

Induction of mouse totipotent stem cells by a defined chemical cocktail

<https://doi.org/10.1038/s41586-022-04967-9>

Received: 24 August 2021

Accepted: 10 June 2022

Published online: 21 June 2022

 Check for updates

Yanyan Hu^{1,2,3}, Yuanyuan Yang^{1,2,3}, Pengcheng Tan^{1,3}, Yuxia Zhang¹, Mengxia Han¹, Jiawei Yu¹, Xin Zhang¹, Zeran Jia¹, Dan Wang¹, Ke Yao¹, Huanhuan Pang¹, Zeping Hu¹, Yingqing Li¹, Tianhua Ma¹✉, Kang Liu¹✉ & Sheng Ding^{1,2}✉

In mice, only the zygotes and blastomeres from 2-cell embryos are authentic totipotent stem cells (TotiSCs) capable of producing all the differentiated cells in both embryonic and extraembryonic tissues and forming an entire organism¹. However, it remains unknown whether and how totipotent stem cells can be established in vitro in the absence of germline cells. Here we demonstrate the induction and long-term maintenance of TotiSCs from mouse pluripotent stem cells using a combination of three small molecules: the retinoic acid analogue TTNPB, 1-azakenpaullone and the kinase blocker WS6. The resulting chemically induced totipotent stem cells (ciTotiSCs), resembled mouse totipotent 2-cell embryo cells at the transcriptome, epigenome and metabolome levels. In addition, ciTotiSCs exhibited bidirectional developmental potentials and were able to produce both embryonic and extraembryonic cells in vitro and in teratoma. Furthermore, following injection into 8-cell embryos, ciTotiSCs contributed to both embryonic and extraembryonic lineages with high efficiency. Our chemical approach to totipotent stem cell induction and maintenance provides a defined in vitro system for manipulating and developing understanding of the totipotent state and the development of multicellular organisms from non-germline cells.

In mammals, intrinsic totipotency is the capacity by a cell to give rise to all the differentiated cells in both embryonic and extraembryonic tissues and form an organism¹. In mice, only the zygote and its immediate descendants, 2-cell-stage blastomeres, are totipotent^{2,3}. As an embryo develops, TotiSCs undergo irreversible differentiation into the first two different lineages: the trophectoderm, which generates the extraembryonic placenta, and the inner cell mass (ICM) which gives rise to the pluripotent epiblast (EPI) and primitive endoderm (PrE), which are precursors of embryonic fetus and extraembryonic yolk sac, respectively. So far, TotiSCs have been generated in vitro only transiently through in vitro fertilization or somatic cell nuclear transfer using germline cells. Therefore, inducing and maintaining authentic TotiSCs in vitro from non-germline represents a key target.

Recent studies have reported observations of cells expressing certain 2-cell-stage genes from mouse pluripotent stem cells⁴ (PSCs), particularly in the presence of specific small molecules, cytokines or transgenes^{5,6}. Around 0.1–1% of cells in mouse embryonic stem (ES) cell and mouse induced pluripotent stem cell cultures express murine endogenous retrovirus with leucine tRNA primer (MERVL) and a few other 2-cell-stage-specific genes (these cells are called 2-cell-like cells (2CLCs)), and possess the limited differentiation potential of 2-cell blastomeres⁴. Subsequent studies have identified several crucial factors determining the 2CLC state, such as DUX^{7,8}, and showed that modulating the expression of these factors was able to induce more 2CLCs. In addition, mouse expanded potential stem cells (mEPSCs) have the potential

to develop into both embryonic and extraembryonic lineages^{9,10}. However, the contribution of mEPSCs to extraembryonic tissues in vivo was challenged by a recent study with stringent criteria³. Notably, the transcriptional profiles of 2CLCs and mEPSCs actually more closely resemble mouse ES cells and post-implantation embryos than zygotes or 2-cell-stage totipotent embryos^{3,11}, therefore explaining their limited real developmental potentials compared with totipotent cells.

It has remained challenging to induce, capture and long-term maintain authentic TotiSCs in vitro from non-germline cells. In this article, we identify a combination of three small molecules that enable the induction and maintenance of ciTotiSCs from mPSCs, which closely resemble mouse totipotent 2-cell-stage embryos in terms of transcriptome, epigenome and metabolome, and developmental potentials into both embryonic and extraembryonic lineages.

Chemical induction of TotiSCs from mouse ES cells

To develop the conditions to induce TotiSCs, we first screened more than 3,000 chemical compounds using mouse ES cells with the 2-cell-stage-specific MERVL-tdTomato reporter and identified 23 compounds that could increase the number of tdTomato⁺ cells by threefold over the control condition (Fig. 1a,b and Extended Data Fig. 1a,b). Of note, we found several RAR agonists among those top hits. Although RAR agonists could induce massive generation of tdTomato⁺ cells in mouse ES cells (Extended Data Fig. 1c), these cells gradually died or

¹School of Pharmaceutical Sciences, Tsinghua University, Beijing, China. ²Tsinghua–Peking Joint Center for Life Sciences, Tsinghua University, Beijing, China. ³These authors contributed equally: Yanyan Hu, Yuanyuan Yang, Pengcheng Tan. ✉e-mail: matianhua@tsinghua.edu.cn; liukang_mark@hotmail.com; shengding@tsinghua.edu.cn

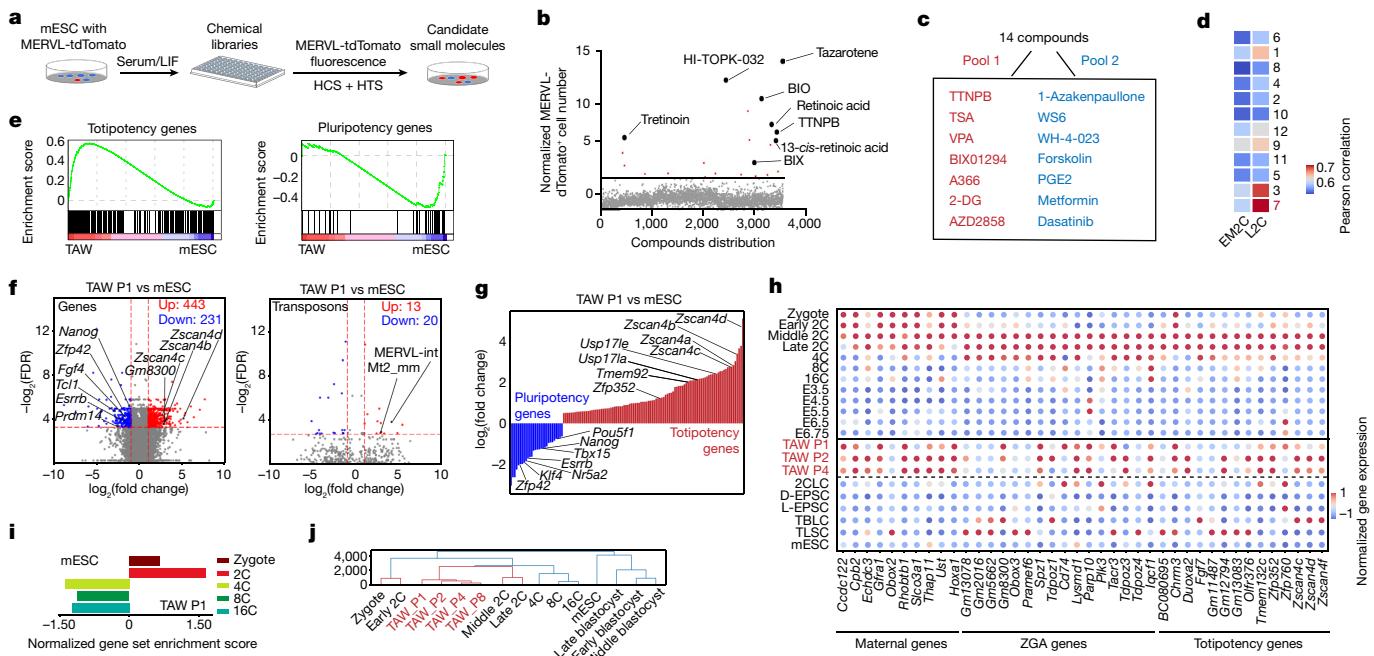


Fig. 1 | A chemical cocktail enables induction and maintenance of mouse totipotent stem cells. **a**, Schematic showing the high-throughput and high-content screening for small molecules capable of inducing the totipotency marker MERVLTdTomato in mouse ES cells (mESCs). **b**, Results of high-throughput and high-content screening. The threshold was set to three times the normalized MERVLTdTomato⁺ cell number of the DMSO control (solid line). Representative hit molecules are labelled. BIO, 6-bromoindirubin-3-oxime; BIX, BIX01294. **c**, Selected compound pools used for the second round of combination tests. 2-DG, 2-deoxy-D-glucose; PGE2, prostaglandin E2; TSA, trichostatin A; VPA, valproic acid. **d**, Transcriptome correlation analysis of cells treated with various combinations of compounds (1–12) and mouse 2-cell-stage embryos. EM2C, early-middle 2-cell embryo; L2C, late 2-cell embryo. **e**, GSEA of totipotency genes (left) and pluripotency genes (right) in mouse ES cells treated with or without TAW for one passage. **f**, Volcano plots showing up-regulated (red) and down-regulated (blue) genes (left; log₂(fold change) > 1, false discovery rate (FDR) < 0.1) and transposons (right; log₂(fold change) > 1),

FDR < 0.15) in mouse ES cells treated with TAW. The Benjamini–Hochberg method was used to control the FDR. **g**, Transcriptional changes of pluripotency-specific (blue) and totipotency-specific (red) genes in mouse ES cells treated with TAW for one passage. **h**, Normalized expression of maternal, zygotic gene activation (ZGA) and totipotency-specific genes by bulk RNA-seq analyses of mouse ES cells, mouse ES cells treated with TAW for 1, 2 or 4 passages (TAW P1, P2 or P4, respectively), TBLCs¹¹, TLSCs¹⁹, L-EPSCs¹⁰, D-EPSCs⁹ and scRNA-seq analyses of mouse embryos at various stages from zygote, 2-, 4-, 8- and 16-cell (2C, 4C, 8C and 16C) stages to E6.75. **i**, GSEA analysis of bulk RNA-seq data from mouse ES cells treated with TAW for one passage by the indicated embryonic stage-specific gene sets. **j**, Hierarchical clustering analysis of mouse ES cells, mouse ES cells treated with TAW for 1, 2, 4 or 8 passages and mouse embryos at various stages from zygote to blastocyst, based on defined gene sets including totipotency, pluripotency, maternal and ZGA genes¹⁹. All datasets in bioinformatic analyses were summarized in Supplementary Table 1.

differentiated and did not achieve the totipotent state in the presence of RAR agonists in long-term culture. This result indicated that activation of RAR alone was not sufficient to establish and maintain cell totipotency¹². To overcome this hurdle, we subsequently examined additional small molecules, especially those enhancing reprogramming, survival or self-renewal of stem cells, in a combinatorial manner. Next, we designed and constructed a 14-compound library in two subgroups. Group 1 contained 7 different hits from the first round of chemical screening, and group 2 included compounds reported to enhance cell survival, proliferation, reprogramming and differentiation inhibition (Fig. 1c). After testing various combinations of these compounds, we identified 11 combinations that achieved induction and long-term maintenance of MERVLTdTomato⁺ cells in vitro (Extended Data Fig. 1d,e). Given that MERVLTdTomato expression alone does not represent an authentic TotiSC state, we performed RNA-sequencing (RNA-seq) analysis to further evaluate these conditions stringently. Condition no. 7 out of the 11 combinations, which comprised TTNPB, 1-azakenpaullone and WS6 (collectively referred to as TAW), induced cells to the state most closely resembling the mouse 2-cell embryo, as revealed by analyses of transcriptome-based correlation and differentially expressed genes (DEGs) (Fig. 1d and Extended Data Fig. 1f).

Although typical MERVLTdTomato⁺ dome-shaped colonies were retained after TotiSC induction and over more than four passages under the TAW conditions, fluorescence-activated cell sorting (FACS) analysis

showed that expression of the pluripotency gene *OCT4* was significantly down-regulated in cells at passage (P)1 and decreased to the lowest level from P2, indicating that most cells were reprogrammed back from pluripotency toward totipotency, which was further confirmed by immunostaining of OCT4 and NANOG (Extended Data Fig. 1g,h). RNA-seq analysis indicated that nearly all the classical totipotency-specific genes, including the *Zscan4* family, *Tcstv3*, *Gm6763*, *Gm8300*, *Usp17la*, *Tmem92*, *Btg1* and MERVLT repeats (MERVL-int and Mt2_mm), were activated or highly expressed, whereas most pluripotency-specific genes, including *Zfp42*, *Pou5f1*, *Sox2*, *Nanog*, *Tbx15*, *Esrbb* and *Fgf4* were silenced in TAW-treated cells from P1 to P8 (Fig. 1e–g and Extended Data Fig. 1j–l). In addition to the classical totipotency-specific genes (such as *Zscan4* family genes), many other essential totipotency transcription factors, especially maternal and ZGA genes—which have important roles in spermatogenesis, oogenesis, early embryo development and zygotic genome activation (including *Tdpoz3*¹³, *Tdpoz4*¹³, *Zfp352*¹⁴, *Spz1*¹⁵, *Gfra1*¹⁶, *Rhobtb1*¹⁷ and *Iqcf1*¹⁸)—were also specifically activated or highly expressed in TAW-treated cells, compared with mouse ES cells, mEPSCs^{8,9} (D-EPSCs and L-EPSCs), 2CLCs⁴ or totipotent blastomere-like cells¹¹ (TBLCs) and totipotent-like stem cells¹⁹ (TLSCs) (Fig. 1h). Gene set enrichment analysis (GSEA) using genes expressed at pre-implantation embryonic stages revealed significant enrichment of zygote and 2-cell-stage signatures in TAW-reprogrammed cells (Fig. 1i, Extended Data Fig. 1m and Supplementary Table 2). Clustering analysis

