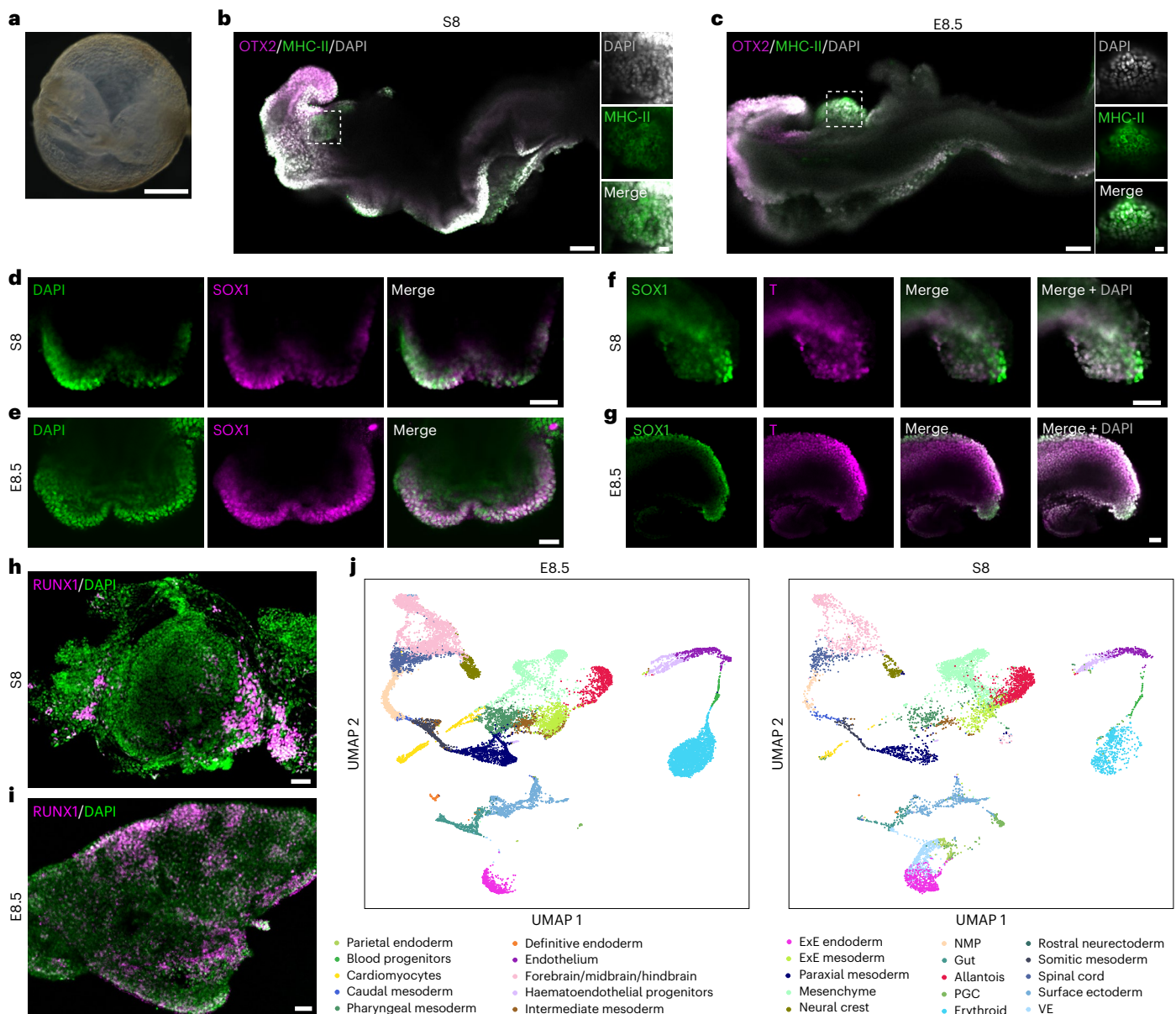


Fig. 7 | Totipotent-like cell-derived embryoids capture hallmarks of early organogenesis. **a**, Representative bright-field image of whole embryo-like structures at S8. $n = 8$ from 5 experiments. Scale bar, 500 μm . **b,c**, Immunofluorescent analysis showing the expression of OTX2 and MHC-II in post-implantation embryo-like structures at S8 (**b**) and E8.5 embryos (**c**). Insets are enlargements of the dashed boxes, indicating the formation of heart region. $n = 3$ S8 embryos from 3 experiments, $n = 2$ E8.5 embryos from 2 experiments. Scale bars, 50 μm . **d,e**, Immunofluorescent analysis showing the expression of SOX1 in post-implantation embryo-like structures at S8 (**d**) and E8.5 embryos (**e**), indicating the formation of the neural folds. $n = 2$ S8 embryos from 2 experiments, $n = 2$ E8.5 embryos from 2 experiments. Scale bars, 100 μm .

f,g, Immunofluorescent analysis showing the expression of SOX1 and T in post-implantation embryo-like structures at S8 (**f**) and E8.5 embryos (**g**), indicating the formation of tail bud. $n = 2$ S8 embryos from 2 experiments, $n = 2$ E8.5 embryos from 2 experiments. Scale bars, 50 μm . **h,i**, Immunofluorescent analysis showing the expression of RUNX1 in dissected yolk sac-like membrane from S8 embryo-like structures (**h**) and E8.5 embryos (**i**). $n = 5$ S8 embryos from 3 experiments, $n = 3$ E8.5 embryos from 3 experiments. Scale bars, 50 μm . **j**, UMAP analysis of transcriptome of cells from embryo-like structures at S8. NMP, neuromesodermal progenitor; PGC, primordial germ cell. These samples were integrated with published single-cell transcriptomic datasets from E8.5 mouse embryos (E-MTAB-6967).

epigenetic modifications, are necessary to elucidate the similarities and differences in developmental trajectories between our model and natural embryos in future studies.

In contrast to strategies that rely on the assembly of distinct stem cell types^{13–17}, our study establishes a totipotent cell-based approach for modelling both pre- and post-implantation development. While the aggregation of naive pluripotent stem cells with transgene-induced extraembryonic stem cells can self-organize into post-implantation embryo-like structures^{18,20–22}, this method does not fully recapitulate the natural progression of early embryogenesis. Specifically, the

mixing of embryonic and extraembryonic stem cell types bypasses the earliest developmental stages, as the resulting coaggregates spontaneously organize into structures resembling peri-implantation embryos around E5.0 (refs. 20–22). In addition to stem cell assembly-based models, recent advances have utilized small molecules to induce 8- to 16-cell-like states from pluripotent stem cells or to enhance the developmental plasticity of pluripotent stem cells^{47–49}. These strategies have successfully produced pre- and post-implantation embryo models. While such models can proceed to the E8.5 stage, they primarily recapitulate specific developmental stages and do not capture the

earliest stages of embryogenesis. By contrast, our model enables a continuous reconstruction of mouse embryogenesis from E1.5 to E7.5, particularly capturing ZGA (Fig. 1 and Extended Data Fig. 2). This extended developmental coverage may offer a more comprehensive and physiologically relevant platform for studying the continuum of early mouse embryogenesis.

Notably, a recent study reported the successful application of mouse totipotent blastomere-like cells (mTBLCs) to model early mouse embryogenesis²⁸. Compared with our study, modelling of pre-implantation development using mTBLCs was primarily conducted through two-dimensional spontaneous differentiation. Furthermore, the induction of blastoids from mTBLCs was not examined using single-cell RNA-seq, leaving it unclear whether mTBLC-derived blastoids accurately follow the developmental trajectories of natural pre-implantation embryos. In addition, although mTBLC-derived blastoids have been shown to give rise to egg-cylinder-like structures in vitro, it remains unknown whether these structures possess the capacity to develop into more advanced stages, particularly those associated with early organogenesis. In comparison, our study provides a comprehensive characterization of embryo-like structures spanning multiple developmental stages, from pre- to post-implantation (S1–S7) (Figs. 1–5 and 7 and Extended Data Figs. 2, 4 and 9). Our single-cell transcriptomic analyses also support the reconstruction of developmental trajectories from E1.5 to E7.5 (Fig. 6e). Moreover, by using a roller culture system, we extended in vitro development to generate advanced embryo-like structures (S8 embryoids) (Fig. 7 and Extended Data Fig. 9).

Although our model recapitulates mouse embryogenesis from E1.5 to E7.5, several differences were noted between the embryo-like structures generated and natural mouse embryos. Compared with their in vivo counterparts, the sizes of embryo-like structures at each stage displayed greater variability and were not comparable to corresponding natural embryos at certain stages (Extended Data Figs. 2b and 6a). Similar size heterogeneity has also been reported in recent studies of post-implantation embryo models^{20,22}. In addition, consistent with prior findings on blastoids, the proportions of EPI and PrE lineages in our blastoids differ from those observed in E4.5 blastocysts (Fig. 3h). Furthermore, although S4 blastoids are capable of inducing implantation in surrogate mice (Extended Data Fig. 5h), post-implantation development in vivo was not observed. This limited developmental potential is consistent with previous reports^{12,23,27}, which is probably attributable to deficiencies in generating functional trophoblast subtypes. Indeed, both our data and earlier studies suggest the absence of certain trophoblast subtypes in post-implantation embryoids^{20,21,28} (Extended Data Fig. 7g). However, the overall developmental efficiency of our embryo model remains lower than that of natural embryos, and the divergence between the embryo-like structures and their natural counterparts become more pronounced at later developmental stages. Although several features indicative of early organogenesis were observed in S8 embryoids (Fig. 7 and Extended Data Fig. 9), these structures remain abnormal and exhibit clear differences from E8.5 mouse embryos. This highlights the need for further optimization to improve the fidelity and robustness of in vitro embryo models.

Overall, our study demonstrates the feasibility of using totipotent-like cells to model mouse embryogenesis from ZGA to gastrulation. Our model offers a powerful tool for investigating the mechanisms underlying embryo development in vitro, as well as for screening environmental factors that affect embryogenesis. In principle, this strategy can be extended to other mammalian species, including humans. This opens a promising path towards continuously recapitulating embryogenesis across diverse mammalian species.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions

and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41556-025-01793-9>.

References

- Handford, C. E., Junyent, S., Jorgensen, V. & Zernicka-Goetz, M. Topical section: embryonic models (2023) for Current Opinion in Genetics & Development. *Curr. Opin. Genet. Dev.* **84**, 102134 (2024).
- Zernicka-Goetz, M. The evolution of embryo models. *Nat. Methods* **20**, 1844–1848 (2023).
- Rossant, J. & Tam, P. P. L. Opportunities and challenges with stem cell-based embryo models. *Stem Cell Rep.* **16**, 1031–1038 (2021).
- Kobayashi, N. & Fu, J. Stem cell-derived embryo models: a frontier of human embryology. *Med. Rev.* **3**, 343–346 (2023).
- Beccari, L. et al. Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature* **562**, 272–276 (2018).
- Rossi, G. et al. Capturing cardiogenesis in gastruloids. *Cell Stem Cell* **28**, 230–240 (2021).
- van den Brink, S. C. et al. Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. *Development* **141**, 4231–4242 (2014).
- Bolondi, A. et al. Reconstructing axial progenitor field dynamics in mouse stem cell-derived embryoids. *Dev. Cell* **59**, 1489–1505 (2024).
- Karvas, R. M. et al. 3D-cultured blastoids model human embryogenesis from pre-implantation to early gastrulation stages. *Cell Stem Cell* **30**, 1148–1165 (2023).
- Kagawa, H. et al. Human blastoids model blastocyst development and implantation. *Nature* **601**, 600–605 (2022).
- Yu, L. et al. Blastocyst-like structures generated from human pluripotent stem cells. *Nature* **591**, 620–626 (2021).
- Li, R. et al. Generation of blastocyst-like structures from mouse embryonic and adult cell cultures. *Cell* **179**, 687–702 (2019).
- 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 (2019).
- Zhang, S. et al. Implantation initiation of self-assembled embryo-like structures generated using three types of mouse blastocyst-derived stem cells. *Nat. Commun.* **10**, 496 (2019).
- Girgin, M. U. et al. Bioengineered embryoids mimic post-implantation development in vitro. *Nat. Commun.* **12**, 5140 (2021).
- Sozen, B. et al. Self-assembly of embryonic and two extra-embryonic stem cell types into gastrulating embryo-like structures. *Nat. Cell Biol.* **20**, 979–989 (2018).
- Amadei, G. et al. Inducible stem-cell-derived embryos capture mouse morphogenetic events in vitro. *Dev. Cell* **56**, 366–382 (2021).
- Li, H. et al. In vitro generation of mouse morula-like cells. *Dev. Cell* **58**, 2510–2527 (2023).
- Lau, K. Y. C. et al. Mouse embryo model derived exclusively from embryonic stem cells undergoes neurulation and heart development. *Cell Stem Cell* **29**, 1445–1458 (2022).
- Tarazi, S. et al. Post-gastrulation synthetic embryos generated ex utero from mouse naive ESCs. *Cell* **185**, 3290–3306 (2022).
- Amadei, G. et al. Embryo model completes gastrulation to neurulation and organogenesis. *Nature* **610**, 143–153 (2022).
- Xu, Y. et al. Derivation of totipotent-like stem cells with blastocyst-like structure forming potential. *Cell Res.* **32**, 513–529 (2022).
- Shen, H. et al. Mouse totipotent stem cells captured and maintained through spliceosomal repression. *Cell* **184**, 2843–2859 (2021).

