Article

Rolling back human pluripotent stem cells to an eight-cell embryo-like stage

https://doi.org/10.1038/s41586-022-04625-0

Received: 7 January 2021

Accepted: 10 March 2022

Published online: 21 March 2022

Check for updates

Md. Abdul Mazid[1,2,14]✉, Carl Ward[1,14], Zhiwei Luo[1,14], Chuanyu Liu[3,14], Yunpan Li[1], Yiwei Lai[1,3], Liang Wu[2,3], Jinxiu Li[2,3], Wenqi Jia[1,2], Yu Jiang[4], Hao Liu[1], Lixin Fu[1,2], Yueli Yang[4], David P. Ibañez[1,2], Junjian Lai[1], Xiaoyu Wei[2,3], Juan An[1,5], Pengcheng Guo[4], Yue Yuan[2,3], Qiuting Deng[2,3], Yang Wang[3], Ying Liu[3], Fei Gao[6], Junwen Wang[7], Shahriar Zaman[8], Baoming Qin[9], Guangming Wu[10], Patrick H. Maxwell[11], Xun Xu[3,12], Longqi Liu[3]✉, Wenjuan Li[1]✉ & Miguel A. Esteban[1,3,4,13]✉

After fertilization, the quiescent zygote experiences a burst of genome activation that initiates a short-lived totipotent state. Understanding the process of totipotency in human cells would have broad applications. However, in contrast to in mice[1,2], demonstration of the time of zygotic genome activation or the eight-cell (8C) stage in in vitro cultured human cells has not yet been reported, and the study of embryos is limited by ethical and practical considerations. Here we describe a transgene-free, rapid and controllable method for producing 8C-like cells (8CLCs) from human pluripotent stem cells. Single-cell analysis identified key molecular events and gene networks associated with this conversion. Loss-of-function experiments identified fundamental roles for DPPA3, a master regulator of DNA methylation in oocytes[3], and TPRX1, a eutherian totipotent cell homeobox (ETCHbox) family transcription factor that is absent in mice[4]. DPPA3 induces DNA demethylation throughout the 8CLC conversion process, whereas TPRX1 is a key executor of 8CLC gene networks. We further demonstrate that 8CLCs can produce embryonic and extraembryonic lineages in vitro or in vivo in the form of blastoids[5] and complex teratomas. Our approach provides a resource to uncover the molecular process of early human embryogenesis.

Understanding human embryogenesis is necessary to develop treatments for infertility and developmental disorders, to clarify the role of heritable epigenetics in our species and to produce optimal differentiated cells for regenerative medicine or in vitro studies. Pluripotent stem cells (PSCs) are an in vitro representation of pluripotent cells in the mammalian inner cell mass (ICM) of the blastocyst and can be used to study early developmental processes. Similar to their mouse counterparts, human PSCs can be maintained in naive or primed states, which reflect the preimplantation and postimplantation ICM, respectively[6].

Mouse naive PSC cultures contain a small proportion (approximately 0.5%) of cells sharing features of the two-cell (2C) embryonic stage, a time when mouse zygotic genome activation (ZGA) occurs and totipotency is achieved. These cells are termed 2C-like cells (2CLCs), and their proportion can be enhanced through manipulation[1,7]. Despite advancements achieved in methods using rodent models, the generation of 8CLCs that transcriptionally and epigenetically match the human 8C embryo, and evaluation of their functional properties, have not yet been described. Only low activation of few 8C-embryo-enriched (totipotency) genes was reported in bulk RNA-sequencing (RNA-seq) of naive induced pluripotent stem (iPS) cells generated through somatic cell reprogramming in 5iLAF medium[8]. Although expanded pluripotent stem cells (EPSCs) that contribute to extraembryonic lineages have been generated[9–11], they transcriptionally resemble primed PSCs[12,13]. The lack of bona fide 8CLCs that can be enriched and purified for further experimentation represents a substantial knowledge gap because human embryonic material is limited and subjected to special ethical considerations.