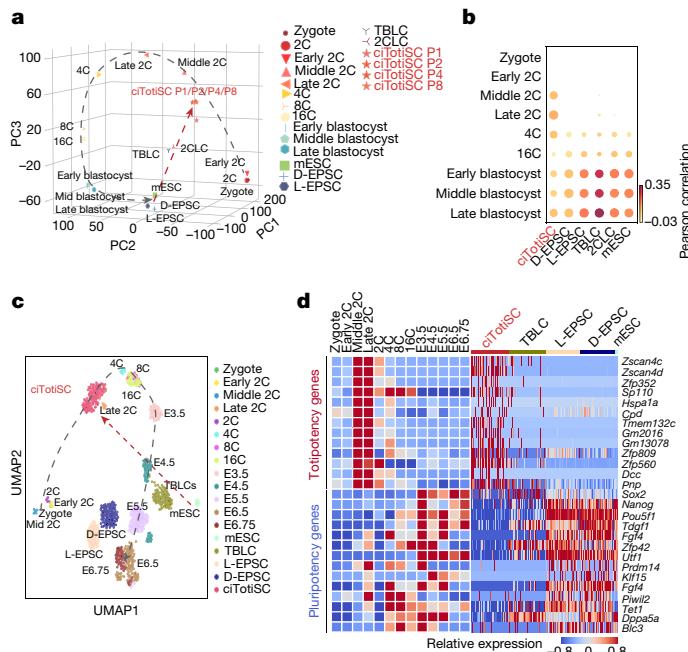


Fig. 2 | ciTotiSCs possess characteristic transcriptomic features resembling totipotent 2-cell embryo blastomeres. **a**, Transcriptome-based PCA of bulk RNA-seq data from ciTotiSCs (P1, P2, P4 and P8) and 2CLCs, and scRNA-seq data from mouse ES cells, L-EPSCs, D-EPSCs, TBLCs and mouse embryos from zygote to late blastocyst. **b**, Correlation analysis of bulk RNA-seq data from 2CLCs, and scRNA-seq data from ciTotiSCs, mouse ES cells, L-EPSCs, D-EPSCs, TBLCs and mouse embryos from zygote to late blastocyst, based on defined totipotency and pluripotency gene sets¹⁹. **c**, Transcriptome UMAP analysis of ciTotiSCs, mouse ES cells, L-EPSCs, D-EPSCs, TBLCs and mouse embryos at the indicated stages at the single-cell level. **d**, Heat map comparing scRNA-seq analysis of pluripotency- or totipotency-specific gene expression in mouse ES cells, L-EPSCs, D-EPSCs, TBLCs, ciTotiSCs (right) and mouse embryos from zygote to E6.75 (left). All datasets in bioinformatic analyses were summarized in Supplementary Table 1.

of TAW-reprogrammed cells from P1 to P8, mouse ES cells and a series of early mouse embryos (from zygote to late blastocyst) indicated that TAW-reprogrammed cells were clustered tightly together with totipotent 2-cell embryos (Fig. 1j).

Together, these results suggested that the combination of TTNPB, 1-azakenpaulone and WS6 enables the generation of ciTotiSCs from mouse ES cells and allows their maintenance in vitro for multiple passages with normal diploid karyotypes (Extended Data Fig. 1i).

Molecular characterization of ciTotiSCs

To thoroughly characterize mouse ciTotiSCs, we stringently analysed their whole transcriptome rather than a panel of specific marker genes. Transcriptome-based principal component analysis (PCA) of ciTotiSCs, mouse ES cells, D-EPSC, L-EPSC, TBLCs, 2CLCs and a series of early mouse embryos (from zygote to E6.5 embryo) indicated that ciTotiSCs (from P1 to P8) exhibited a complete transcriptional shift from mouse ES cells to the state of 2-cell embryo, whereas TBLCs and 2CLCs appeared to represent intermediate stages of pluripotency-to-totipotency transition (Fig. 2a). Pearson correlation analysis also consistently revealed that ciTotiSCs were similar to 2-cell embryos (Fig. 2b).

Next, we sought to evaluate the cell heterogeneity of ciTotiSCs at single-cell level by single-cell RNA-seq (scRNA-seq), as MERVL was expressed heterogeneously in ciTotiSCs (Extended Data Fig. 2a). Uniform manifold approximation and projection (UMAP) analysis identified three clusters (A to C). Further analysis revealed that totipotency

genes were all expressed in those three clusters, whereas some XEN-like cell genes and pluripotency genes were expressed at very low level in cluster B (Extended Data Fig. 2b,c). UMAP and single-cell transcriptome-based PCA analysis together with mouse embryos indicated that ciTotiSCs were most consistently correlated with the 2-cell mouse embryo, whereas D-EPSCs⁹ and L-EPSCs¹⁰ were more similar to E5.5–E6.75 embryos, as previously reported³ (Fig. 2c and Extended Data Fig. 2d). Furthermore, UMAP analysis at single-cell level showed that a number of TBLCs resembled mouse ES cells and embryonic day (E)4.5 embryos rather than 2-cell embryos (Fig. 2c and Extended Data Fig. 2d). In addition, analysis of representative marker genes at single-cell level indicated that totipotent embryos and ciTotiSCs show high expression of totipotency-specific genes, such as *Zscan4c*, *Zscan4d*, *Zfp352* and *Sp110*, and suppress the expression of pluripotency genes, such as *Pou5f1*, *Tdgp1*, *Nanog*, *Zfp42* and *Sox2*, compared with mouse ES cells and mouse embryos at later stages. By contrast, no significant or partial changes in expression of these genes were observed in L-EPSC, D-EPSC and TBLCs (Fig. 2d and Extended Data Fig. 2e). Together, these transcriptome analyses consistently revealed that ciTotiSCs were distinguishable from mEPSCs and TBLCs, and closely resemble mouse totipotent 2-cell blastomeres.

To further characterize ciTotiSCs, we used assay for transposase-accessible chromatin sequencing (ATAC-seq) to interrogate genome-wide chromatin accessibility in ciTotiSCs. When comparing with mouse ES cells, ciTotiSCs gained 7,034 open loci that resemble those in 2-cell embryos, and 4,361ICM open loci specific to blastocyst were lost²⁰ (Fig. 3a and Extended Data Fig. 3a). As expected, in ciTotiSCs, 2-cell-specific genes, including the *Zscan4* family, *Zfp352*, *Gm13078*, *Usp17la* and specific retroviral repeats (*ERV1*, *ERVK*, *MERVL-int* and *Mt2-mm*) were located within open chromatin loci, whereas pluripotency genes, such as *Nanog*, *Zfp42*, *Fgf4*, *Tdgp1*, and *Esrbb*, were located in closed chromatin loci (Fig. 3b,c and Extended Data Fig. 3b,c). These ATAC-seq analyses indicated that ciTotiSCs possess genome-wide totipotency-specific chromatin accessibility.

During early embryogenesis, genomic DNA methylation patterns were highly dynamic and developmental stage-dependent. We performed reduced representation bisulfite sequencing (RRBS) to analyse the DNA methylome of ciTotiSCs. We found that the genome of ciTotiSCs was hypomethylated, similar to zygotes or 2-cell embryos²¹ (Fig. 3d,f). Specifically, we observed DNA hypomethylation in genomic loci including 2-cell-specific genes, retroviral repeats and on the X chromosome (Extended Data Fig. 3d–f). Furthermore, global DNA methylome-based PCA analysis showed that ciTotiSCs were closely correlated with totipotent zygotes and 2-cell embryos, and were distinct from their parental mouse ES cells (Fig. 3e). These RRBS analyses consistently demonstrated that ciTotiSCs shared similar genomic DNA methylation patterns with mouse totipotent embryos.

We next examined the metabolome of ciTotiSCs on the basis that totipotent cells exhibit key metabolic features²². Consistently, we found that the top differentially detected metabolites between ciTotiSCs and mouse ES cells closely resembled those between totipotent 2-cell embryos and blastocysts (Fig. 3g). Further analysis demonstrated that totipotent ciTotiSCs and 2-cell embryos favour one-carbon metabolism and pathways related to a more reductive state, whereas pluripotent mouse ES cells and blastocysts possess higher levels of metabolites in purine metabolism and mitochondrial tricarboxylic acid cycle, representing a more oxidative state (Extended Data Fig. 3g).

Together, these results collectively demonstrated that ciTotiSCs generated in vitro highly resemble mouse early totipotent embryo in vivo at transcriptome, epigenome and metabolome level.

Differentiation potential of ciTotiSCs

In contrast to PSCs that can contribute only to the fetus, mammalian totipotent stem cells can give rise to both embryonic and extraembryonic

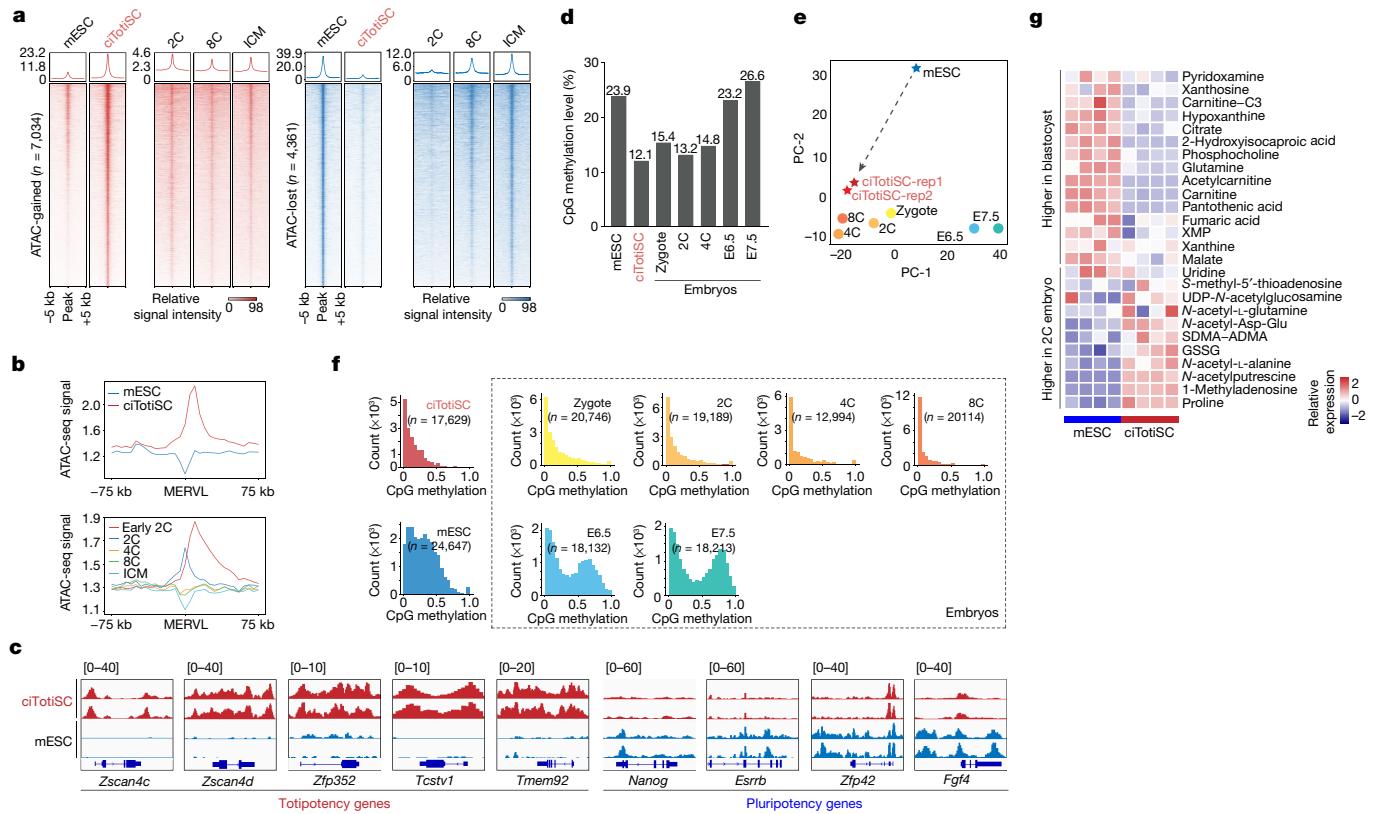


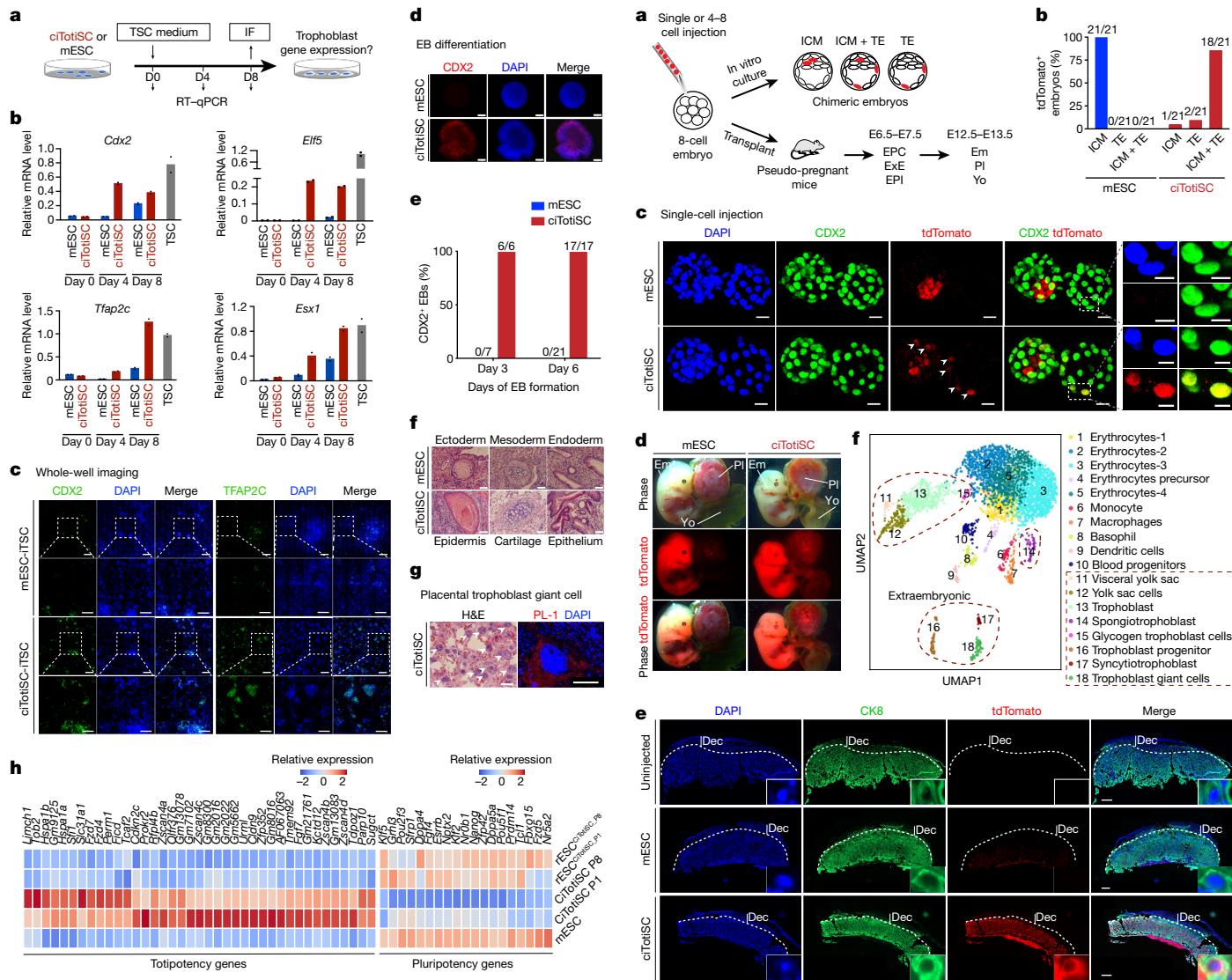
Fig. 3 | The epigenome and metabolome of ciTotiSCs are similar to those of mouse totipotent blastomeres. **a**, ATAC-seq analysis of ciTotiSCs, mouse ES cells, 2-cell or 8-cell embryos and the ICM. Heat maps show the landscapes of open (red) and closed (blue) peaks around the peak centre. **b**, Meta plot of ATAC-seq signals in the 75-kb region around whole-genome-wide MERVLs. **c**, Integrative Genomics Viewer (IGV) analysis showing ATAC-seq signals near representative totipotency and pluripotency genes. **d**, Histogram showing CpG methylation levels in mouse ES cells, ciTotiSCs and embryos at the indicated stages, calculated using RRBS data. **e**, Whole-genome DNA methylation PCA analysis of ciTotiSCs, mouse ES cells and mouse embryos at the indicated stages. Two replicates for ciTotiSC, rep1 and rep2. **f**, RRBS analysis of DNA CpG methylation on 100-kb tiles in ciTotiSCs, mouse ES cells and mouse embryos at the indicated stages. *n* represents the number of tiles in each sample. **g**, Heat map showing the relative abundance of differentially exhibited metabolites in mouse ES cells and ciTotiSCs. *n* = 4 biological replicates. All datasets in bioinformatic analyses were summarized in Supplementary Table 1.