25. Yang, M. et al. Chemical-induced chromatin remodeling reprograms mouse ESCs to totipotent-like stem cells. *Cell Stem Cell* **29**, 400–418 (2022).
26. Hu, Y. et al. Induction of mouse totipotent stem cells by a defined chemical cocktail. *Nature* **617**, 792–797 (2023).
27. Zhang, P. et al. Highly efficient generation of blastocyst-like structures from spliceosomes-repressed mouse totipotent blastomere-like cells. *Sci. China Life Sci.* **66**, 423–435 (2023).
28. Peng, B., Wang, Q., Zhang, F., Shen, H. & Du, P. Mouse totipotent blastomere-like cells model embryogenesis from zygotic genome activation to post implantation. *Cell Stem Cell* **32**, 391–408 (2025).
29. Yang, Y. et al. Derivation of pluripotent stem cells with in vivo embryonic and extraembryonic potency. *Cell* **169**, 243–257 (2017).
30. Macfarlan, T. S. et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).
31. 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).
32. Domon, M. A changing pattern of the cell cycle during the first two cleavage divisions of the mouse: (cell cycle/mouse embryos/ DNA synthesis). *Dev. Growth Differ.* **25**, 537–545 (1983).
33. Lee, Y. S., Thouas, G. A. & Gardner, D. K. Developmental kinetics of cleavage stage mouse embryos are related to their subsequent carbohydrate and amino acid utilization at the blastocyst stage. *Hum. Reprod.* **30**, 543–552 (2015).
34. Li, L. et al. Lineage regulators TFAP2C and NR5A2 function as bipotency activators in totipotent embryos. *Nat. Struct. Mol. Biol.* **31**, 950–963 (2024).
35. Zhu, M. et al. Tead4 and Tfap2c generate bipotency and a bistable switch in totipotent embryos to promote robust lineage diversification. *Nat. Struct. Mol. Biol.* **31**, 964–976 (2024).
36. 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).
37. Marikawa, Y. & Alarcon, V. B. RHOA activity in expanding blastocysts is essential to regulate HIPPO–YAP signaling and to maintain the trophectoderm-specific gene expression program in a ROCK/actin filament-independent manner. *Mol. Hum. Reprod.* **25**, 43–60 (2019).
38. Iturbide, A. et al. Retinoic acid signaling is critical during the totipotency window in early mammalian development. *Nat. Struct. Mol. Biol.* **28**, 521–532 (2021).
39. Kono, K., Tamashiro, D. A. & Alarcon, V. B. Inhibition of RHO–ROCK signaling enhances ICM and suppresses TE characteristics through activation of Hippo signaling in the mouse blastocyst. *Dev. Biol.* **394**, 142–155 (2014).
40. Bedzhov, I., Leung, C. Y., Bialecka, M. & Zernicka-Goetz, M. In vitro culture of mouse blastocysts beyond the implantation stages. *Nat. Protoc.* **9**, 2732–2739 (2014).
41. Christodoulou, N. et al. Morphogenesis of extra-embryonic tissues directs the remodelling of the mouse embryo at implantation. *Nat. Commun.* **10**, 3557 (2019).
42. Guzman-Ayala, M., Ben-Haim, N., Beck, S. & Constam, D. B. Nodal protein processing and fibroblast growth factor 4 synergize to maintain a trophoblast stem cell microenvironment. *Proc. Natl Acad. Sci. USA* **101**, 15656–15660 (2004).
43. Rossant, J. & Cross, J. C. Placental development: lessons from mouse mutants. *Nat. Rev. Genet.* **2**, 538–548 (2001).
44. Sozen, B., Demir, N. & Zernicka-Goetz, M. BMP signalling is required for extra-embryonic ectoderm development during pre-to-post-implantation transition of the mouse embryo. *Dev. Biol.* **470**, 84–94 (2021).
45. Li, Y. et al. Inhibition of Wnt activity improves peri-implantation development of somatic cell nuclear transfer embryos. *Natl Sci. Rev.* **10**, nwad173 (2023).
46. Aguilera-Castrejon, A. et al. Ex utero mouse embryogenesis from pre-gastrulation to late organogenesis. *Nature* **593**, 119–124 (2021).
47. Li, H. et al. A complete model of mouse embryogenesis through organogenesis enabled by chemically induced embryo founder cells. *Cell* <https://doi.org/10.1016/j.cell.2025.07.018> (2025).
48. Yilmaz, A. et al. Transgene-free generation of mouse post-gastrulation whole embryo models solely from naive ESCs and iPSCs. *Cell Stem Cell* **32**, 1545–1562.e12 (2025).
49. Liu, K. et al. Modeling post-gastrula development via bidirectional pluripotent stem cells. *Cell Res.* <https://doi.org/10.1038/s41422-025-01172-x> (2025).

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) 2025

¹MOE Key Laboratory of Cell Proliferation and Differentiation, School of Life Sciences and MOE Engineering Research Center of Regenerative Medicine, School of Basic Medical Sciences, State Key Laboratory of Natural and Biomimetic Drugs, Peking University Health Science Center, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China. ²Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China.

³Department of Cell Biology, School of Basic Medical Sciences, Peking University Stem Cell Research Center, Peking University Health Science Center, Peking University, Beijing, China. ⁴Department of Biochemistry and Molecular Biology, College of Basic Medicine, Key Laboratory of Neural and Vascular Biology of Ministry of Education, Key Laboratory of Vascular Biology of Hebei Province, Hebei Medical University, Shijiazhuang, China. ⁵School of Life Sciences, Center for Bioinformatics, Center for Statistical Science, Peking University, Beijing, China. ⁶State Key Laboratory of Natural and Biomimetic Drugs, Department of Molecular and Cellular Pharmacology, School of Pharmaceutical Sciences, Peking University, Beijing, China. ⁷Changping Laboratory, Beijing, China. ⁸These authors contributed equally: Yixuan Ren, Xuyang Wang, Haiyin Liu, Yaxing Xu, Ruoqi Cheng, Shengnan Ren.

✉ e-mail: guanjiangyang@hsc.pku.edu.cn; cheng_li@pku.edu.cn; hongkui_deng@pku.edu.cn; jun_xu@bjmu.edu.cn

Methods

Ethics statement

All animal experiments were performed in accordance with the NIH guidelines, and all mouse experiments were approved by the Institutional Animal Care and Use Committee of Peking University (LSC-DengHK-14).

Mice

The mouse strain B6-Tg (C57BL/6-tdTomato) aged 25–28 days for establishing EPS cell lines was bought from Beijing Vitalstar Biotechnology. The other mouse strain ICR was purchased from Peking University Health Science Center Department of Laboratory Animal Science. ICR mice used for collecting eight-cell embryos were 25–28 days old; recipient females were 7–8 weeks old; and vasectomized males were under 1 year of age. The mice were housed in a temperature-controlled room ($22 \pm 1^\circ\text{C}$) with 40–60% humidity, under a 12-h light/dark cycle between 06:00 and 18:00. For chimeric experiments, mice and embryos were randomly distributed among groups at the start of each experiment. For in vitro experiments using cells, cells were randomly assigned to either experiment or control groups.

Cell lines

The following cell lines were used in this study: mc6-1/Mu-mClover3-10# and C1-EPS-20#, established as previously described²⁹. All cell lines were cultured under standard conditions of 20% O₂ and 5% CO₂ at 37 °C. Mouse EPS cells were cultured on feeder cells in LCDM medium, with the culture medium refreshed daily. Mouse EPS cells were passaged every 3 days using 0.05% trypsin–EDTA (Gibco, 25300-062) and seeded at a split ratio ranging from 1:8 to 1:15. LCDM medium consisted of N2B27 basal supplemented with Recombinant Human LIF (10 ng ml⁻¹, Novoprotein, C017), CHIR-99021 (3 µM, Selleck, S1263), (S)-(+)-dimethindene maleate (2 µM, Tocris, 1425) and minocycline hydrochloride (2 µM, Tocris, 3268). N2B27 medium was prepared by combining 50% DMEM/F12 (Gibco, 11330032), 50% Neurobasal (Gibco, 21103049), 0.05% N2 supplement (Gibco, 17502048), 1% B27 supplement (Gibco, 12587010), 1% GlutaMAX (Gibco, 35050061) and 1% MEM non-essential amino acids solution (Gibco, 11140050).

Identification of small molecules that induce transient totipotent-like cells

tdTomato-labelled mouse EPS cells carrying the *MuERV-L*-mClover3 reporter were seeded onto 48-well plates coated with Matrigel at a density of 5,000 cells per well. After seeding the cells, the culture medium was changed into the 1640 basal medium: RPMI Medium 1640 basic (Gibco, 22400-089) supplemented with 5% knockout serum replacement (Gibco, 10828-028), 1% N2 supplement (Gibco, 17502-048), 2% B27 supplement (Gibco, 12587-010), 1% GlutaMAX (Gibco, 35050-061), 1% non-essential amino acids (Gibco, 11140-050), 1% sodium pyruvate (Gibco, 11360-070), 0.14% sodium DL-lactate solution (Sigma, L7900), 0.2% chemically defined lipid concentrate (Gibco, 11905-031), 50 µg ml⁻¹ bovine serum albumin (Sigma, A1933), 0.1% 2-mercaptoethanol (Gibco, 21985023) and 1% penicillin–streptomycin (Gibco, 15140163). CHIR-99021 (3 µM, Selleck, S1263) was also added into the culture medium. Small molecules were added individually into each well (Supplementary Table 1). After 3 days of treatment, the fluorescence of cells was photographed using fluorescence microscope. Both the fluorescence of *MuERV-L*-mClover3 reporter and tdTomato were photographed. For each sample, three different fields were randomly photographed. Total cell area was calculated based on the fluorescent area of tdTomato, and the percentages of *MuERV-L*⁺ cells were determined using ImageJ software.

Formation of S1–S8 embryos

To generate embryo models that sequentially recapitulate different stages, from ZGA to the gastrulation, we developed a stepwise protocol.

At S1, mouse EPS cells were digested by 0.05% trypsin–EDTA (Gibco, 25300-062) into single cells; then, 9,600 mouse EPS cells per well were plated in the pretreated Aggrewell (Stemcell Technologies, 34415). One millilitre of totipotency medium was added to each well and the cells were treated for 2 days. The totipotency medium consisted of 1640 basal medium supplemented with elvitegravir (1 µM, MCE, HY-14740), CHIR-99021 (3 µM, Selleck, S1263), CD1530 (0.2 µM, MCE, HY-108527) and PD0325901 (0.5 µM, Selleck, S1036). The 1640 basal medium consisted of RPMI Medium 1640 basic (Gibco, 22400-089) supplemented with 5% knockout serum replacement (Gibco, 10828-028), 1% N2 supplement (Gibco, 17502-048), 2% B27 supplement (Gibco, 12587-010), 1% GlutaMAX (Gibco, 35050-061), 1% non-essential amino acids (Gibco, 11140-050), 1% sodium pyruvate (Gibco, 11360-070), 0.14% sodium DL-lactate solution (Sigma, L7900), 0.2% chemically defined lipid concentrate (Gibco, 11905-031), 50 µg ml⁻¹ bovine serum albumin (Sigma, A1933) and 1% penicillin–streptomycin (Gibco, 15140163). To accommodate the varying proliferation rates of different cell lines, the addition and concentration titration of 2-mercaptoethanol (Gibco, 21985023) could be considered.

At S2, 1 ml of fresh 4-cell to 64-cell medium was added to each well for 1 day. At stage S3, cells were collected after an additional day of treatment under the same conditions. The 4-cell to 64-cell medium consisted of 1640 basal medium supplemented with CHIR-99021 (10 µM, Selleck, S1263), birabresib (0.05 µM, Selleck, S7360), CD1530 (0.2 µM, MCE, HY-108527), 8-Br-cAMP (1 mM, Selleck, S7857) and 0.05% 2-mercaptoethanol (Gibco, 21985023). To accommodate different cell lines, the knockout serum replacement in 1640 basal medium could be replaced by FBS (fetal bovine serum, VISTECH, SE200-ES), and the concentration of 2-Mercaptoethanol (Gibco, 21985023) could be titrated.

At S4, 1 ml of fresh blastoid medium was added to each well and the cells were cultured for 3 days to generate blastoids. Blastoid medium consisted of 50% KSOM (LIFE-iLAB, CE002), 25% N2B27 basal and 25% TSC basal, supplemented with CHIR-99021 (3 µM, Selleck, S1263), GA-017 (10 µM, Selleck, E1145), recombinant human BMP-4 (20 ng ml⁻¹, Novoprotein, C093) and recombinant human FGFb (157AA) (20 ng ml⁻¹, Novoprotein, C046). TSC basal consisted of 80% RPMI Medium 1640 basic (Gibco, 22400-089) and 20% FBS (VISTECH, SE200-ES), supplemented with 1% GlutaMAX (Gibco, 35050061) and 1% sodium pyruvate (Gibco, 11360070). To accelerate the kinetics of blastoids formation, cell aggregates could be transferred to six-well non-adherent suspension culturing plates (BEAVER, 40406). For the acquisition of post-implantation embryoids, early blastoids of 1-day treatment should be retained in the Aggrewell (Stemcell Technologies, 34415) for subsequent processing.