Generation of 8CLCs from human PSCs

We conducted a screen with primed human embryonic stem (ES) cells for inhibitors that target signalling and epigenetic pathways to

[1]Laboratory of Integrative Biology, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China. [2]University of Chinese Academy of Sciences, Beijing, China. [3]BGI-Shenzhen, Shenzhen, China. [4]Jilin Provincial Key Laboratory of Animal Embryo Engineering, Key Laboratory of Zoonoses Research, Ministry of Education, College of Veterinary Medicine, Jilin University, Changchun, China. [5]School of Life Sciences, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China. [6]Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China. [7]E-GENE, Shenzhen, China. [8]Department of Genetic Engineering and Biotechnology, Faculty of Life and Earth Sciences, University of Rajshahi, Rajshahi, Bangladesh. [9]Laboratory of Metabolism and Cell Fate, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China. [10]Guangzhou Laboratory, Guangzhou, China. [11]Cambridge Institute for Medical Research, Department of Medicine, University of Cambridge, Cambridge, UK. [12]Guangdong Provincial Key Laboratory of Genome Read and Write, Shenzhen, China. [13]Institute of Stem Cells and Regeneration, Chinese Academy of Sciences, Beijing, China. [14]These authors contributed equally: Md. Abdul Mazid, Carl Ward, Zhiwei Luo, Chuanyu Liu. ✉e-mail: mazid@gibh.ac.cn; liulongqi@genomics.cn; li_wenjuan@gibh.ac.cn; miguel@gibh.ac.cn