indicated stages, calculated using RRBS data. **e**, Whole-genome DNA methylation PCA analysis of ciTotiSCs, mouse ES cells and mouse embryos at the indicated stages. Two replicates for ciTotiSC, rep1 and rep2. **f**, RRBS analysis of DNA CpG methylation on 100-kb tiles in ciTotiSCs, mouse ES cells and mouse embryos at the indicated stages. *n* represents the number of tiles in each sample. **g**, Heat map showing the relative abundance of differentially exhibited metabolites in mouse ES cells and ciTotiSCs. *n* = 4 biological replicates. All datasets in bioinformatic analyses were summarized in Supplementary Table 1.

lineages. We therefore compared the differentiation potentials of mouse ES cells and ciTotiSCs by inducing their differentiation to extraembryonic trophoblast stem cells (TSCs) in the presence of FGF4 and heparin and examining the expression of TSC-specific markers at different time points (Fig. 4a). As expected, analysis by quantitative PCR with reverse transcription (RT-qPCR) confirmed that typical trophoblast specific marker genes (such as *Cdx2*, *Eif5*, *Tfap2c* and *Esx1*) were gradually induced in the differentiating ciTotiSCs cells in a time-dependent manner (Fig. 4b). In addition, robust induction of CDX2 and TFAP2C expression was observed during ciTotiSC differentiation by immunostaining and unbiased whole-well imaging analysis, but was rare in mouse ES cell-derived cells (Fig. 4c). ciTotiSCs could also gradually differentiate into pluripotent embryonic stem cells after transfer to mouse ES cell medium (2i/LIF medium), thus mimicking the normal embryonic development process (Fig. 4h). To further evaluate the intrinsic differentiation potential of ciTotiSCs into extraembryonic lineages, we performed a spontaneous embryoid body differentiation assay. On day 3 and day 6 of embryoid body differentiation, a high percentage of CDX2⁺ cells arose in all ciTotiSCs-derived embryoid bodies, whereas little CDX2 signal could be detected in embryoid bodies from mouse ES cells (Fig. 4d,e). When ciTotiSCs were transplanted into immunodeficient mice, the resulting ciTotiSC-derived teratomas contained cells from the three embryonic germ layers (Fig. 4f) as well as extraembryonic lineages, such as trophoblast giant cells, characterized by enlarged nuclei, large cell volume, proximity to internal haemorrhages and expression of PL-1 (Fig. 4g). These results demonstrated that ciTotiSCs are capable of differentiating into both embryonic and extraembryonic lineages spontaneously in vitro and in teratoma.

Lineage contribution of ciTotiSCs in vivo

One of the most stringent criteria for totipotency is the contribution to both embryonic and extraembryonic tissues in chimeric mice. We therefore performed chimerism experiments using ciTotiSCs and analysed lineage contributions at different time points during pre-implantation and post-implantation embryonic development (Fig. 5a). In brief, single or 4–8 tdTomato-labelled mouse ES cells or ciTotiSCs were injected into host mouse 8-cell embryos. After 48 h of in vitro culture, the chimeric embryos developed into late blastocyst stage (E4.5). As expected, mouse ES cells contributed only to embryonic ICM. By contrast, 18 out of the 21 embryos injected with ciTotiSCs (85.7%) had a large proportion of tdTomato⁺ cells in both the ICM and trophectoderm, 2 embryos (9.5%) had tdTomato⁺ cells only in trophectoderm and 1 embryo (4.8%) had tdTomato⁺ cells only in the ICM (Fig. 5b and Extended Data Fig. 4a). Immunostaining of the trophectoderm-specific marker CDX2 confirmed that tdTomato-labelled ciTotiSCs could indeed differentiate into functional trophectoderm lineages with co-expression of CDX2 (Fig. 5c and Extended Data Fig. 4b). These results demonstrated that ciTotiSCs possess robust potentials to develop into both early embryonic and extraembryonic lineages in pre-implantation blastocyst. For the post-implantation test, we transplanted 8-cell embryos injected with single or multiple tdTomato-labelled ciTotiSCs or mouse ES cells into the oviducts of pseudo-pregnant female mice and further examined chimeras around E6.5–E7.5 and E12.5–E13.5. In E7.5 chimeric embryos, we observed that single ciTotiSCs contributed to a large



embryos, whereas mouse ES cells could develop into embryonic parts only (Fig. 5d and Extended Data Fig. 4d,e,h).

Next, we rigorously examined whether ciTotiSCs were able to contribute to the extraembryonic trophoblast compartment of E13.5 placenta. The placenta has a complex structure that includes three layers. Aside from decidua (the maternal part of placenta) and the junctional zone, the labyrinth layer contains both extraembryonic trophoblast and embryo-derived cells. We observed that tdTomato-labelled ciTotiSC-derived cells were localized in the labyrinth and the junctional zone of the placenta and showed expression of the trophoblast marker CK8 and the trophoblast giant cell marker proliferin, whereas mouse ES cell-derived cells were found only in labyrinth, particularly within the embryonic compartment, and did not stain for CK8 and proliferin (Fig. 5e and Extended Data Fig. 4i). This result further confirmed that ciTotiSCs were able to develop into the functional extraembryonic compartment of placenta *in vivo*. Moreover, to fully characterize descendent cell types from ciTotiSCs in extraembryonic tissues around E13.5, we used scRNA-seq to analyse ciTotiSC-derived tdTomato⁺ cells in chimeric placenta and yolk sac. After clustering with extraembryonic lineage-specific genes, we revealed that ciTotiSC-derived cells could develop into representative extraembryonic trophoblast and yolk sac cell types, such as visceral yolk sac cells (*Apoa4*⁺, *Fxyd2*⁺ and *Entpd2*⁺), spongiotrophoblast cells (*Tpbpa*⁺ and *Rhox9*⁺) and syncytiotrophoblast cells (*Itm2a*⁺), and cells of embryonic origin including erythrocytes, macrophages and monocytes (Fig. 5f and Extended Data Fig. 4j). Moreover, ciTotiSCs retained the ability to produce germline chimera and healthy chimeric offspring (Extended Data Fig. 4f,g).

Collectively, these results demonstrate that ciTotiSCs are capable of generating all the differentiated cell types in both embryonic and extraembryonic tissues of an organism with high efficiency, functionally resembling totipotent blastomere³.

Discussion

In this study, we successfully generated ciTotiSCs with molecular and functional characteristics of mouse 2-cell embryos from mouse ES cells using a defined chemical cocktail *in vitro*. With limitations by different sequencing and bioinformatic methods, and/or the marker gene-sets selected for comparison in mind, our thorough analyses consistently revealed that ciTotiSCs specifically expressed nearly all recognized totipotency genes and showed almost complete silencing of pluripotency genes. More importantly, ciTotiSCs possess totipotent developmental potential and enable highly efficient generation of embryonic and extraembryonic cells both *in vitro* and *in vivo*. Therefore, our rigorous molecular and functional analyses indicate that ciTotiSCs closely resemble authentic mouse totipotent 2-cell-stage blastomeres.

All three components of TAW are indispensable for ciTotiSC induction and maintenance (Extended Data Fig. 5a). Activation of the retinoic acid receptor by TTNPB is essential for the induction of cell totipotency, consistent with a previous report¹² (Extended Data Fig. 5b–g). 1-Azakenpaualone appears to function in part by inhibiting differentiation and other negative effects of TTNPB in long-term culture, and by promoting self-renewal of ciTotiSCs through simultaneous specific activation of Wnt signalling and G2/M staging (Extended Data Fig. 5h,i). Notably, WS6 may act to promote and stabilize ciTotiSCs by inhibiting the NF-κB-mediated immune response triggered by ERV activation^{23–26} (Extended Data Fig. 5k,j). In addition, we found that *Dux* is essential for ciTotiSC induction, which is also partially dependent on *Tp53*, as revealed by *Dux*- or *Tp53*-knockout experiments (Extended Data Fig. 6a–d).

In summary, these studies demonstrate the feasibility of generating mouse totipotent stem cells *in vitro* from non-germline cells, and

provide a cellular system to better manipulate and understand the initial cell fate determination in mammal development and investigate basic questions concerning the development of multicellular organisms.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-022-04967-9>.

1. Tarkowski, A. K. Experiments on the development of isolated blastomeres of mouse egg. *Nature* **184**, 1286–1287 (1959).
2. Solter, D. From teratocarcinomas to embryonic stem cells and beyond: a history of embryonic stem cell research. *Nat. Rev. Genet.* **7**, 319–327 (2006).
3. Posfai, E. et al. Evaluating totipotency using criteria of increasing stringency. *Nat. Cell Biol.* **23**, 49–60 (2021).
4. Macfarlan, T. S. et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57 (2012).
5. Iturride, A. & Torres-Padilla, M. E. A cell in hand is worth two in the embryo: recent advances in 2-cell like cell reprogramming. *Curr. Opin. Genet. Dev.* **64**, 26–30 (2020).
6. Riveiro, A. R. & Brickman, J. M. From pluripotency to totipotency: an experimentalist's guide to cellular potency. *Development* **147**, dev.189845 (2020).
7. De laico, A. et al. DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* **49**, 941 (2017).
8. Hendrickson, P. G. et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925 (2017).
9. Yang, Y. et al. Derivation of pluripotent stem cells with *in vivo* embryonic and extraembryonic potency. *Cell* **169**, 243–257.e225 (2017).
10. Yang, J. et al. Establishment of mouse expanded potential stem cells. *Nature* **550**, 393 (2017).
11. Shen, H. et al. Mouse totipotent stem cells captured and maintained through spliceosomal repression. *Cell* **184**, 2843–2859.e20 (2021).
12. Ituride, A. et al. Retinoic acid signaling is critical during the totipotency window in early mammalian development. *Nat. Struct. Mol. Biol.* **28**, 521–532 (2021).
13. Huang, C. J., Chen, C. Y., Chen, H. H., Tsai, S. F. & Choo, K. B. TDPOZ, a family of bipartite animal and plant proteins that contain the TRAF (TD) and POZ/BTB domains. *Gene* **324**, 117–127 (2004).
14. Liu, T. Y., Chen, H. H., Lee, K. H. & Choo, K. B. Display of different modes of transcription by the promoters of an early embryonic gene, Zfp352, in preimplantation embryos and in somatic cells. *Mol. Reprod. Dev.* **64**, 52–60 (2003).
15. Hsu, S. H., Hsieh-Li, H. M. & Li, H. Dysfunctional spermatogenesis in transgenic mice overexpressing bHLH-Zip transcription factor, Spz1. *Exp. Cell. Res.* **294**, 185–198 (2004).
16. Kawamura, K. et al. Completion of meiosis I of preovulatory oocytes and facilitation of preimplantation embryo development by glial cell line-derived neurotrophic factor. *Dev. Biol.* **315**, 189–202 (2008).
17. Xu, Y. et al. Gene expression profiles in mouse cumulus cells derived from *in vitro* matured oocytes with and without blastocyst formation. *Gene Expr. Patterns* **25–26**, 46–58 (2017).
18. Fang, P. et al. A novel acrosomal protein, IQCF1, involved in sperm capacitation and the acrosome reaction. *Andrology* **3**, 332–344 (2015).
19. Yang, M. et al. Chemical-induced chromatin remodeling reprograms mouse ESCs to totipotent-like stem cells. *Cell Stem Cell* **29**, 400–418.e13 (2022).
20. Wu, J. et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652–657 (2016).
21. Smith, Z. D. et al. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339–344 (2012).
22. Zhao, J. et al. Metabolic remodelling during early mouse embryo development. *Nat. Metab.* **3**, 1372–1384 (2021).
23. Lima-Junior, D. S. et al. Endogenous retroviruses promote homeostatic and inflammatory responses to the microbiota. *Cell* **184**, 3794–3811.e3719 (2021).
24. Chiappinelli, K. B. et al. Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell* **162**, 974–986 (2015).
25. Nishikimi, A., Mukai, J. & Yamada, M. Nuclear translocation of nuclear factor kappa B in early 1-cell mouse embryos. *Biol. Reprod.* **60**, 1536–1541 (1999).
26. Paciolla, M. et al. Nuclear factor-kappa-B-inhibitor alpha (NFKBA) is a developmental marker of NF-κB/p65 activation during *in vitro* oocyte maturation and early embryogenesis. *Hum. Reprod.* **26**, 1191–1201 (2011).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2022

Article

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Mouse strains

All animal experiments in this study were performed in accordance with the guidelines and regulations of IACUC (Institutional Animal Care and Use Committee) of Tsinghua University, Beijing, China. All the animal protocols used in this study have been approved by IACUC (Institutional Animal Care and Use Committee) of Tsinghua University. All mice were housed in individually ventilated cages (maximal six mice per cage) at Tsinghua University. The mice were maintained on a 12/12-h light and dark cycle, 22–26 °C with sterile pellet food and water ad libitum. ICR mice purchased from Beijing Vital River Laboratory Animal Technology were used as embryo donor and the pseudo-pregnant recipients for embryo injection. For chimeric experiments, mouse 8-cell embryos were collected by flushing the uterus of 6–8 weeks old pregnant ICR mouse with M2 medium (Sigma, M7167) and cultured in KSOM medium (Millipore, MR-101) drops covered by mineral oil (Sigma, M8410) under 5% CO₂ at 37 °C.