At S5, 1 ml of fresh peri-implantation medium was gently added to each well. Following a 3-day treatment, structures resembling early egg cylinder-like formations surrounded by a layer of VE-like cells began to emerge. Peri-implantation medium consisted of 50% KSOM (LIFE-iLAB, CE002), 25% N2B27 basal and 25% TSC basal, supplemented with recombinant mouse FGF-4 (25 ng ml⁻¹, Novoprotein, CR66), heparin (1 µg ml⁻¹, Macklin, H811552), recombinant human/mouse/rat Activin A (20 ng ml⁻¹, Novoprotein, C687), XAV939 (2 µM, Selleck, S1180) and recombinant human BMP-4 (5 ng ml⁻¹, Novoprotein, C093).

At S6 and S7, S5 embryoids were transferred to six-well non-adherent suspension culturing plates (BEAVER, 40406). A total of 3–4 ml of post-implantation medium (modified IVC1 medium) was added to each well. Following 1 day of treatment, structures morphologically resembling E6.5 embryos with two lumens were observed. After an additional day of treatment, the structures increased in size, and some were found to develop three cavities, resembling E7.5 embryos. Post-implantation medium/modified IVC1 medium consisted of Advanced DMEM/F12 (Gibco, 12634010) supplemented with 20% FBS (VISTECH, SE200-ES), 1% mM GlutaMAX (Gibco, 35050061), 1% ITS-X supplement (Gibco, 51500056), 1% penicillin–streptomycin (Gibco, 15140163), 8 nM β-oestradiol (Sigma, E8875), 200 ng ml⁻¹ progesterone

(Sigma, V900699), 25 mM *N*-acetyl-L-cysteine (Sigma-Aldrich, A9165), 100 nM T3 (3,3',5-triiodo-L-thyronine sodium) (MCE, HY-A0070), 1 mg ml⁻¹ D(+)-glucose (MCE, HY-B0389).

At S8, S7 embryoids were selected under dissection microscope for further culture with the roller culture system. We selected elongated egg cylinder with a thick EPI-like layer that resembled the EPI in natural mouse embryos and transferred them to post-implantation medium/modified IVC2 medium. Post-implantation medium/modified IVC2 medium consisted of 25% DMEM (Gibco, 11054) supplemented with 50% CD female rat serum, 25% human cord serum, 1% mM GlutaMAX (Gibco, 35050061), 1% penicillin–streptomycin (Gibco, 15140163), 1% sodium pyruvate (Gibco, 11360070) 11 mM HEPES (Gibco, 15630056) and 4 mg ml⁻¹ D(+)-glucose (MCE, HY-B0389). Rat serum and human cord serum were thawed at room temperature and heat-inactivated for 30 min at 56 °C. Up to 50 S7 embryoids were cultured in each bottle, filled with 2 ml post-implantation medium/modified IVC2 medium.

The detailed protocol⁵⁰ of this study can be found at <https://doi.org/10.17504/protocols.io.261gekxdjg47/v1>.

Long-term culture of S1 totipotent like cells

All cell lines were cultured under 20% O₂ and 5% CO₂ at 37 °C. S1 totipotent-like cells were cultured in totipotency medium supplemented with 0.1% 2-mercaptoethanol (Gibco, 21985023) on feeder cells. The culture medium was changed every day. Cells were passaged every 3 days with 0.05% trypsin–EDTA (Gibco, 25300-062) and seeded at a split ratio ranging from 1:8 to 1:15.

Chimeric assay

Cells were digested by 0.05% trypsin–EDTA (Gibco, 25300-062), and the digested cells were filtered through a 40-µm cell strainer and centrifuged at 160–250g for 3 min at room temperature. The supernatant was removed, and the cells were suspended using culture medium and placed on ice before injection. After being placed on ice, the digested cells should be injected within 1 h; otherwise, another batch of cells is digested for the remaining injections.

Single cells or multiple cells were microinjected into eight-cell ICR diploid mouse embryos. For the generation of chimeric blastocysts, the embryos were transferred into the KSOM medium (Merck, MR-106-D). Injected embryos were transferred to uterine horns of 0.5 post-coitum pseudo-pregnant females to recover embryos at the E6.5 developmental stage.

Immunofluorescence of S1–S3 cell aggregates, blastoids and chimeric blastocysts

Cell aggregates, blastoids and chimeric blastocysts were washed three times in phosphate-buffered saline (PBS) droplets, then fixed in 4% paraformaldehyde (DingGuo, AR-0211) droplets at room temperature for 20 min. PBS (Corning, 21-485 040-CVR) that contained 0.2% Triton X-100 (Sigma-Aldrich, T8787) and 3% normal donkey serum (Jackson ImmunoResearch, 017-000-121) was used to block cell aggregates, blastoids and chimeric blastocysts at room temperature for 1 h. Primary antibodies were diluted with blocking solution, then embryos were incubated at 4 °C overnight. Cell aggregates, blastoids and chimeric blastocysts were washed three times in PBS droplets, then incubated with secondary antibodies (Jackson ImmunoResearch) at room temperature for 1 h. After final washes, cell aggregates, blastoids and chimeric blastocysts were transferred to a confocal dish in PBS droplets covered with paraffin for imaging. Confocal microscope imaging was performed using Olympus FV3000 and Leica TCS-SP8. The following primary antibodies were used: anti-ZSCAN4 (1:5000; Millipore-Sigma, AB4340), anti-MuERV-L-Gag (1:500; Epigentek, A-2801-100), anti-OCT4 (1:500; Abcam, ab181557), anti-TFAP2C (1:500; Abcam, ab218107), anti-SOX17 (1:200; R&D, AF1924), anti-CDX2 (1:500; BioGenex, MU392A), anti-NR5A2 (1:500; R&D, PP-H2325), anti-SOX2 (1:200; MilliporeSigma, AB5603) and anti-TFAP2C (1:200, R&D, AF5059).

Immunofluorescence of and S5–S7 embryoids

S5–S7 embryoids were fixed in 4% PFA at room temperature for 20 min. After washing three times with PBST (PBS/0.5% Tween-20). The embryoids were permeabilized with blocking solution (PBS, 0.03% Triton X-100 and 0.1% glycine) at room temperature for 1 h. Primary antibodies were diluted at a concentration of 1:200/1:500 in antibody buffer (PBS, 10% FBS and 1% Tween-20). Then, the permeabilized embryoids were incubated with the primary antibodies overnight at 4 °C. On the following day, embryoids were washed in PBST three times, then incubated with secondary antibodies diluted in antibody solution for 1 h at room temperature. The embryoids were then washed in PBST 5 min three times. The nuclei were stained with DAPI (Roche Life Science, 10236276001) at room temperature for 30 min and washed with PBST three times. Confocal microscope imaging was performed using Leica TCS-SP8 and Olympus FV3000. The following primary antibodies were used: anti-tdTomato (1:2,000; SICGEN, AB8181-200), anti-OCT4 (1:500; Abcam, ab181557), anti-CDX2 (1:500; Abcam, ab76541), anti-CDX2 (1:500; BioGenex, MU392A), anti-SOX17 (1:200; R&D, AF1924), anti-T/Bra (1:500; Abcam, ab209665), anti-T/Bra (1:500; R&D, AF2085), anti-FOXA2 (1:500; Abcam, ab108422), anti-CER1 (1:500; R&D, AF1986), anti-E-CADHERIN (1:500; Abcam, ab231303), anti-RUNX1 (1:500; Abcam, ab92336), anti-OTX2 (1:200; R&D, AF1979), anti-MHC-II (1:500; R&D, MAB4470) and anti-SOX1 (1:200; R&D, AF3369).

S4 blastoids transfer

S4 blastoids were manually picked up under a stereomicroscope and transferred into KSOM droplets using a mouth pipette. The uterine horn of surrogate at 2.5 dpc was exposed surgically. After three washes in KSOM, S4 blastoids were loaded into the pipette with an air bubble and transferred to the uterine horn. Around 15 S4 blastoids were transferred into each uterine horn. The uterus was dissected out at 6.5 or 7.5 dpc.

Quantitative PCR analysis

Total RNA was isolated using the Direct-zol RNA Kits (ZYMO Research, R2052). RNA was converted to cDNA using TransScript FirstStrand cDNA Synthesis SuperMix (Accurate Biology, AG11728). Quantitative PCR analysis was conducted using the KAPA SYBR FAST qPCR Kit (Accurate Biology, AG11701) with the Bio-Rad CFX Connect Real-Time System. The primers that were used for quantitative PCR analysis were generated from PrimerBank and are listed in Supplementary Table 9. The data were analysed using the delta–delta CT method.

Doubling time calculation

Cells were digested and seeded at a density of 4×10^4 per well of a 24-well plate. After 60 h, the cell numbers were counted. The formula used for calculating doubling time was: $DT = 60 \times [\log_2 / (\lg - \lg(\text{number of cells at 60 h}) - \lg(\text{number of cells at 0 h}))]$.

Bulk RNA-seq data processing

All raw reads were first trimmed by Trimmomatic (version 0.39) to remove adapters and low-quality reads. Then, cleaned data were mapped to the mouse reference genome (mm10) using STAR (version 2.7.10b) with default parameters. The count matrix of gene expression in each sample was generated by featureCounts (version 2.0.6). The gene expression level was normalized by all mapped reads of each sample (count per million). The DEGs were identified using DESeq2 (version 1.42.1) R package with *P* value <0.05 and $|\log_2 \text{FC}| > 1$. FC, fold change.

Single-cell RNA-seq data processing

The single-cell RNA-seq data were collected and mapped to mouse reference genome (mm10) using Cell Ranger (version 5.0.1) for all our samples. We performed preprocessing using Seurat (version 4.3.0.1) R package. In detail, quality control was first performed to remove the cells with (1) total unique molecular identifier (UMI) counts <2,000, (2) detected gene number <1,500 or (3) mitochondrial UMI counts/

total UMI counts >10%. The normalization was performed using the `NormalizeData` function with default parameters. We performed the standard Seurat clustering pipeline using the following functions: `FindVariableFeatures` with 2000 genes, `ScaleData`, `RunPCA`, `FindNeighbors` with the first 25 principal components, and `FindClusters` with variable resolutions from 0.1 to 1.5. The UMAP dimension reduction was performed with first 25 principal components using the `RunUMAP` function. The DEGs were analysed using the `FindAllMarkers` function based on normalized gene expression.

For integration analysis, we collected six public datasets as references ([GSE45719](#) from Deng et al.⁵¹; [GSE136714](#) from Wang et al.⁵²; [GSE84892](#) from Posfai et al.⁵³; [GSE123046](#) from Nowotschin et al.⁵⁴; and [E-MTAB-6967](#) from Pijuan-Sala et al.⁵⁵). The preprocessing was performed using the same procedures as described above. We randomly selected cells in each group and adjusted the number of cells from different sample sources to the same level. Next, the canonical correlation analysis method was used to integrate our data and published data. The `SelectIntegrationFeatures` function, followed by the `FindIntegrationAnchors` and `IntegrateData` functions, was used to perform integration analysis. Then, the integrated data were used to perform the standard Seurat clustering pipeline as mentioned above. The median UMAP coordinates of cells of the same type in S1 and S2 were used to represent that cell type for visualization. The median of UMAP embeddings was used for hierarchical clustering from S1 to S3 using the `hclust` function.

Pseudotime analysis

For S1, S2 and S3, pseudotime analysis was performed using the `monocle2` (version 2.30.0) R package with default parameters as recommended by the developers (<https://github.com/cole-trapnell-lab/monocle-release>). After ordering genes, the `DDRTree` method was used to reduce dimensionality and build the tree.

When S5 and S7 were added, due to the complexity of cell types, pseudotime analysis was performed using the `monocle3` (version 1.3.1) R package. `Monocle3` is a powerful tool capable of handling complex single-cell RNA-seq data. It supports multistart trajectories, loop structures and automatic partitioning algorithms, thereby more accurately reflecting the developmental processes of cells. We conducted the analysis following the workflow recommended by the developers (<https://cole-trapnell-lab.github.io/monocle3/>).

Gene set enrichment analysis

Gene set enrichment analysis was performed using the `ClusterProfiler` (version 4.10.0) R package. The `enrichKEGG` function was used for KEGG annotation, and the `enrichGO` function was used for GO annotation.

Gene set score calculation

For single-cell RNA-seq data, the scores for gene sets were calculated in each cell using the `AddModuleScore` function of the Seurat package.

For bulk RNA-seq data, the scores for gene sets were calculated in each sample using the `GSVA` (version 1.50.0) R package.

The ZGA gene set was obtained from genes highly expressed in the late two-cell stage compared with the zygote using the `FindMarkers` function of the Seurat package.

The pluripotent gene set was obtained from the top 100 blastocyst-specific genes compared with early stages.

Genes regulated by *Nr5a2* or *Tfap2c* were obtained from genes downregulated in *Nr5a2*- or *Tfap2c*-knockout embryos compared with normal eight-cell embryos using data from [GSE216256](#) ref. 34 and [GSE229740](#) ref. 56 for DEG analysis (cut-off: P value < 0.05 and $\log_2FC$ > 1).

The E5.5 three-germ layer-specific genes were obtained from the top 100 genes of each germ layer using the `FindAllMarkers` function of the Seurat package.

Stage-specific genes in natural embryos—2-cell, 4- to 16-cell, 32- to 64-cell, E4.5, E5.5 and E7.5 mouse embryos—were identified and selected using the `FindAllMarkers` function of the Seurat package.

Correlation analysis between S7, S8 and natural embryos

The average expression of the 2,000 variable genes identified by `FindVariableFeatures` function for each cell type was calculated using the integrated matrix using the `AverageExpression` function. The correlation was calculated using the Pearson method. The data met the assumptions of the statistical tests used; normality and equal variances were formally tested.

Inclusion criteria of S1–S8 embryoids

As S1–S3 cell aggregates showed no morphological differences, aggregates for analysis were collected without morphological selection. All S4–S8 embryoids' morphologies were examined under a dissection microscope. S4 blastoids were selected for analyses or transfer if (1) the expanded blastocoel-like cavities approached the size of natural blastocysts, and (2) ICM-like aggregates within cavities exhibited volumetric equivalence to *in vivo* counterparts. S5 embryoids were selected for analyses if (1) there were two distinct cellular compartments enclosed by a thin outer cell layer, and (2) there was a clear epithelialized EPI-like compartment with a lumen. S6 embryoids were selected for analyses if (1) there were two distinct cellular compartments with lumen enclosed by a thin outer cell layer, and (2) there were a clear and thick epithelialized EPI-like layer. For S7 embryoids, we selected elongated egg cylinders with a thick EPI-like layer that resembled the EPI in natural mouse embryos. Subcompartments resembling the ExC and EC might be observed in the ExE-like compartment. S8 embryoids exhibiting developed headfold-like and tail bud-like structures within a yolk sac-like membrane were selected for further analyses.

Statistics and reproducibility

The number of biological replicates and the exact P values are described in the figure legends. Sample sizes were determined based on preliminary experiments and commonly used sample sizes reported in comparable publications within this field. All analyses were performed with GraphPad Prism software. Statistical methods are described in the figure legends. In box-and-whisker plots, the box spans from the first to the third quartile with a line at the median, and the whiskers extend to the farthest point within $1.5 \times$ the interquartile range.

Data collection and analysis were not performed blind to the conditions of the experiments. No data points were excluded from the analyses.

Reporting summary

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

Data availability

Sequencing data that support the findings of this study are available via Gene Expression Omnibus (GEO) under accession codes [GSE287909](#) (single-cell RNA-seq) and [GSE287908](#) (bulk RNA-seq). Previously published data that were reanalysed here are available under accession codes [GSE45719](#), [GSE136714](#), [GSE84892](#), [GSE100597](#), [GSE123046](#) and [E-MTAB-6967](#) (developmental stage ranging from zygote to E8.5); and [GSE229740](#) and [GSE216256](#) (bulk RNA-seq data for normal and *Nr5a2*/*Tfap2c*-knockdown embryos). Source data are provided with this study. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

References

- Ren, Y. & Xu, J. Stepwise Protocol of establishing embryo model from a continuous totipotent-like cell-based embryo model recapitulates mouse embryogenesis from zygotic genome activation to gastrulation. *protocols.io* <https://www.protocols.io/view/stepwise-protocol-of-establishing-embryo-model-fro-haj6b2crf> (2025).

51. Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196 (2014).
52. Wang, Y. et al. Single-cell multiomics sequencing reveals the functional regulatory landscape of early embryos. *Nat. Commun.* **12**, 1247 (2021).
53. Posfai, E. et al. Position- and Hippo signaling-dependent plasticity during lineage segregation in the early mouse embryo. *Elife* **6**, e22906 (2017).
54. Nowotschin, S. et al. The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* **569**, 361–367 (2019).
55. Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).
56. Lai, F. et al. NR5A2 connects zygotic genome activation to the first lineage segregation in totipotent embryos. *Cell Res.* **33**, 952–966 (2023).

Acknowledgements

We thank Peking University Medical and Health Analysis Center for assistance with confocal imaging. We thank Y. N. Liu for assistance with scientific illustration. This work was supported by the National Key R&D Program of China (grant no. 2021YFA1100300 to J.X. and C.L.), the National Natural Science Foundation of China (grant nos. 32288102 to H.D. and C.L., 32025006 to C.L., 32370843 to J.X., 32422022 to J.G. and 32400656 to Y.X.), the Beijing Natural Science Foundation (grant nos. JQ24042 to J.X., Z230012 to J.G.) and the Natural Science Foundation of Hebei Province (grant no. C2024206031 to Y.X.). Part of the data analysis was performed on the High Performance Computing Platform of the Center for Life Sciences, Peking University.

Author contributions

J.X., H.D. and Y.R. conceived this project. Y.R., X.W., H.L. and Y.X. performed the experiments. R.C. performed bioinformatic analyses, and C.L. provided guidance. S.R. performed independent validation experiments, and J.G. provided guidance. Z.L., Y.H. and B.L. helped with microinjections and mouse breeding. J.X. and H.D. wrote the manuscript with input from other coauthors.

Competing interests

The authors declare no competing interests.

Additional information

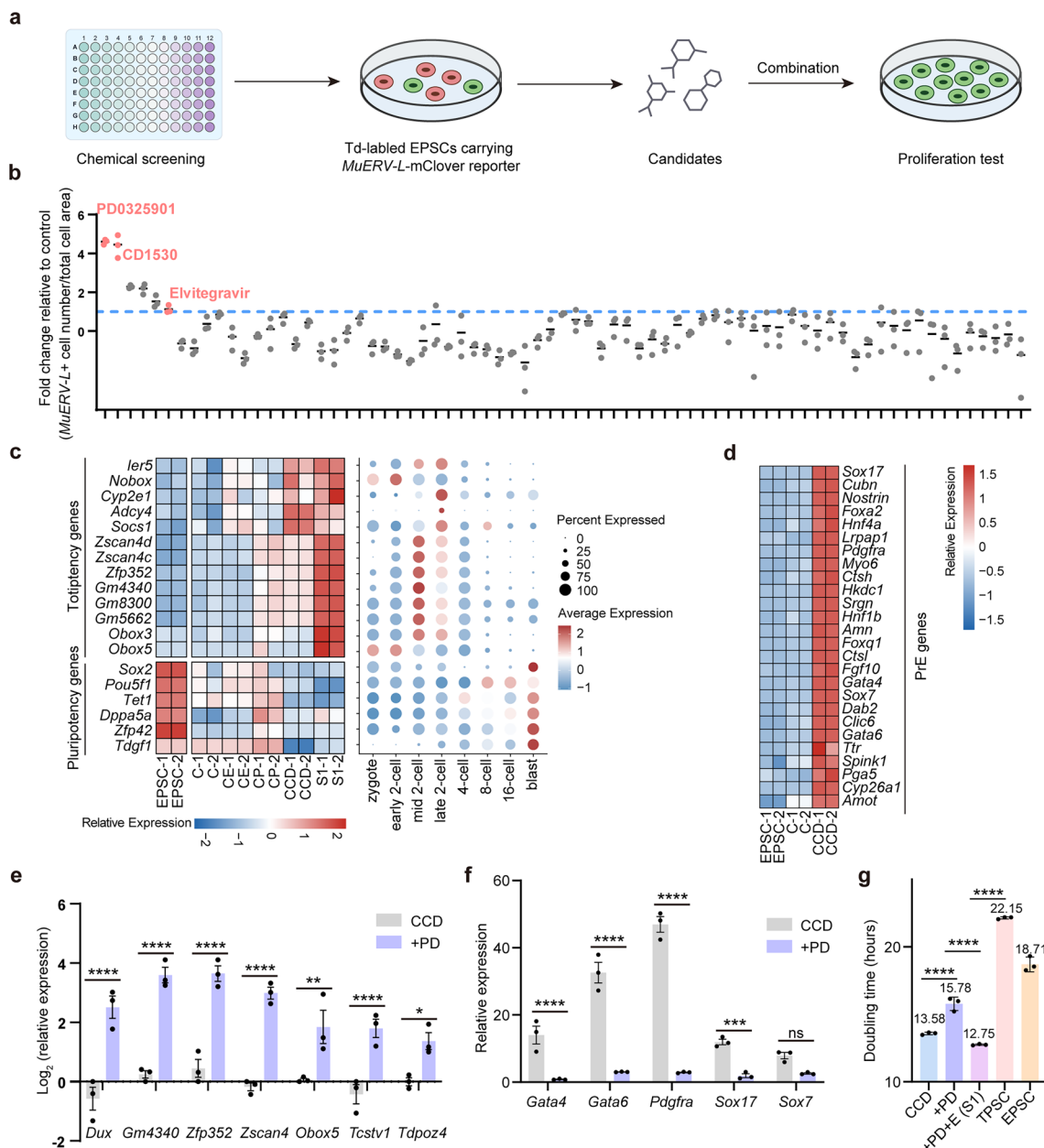
Extended data is available for this paper at <https://doi.org/10.1038/s41556-025-01793-9>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41556-025-01793-9>.

Correspondence and requests for materials should be addressed to Jingyang Guan, Cheng Li, Hongkui Deng or Jun Xu.

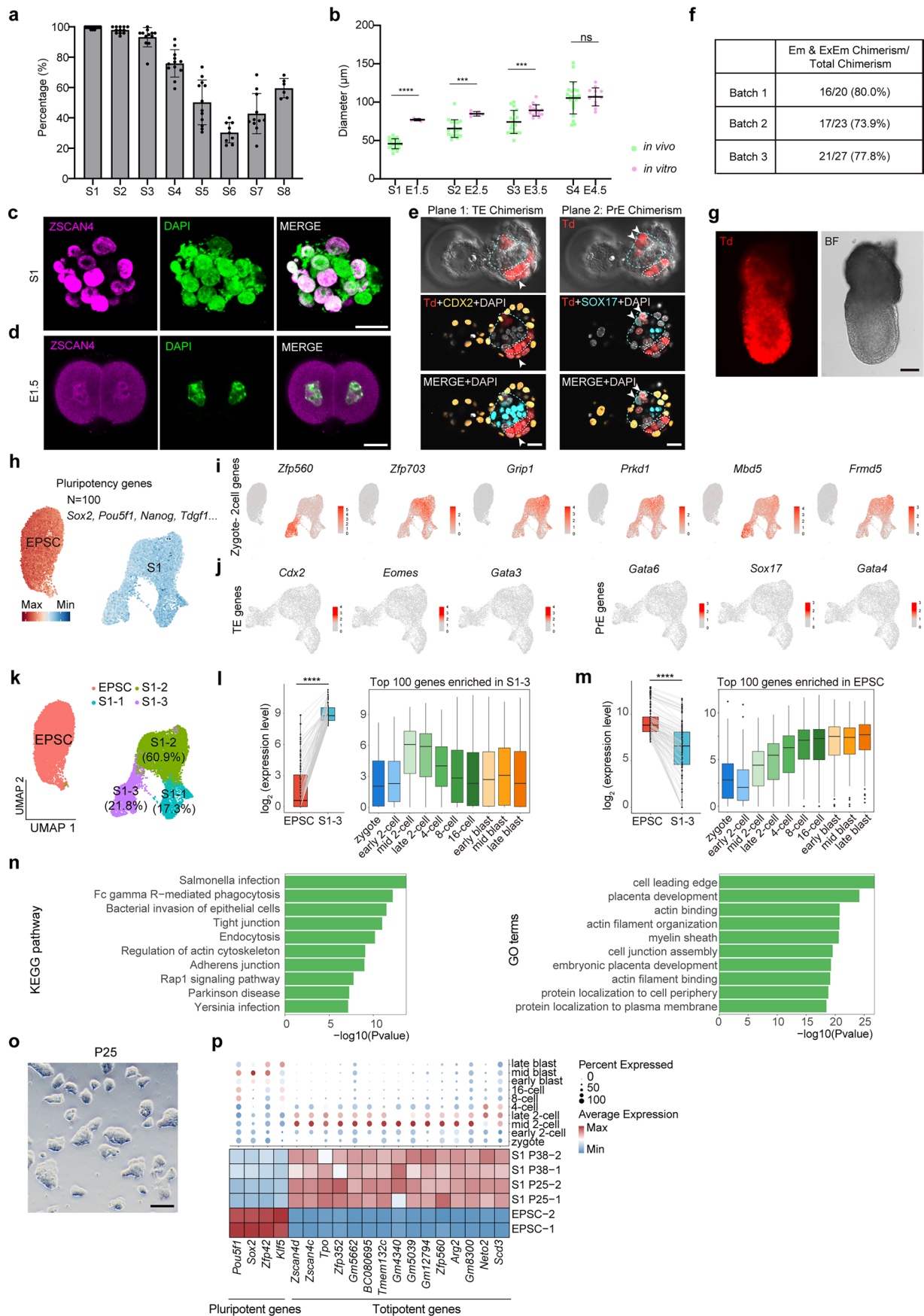
Peer review information *Nature Cell Biology* thanks Alfonso Martinez Arias, Baojiang Wu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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