Cell culture

Mouse ES cell lines including R1-mESCs, OG2-mESCs and tdTomato-labelled mouse ES cells (derived from E3.5 ICM of C57BL/6-Gt(ROSA)26Sor^{tm1(CAG-tdTomato)}/Vst, Vitalstar) were cultured on inactivated mouse embryonic fibroblast (MEF) feeder layers pre-coated with 0.3% gelatin (Sigma, G1890), using Serum/LIF/2i medium composed of KnockOut DMEM (GIBCO, 10829018) supplemented with 15% FBS (GIBCO, 1009914), 1× L-glutaMAX (GIBCO, 35050061), 1× penicillin-streptomycin (GIBCO, 10378016), 1× non-essential amino acids (NEAA) (GIBCO, 11140050), 1× sodium pyruvate (GIBCO, 11360070), 55 μM 2-mercaptoethanol (ThermoFisher, 21985023), 1,000 U ml⁻¹ mouse leukaemia inhibitory factor (mLIF) (Millipore, ESG1107), 3 μM CHIR-99021 (Selleck, S2924) and 1 μM PD0325901 (Selleck, S1036). All cell lines were mycoplasma negative.

To generate the MERVL-tdTomato reporter cell line, R1-mESCs or OG2-mESCs were transfected with MERVL-tdTomato (MERVL-tdTomato) reporter plasmid (Addgene plasmid #40281) using Lipofectamine 3000 (Thermo Fisher Scientific, L3000). Transfected cells were selected using 150 μg ml⁻¹ hygromycin (ThermoFisher, 10687010) for 7 days and individual colonies were picked and expanded.

To convert mouse ES cells into ciTotiSCs, mouse ES cells were cultured in KSR basal medium composed of KnockOut DMEM, 5% KSR (Gibco, 10828010), CDL (CD lipid concentrate, 500×, Gibco, 11905-031), 1% N2 (ThermoFisher, A1370701), 1× L-glutaMAX, 1× penicillin-streptomycin, 1× non-essential amino acids (NEAA), 1× sodium pyruvate, 55 μM 2-mercaptoethanol and 1,000 U ml⁻¹ mLIF, 50 ng ml⁻¹ sodium L-ascorbyl-2-phosphate (Selleck, S5115) supplemented with 2.5 μM 1-azakenpau lone (Selleck, S7193), 0.5 μM WS6 (Selleck, S7442) and 0.2 μM TTNPB (Selleck, S4627). Note that it was recommended to test 50 nM–0.2 μM TTNPB in different cell lines. Cells were passaged at high densities (1:3–1:5) using 0.05% Trypsin-EDTA on inactivated MEF feeder layers. A step-by-step protocol is available on the Protocol Exchange²⁷.

High-content compound screening

To perform high-content chemical screen, MERVL-tdTomato reporter R1-mESCs were trypsinized and replated into MEF feeder pre-plated 384-well plates at a density of 800 cells per well in 40 μl mouse ES cell medium (without CHIR-99021 and PD0325901). Eight hours after plating, cells were treated with single compounds from the Sigma-LOPAC and Selleck-FDA approved drug libraries at 5 μM by Echo550. Meanwhile, DMSO was used as the negative control. After 72 h of treatment, plates were imaged by confocal Opera phoenix and analysed by HARMONY analysis system and PhenoLOGIC machine learning algorithms

(PerkinElmer). Compounds that increased the number of tdTomato⁺ cells by more than threefold over DMSO were selected as candidate hits.

RNA extraction and RT-qPCR analysis

Total RNAs were extracted using TRIzol reagent (Thermo Fisher Scientific, 15596026) according to the manufacturer's instructions. Next, cDNAs were reverse-transcribed with iScript cDNA synthesis kit (Bio-Rad Laboratories, 1708891). For RT-qPCR, samples were amplified in a CFX96 system (Bio-Rad) using iTaq Universal SYBR Green qPCR Supermix (Bio-Rad Laboratories, 1725121). Gene expression was normalized to *Gapdh*.

Embryoid body differentiation

R1-mESCs and R1-ciTotiSCs were digested with 0.05% Trypsin-EDTA and resuspended at 10⁵ cells per ml in EB medium (mouse ES cell culture medium without mLIF, CHIR-99021 and PD0325901) after removal of feeder cells by gelatin. To form hanging drops, 20 μl cells per suspension drop was cultured below the lid of 10 cm petri dish under 5% CO₂ at 37 °C. After 2 days, the aggregated embryoid bodies were washed off and further cultured in ultra-low attachment 6-well plates in EB medium. Embryoid bodies were collected at day 0, 3 or 6 for immunostaining.

Cell conversion to TSCs

To convert OG2-mESCs and OG2-ciTotiSCs into TSCs, cells were trypsinized, followed by resuspension in TSC medium containing RPMI1640 (Gibco, C11875500BT), 20% FBS, 1× glutaMAX, 1× sodium pyruvate (Gibco, 11360070), 1× penicillin-streptomycin, 55 μM 2-mercaptoethanol, 25 ng ml⁻¹ FGF4 (R&D systems, 235-F4) and 1 μg ml⁻¹ heparin (Sigma-Aldrich, H3149). Then cells were replated at a density of 1 × 10⁵ cells on 12-well plates pre-coated with 0.3% gelatin and cultured in TSC medium. The culture medium was changed daily and cells were collected at days 0, 4, 8 for RT-qPCR and day 8 for immunostaining.

Metabolite extraction and metabolomics

Metabolite extraction was performed as previously described²². In brief, 500,000 cells were extracted using 1 ml ice-cold 80% methanol. After vortexed, the cells were centrifuged at 14,000 g for 15 min at 4 °C, and the supernatant was transferred to a new tube and evaporated with a speed vacuum. Dried metabolite pellets were kept at –80 °C until LC-MS analysis.

Metabolomics analysis was performed according to the previous described method²². Dried metabolites were reconstituted in 0.03% formic acid in water, vortexed, centrifuged at 14,000g for 15 min at 4 °C and the supernatant was analysed using liquid chromatography-tandem mass spectrometry (LC-MS/MS). A UHPLC system (Nexera X2 LC-30A, Shimadzu) was used for liquid chromatography, equipped with an ACQUITY UPLC HSS-T3 column (150 × 2.1 mm, 1.8 μm, Waters). Mobile phase A was 0.03% formic acid in water and mobile phase B was 0.03% formic acid in acetonitrile. The elution gradient was as follows: 0–3 min 1% mobile phase B; 3–15 min 1–99% B; 15–17 min 99% B; 17–17.1 min 99–1% B; and 17.1–20 min 1% B. The flow rate was 0.25 ml min⁻¹, the column was kept at 35 °C, and the samples in the autosampler were maintained at 4 °C. Mass spectrometry was performed with a triple-quadrupole mass spectrometer (QTRAP 6500+, SCIEX) in multiple-reaction monitoring mode. Chromatogram review and peak area integration were performed using MultiQuant software v.3.0 (SCIEX). Each detected metabolite was normalized to the sum of total peak area from all detected metabolites within that sample to correct any variations introduced from sample handling through instrument analysis. Normalized peak areas were used as variables for multivariate and univariate statistical data analyses.

Mouse chimeric assay

Before microinjection, tdTomato-labelled mouse ES cells and tdTomato-labelled ciTotiSCs were digested to single cells with 0.05% Trypsin-EDTA and plated on 0.3% gelatin for 30 min to remove

feeders. To generate chimeric E4.5 blastocysts, single or 4–8 cells of tdTomato-labelled mouse ES cells and tdTomato-labelled ciTotiSCs were injected into ICR mouse 8-cell embryos and cultured in KSOM medium for 48 h. Then, the chimeric embryos were imaged and fixed for immunostaining. For generating chimeric E7.5 or E13.5 embryos, single or 4–8 cells were also injected into ICR mouse 8-cell embryos and recovered in KSOM medium for 1–2 h. Injected embryos were then transplanted into uterine horns of 0.5 or 2.5 days post-coitum pseudo-pregnant ICR females. For E7.5 chimeras, the conceptuses were dissected and fixed for immunostaining. For E13.5 chimeras, the conceptuses were also dissected, imaged and further analysed. The placentas of E13.5 chimeras were isolated, fixed and sectioned for immunohistochemistry.

Immunofluorescence

For monolayer cell immunostaining, cells were washed with DPBS 3 times and fixed in 4% paraformaldehyde for 30 min at room temperature. Donkey serum (2.5% in DPBS) was used for blocking for 1 h at room temperature or overnight at 4 °C. Triton X-100 (0.2%) was added for permeabilization during blocking. Next, cells were incubated with primary antibody in blocking buffer overnight at 4 °C. After washing three times with DPBS, samples were incubated with secondary antibody for 1 h at room temperature. Following DPBS washes, cells were stained with DAPI for 5 min. Antibodies used: mouse anti-MERVL-gag (1:200, Epigentek, A-2801), mouse anti-CDX2 (1:200, BioGenex, MU392A-UC), rabbit anti-TFAP2C (1:100, Santa Cruz, Sc-8977), goat anti-mouse IgG Alexa Fluor 647 (1:500, Life Technologies, A-21241), donkey anti-mouse IgG Alexa Fluor 488 (1:500, Life Technologies, A-21202), donkey anti-rabbit IgG Alexa Fluor 488 (1:500, Life Technologies, A-21206).

For E4.5 chimeras, embryos were fixed in 4% paraformaldehyde for 30 min and permeabilized in 0.5% TritonX-100 for 15 min at room temperature. Next, embryos were blocked with blocking buffer (PBST (DPBS + 0.1% Tween-20), 2% BSA and 5% donkey serum) for 2 h at room temperature, followed by incubation in primary antibody (mouse anti-CDX2) overnight at 4 °C. After washing with PBST 3 times, embryos were incubated in diluted secondary antibody for 4–6 h at room temperature. Next, embryos were washed three times in PBST and stained for nuclei using DAPI for 5 min. Embryos were transferred into glass-bottom dishes for confocal microscope imaging.

For E7.5 chimeras, dissected embryos were also fixed in 4% paraformaldehyde for 30 min at room temperature. After washing three times, embryos were permeabilized with 0.5% TritonX-100 for 20 min, followed by blocking in PBST buffer with 10% BSA and 5% donkey serum for 8 h at room temperature or overnight at 4 °C. Subsequently, embryos were incubated in diluted primary antibody overnight at 4 °C. Next, embryos were washed three times and incubated with secondary antibody for 4–6 h at room temperature. After washing three times, the nuclei of embryos were also labelled with DAPI for 5 min at room temperature. The following antibodies were used: mouse anti-ELF5 (1:200; Santa Cruz, sc-166653), rabbit anti-OCT4 (1:200; Abcam, ab19857), goat anti-tdTomato (1:100, Biorbyt, orb182397), donkey anti-mouse IgG Alexa Fluor 488 (1:500, Life Technologies, A-21202), goat anti-rabbit IgG Alexa Fluor 647 (1:500, Life Technologies, A-21244), donkey anti-goat IgG Alexa Fluor 555 (1:500, Life Technologies, A-21432). Confocal images were taken by LSM710 Confocal Microscope (Zeiss). Immunofluorescence staining images were adjusted with ImageJ 1.53a.

E13.5 chimera imaging and placental section immunohistochemistry

For E13.5 chimeras, the fetus, yolk sac, placenta and gonad were separated on ice and imaged under stereo fluorescence microscope to observe tdTomato⁺ cell localization. Then, placentas were washed in cold DPBS and fixed in 4% paraformaldehyde overnight at 4 °C. After dehydration by gradient sucrose solutions, placentas were embedded

in OCT (SAKURA, 4583), frozen and cut into 10 µm sections for immunohistochemistry. After briefly washing with PBST, the slides were blocked and permeabilized in 2.5% donkey serum in PBST (0.2% Triton X-100 in DPBS) for 1 h at room temperature. The slides were incubated in diluted primary antibodies overnight at 4 °C, followed by washing 3 times. Next, the slides were incubated in diluted secondary antibodies for 1 h at room temperature, washed three times and eventually stained with DAPI. The antibodies were used: rat anti-CK8 (1:10, Developmental Studies Hybridoma Bank, TROMA-1, AB_531826), mouse anti-proliferin (1:50, Santa Cruz, sc-271891), goat anti-tdTomato (1:100, Biorbyt, orb182397), donkey anti-mouse IgG Alexa Fluor 488 (1:500, Life Technologies, A-21202) or donkey anti-goat IgG Alexa Fluor 555 (1:500, Life Technologies, A-21432).

Flow cytometry

Cells were trypsinized and resuspended in appropriate medium, followed by filtering through 40 µm cell strainers. Then, samples were analysed on Aria II and Aria III Flow Cytometer Systems (BD Biosciences). Data analysis was performed by FlowJo (v9.3.2 and v10.4) software. Flow cytometry gating strategies are shown in Supplementary Figure 1.

To assess the contribution of tdTomato⁺ cells to conceptuses of E13.5 chimera, embryos, yolk sacs and placentas were cut into small pieces (~1 mm diameter) on ice. For placentas, samples were digested with accutase (Gibco, A1110501) for 10 min at 37 °C with shaking. For embryos and yolk sacs, samples were digested with collagenase IV (addition of 1 U ml⁻¹ DNase) for 30 min or 5 min at 37 °C, respectively, followed by incubation with TrypLE (Gibco, 12605028) for 3–5 min. After digestion, all tissues were dissociated into single cells by pipetting and filtered through 70 µm cell strainers. Then, cells were centrifuged, resuspended and sorted by flow cytometry.