Extended Data Fig. 1 | Identification of a chemical cocktail that enables rapid induction of totipotent-like cells with robust proliferation ability. **a**, Schematic of the strategy used for identifying chemical cocktails that supports generation of transiently induced totipotent-like cells. Td, tdTomato. **b**, Results of the primary chemical screen using mouse EPSCs carrying the *MuERV-L-mClover3* reporter. The dashed line indicates the threshold, which is set to two-fold changes. The values were normalized based on log2. $n = 3$ biological replicates. **c**, Heatmaps showing the upregulation of totipotent marker genes in mouse EPSCs under different chemical combinations. Samples were treated for three days before their collection. C, CHIR-99021; CE, CHIR-99021+Elvitegravir; CP, CHIR-99021 + PD0325901; CCD, CHIR-99021 + CD1530; SI, CD1530 + PD0325901+Elvitegravir+CHIR-99021. **d**, Heatmaps showing the upregulation of primitive endoderm genes in mouse EPSCs under different chemical combinations. Samples were treated for three days before their collection. C, CHIR-99021; CCD, CHIR-99021 + CD1530. **e**, q-PCR analysis showing the effect of PD0325901 on the expression levels of classical totipotent

marker genes. Samples were treated for three days before their collection. CCD, CHIR-99021 + CD1530; +PD, CHIR-99021 + CD1530 + PD0325901. Data are presented as mean $\pm$ SEM, $n = 3$ biological replicates. **** $p < 0.0001$; ns, non-significant (Two-way ANOVA). p values from left to right: 3.31×10^{-7} , 6.67×10^{-8} , 1.64×10^{-7} , 2.04×10^{-7} , 1.40×10^{-3} , 7.98×10^{-5} , 1.89×10^{-2} . **f**, q-PCR analysis showing the effect of PD0325901 on the expression levels of classical primitive endoderm genes. Samples were treated for three days before their collection. CCD, CHIR-99021 + CD1530; +PD, CHIR-99021 + CD1530 + PD0325901. Data are presented as mean $\pm$ SEM, $n = 3$ biological replicates. **** $p < 0.0001$; ns, non-significant (Two-way ANOVA). p values from left to right: 3.29×10^{-5} , 8.05×10^{-11} , 4.60×10^{-14} , 8.70×10^{-4} , 0.117. **g**, Cell doubling times of CCD, +PD, +PD + E(S1), EPSCs and TPSC. CCD, CHIR-99021 + CD1530; +PD, CHIR-99021 + CD1530 + PD0325901; +PD + E(S1), CHIR-99021 + CD1530 + PD0325901+Elvitegravir. Data are presented as mean $\pm$ SD, $n = 3$ biological replicates. **** $p < 0.0001$ (One-way ANOVA). p values from left to right: 3.35×10^{-5} , 1.85×10^{-6} , 3.10×10^{-11} .

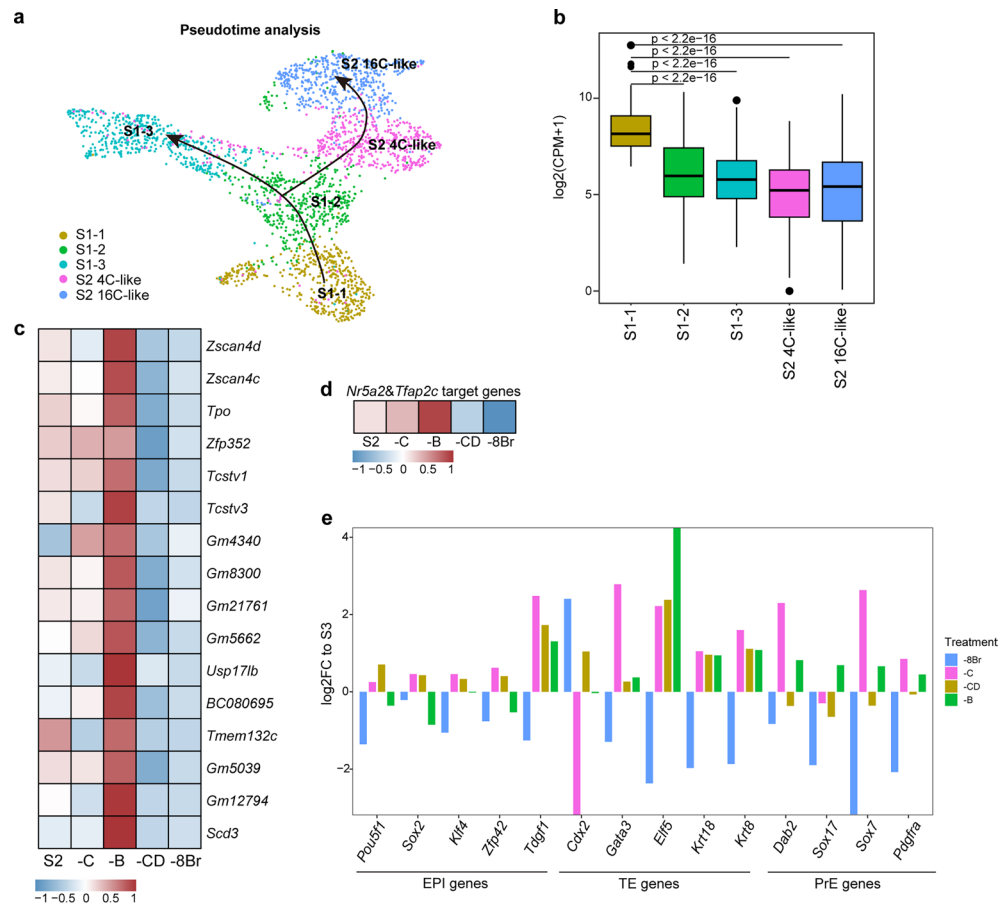


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Further characterization of the transcriptomic features and chimeric ability of chemically induced totipotent-like cells.

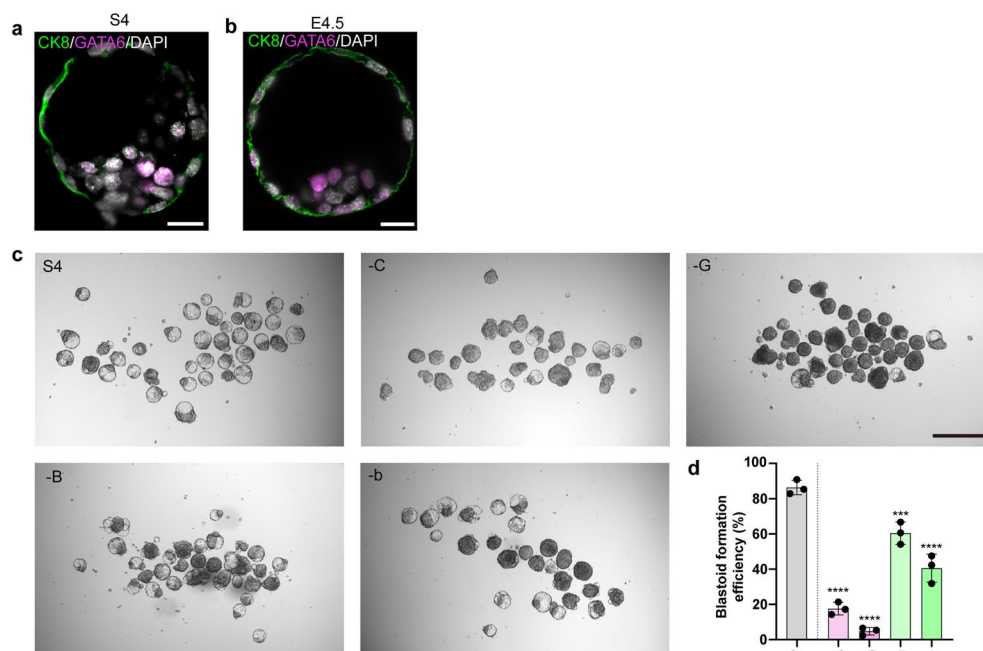
a, Quantification of embryoid formation efficiency from S1 to S8. Data are presented as mean $\pm$ SD. $n = 294$ for S1 from 12 experiments, 292 for S2 from 12 experiments, 225 for S3 from 12 experiments, 286 for S4 from 12 experiments, 294 for S5 from 12 experiments, 158 for S6 from 7 experiments, 189 for S7 from 11 experiments, 300 for S8 from 6 experiments. Each dot indicates the efficiency of one batch of experiment. **b**, Size comparison of S1-S4 embryoids and stage matched embryos. Data are presented as mean $\pm$ SD. $n = 6$ for E1.5 embryos, 22 for S1 cell aggregates, 6 for E2.5 embryos, 18 for S2 cell aggregates, 16 for E3.5 embryos, 18 for S3 cell aggregates, 12 for E4.5 embryos, 21 for S4 blastoids. $***p < 0.001$, $****p < 0.0001$, ns, non-significant (two-sided Student's t -test). Each dot indicates the diameters of one cell aggregate or embryo. p values from left to right: 8.59×10^{-12} , 0.006, 0.009, 0.8326. **c, d**, Immunofluorescent analysis showing the expression of ZSCAN4 in induced totipotent-like cells at S1 (**c**) and E1.5 embryos (**d**). $n = 5$ S1 cell aggregates from 2 experiments, $n = 3$ E1.5 2-cell embryo from 2 experiments. Scale bars of (**c**) and (**d**), 20 μm . **e**, Bright field and immunofluorescent image showing the presence of derivatives of induced totipotent-like cells in E4.5 mouse blastocysts. Td, tdTomato. Td⁺ cells were found to be presented in the EPI (white dashed line), PrE (blue dashed line) and TE (white arrowhead) region. $n = 3$ from 3 experiments. Scale bars of left and right panels, 20 μm . **f**, Summary of chimeric assays in mouse E4.5 blastocysts. **g**, Representative image showing the presence of derivatives of induced totipotent-like cells in E6.5 mouse embryo. Scale bar, 100 μm . Td, tdTomato.

Similar images were obtained in at least 3 independent experiments. **h**, Feature plot showing the expression of pluripotency genes in EPSCs and cell aggregates at S1. **i**, Feature plots showing the expression of representative genes expressed from zygote to 2-cell embryo in EPSCs and cell aggregates at S1. **j**, Feature plots showing the expression of marker genes for trophoctoderm and primitive endoderm in cell aggregates at S1. **k**, UMAP analysis showing the presence of three subpopulations in cell aggregates at S1. **l, m**, Analysis of differentially expressed genes of subpopulation 3 at S1 (S1-3) and EPSCs. Box plots show the expression of top 100 genes enriched in S1-3 (**l**) or EPSCs (**m**) at different developmental stages during pre-implantation development. Each box depicts the median level along with the first and third quartiles. The whiskers represent the minimum and maximum values. Points indicate outliers, which are defined as values exceeding 1.5 times the interquartile range from the box. The dots in the boxplot represent genes. The left plot shows the results of a two-sided Wilcoxon rank-sum test, $n = 100$ (number of genes). The right plot presents the results of a two-sided Kruskal Wallis test, $n = 100$ (number of genes). p values of **l** from left to right: 5.67×10^{-34} , 4.49×10^{-12} . p values of **m** from left to right: 2.46×10^{-15} , 3.18×10^{-54} . **n**, GO and KEGG analysis of specifically expressed genes in subpopulation 1 at S1 (S1-1). p value was calculated using the one-sided hypergeometric test. **o**, Representative bright field image of long-term cultured S1 cells. Scale bar, 100 μm . Similar images were obtained in at least 3 independent experiments. **p**, Heatmaps showing the upregulation of totipotent marker genes in long-term cultured S1 cells.