Teratoma formation

OG2-mESCs and OG2-ciTotiSCs (TAW-treated for 2.5 d) were digested into single cells by 0.05% Trypsin-EDTA and resuspended in DPBS. About 10⁶ cells per cell line were injected subcutaneously into both dorsal flanks of 6–8 weeks old female severe combined immunodeficiency (SCID) mice. 3–4 weeks later, the teratomas were dissected, fixed in 4% paraformaldehyde and embedded in paraffin before sectioning. Next, tissue sections were analysed by H&E staining and immunohistochemistry. The antibodies were used: goat anti-placental lactogen I (1:75, Santa Cruz, Sc-34713), donkey anti-goat IgG Alexa Fluor 555 (1:500, Life technologies, A-21432).

RNA-seq library preparation

To perform smart-seq2 for chemical combination screen, MERVL-tdTomato reporter OG2-mESCs were treated with different compound combinations for 3 days, and then the tdTomato⁺ cells were sorted by FACS. The RNA-seq libraries were generated from these samples according to smart-seq2 library preparation protocol with minor modification. Firstly, cells were lysed in lysis buffer. RNAs were captured with 25 nt oligo (dT) primers and reversed into cDNAs. After amplification and purification, cDNAs were sheared to approximately 300 bp by CovarisS2 and captured by DynabeadsR MyOne Streptavidin C1 beads. All libraries were constructed using a Kapa Hyper Prep Kit (Kapa Biosystems) and sequenced on 150 bp paired-ends Illumina Novaseq 6000 platform.

For bulk RNA-seq library preparation, OG2-ciTotiSCs (P1/P2/P4/P8) were digested into single cells and suspended in culture medium. After filtering through 40 µm cell strainers, feeder cells were removed by FACS. Then total RNAs were extracted using TRIZol reagent and used for generating mRNA-seq library using NEBNext® UltraTM RNA Library Prep Kit for Illumina® (#E7530L, NEB, USA) according to the manufacturer's protocol. Bulk RNA-seq libraries were sequenced using 150 bp paired-ends Illumina Novaseq 6000 platform.

Article

RRBS library preparation

To perform RRBS, genomic DNAs were extracted from OG2-ciTotiSCs (TAW-treated for 2.5 d) using DNeasy Blood and Tissue Kits (QIAGEN) and digested by Mspl (NEB). The digested DNAs were filled in and tailed. Then, bisulfite conversion was performed using MethylCode bisulfite conversion kit (Invitrogen). After amplification and purification, all libraries were constructed and subjected to 150 bp paired-ends Illumina Novaseq 6000 platform.

Single-cell RNA-seq library preparation

For sample preparation, about 10,000 living OG2-ciTotiSCs (TAW-treated for 2.5 d) were used to generate 10X Genomic's scRNA-seq library using single-cell 3' library and Gel Bead kit V2 according to the manufacturer's manual. For placenta and yolk sac of E13.5 chimera, digested cells were filtered, centrifuged and resuspended in PBS with 10% FBS. Around 10,000 FACS sorted tdTomato⁺ living cells were used for generating 10X Genomic's scRNA-seq library. All libraries were pooled and subjected to the Illumina Novaseq system.

Smart-seq2 data analysis

For smart-seq2 data, after demultiplexed and trimmed adaptors, reads were aligned to the mm10 reference genome using TopHat2 (v2.0.13). Reads aligned to genes are counted by featureCounts (v1.6.5).

Bulk RNA-seq data analysis

For bulk paired-end RNA-seq raw data, TrimGalore software (with Cutadapt v1.18) was used to trim the adaptor and first 10 bp of both reads. TopHat2 (v2.1.1) was used for the alignment to the mm10 reference genome with default parameters. The gene expression count matrix and fragments per kilobase of transcript per million mapped reads (FPKM) value were analysed by Cufflinks (v2.2.1) default pipeline using mm10 RefSeq gene annotation downloaded from UCSC table browser. To obtain the expression level of repeats, the trimmed reads were mapped to the mm10 repeats (RepeatMasker) reference downloaded from USCS table browser and Cufflinks was used to count the number and calculate FPKM of repeats. DEGs were considered significant if $\log_2(\text{fold change}) > 1$ and Benjamini–Hochberg FDR value < 0.1 . Differentially expressed repeats were considered significant if $\log_2(\text{fold change}) > 1$ and Benjamini–Hochberg FDR value < 0.15 .

Bulk RNA-seq reference data integrative analysis

Bulk RNA-seq data of 2CLC, D-EPSC, L-EPSC, TLSCs and TBLCs (downloaded from SRR385620, GSM2135530, GSM4322195, GSM5603636, and GSM5160100) were obtained from published resources. Bulk RNA-seq data of 2-cell embryos were downloaded from GSM1625862 and GSM1625863. Raw sequencing data was analyzed using the same process of bulk RNA-seq data. Reference data were merged with ciTotiSCs bulk RNA-seq data. The $\log_2(FPKM + 1)$ value of totipotency genes and pluripotency genes of different samples and different stages were calculated and z-score normalized before shown in heatmap.

Single-cell RNA-seq data analysis

For scRNA-seq data, Cell Ranger (v6.0.2) was used to align and quantify the ciTotiSCs scRNA-seq data and scRNA-seq data of tdTomato⁺ cells from placenta and yolk sac against the mm10 reference genome. Scanpy (v1.4.5) python package was used to preprocess, merge, and perform UMAP analysis of ciTotiSCs scRNA-seq data with scRNA-seq data of different embryonic stages, mouse ES cells, D-EPSC, L-EPSC, and TBLCs (downloaded from GSE100597, GSE45719, GSE78140, GSE145609, and GSM5195025). Only cells with more than 300 detected genes and genes with more than 50 cells detected across all stages and samples were remained for further analysis. Total counts per

cell were normalized to 10,000, and counts were log transformed. Top 1,000 highly variable genes were used to perform PCA analysis. The normalized expression data was regressed out and scaled, and scRNA-seq data of ciTotiSCs and TBLCs were downsampled using random.choice function in python to 200 cells before performing PCA and UMAP analysis. As for comparison with bulk RNA-seq data, the scRNA-seq data of specific embryo stages and samples were merged by means. The same dataset of expression levels of genes of different embryo stages and samples were z-score normalized before shown in heatmap. scRNA-seq data of tdTomato⁺ cells from placenta and yolk sac were preprocessed by Scanpy. The preprocessing method was the same as ciTotiSCs scRNA-seq data before performing Louvain clustering. As for the cell heterogeneity analysis, Gm42418 and AY036118 were removed from datasets and only cells with more than 2,000 detected genes and genes with more than 50 cells detected were remained. Cells with mitochondrial genes expression > 0.05 were also removed before preprocessing by Scanpy. Total counts per cell were normalized to 1 M, and counts were log transformed. 1,000 highly variable genes were chosen to perform UMAP analysis and Louvain clustering.

Principal component analysis and unsupervised hierarchical clustering analysis

PCA analysis was performed by the PCA function in sklearn.decomposition python package. Unsupervised hierarchical clustering analysis was performed by linkage and shown by dendrogram function from scipy.cluster.hierarchy package. Gene expression levels of replicates were merged by mean and FPKM values were \log_2 -transformed before PCA analysis or unsupervised hierarchical clustering analysis. Gene set of the 2-cell stage was acquired from published data^{3,11}.

Gene set enrichment analysis

GSEA was performed using the GSEA (4.1.0) with default parameters. Gene sets of different embryonic stages were downloaded from published data^{3,11,19}. Previous integrated bulk RNA-seq dataset was used to perform GSEA.

Motif enrichment analysis

RAR binding sites data across the whole genome were downloaded from MotifMap^{28,29}. Maternal gene sets, totipotency gene sets, and pluripotency gene sets were downloaded from a published paper¹⁹. Enrichment P-value was calculated by Chi-square test with python scipy.stats.chi2_contingency package.

ATAC-seq library preparation and sequencing

The ATAC-seq library preparation protocol was modified from previously published methods³⁰. In brief, after cells were collected by flow cytometry as described before, 100,000 cells were centrifuged at 500g for 5 min in a 4 °C pre-chilled fixed-angle centrifuge and washed by ice-cold PBS. After centrifugation, supernatant was removed with two pipetting steps to avoid the cell pellet. Cell pellets were then resuspended in 50 µl of ATAC-seq RSB containing 0.1% IGEPAL-630, 0.1% Tween-20, and 0.01% digitonin and incubated on ice for 10 min. After lysis, 1 ml of ATAC-seq RSB containing 0.1% Tween-20 (without IGEPAL-630 or digitonin) was added, and the tubes were inverted for six times to mix. Nuclei were then centrifuged for 10 min at 500g in a 4 °C pre-chilled fixed-angle centrifuge. Supernatant was removed and nuclei were then incubated with the Tn5 transposome and tagmentation buffer at 37 °C for 30 min (Novoprotein, N248). After the tagmentation, the stop buffer was added directly into the reaction to end the tagmentation. Reactions were cleaned up with 2.2x VAHTS DNA Clean Beads (Vazyme, N411). PCR was performed to amplify the library for 8 cycles by KAPA HiFi PCR Kit (Kapa Biosystems, kk2102) using the following PCR conditions: 72 °C for 5 min; 98 °C for 3 mins; and thermocycling at 98 °C for 20 s, 63 °C for 30 s and 72 °C for 3 min.

After the PCR reaction, libraries were purified with the 1.8x VAHTS DNA Clean Beads.

ATAC-seq data analysis

All ATAC-seq reads were first aligned to the GRCm38 genome using Bowtie2 (version 2.3.3.1) using the parameters “-X-mm -t-q -N1 -L 25 --no-mixed --no-discordant”. For the comparison of chromatin accessibility with mouse embryos (downloaded from GSE66581), only paired-end reads were used, and the aligning parameters are identical with shown above. All unmapped reads, non-uniquely mapped reads and PCR duplicates were removed. The resulted bam was then subjected to the removal of repeat regions using the blacklist mask. Tn5 insertion site was determined by shift mate pair ends towards mate centre by 4 bp, and 250 bp regions around the insertion size were considered as open sites. For downstream analysis, we normalized the read counts by computing the number of reads per kilobase of bin per million of reads (RPKM). To visualize the ATAC-seq signal, we generated bigwig tracks from the open region defined above, and the bin size is 100 bp. After validating the high correlation between each biological replicate, we merged bigwig files of each replicate of the same condition. The merged bigwig files were then used in all plots related to chromatin accessibility. P-values of repeat elements violin plots are calculated with paired Wilcoxon test, and adjusted using false discovery rate.

RRBS data analysis

For RRBS data, raw reads were trimmed adaptors, and were mapped to the reference genome mm10 using BS-Seeker2 (v2.1.1) (parameter: -m 0.06) in mode of bowtie. Two mates were mapped separately for the paired-end sequencing data. Average CG methylation levels were profiled in 100 kb bins in each sample using CGmapTools and were used to perform PCA analysis using PCA function in sklearn.decomposition python package. Methylation level of repeats were calculated by bedtools (v2.29.2) intersect function based on the 100 kb bin methylation level. The reference RRBS data of embryos and mouse ES cells were downloaded from GEO database (embryo: GSE34864; mouse ES cells: GSE133926).

Alignment track visualization

The visualization of ATAC-seq and RRBS tracks was performed using IGV.

Statistics and reproducibility

The experiments in this study were performed with at least three biological replicates unless specified. All statistical analyses were performed with Graphpad prism 8.0 software or R Bioconductor. The tests used and the P-values are listed on figures and figure legends. Details of statistical tests are outlined within figures and figure legends. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

Reporting summary

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

Data availability

RNA-seq, scRNA-seq, ATAC-seq and RRBS data from this study have been deposited at the NCBI Gene Expression Omnibus (GEO) under accession number GSE185005. Source data are provided with this paper.

Code availability

The code for the analyses can be found at <https://github.com/pengchengtan/Hu-et-al.-2022-ciTotiSC>.

27. Yanyan, H. et al. Induction and maintenance of mouse totipotent stem cells by a defined chemical cocktail. *Protoc. Exch.* <https://doi.org/10.21203/rs.3.pex-1927/v1> (2022).
28. Xie, X., Rigor, P. & Baldi, P. MotifMap: a human genome-wide map of candidate regulatory motif sites. *Bioinformatics* **25**, 167–174 (2009).
29. Daily, K., Patel, V. R., Rigor, P., Xie, X., & Baldi, P. MotifMap-integrative genome-wide maps of regulatory motif sites for model species. *BMC Bioinf.* **12**, 495 (2011).
30. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).

Acknowledgements We thank Tsinghua University Center of Pharmaceutical Technology for assistance with chemical screening, the Imaging Core Facility and the Center of Biomedical Analysis for assistance in confocal microscopy and flow cytometry analysis, and the Laboratory Animal Research Center for assistance in mouse embryo microinjection and transplantation. We thank J. Yong for technical assistance of Smart seq2 RNA-seq and Y. Li for early exploration of RNA-seq data analysis. This work is supported by the National Key R&D Program of China (2017YFA0104001 to S.D.), the National Natural Science Foundation of China (31771530 to T.M. and 32030031 and 31530025 to S.D.), Center for Life Sciences (to S.D.) and Tsinghua University Initiative Scientific Research Program (to T.M.).

Author contributions Y.H., YY., T.M., K.L. and S.D. designed the study, Y.H. and YY. performed most of the experiments, and Y.H., YY., T.M., K.L. and S.D. interpreted the data. P.T. performed the bioinformatics analyses of RNA-seq, scRNA-seq and RRBS data. Y.Z. performed the mouse embryo microinjection experiments and mouse breeding. M.H. assisted with parts of the experiments. J.Y., X.Z. and Z.J. assisted in ATAC-seq library preparation and data analysis, Y.L. supervised this part of the work. K.Y. and H.P. assisted in metabolomics assay and data analysis, Z.H. supervised this part of the work. D.W. assisted with the reagents preparation. Y.H., YY., S.D., K.L. and T.M. conceived this project. T.M., K.L. and S.D. supervised the study. Y.H., YY., T.M., K.L. and S.D. wrote the manuscript with input from all authors. S.D. is the lead contact.

Competing interests S.D., K.L., Y.H., YY., P.T. and T.M. are listed as inventors on the priority patent application CN202110989429.8 (*Induced Totipotent Stem Cell and the Preparation Method Thereof*) filed by Tsinghua University, Beijing, on 26 August 2021. The other authors declare no competing interests.

Additional information

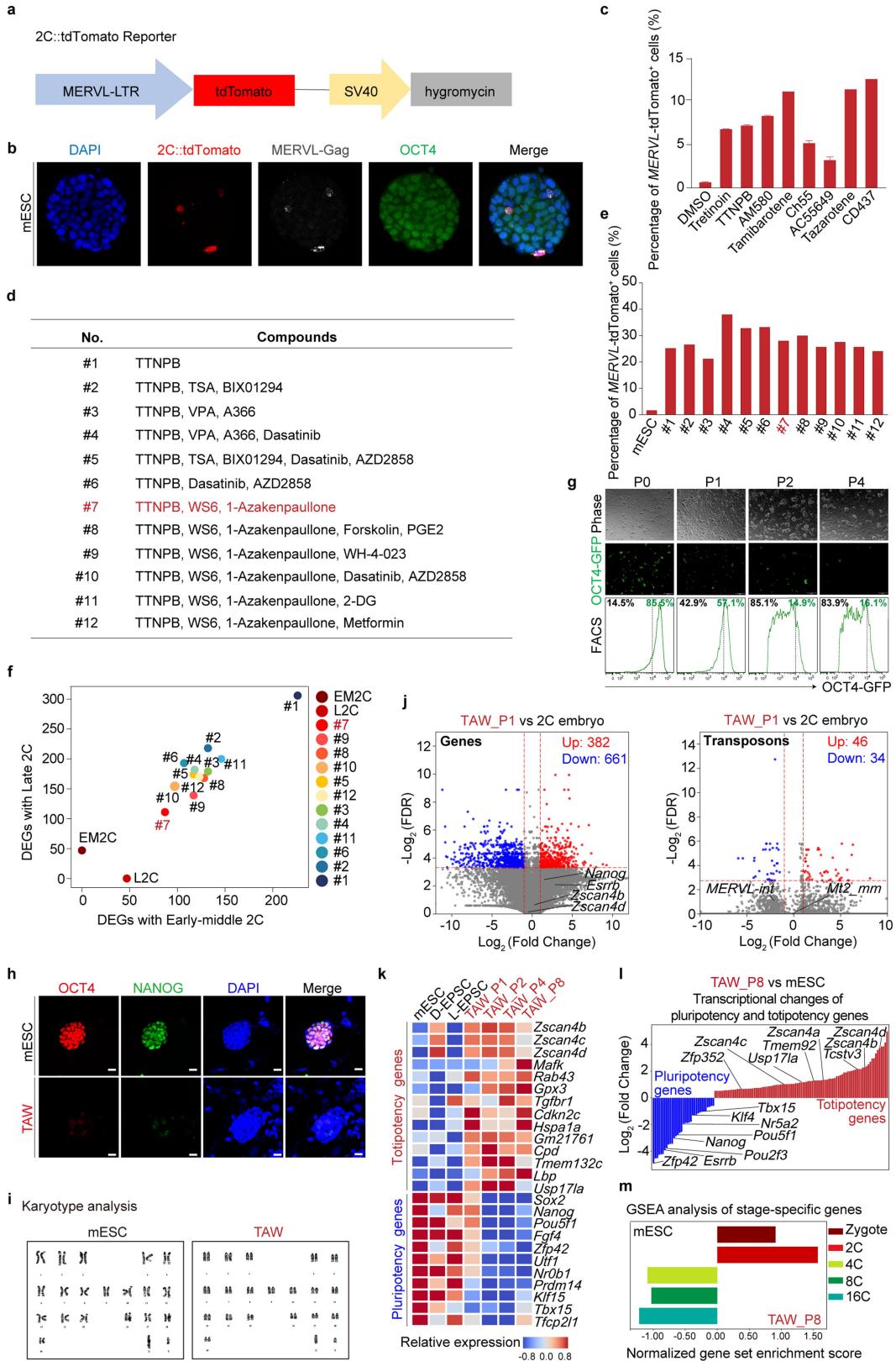
Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-022-04967-9>.

Correspondence and requests for materials should be addressed to Tianhua Ma, Kang Liu or Sheng Ding.

Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work.

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

Article

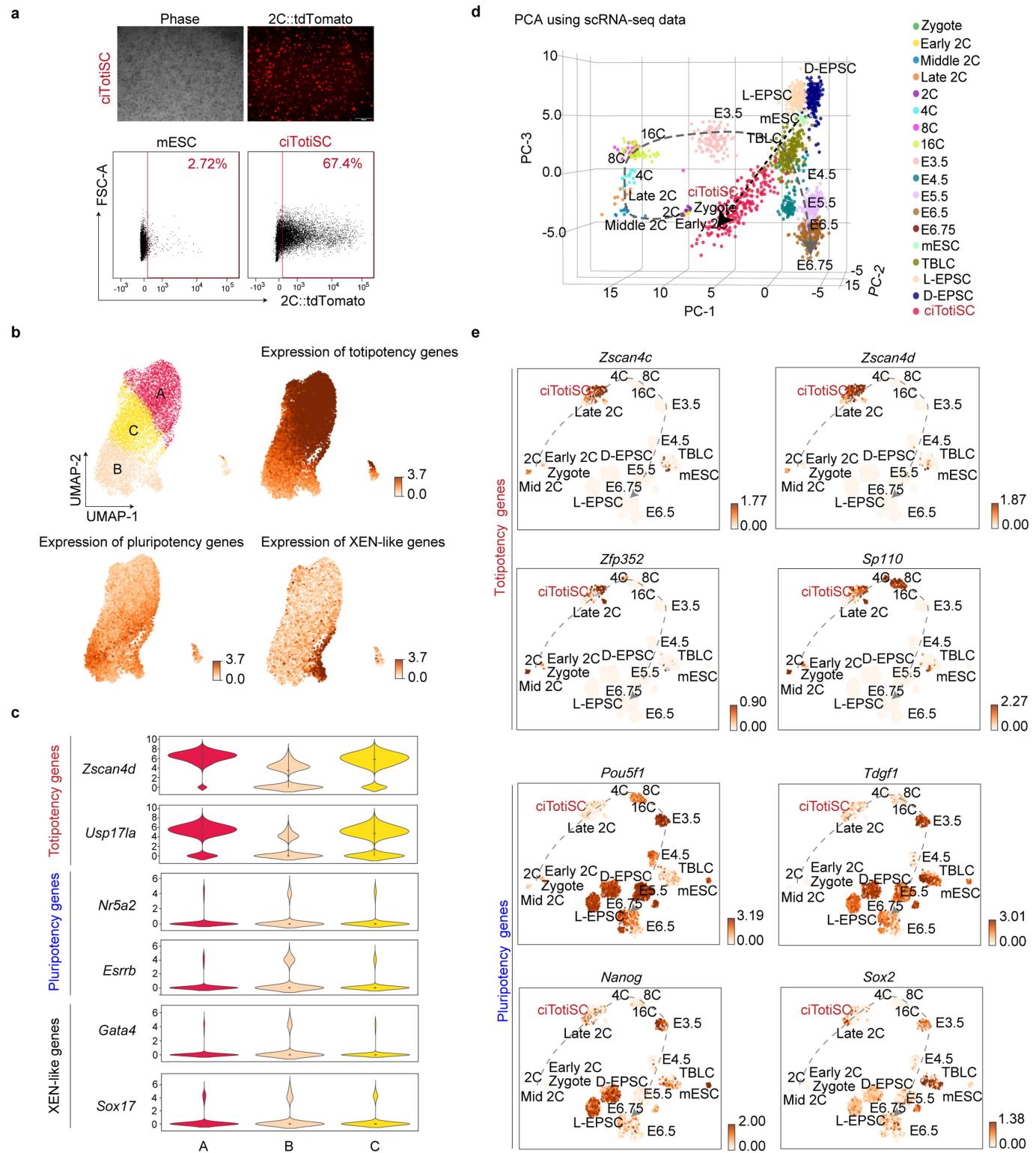


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Screening of chemical compounds enabling induction and maintenance of mouse totipotent stem cells. **a**, Schematic diagram of MERVL-tdTomato reporter. **b**, Immunostaining of MERVL-Gag in mouse ES cells with MERVL-tdTomato & OCT4-GFP dual-reporter mouse ES cells under basal 2i/LIF medium. **c**, Bar graph showing the percentage of MERVL-tdTomato⁺ cells generated by treatment with indicated RAR agonists. **d**, Detailed list of compound combinations. **e**, Bar graph showing the percentage of MERVL-tdTomato⁺ cells generated by treatment with different compound combinations. **f**, The number of DEGs (differentially expressed genes, *p*-value < 0.001) between each of the conditions to mouse 2-cell embryos. L2C: Late 2C embryo; EM2C: Early-Middle 2C embryo. **g**, OCT4-GFP changes in cells under sustained TTNPB, 1-Azakenpaulone and WS6/TAW treatment over 4 passages. Up: imaging of colony morphology. Middle: imaging of OCT4-GFP reporter. Low: FACS analysis of OCT4-GFP⁺ cells. **h**, Immunostaining of pluripotency markers OCT4 and NANOG in mouse ES

cells treated with or without TAW. Scale bar: 20 μ m. **i**, The karyotype analysis of mouse ES cells treated with or without TAW. **j**, Volcano plots showing up- (red) and down- (blue) regulated genes (left, $\log_2(\text{FC}) > 1$, $\text{FDR} < 0.1$) and transposons (right, $\log_2(\text{FC}) > 1$, $\text{FDR} < 0.15$) in TAW_P1 versus mouse 2C embryo. Benjamini-Hochberg method was used to control the false discovery rate. Some totipotency genes/transposons and pluripotency genes were labeled. **k**, Transcriptional changes of pluripotency (blue) and totipotency (red) specific genes among mouse ES cells, mouse ES cells treated with TAW for 1/2/4/8 passages, D-EPSC and L-EPSC. **l**, Transcriptional changes of pluripotency (blue) and totipotency (red) specific genes in mouse ES cells treated with TAW for 8 passages. **m**, GSEA analysis of bulk RNA-seq data of mouse ES cells treated with TAW for 8 passages by the indicated embryonic stage-specific gene sets. All datasets in bioinformatic analyses were summarized in Supplementary Table 1.

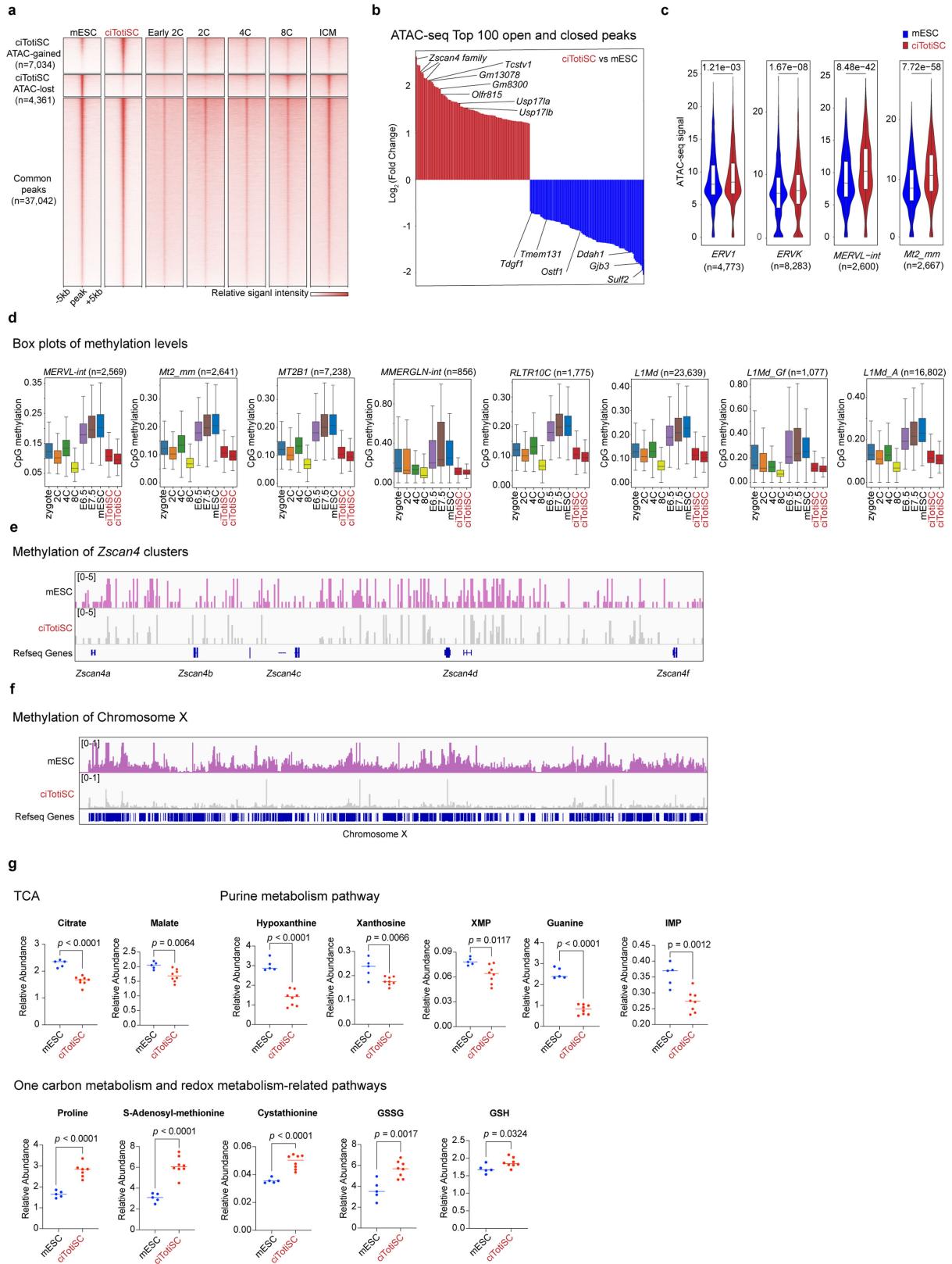
Article



Extended Data Fig. 2 | ciTotiSCs exhibit characteristic transcriptome features close to totipotent blastomeres at the single-cell level.

a, Representative images and flow cytometry analysis of MERVL-tdTomato of ciTotiSCs. Scale bars, 500 μ m. b, UMAP plot from scRNA-seq displaying three clusters (A-C) identified in ciTotiSCs culture. The expression of representative 51 totipotency genes, 47 pluripotency genes and 6 primitive endoderm genes were shown in UMAP plot. c, Violin plots showing the expression distribution

of specific marker genes in each cluster shown in (b). d, Transcriptome PCA analysis of ciTotiSCs, mouse ES cells, L-EPSC, D-EPSC, TBLCs and mouse embryos from zygote to E6.75 at the single-cell level. e, FeaturePlots projecting expression of representative pluripotency and totipotency genes, overlapping Fig. 2c UMAP. All datasets in bioinformatic analyses were summarized in Supplementary Table 1.