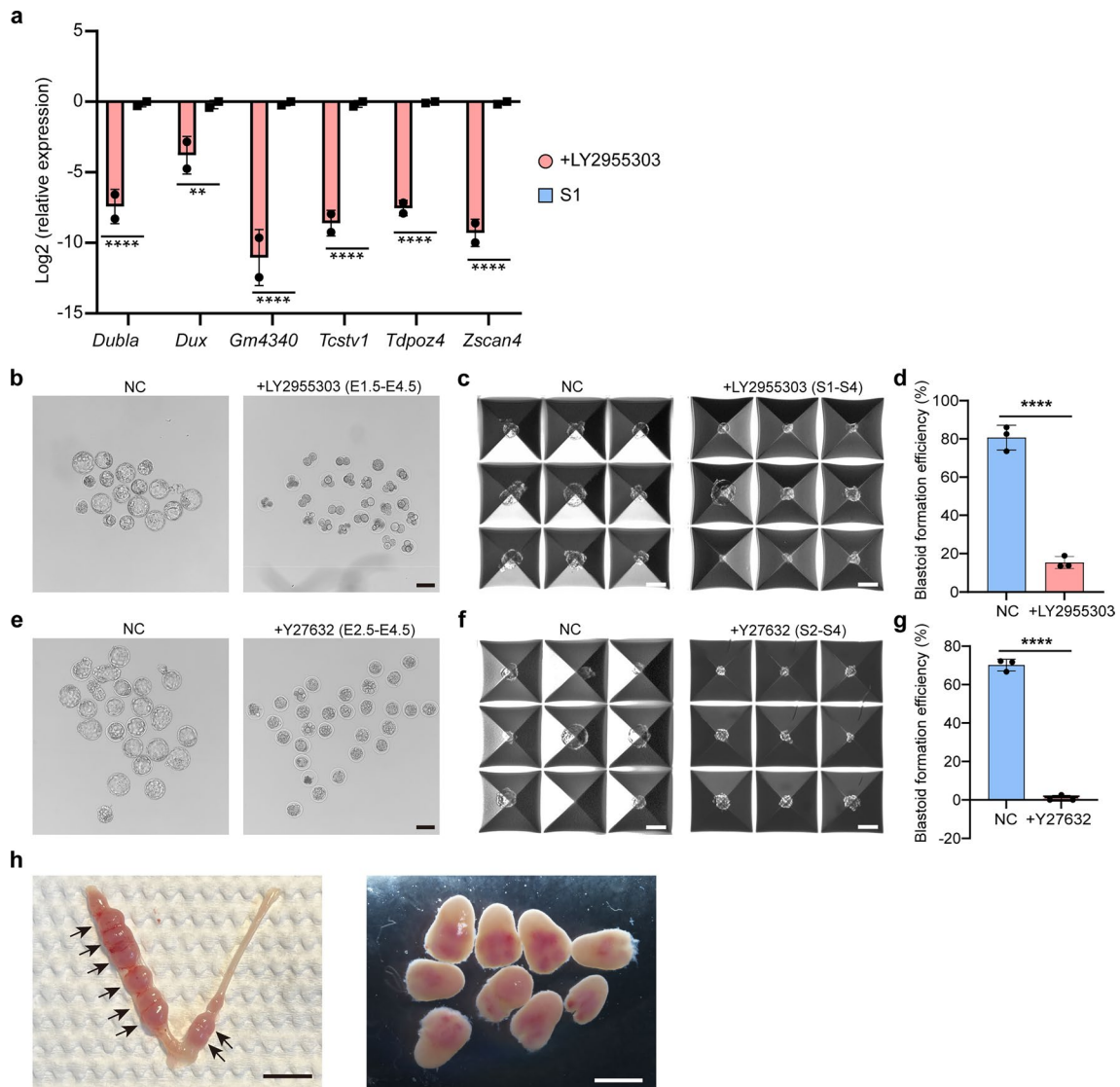
Extended Data Fig. 3 | Transcriptomic analysis of the transcriptional features of S2 and S3 cells and the effect of compound from S2-S3 base. **a**, Pseudotime trajectory reconstruction of single cells from subpopulations at S1 and S2. **b**, Box plots showing the expression of genes enriched in subpopulation 1 at S1 (S1-1) across subpopulations from S1 and S2. Each box depicts the median level along with the first and third quartiles. The whiskers represent the minimum and maximum values. Points indicate outliers, which are defined as values exceeding 1.5 times the interquartile range from the box. The dots in the boxplot represent genes. p value was calculated using the two-sided Wilcoxon rank-sum test, $n = 100$ (number of genes). p values from left to right: 2.96×10^{-17} , 3.04×10^{-22} , 2.16×10^{-27} , 7.28×10^{-24} . **c**, Heatmaps showing the expression of totipotency marker genes in S2 cell aggregates under different chemical combinations.

S2, CHIR-99021+Birabresib+CD1530 + 8-Br-cAMP; -C, Birabresib+CD1530 + 8-Br-cAMP; -B, CHIR-99021 + CD1530 + 8-Br-cAMP; -CD, CHIR-99021+Birabresib+8-Br-cAMP; -8Br, CHIR-99021+Birabresib+CD1530. **d**, Heatmaps showing the expression of *Nr5a2* and *Tfap2c* target genes in S2 cell aggregates under different chemical combinations. S2, CHIR-99021+Birabresib+CD1530 + 8-Br-cAMP; -C, Birabresib+CD1530 + 8-Br-cAMP; -B, CHIR-99021 + CD1530 + 8-Br-cAMP; -CD, CHIR-99021+Birabresib+8-Br-cAMP; -8Br, CHIR-99021+Birabresib+CD1530. **e**, Expression of marker genes for epiblast, trophectoderm and primitive endoderm in S3 cell aggregates under different chemical combinations. S3, CHIR-99021+Birabresib+CD1530 + 8-Br-cAMP; -C, Birabresib+CD1530 + 8-Br-cAMP; -B, CHIR-99021 + CD1530 + 8-Br-cAMP; -CD, CHIR-99021+Birabresib+8-Br-cAMP; -8Br, CHIR-99021+Birabresib+CD1530.



Extended Data Fig. 4 | Further characterization of blastoid at S4 and the effect of compound from S4 base. a, b, Immunofluorescent analysis showing the expression of CK8 and GATA6 in blastoids at S4 (**a**) and E4.5 blastocysts (**b**). $n = 5$ S4 blastoids from 2 experiments, $n = 3$ E4.5 blastocysts from 2 experiments. Scale bars, 20 μm . **c, d,** Bright field images (**c**) and formation efficiency (**d**) of S4 blastoids under different chemical combinations.

S4, CHIR-99021 + GA017 + BMP4+bFGF; -C, GA017 + BMP4+bFGF; -G, CHIR-99021 + BMP4+bFGF; -B, CHIR-99021 + GA017+bFGF; -b, CHIR-99021 + GA017 + BMP4. Data are presented as mean $\pm$ SD. $n = 3$, *** $p < 0.001$, **** $p < 0.0001$ (One-way ANOVA). p values from left to right: 2.94×10^{-6} , 4.14×10^{-4} , 1.19×10^{-8} , 6.30×10^{-8} . Scale bar, 500 μm .



Extended Data Fig. 5 | Evaluation of developmental relevance of S4 blastoids.

a, Q-PCR analysis showing the effect of LY2955303 on the expression levels of classical totipotency marker genes at S1. Data are presented as mean $\pm$ SD, $n = 3$ biological replicates. $^{**}p < 0.01$; $^{****}p < 0.0001$ (Two-way ANOVA). p values from left to right: 1.75×10^{-5} , 9.89×10^{-3} , 2.21×10^{-7} , 3.73×10^{-6} , 1.32×10^{-5} , 1.45×10^{-6} .

b, Representative bright-field images showing the effect of continuous LY2955303 treatment (10 μ M, E1.5 to E4.5) on blastocyst formation. Scale bar, 100 μ m. Similar images were obtained in at least 3 independent experiments.

c, Representative bright-field images showing the effect of continuous LY2955303 treatment (10 μ M, S1 to S4) on blastoid formation. Scale bar, 100 μ m.

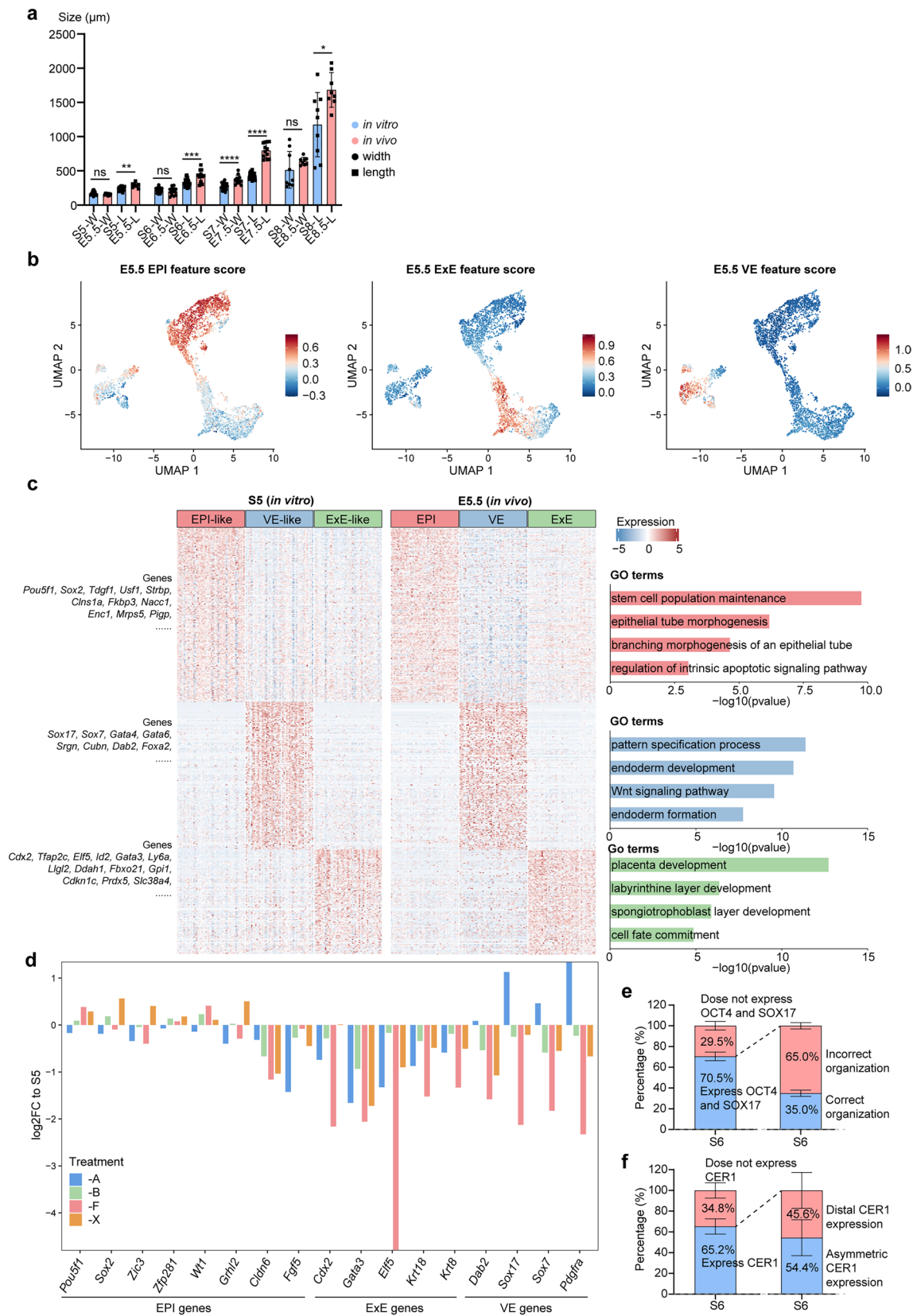
d, Histograms showing the effect of continuous LY2955303 treatment (10 μ M, S1 to S4) on blastoid formation. Data are presented as mean $\pm$ SD,

$n = 3$ biological replicates, $^{****}p < 0.0001$ (two-sided Student's t -test).

$p = 9.23 \times 10^{-5}$. **e**, Representative bright-field images showing the effect of continuous Y27632 treatment (20 μ M, E2.5 to E4.5) on blastocyst formation. Scale bar, 100 μ m. Similar images were obtained in at least 3 independent experiments.

f, Representative bright-field images showing the effect of continuous Y27632 treatment (20 μ M, S2 to S4) on blastoid formation. Scale bar, 100 μ m. **g**, Histograms showing the effect of continuous Y27632 treatment (20 μ M, S2 to S4) on blastoid formation. Data are presented as mean $\pm$ SD, $n = 3$ biological replicates, $^{****}p < 0.0001$ (two-sided Student's t -test).

$p = 3.80 \times 10^{-6}$. **h**, Representative images showing the formation of decidua in the mouse uterus after S4 blastoids were transferred *in vivo*. Scale bar, 1 cm (left); 200 μ m (right). This experiment was repeated 3 times.