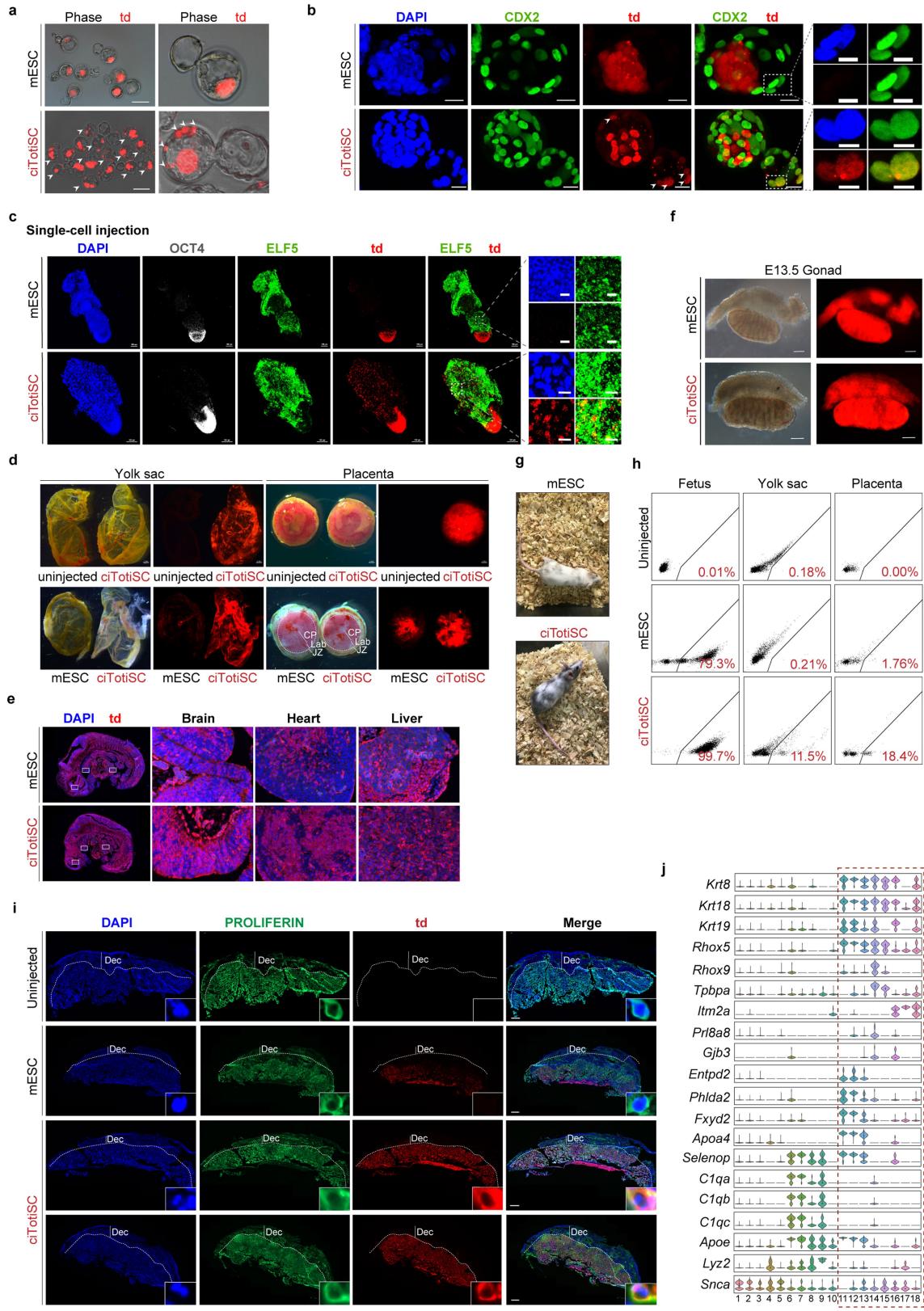


Extended Data Fig. 3 | See next page for caption.

Article

Extended Data Fig. 3 | The epigenomic and metabolic features of ciTotiSCs are similar to totipotent 2C-embryo blastomeres. **a**, Comparison of the chromatin accessibility among mouse ES cells, ciTotiSCs and mouse embryos at the indicated developmental stages by ATAC-seq. **b**, Chromatin accessibility changes after chemical induction of TotiSC. Totipotency and pluripotency genes located in top open (red) and closed (blue) peaks were indicated. **c**, Chromatin accessibility (\log_{10} (RPKM+1) transformed value) of 2C specific retrotransposon elements in ciTotiSCs and mouse ES cells. The central line corresponds to the median, the boxes indicate the lower and upper quartiles. *P* values were determined using two-sided student's *t*-test, and then adjusted using Holm's method. **d**, Boxplots of DNA methylation levels on 2C specific

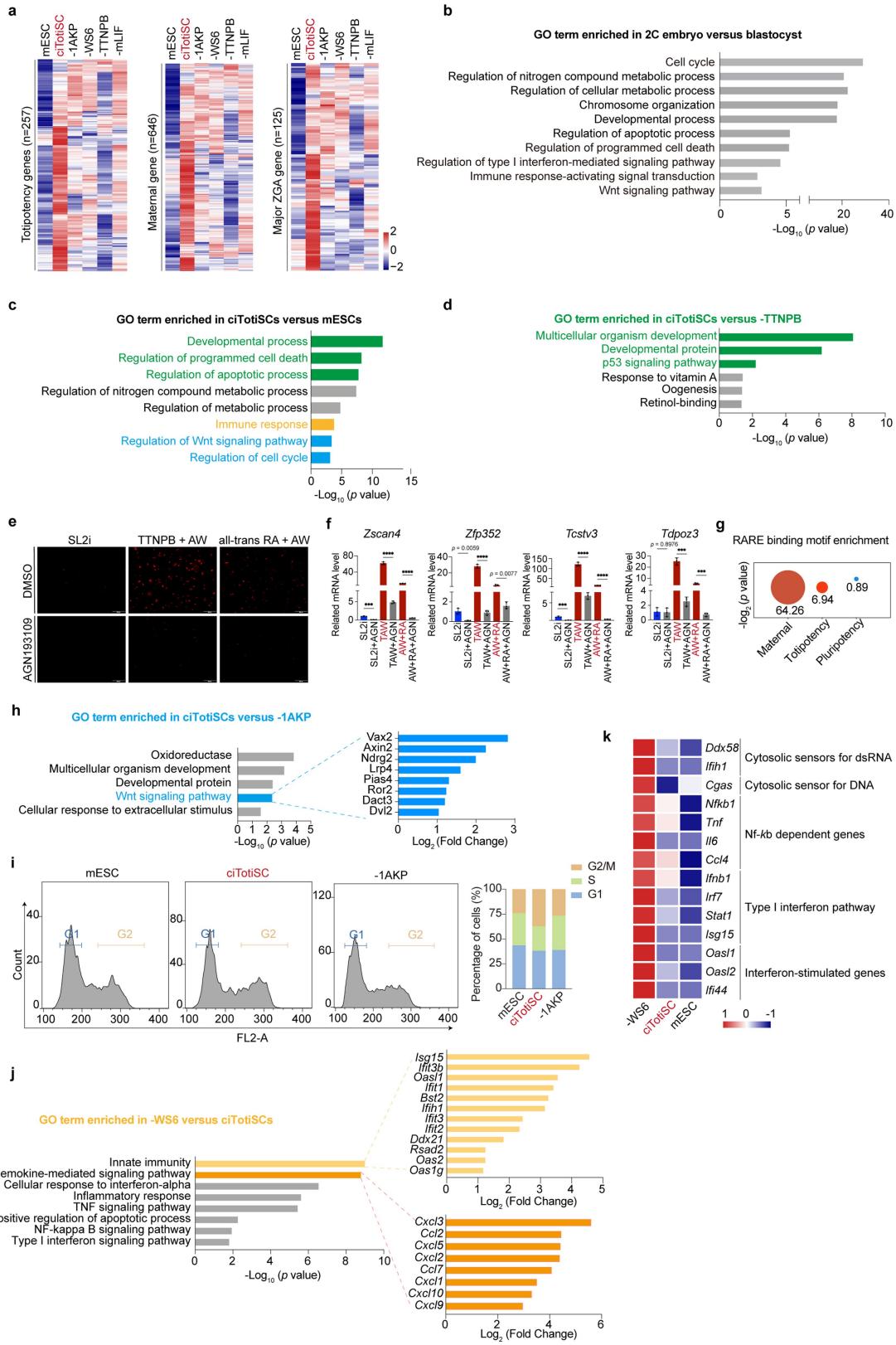
retrotransposon elements in mouse ES cells, ciTotiSCs and mouse embryos at the indicated stages. The central line is the median, the boxes indicate the lower and upper quartiles. **e**, Different CpG methylation pattern of *Zscan4* clusters in mouse ES cells and ciTotiSCs. **f**, Different CpG methylation pattern of X chromosome in mouse ES cells and ciTotiSCs. **g**, Abundance of metabolites involved in the TCA cycle, purine metabolism pathway, one-carbon metabolism and redox metabolism-related pathway. Mean relative fold-change and error bar were calculated from $n = 5$ or 8 biological experiments. *P* values determined by two-sided Student's *t*-test. All datasets in bioinformatic analyses were summarized in Supplementary Table 1.



Extended Data Fig. 4 | See next page for caption.

Article

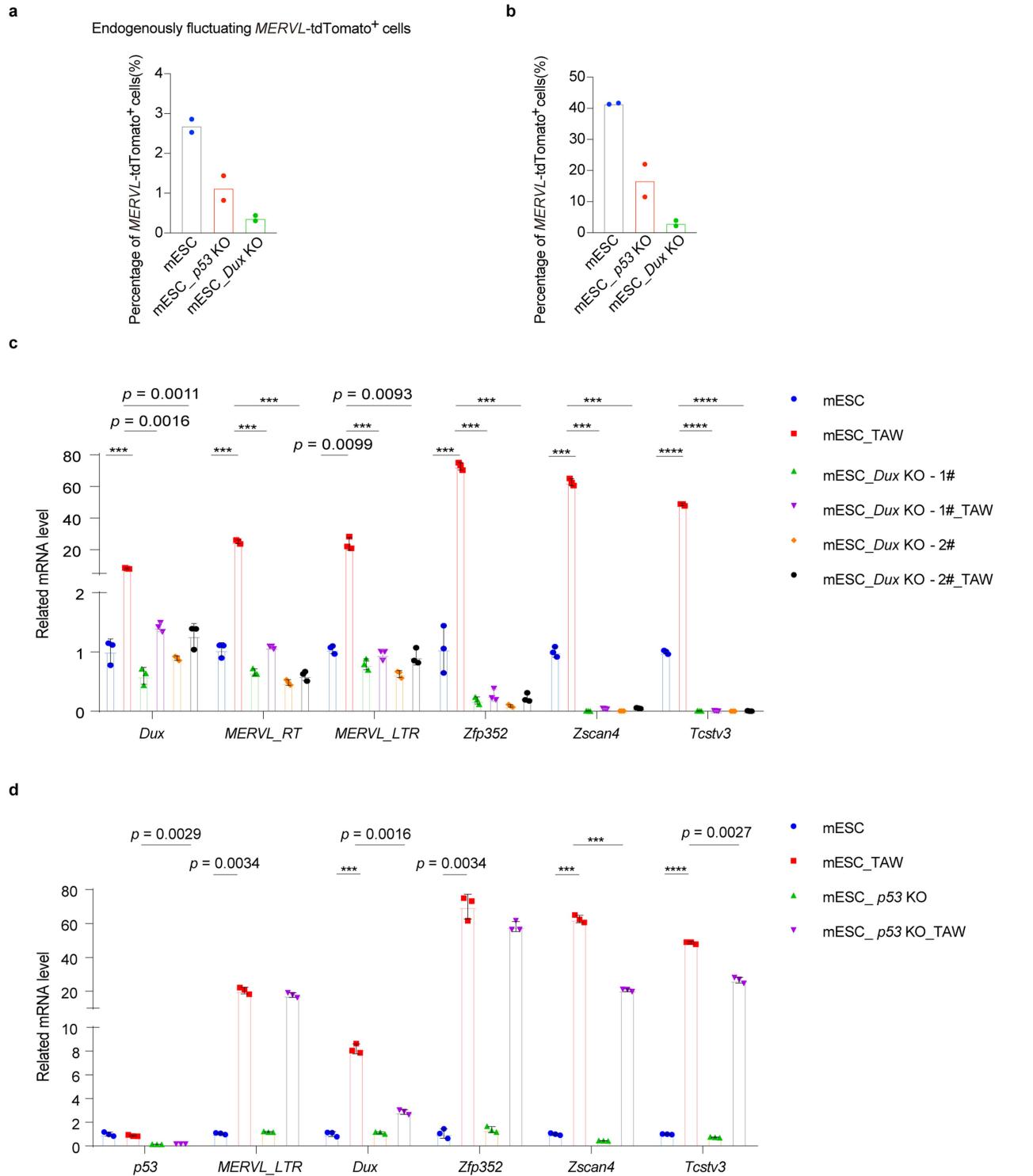
Extended Data Fig. 4 | Characterization of ciTotiSCs' chimerism potential *in vivo*. **a**, Live images of chimeras at E4.5, which were developed from 8-cell embryos injected with tdTomato⁺ ciTotiSCs or mouse ES cells. Embryos with tdTomato⁺ cells integrated trophectoderm were pointed by white arrows. Scale bars: 100 μm. **b**, Representative images showing expression of CDX2 in chimeric blastocysts at E4.5 *in vitro*, developed from 8-cell embryos injected with multiple tdTomato⁺ ciTotiSCs or mouse ES cells. Scale bars: 20 μm (left), 10 μm (right). **c**, Representative images showing expression of ELF5 and OCT4 in chimeras at E7.5 *in vivo*, developed from 8-cell embryos injected with single tdTomato⁺ ciTotiSC or mouse ES cells. Scale bars: 100 μm (left), 25 μm (right). **d**, Representative images of extraembryonic tissues from E13.5 chimeric conceptuses derived from uninjected control 8-cell embryos or 8-cell embryos injected with tdTomato-labeled ciTotiSCs or mouse ES cells. JZ: junctional zone; Lab: labyrinth; CP: chorionic plate. **e**, Representative images of the multiple ciTotiSCs-derived E13.5 chimeric embryo sections, derived from 8-cell embryos injected with tdTomato-labeled ciTotiSCs or mouse ES cells. **f**, A representative image of E13.5 gonads from chimera contributed by ciTotiSCs and mouse ES cells. **g**, ciTotiSCs or mouse ES cells-derived chimeric mice. **h**, FACS analysis of developmental contribution of tdTomato⁺ cells in fetus, yolk sac and placenta of E13.5 chimeric conceptuses, derived from uninjected control 8-cell embryos or 8-cell embryos injected with tdTomato-labeled ciTotiSCs or mouse ES cells. **i**, Representative images of placenta sections from E13.5 chimera, derived from uninjected control 8-cell embryos or 8-cell embryos injected with tdTomato-labeled ciTotiSCs or mouse ES cells co-immunostained with trophoblast cell marker PROLIFERIN. The insets showed enlarged images of single cells. Scale bars: 500 μm. **j**, Violin plots showing relative expression distribution of specific marker genes for each cluster shown in (Fig. 5f).



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

Article

Extended Data Fig. 5 | The derivation of ciTotiSCs is dependent on TAW cocktail. **a**, Heatmaps revealing expression changes of totipotency genes, ZGA genes and maternal genes after the removal of individual molecule from the TAW condition. **b**, Analysis of Gene Ontology (GO) terms enriched in 2C embryo versus blastocyst. *P* values determined by two-sided Student's *t*-test. **c**, GO analysis of terms enriched in ciTotiSCs versus mouse ES cells. Specific pathways of interest were colored. *P* values determined by two-sided Student's *t*-test. **d**, GO analysis of cells cultured with AW (-TTNPB) versus ciTotiSCs. *P* values determined by two-sided Student's *t*-test. **e**, Representative fluorescence images of cells with MERV1-tdTomato reporter, treated with (TTNPB + AW) or (all-trans RA + AW) in the presence or absence of RAR antagonist AGN193109 for 72 h. Scale bars: 100 μ m. **f**, RT-qPCR analysis of representative totipotency genes in mouse ES cells treated with (TTNPB + AW) or (all-trans RA + AW) in the presence or absence of RAR antagonist AGN193109 for 72 h. Expression levels are relative to *Gapdh*. Data are mean \pm s.d. ($n = 3$). **g**, The enrichment of RARE binding motif in maternal, totipotency and pluripotency gene regulatory regions. Dot size: -log₁₀ (*p* value). *P* values determined by two-sided Student's *t*-test. **h**, GO analysis of cells cultured with TW (-1AKP) versus ciTotiSCs. Upregulated genes are shown for specific Wnt signaling pathway. *P* values determined by two-sided Student's *t*-test. **i**, Flow cytometry analysis (left) and quantification (right) of cell cycle distribution of mouse ES cells, ciTotiSCs, and ciTotiSCs cultured with TW (-1AKP). **j**, GO analysis of cells cultured with TA (-WS6) versus ciTotiSCs. Upregulated genes are shown for two pathways of interest. *P* values determined by two-sided Student's *t*-test. **k**, The expression of genes involved in NF- κ B-mediated signaling in mouse ES cells, ciTotiSCs and cells cultured with TA (-WS6).



Extended Data Fig. 6 | Dux or p53 knockout impairs ciTotiSC generation.

a, Percentage of endogenously fluctuating MERV_L-tdTomato⁺ cells in WT, Dux and p53 knockout mouse ES cells, analyzed by flow cytometry. $n = 2$ biological replicates. **b**, Percentage of MERV_L-tdTomato⁺ cells in WT, Dux and p53 knockout mouse ES cells treated with TAW for 1 passage, analyzed by flow cytometry. $n = 2$ biological replicates. **c**, Expression of representative

totipotency MERV_L repeats and genes in WT and Dux knockout mouse ES cells treated with or without TAW for 1 passage, detected by RT-qPCR. Data are mean \pm s.d. ($n = 3$). Pvalues determined by two-sided Student's t-test. **d**, Expression of representative totipotency MERV_L repeats and genes in WT and p53 knockout mouse ES cells treated with or without TAW for 1 passage, detected by RT-qPCR. Data are mean \pm s.d. ($n = 3$). Pvalues determined by two-sided Student's t-test.

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

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

Software and code

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

Data collection	In high-content compound screening, drug libraries were handled by Echo550 (Labcyte) and cells were imaged by confocal Opera Phenix (PerkinElmer). Real-time qPCR were performed using CFX96 and CFX384 (BIO-RAD). Confocal images were taken by LSM710 Confocal Microscope (Zeiss). Flow cytometry samples were analyzed on Aria II and Aria III Flow Cytometer System (BD Biosciences). RNA-seq, ATAC-seq and RRBS libraries were sequenced using Illumina Novaseq platform. Metabolites were extracted by 80% methanol and were analyzed by LC-MS/MS (Nexera X2 LC-30A, Shimadzu and QTRAP 6500+, SCIEX)
Data analysis	Data from high-content compound screening were analyzed by HARMONY (v4.9) analysis system and PhenoLOGIC machine learning algorithms (PerkinElmer). Analysis of flow cytometry data was performed by FlowJo (v9.3.2 and v10.4). Statistic figure were analyzed and drawn by Prism 8. Immunofluorescence staining images were adjusted by imageJ 1.53a. RNA-seq, RRBS and ATAC seq datasets were analyzed using the following tools: TopHat2 (v2.0.13), TopHat2 (v2.1.1), featureCounts (v1.6.5), TrimGalore software (with Cutadapt v1.18), Cufflinks (v2.2.1), Cell Ranger (v6.0.2), Scanpy python package (v1.4.5), Scanorama (v1.7.1), Multicore-tsne (v8.0.2), GSEA (v4.1.0), Bowtie2 (v2.3.3.1), BS-Seeker2 (v2.1.1), bedtools (v2.29.2), IGV_2.9.4, R (v 4.1.1), Python (v 3.8.2), Cell Ranger (v6.0.2). Metabolomics were analyzed by MultiQuant software (v.3.0)(SCIEX).

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](#)

RNA-seq, scRNA-seq, ATAC-seq and RRBS raw data in this study were deposited at the NCBI Gene Expression Omnibus (GEO) repository under the accession number GSE185005. Reference data from other works can be accessed in NCBI GEO under their GSE numbers. scRNA-seq data of different embryonic stages, mESCs, D-EPSC, L-EPSC and TBLCs were downloaded from GSE100597, GSE45719, GSE78140, GSE145609 and GSM5195025. For the comparison of chromatin accessibility with mouse embryos were downloaded from GSE66581. The reference RRBS data of embryos and mESCs were downloaded from GEO database (embryo: GSE34864; mESCs: GSE133926). Source data are provided with this paper.

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	Sample size was determined based on previously published studies and ensure reproducibility. No statistical method was used to pre-determine sample size. The numbers of independent experiments for all the statistical results were described in the figure legend.
Data exclusions	No data was excluded from the analysis.
Replication	The experiments in this study were performed with at least three biological replicates unless specified. All attempts at replication were successful.
Randomization	No randomization methods were used as all samples were randomly collected from their populations.
Blinding	During embryo injection, the investigator injected the cells to the 8-cell stage embryos without knowing the cell types. In the other experiments, the investigators were not blinded to group allocation during data collection because no bias could have been removed by blinding.

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
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

Primary antibodies:
 Mouse anti-MERVL-gag (Epigentek, A-2801), 1:200 dilution.
 Mouse anti-CDX2 (BioGenex, MU392A-UC), 1:200 dilution.
 Mouse anti-Elf5 (Santa-Cruz, sc-166653), 1:100 dilution.

Rabbit anti-Oct4 (Abcam, ab19857), 1:200 dilution.
 Goat anti-tdTomato (Biorbyt, orb182397), 1:100 dilution.
 Rat anti-Cytokeratin 8 (Developmental Studies Hybridoma Bank, TROMA-1, AB_531826), 1:10 dilution.
 Mouse anti-Proliferin (Santa Cruz, sc-271891), 1:50 dilution.
 Goat anti-Placental lactogen I (Santa Cruz, Sc-34713), 1:75 dilution.
 Rabbit anti-Tfap2c (Santa cruz, Sc-8977), 1:50 dilution.
 Secondary antibodies:
 Donkey anti-Mouse IgG 488 (Life technologies, A-21202), 1:500 dilution.
 Donkey anti-Mouse IgG 555 (Life technologies, A-31570), 1:500 dilution.
 Goat anti-Mouse IgG 647 (Life technologies, A-21241), 1:500 dilution.
 Goat anti-Rabbit IgG 647 (Life technologies, A-21244), 1:500 dilution.
 Donkey anti-Rabbit IgG 488 (Life technologies, A-21206), 1:500 dilution.
 Donkey anti-Goat IgG 555 (Life technologies, A-21432), 1:500 dilution.
 Donkey anti-Rat IgG 488 (Life technologies, A-21208), 1:500 dilution.

Validation

All the antibodies are commercially available and have been validated by the manufacturer.
 Mouse anti-MERVL-gag (Epigentek, A-2801), <https://www.epigentek.com/catalog/muervl-gag-polyclonal-antibody-p-5383.html>
 Mouse anti-CDX2 (BioGenex, MU392A-UC), <https://biogenex.com/product/anti-cdx-2/?v=1c2903397d88>
 Mouse anti-Elf5 (Santa-Cruz, sc-166653), <https://www.scbt.com/p/elf-5-antibody-g-2?requestFrom=search>
 Rabbit anti-Oct4 (Abcam, ab19857), <https://www.abcam.cn/oct4-antibody-ab19857.html>
 Goat anti-tdTomato (Biorbyt, orb182397), <https://www.biorbyt.com/tdtomato-antibody-orb182397.html>
 Rat anti-Cytokeratin 8 (Developmental Studies Hybridoma Bank, TROMA-1, AB_531826), <https://dshb.biology.uiowa.edu/TROMA-1>
 Mouse anti-Proliferin (Santa Cruz, sc-271891), <https://www.scbt.com/p/proliferin-antibody-e-10>
 Goat anti-Placental lactogen I (Santa Cruz, sc-34713), <https://www.scbt.com/p/placental-lactogen-i-antibody-p-17?requestFrom=search>
 Rabbit anti-Tfap2c (Santa cruz, Sc-8977), <https://www.scbt.com/p/ap-2gamma-antibody-h-77?requestFrom=search>
 Donkey anti-Mouse IgG 488 (Life technologies, A-21202), <https://www.thermofisher.cn/cn/zh/antibody/product/Donkey-anti-Mouse-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-21202>
 Donkey anti-Mouse IgG 555 (Life technologies, A-31570), <https://www.thermofisher.cn/cn/zh/antibody/product/Donkey-anti-Mouse-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-31570>
 Goat anti-Mouse IgG 647 (Life technologies, A-21241), <https://www.thermofisher.cn/cn/zh/antibody/product/Goat-anti-Mouse-IgG2a-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-21241>
 Donkey anti-Rabbit IgG 488 (Life technologies, A-21206), <https://www.thermofisher.cn/cn/zh/antibody/product/Donkey-anti-Rabbit-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-21206>
 Donkey anti-Goat IgG 555 (Life technologies, A-21432), <https://www.thermofisher.cn/cn/zh/antibody/product/Donkey-anti-Goat-IgG-H-L-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-21432>
 Donkey anti-Rat IgG 488 (Life technologies, A-21208), <https://www.thermofisher.cn/cn/zh/antibody/product/Donkey-anti-Rat-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-21208>

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

tdTomato-labeled mESCs were derived from E3.5 inner cell mass (ICM) of C57BL/6-Gt(ROSA)26Sor tm1(CAG-tdTomato)/Vst (Vitalstar) mouse.
 OG2-mESCs were derived from E3.5 inner cell mass (ICM) of C57BL/6-CBA-Tg(Pou5f1-EGFP)2Mnn/J (The Jackson Laboratory, Cat# 004654; RRID: IMSR_JAX:004654) mouse.
 R1-mESCs (ATCC, Cat# SCRC-1011; RRID: CVCL_2167)
 ciTotiSCs were derived from mESCs.

Authentication

All cell lines were confirmed with normal morphology and expression of cell specific genes. OG2-mESCs were validated by RNA-seq. tdTomato-labeled mESCs and OG2-mESCs were authenticated by chimera experiments (inject the cells into 8-cell stage mouse embryos). R1-mESCs were purchased from ATCC. ciTotiSCs were well characterized with the details listed in the manuscript.

Mycoplasma contamination

All cell lines were tested negative for mycoplasma contamination.

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

No commonly misidentified cell lines were used herein.

Animals and other organisms

Policy information about [studies involving animals; ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

All mice were housed in individually ventilated cages (maxima six mice per cage) in barrier facility at Tsinghua University. The mice were maintained on a 12/12-hour light and dark cycle, 22–26°C with sterile pellet food and water ad libitum.
 ICR mice purchased from Beijing Vital River Laboratory Animal Technology were used as embryo donor (6–8 weeks old) and the pseudopregnant recipients (4–6 weeks old) for embryo transplantation.

Wild animals

No wild animals were used in the study.

Field-collected samples

No field-collected samples were used in the study.

Ethics oversight

All the animal experiments in this study were performed in accordance with the guidelines and regulations of IACUC (Institutional

Ethics oversight

Animal Care and Use Committee) of Tsinghua University, Beijing, China. All the animal protocols used in this study have been approved by IACUC (Institutional Animal Care and Use Committee) of Tsinghua University.

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

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Cultured cells were trypsinized and resuspended in appropriate medium, followed by filtering through 40 µM cell strainers. To assess the contribution of tdTomato+ cells to different parts of conceptuses of E13.5 chimera, embryos, yolk sacs and placentas were cut into small pieces (~1 mm diameter) on ice. For placentas, samples were digested with accutase (Gibco, A1110501) for 10 min at 37 °C on the shaker. For embryos and yolk sacs, samples were digested with collagenase IV (addition of 1 U/mL DNase) for 30 min or 5 min at 37 °C, respectively, followed by incubation with TrypLE (Gibco, 12605028) for 3-5 min. After digestion, all tissues were dissociated into single cells by pipetting and filtered through 70 µm cell strainers. Then, cells were centrifuged, resuspended and sorted by flow cytometer.

Instrument

Aria II and Aria III Flow Cytometer System (BD Biosciences).

Software

Data analysis was performed by FlowJo 10 software.

Cell population abundance

Sample purity of sorted cells were verified by reload and FACS analyse.

Gating strategy

Preliminary FSC/SSC gates were used to remove debris, doublets and other aggregated particles. The boundary between positive and negative was defined according to the negative control.

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