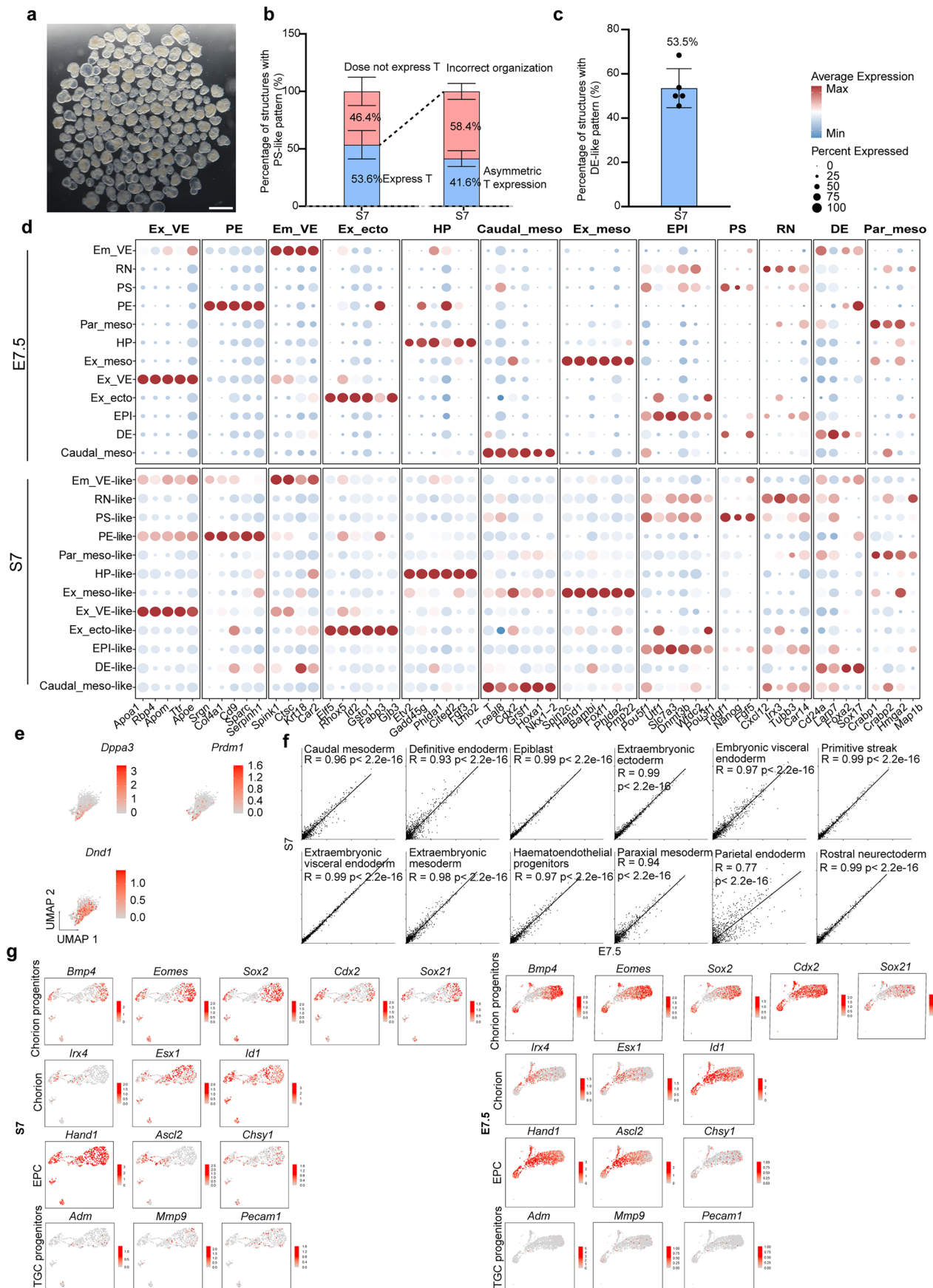


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Further characterization of embryoids at S5 and S6.

a, Size comparison of S5-S8 embryoids and stage matched embryos. Data are presented as mean $\pm$ SD, $n = 9$ for E5.5 embryos, $n = 16$ for S5 embryoids, $n = 13$ for E6.5 embryos, $n = 27$ for S6 embryoids, $n = 12$ for E7.5 embryos, $n = 25$ for S7 embryoids, $n = 8$ for E8.5 embryos, $n = 9$ for S8 embryoids. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, ns, non-significant (two-sided Student's t -test). p values from left to right: 0.185, 0.017, 0.234, 0.0001, 7.30×10^{-6} , 8.42×10^{-17} , 0.210, 0.0157. **b**, Feature plots showing the expression of genes enriched in EPI (left), ExE (middle) and VE (right) in EPI-, ExE, and VE-like cells from embryoids at S5. Specific gene sets for EPI, ExE, and VE in E5.5 embryos were identified using published single-cell transcriptomic datasets (GSE123046). **c**, Heatmaps showing the expression of genes enriched in EPI, ExE and VE in S5 embryoids

and natural E5.5 embryo. GO terms are shown on the right. Specific gene sets for EPI, ExE, and VE in E5.5 embryos were identified using published single-cell transcriptomic datasets (GSE123046). p value was calculated using the one-sided hypergeometric test. **d**, Bar charts showing expression of marker genes for EPI, ExE and VE in S5 embryoids under different chemical combinations. S5, Activin A + BMP4 + FGF4 + XAV939; -A, BMP4 + FGF4 + XAV939; -B, Activin A + FGF4 + XAV939; -F, Activin A + BMP4 + XAV939; -X, Activin A + BMP4 + FGF4. **e**, Percentage of embryoids at S6 expressing OCT4 and SOX17 with correct spatial organization. Data are presented as mean $\pm$ SD, $n = 75$ for S6 embryoids. **f**, Percentage of embryoids at S6 expressing CER1 with asymmetric expression pattern. Data are presented as mean $\pm$ SD, $n = 47$ for S6 embryoids.

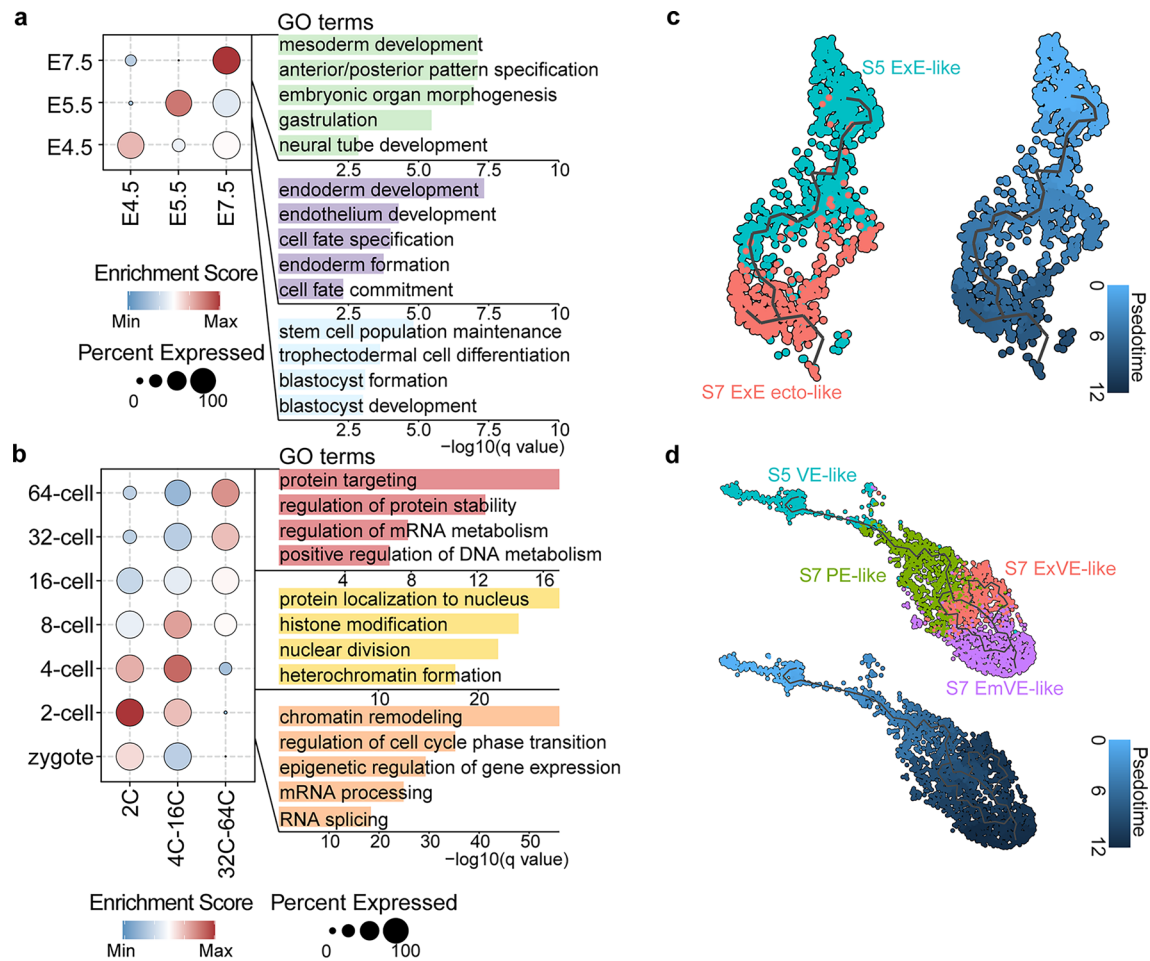


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Further characterization of embryoids at S7.

a, A representative bright field image of embryo-like structures at S7. Scale bar, 500 μ m. Similar images were obtained in at least 3 independent experiments. **b**, Percentage of embryoids at S7 expressing T with correct spatial organization. Data are presented as mean $\pm$ SD, $n = 48$ for S7 embryoids. **c**, Percentage of embryoids at S7 expressing FOXA2 and SOX17 with correct spatial organization. Data are presented as mean $\pm$ SD, $n = 67$ for S7 embryoids. **d**, Dot plots showing the average levels and proportion of cells expressing cell type-specific marker genes in S7 embryoids. Cell type-specific marker genes for different lineages in E7.5 mouse embryos were identified using published single-cell transcriptomic datasets ([E-MTAB-6967](#)). Expressions of these marker genes in different lineages from E7.5 mouse embryos are shown as the control. Em_VE, Embryonic visceral endoderm; RN, Rostral neur ectoderm; PS, Primitive Streak;

PE, Parietal endoderm; Par_meso, Paraxial mesoderm; HP, Haematoendothelial progenitors; Ex_meso, Extraembryonic mesoderm; Ex_VE, Extraembryonic visceral endoderm; Ex_ecto, Extraembryonic ectoderm; EPI, Epiblast; DE, Definitive endoderm; Caudal_meso, Caudal mesoderm. **e**, Feature plots showing the expression of PGC classical genes. **f**, Correlation of gene expression for each identified lineage between embryo-like structures at S7 and natural E7.5 embryos. Correlation coefficients and p values are indicated. p values were calculated using the two-sided t-test. **g**, Feature plots showing expression of indicated lineage markers in S7 embryoids (left) and natural E7.5 embryos (right). *Bmp4*, *Eomes*, *Sox2*, *Cdx2* and *Sox21* mark chorion progenitors; *Irx4*, *Esx1* and *Id1* mark chorion; *Hand1*, *Ascl2* and *Chsy1* mark EPC (ectoplacental cone cell); *Adm*, *Mmp9* and *Pecam1* mark TGC (Trophoblast Giant Cell) progenitors ([E-MTAB-6967](#)).



Extended Data Fig. 8 | Further characterization of the developmental trajectory of totipotent-like cell-derived embryo models from S1 to S7. **a, b**, Dot plots showing the average levels and proportion of cells expressing stage-specific gene sets in natural embryos from 2-cell embryo stage to E7.5 (a, E4.5 to E7.5; b, 2-cell embryo to 64-cell embryo). Stage-specific gene sets for 2C, 4C-16C, 32C-64C, E4.5, E5.5, and E7.5 mouse embryos were identified using

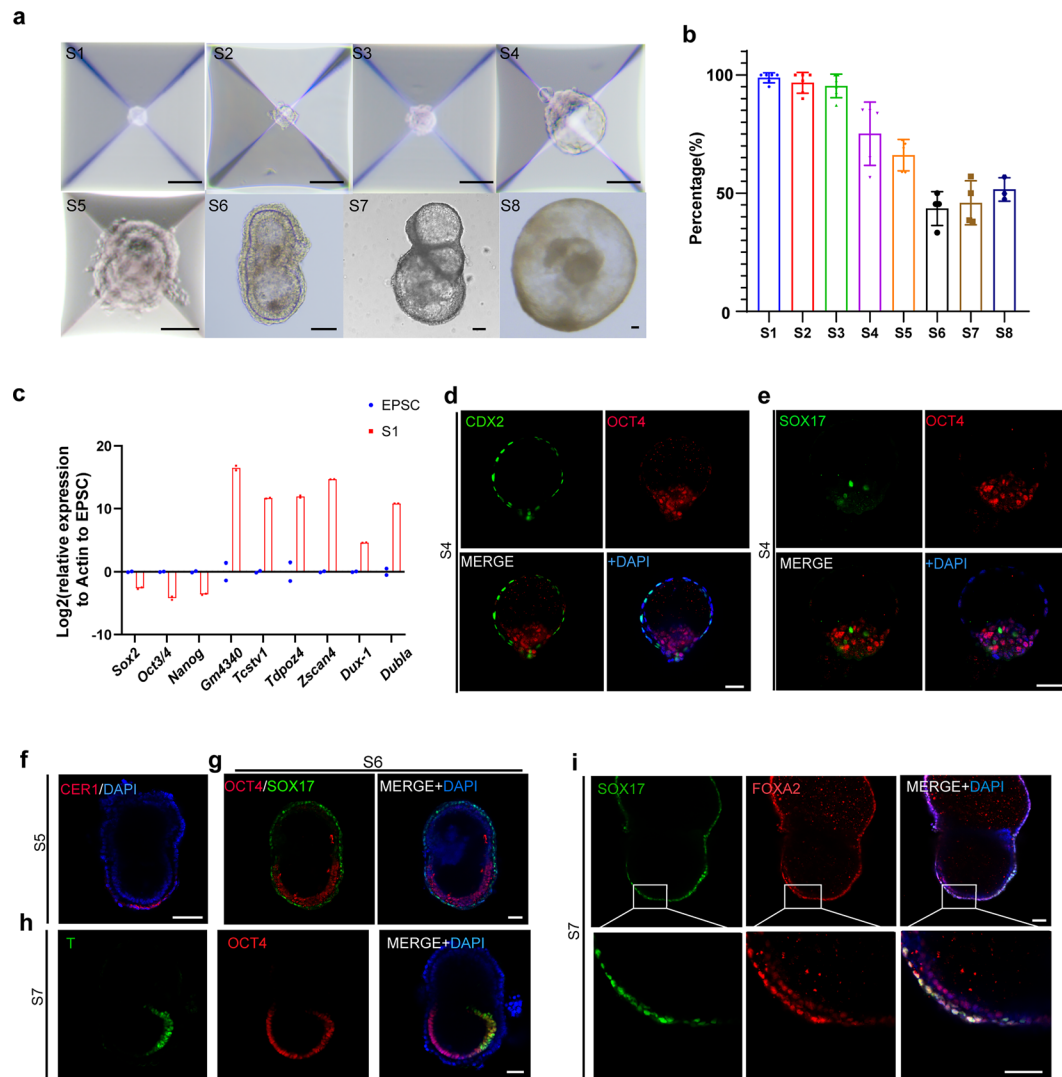
published single-cell transcriptomic datasets (GSE136714, GSE45719, GSE84892, GSE100597, GSE123046, E-MTAB-6967). GO terms that are specifically enriched in each stage are shown, which are also shared by their corresponding *in vitro* counter parts. **c, d**, Pseudotime trajectory reconstruction of ExE (c) and VE (d) lineages between S5 and S7.



Extended Data Fig. 9 | Further characterization of embryoids at S8.

a, Representative bright field images of abnormally developed S8 embryoids. Scale bars, 500 μm . Similar images were obtained in at least 3 independent experiments. **b, c**, Immunofluorescent analysis showing the expression of SOX1 in S8 embryoids (**b**) and E8.5 embryos (**c**), highlighting formation of the neural tube. $n = 2$ S8 embryoids from 2 experiments, $n = 2$ E8.5 embryos from 2 experiments. Scale bars, 100 μm . **d, e**, Immunofluorescent analysis showing the expression of GATA6 and MHC-II in S8 embryoids (**d**) and E8.5 embryos (**e**), highlighting initiation of heart formation. $n = 3$ S8 embryoids from 2 experiments, $n = 2$ E8.5 embryos from 2 experiments. Scale bars, 100 μm . **f**, Dot plots showing the average levels and proportion of cells expressing cell type-specific marker genes in embryo-like structures at S8. Cell type-specific marker genes for different lineages in E8.5 mouse embryos were identified using

published single-cell transcriptomic datasets ([E-MTAB-6967](#)). Expressions of these marker genes in different lineages from E8.5 mouse embryos are shown as the control. VE, Visceral endoderm; SE, Surface ectoderm; SC, Spinal cord; SM, Somitic mesoderm; RN, Rostral neurectoderm; PhM, Pharyngeal mesoderm; PGC, Primordial germ cell; PE, Parietal endoderm; PxM, Paraxial mesoderm; NMP, Neuromesodermal progenitor; NC, Neural crest; Mes, Mesenchyme; IM, Intermediate mesoderm; HP, Haematoendothelial progenitors; FB, Forebrain; ExE_M, Extraembryonic mesoderm; ExE_E, Extraembryonic endoderm; Eryth, Erythroid; Endo, Endothelium; DE, Definitive endoderm; CdM, Caudal mesoderm; CMs, Cardiomyocytes; BP, Blood progenitors; Al, Allantois. **g**, Correlation of gene expression for each identified lineage between embryo-like structures at S8 and natural E8.5 embryos. Correlation coefficients are indicated. Abbreviations are as defined in (**f**).



Extended Data Fig. 10 | Independent validation of the stepwise induction protocol by Guan laboratory. **a**, Bright field images of totipotent-like cell-derived embryo-like structures at different stages. Scale bars, 100 μ m. Similar images were obtained in at least 3 independent experiments. **b**, Quantification of embryo formation efficiency from S1 to S8. Data are mean $\pm$ SD, n = 240 for S1 from 6 experiments, 240 for S2 from 6 experiments, 240 for S3 from 6 experiments, 200 for S4 from 5 experiments, 200 for S5 from 5 experiments, 47 for S6 from 4 experiments, 67 for S7 from 4 experiments, 90 for S8 from 3 experiments. **c**, Q-PCR analysis showing the effect of S1 on the expression levels of classical totipotent marker genes. Samples were treated for three days before their collection. n = 2 biological replicates. **d**, Immunofluorescent analysis

showing the expression of CDX2 and OCT4 in blastoids at S4. n = 3 S4 blastoids from 2 experiments. Scale bar, 50 μ m. **e**, Immunofluorescent analysis showing the expression of SOX17 and OCT4 in blastoids at S4. n = 3 S4 blastoids from 2 experiments. Scale bar, 50 μ m. **f**, Immunofluorescent analysis showing the expression of CER1 at S5. n = 4 from 2 experiments. Scale bar, 50 μ m. **g**, Immunofluorescent analysis showing the expression of OCT4 and SOX17 at S6. n = 3 from 2 experiments. Scale bar, 50 μ m. **h**, Immunofluorescent analysis showing the expression of OCT4 and T at S7. n = 4 from 2 experiments. Scale bar, 50 μ m. **i**, Immunofluorescent analysis showing the expression of SOX17 and FOXA2 at S7. n = 3 from 2 experiments. Scale bar, 50 μ m.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

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

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

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	CFX384 RealTime System Nikon SMZ18 Nikon ECLIPSE Ti Olympus FV3000 Leica TCS-SP8 Leica Stellaris 8 Cell Ranger 5.0.1
Data analysis	Graphpad Prism 10.1.2 R version 4.3.2 Seurat 4.3.0.1 ClusterProfiler 4.10.0 FastQC 0.11.9 Trimomatic 0.39 featureCounts 2.0.6 STAR 2.7.10b pheatmap 1.0.12 Monocle 2.30.0 Monocle3 1.3.1

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

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Sequencing data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE287909 (scRNAseq) and GSE287908 (bulk RNA-seq).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

N/A

Reporting on race, ethnicity, or other socially relevant groupings

N/A

Population characteristics

N/A

Recruitment

N/A

Ethics oversight

N/A

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

Field-specific reporting

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

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

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

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

Sample size

Sample sizes were determined based on preliminary experiments, commonly used sample sizes in comparable publications within this field. We independently performed all experiments at least 3 times, allowing to judge the degree of variability between replicates. The sample sizes and number of repeats are defined in each figure legends.

Data exclusions

No data was excluded from the analysis.

Replication

All experiments presented here been repeated several times with similar results, the number of repeated experiments is indicated in the figure legends.

Randomization

For chimeric experiments, mice and embryos were randomly distributed amongst groups at the start of each experiment. For in vitro experiments using cells, cells were randomly assigned to either experiment or control groups.

Blinding

Blinding was not used in this study.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Antibodies

Antibodies used

Primary antibodies:

Rabbit Anti-Oct4 Monoclonal Antibody, Abcam; Cat#ab181557; RRID:AB_2687916; Clone:EPR17929; 1:500
 Mouse Anti-Cdx2 Monoclonal Antibody, BioGenex; Cat#MU392A-UC; RRID:AB_3101998; Clone:CDX2-88; 1:200
 Rabbit Anti-CDX2 Monoclonal Antibody, Abcam; Cat#ab76541; RRID:AB_1523334; Clone:EPR2764Y; 1:500
 Goat Anti-Human Sox17 Polyclonal Antibody; R and D Systems; Cat#AF1924; RRID:AB_355060; 1:200
 Rabbit Anti-AP2 gamma/TFAP2C Monoclonal Antibody; Abcam; Cat#ab218107; RRID:AB_2891087; Clone:EPR20331; 1:500
 Rabbit Anti-Brachyury / Bry Monoclonal Antibody; Abcam; Cat#ab209665; RRID:AB_2750925; Clone: EPR18113; 1:500
 Rabbit MuERV-L-Gag Polyclonal Antibody; Epigentek; Cat#A-2801; RRID:AB_3097813; 1:500
 Rabbit Anti-TBR2 / Eomes Monoclonal Antibody; Abcam; Cat#ab183991; RRID:AB_2721040; Clone: EPR19012; 1:500
 Gost anti-tdTomato Polyclonal Antibody; SICGEN; Cat#AB8181; RRID:AB_2722750; 1:2000
 Rabbit Anti-Cytokeratin 8 Monoclonal Antibody; Abcam; Cat#ab53280; RRID:AB_869901; Clone: EP1628Y; 1:500
 Rabbit Anti-Zscan4 Antibody Polyclonal Antibody; Millipore; Cat#AB4340; RRID:AB_2827621; 1:5000
 Rabbit Anti-RUNX1 / AML1 Monoclonal Antibody; Abcam; Cat#ab92336; RRID:AB_2049267; Clone: EPR3099; 1:200
 Rabbit Anti-FOXA2 Monoclonal Antibody; Abcam; Cat#ab108422; RRID:AB_11157157; Clone: EPR4466; 1:200
 Goat Anti-Cerberus 1 Polyclonal Antibody; R and D Systems; Cat#AF1986; RRID:AB_2077219; 1:200
 Mouse Anti-Human LRH-1/NR5A2 Monoclonal Antibody; R and D Systems; Cat#PP-H2325; RRID:AB_2154053; Clone:# H2325; 1:200
 Rabbit Anti-SOX2 Polyclonal Antibody; MilliporeSigma; Cat#AB5603; RRID:AB_2286686; 1:200
 Goat Anti-TFAP2C Polyclonal Antibody; R and D Systems; Cat#AF5059; RRID:AB_2255891; 1:200
 Goat anti-OTX2 Polyclonal Antibody; R and D Systems; Cat#AF1979; RRID:AB_2157172; 1:200
 Mouse anti-MHC-II Monoclonal Antibody; R and D Systems; Cat#MAB4470; RRID:AB_1293549; 1:500
 Goat anti-SOX1 Polyclonal Antibody; R and D Systems; Cat#AF3369; RRID:AB_2239879; 1:500

Secondary antibodies:

Cy3-AffiniPure Donkey Anti-Mouse IgG (H+L); Jackson ImmunoResearch Labs; Cat#715-165-151; RRID:AB_2315777; 1:200
 Alexa Fluor 488-AffiniPure Donkey Anti-Mouse IgG (H+L); Jackson ImmunoResearch Labs; Cat#715-545-150; RRID:AB_2340846; 1:200
 Alexa Fluor 647-AffiniPure Donkey Anti-Rabbit IgG (H+L); Jackson ImmunoResearch Labs; Cat#711-605-152; RRID:AB_2492288; 1:200
 Alexa Fluor 488-AffiniPure Donkey Anti-Rabbit IgG (H+L); Jackson ImmunoResearch Labs; Cat#711-545-152; RRID:AB_2313584; 1:200
 Cy3-AffiniPure Donkey Anti-Goat IgG (H+L); Jackson ImmunoResearch Labs; Cat#705-165-147; RRID:AB_2307351; 1:200
 Alexa Fluor 488-AffiniPure Donkey Anti-Goat IgG (H+L); Jackson ImmunoResearch Labs; Cat#705-545-147; RRID:AB_2336933; 1:200
 Alexa Fluor 647-AffiniPure Donkey Anti-Goat IgG (H+L); Jackson ImmunoResearch Labs; Cat#705-605-147; RRID:AB_2340437; 1:200

Validation

All the antibodies have been validated by the company that sells them. Details of the validation statement, antibody profiles and relevant citations can be found on the manufacturer's website. In addition to that, we validated all antibodies used in this study by immunofluorescence in natural mouse embryos and cells, and cross-checked them against their well-described expression pattern in reported studies.

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

Cell lines of EPSCs, mc6-1/Mu-mClover3-10# and C1-EPS-20# were used to generate S1-S8 embryoids. Both cell lines were generated in-house and cultured in LCDM condition.

Authentication

All cell lines in this study were authenticated via gene expression analysis.

Mycoplasma contamination

All cultured cells were tested negative for mycoplasma contamination by PCR.

Commonly misidentified lines
(See [ICLAC](#) register)

No commonly misidentified lines were used in this study.

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	Mice was used in this study. The male and female ICR mice were 6-8 weeks old. The mice were housed with a 12 hr light/dark cycle between 06:00 and 18:00 in a temperature controlled room ($22 \pm 1^\circ\text{C}$) with free access to water and food.
Wild animals	This study did not involve wild animals.
Reporting on sex	Both male and female embryos were used in this study; the embryos were not genotyped to determine the sex.
Field-collected samples	This study did not involve field-collected samples.
Ethics oversight	The Ethics Committee of Peking University Health Science Center approved the protocols used in this study.

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

Plants

Seed stocks	<i>Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.</i>
Novel plant genotypes	<i>Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.</i>
Authentication	<i>Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.</i>