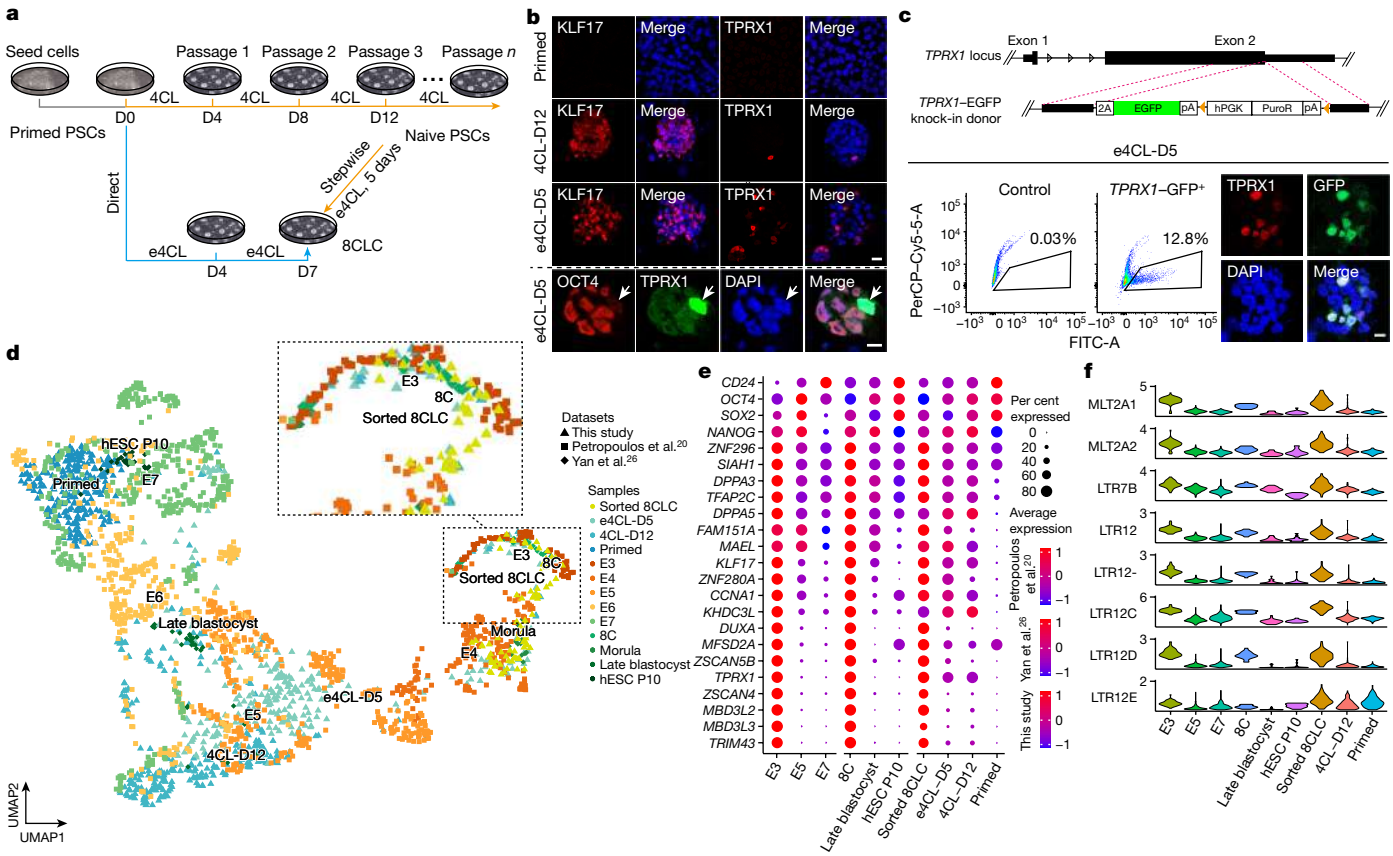


Fig. 1 | Generation of 8CLCs from human PSCs. a, Schematic depicting the generation of naive PSCs and 8CLCs from primed PSCs with 4CL, stepwise e4CL or direct e4CL medium. D, day. **b**, Top, representative images of immunostaining for KLF17 and TPRX1 in primed H9 ES cells untreated or converted by 4CL-D12 or stepwise e4CL-D5. Bottom, representative co-immunostaining images of stepwise e4CL-D5 cells showing mutual exclusivity between OCT4 and TPRX1 (arrows). Representative of three independent experiments. **c**, Top, schematic representing EGFP insertion into the *TPRX1* locus to generate *TPRX1*–EGFP reporter cell lines. Bottom, *TPRX1*–EGFP knock-in H9 ES cells were validated by culture in stepwise e4CL-D5 and FACS analysis (left) and immunostaining with anti-TPRX1 (right). Representative of three independent experiments. **d**, UMAP comparing the human E7 to E3 stage[20], human ES cell passage 10 (hESC P10) to 8C embryo[26] with primed ES cells, 4CL-D12 naive ES cells, stepwise e4CL-D5 cells and sorted 8CLCs from stepwise e4CL-D5 cells profiled using SMART-seq2. All reference datasets used in this study are summarized in Supplementary Table 8. **e**, Bubble plot representing the frequency of expression and average expression of representative pluripotency and totipotency genes in early human embryonic stages and hESC P10 compared to primed ES cells, 4CL-D12 naive ES cells and sorted 8CLCs from **d**. **f**, Violin plot showing the log-normalized expression of representative embryo-enriched TEs in early human embryonic stages and hESC P10 compared to primed ES cells, 4CL-D12 naive ES cells and sorted 8CLCs from **d**. Scale bars, 10 μm (**b** (bottom), **c**) or 20 μm (**b**, top).

gradually define permissive culture conditions for producing 8CLCs (Extended Data Fig. 1a). In the first round, combining the MEK inhibitor PD0325901, the tankyrase inhibitor IWR1 and human leukaemia inhibitory factor (LIF) in a chemically defined medium (N2B27, activin A and vitamin C[14]) on feeder cells produced PSC colonies with a naive PSC-like colony morphology and the highest levels of naive pluripotency gene expression, as assessed by quantitative PCR with reverse transcription (RT–qPCR) (Extended Data Fig. 1b, c). In round two, we tested 14 additional compounds added to this basal medium (Extended Data Fig. 1d). 3-Deazaneplanocin A (DZNep), an inhibitor of *S*-adenosylhomocysteine hydrolase, which is commonly known for its role in suppressing the histone 3 lysine 27 (H3K27) methyltransferase EZH2 (ref. [15]), and the class I histone deacetylase (HDAC) inhibitor trichostatin A (TSA) substantially increased the expression of naive pluripotency genes, particularly in combination (Extended Data Fig. 1e, f). The conversion took about 12 days (three passages) and there was no obvious cell death (Fig. 1a). We termed this naive medium 4CL (4 chemicals+LIF) and validated it in multiple male and female ES cell and iPS cell lines (Extended Data Fig. 1g). We observed that 4CL supports long-term maintenance (tested for up to 15 passages) without notable karyotype abnormalities (Extended Data Fig. 1h and Supplementary Fig. 1). PSCs cultured in 4CL could be passaged as single cells on feeder cells, on surfaces coated with extracellular matrix or in suspension (Extended Data Fig. 1i–k). We also noted that removing vitamin C reduced the expression of naive pluripotency genes in the 4CL-mediated conversion process (Extended Data Fig. 1l). Moreover, TSA could be effectively substituted by other HDAC inhibitors such as valproic acid or sodium butyrate, and DZNep by a related EZH2 inhibitor, CPI-1205 (Extended Data Fig. 1m).

Bulk RNA-seq confirmed that PSCs grown in 4CL display transcriptional similarity with ICM cells[16] and with naive PSC cultures grown in 5iLAF[17] or tt2iLGö medium (Extended Data Fig. 2a and Supplementary Table 1). We then tested the effect of increasing the dose of TSA (from 5 to 20 nM) and DZNep (from 10 to 50 nM) on naive PSCs generated with 4CL. After 5 days, this enhanced 4CL medium (e4CL) promoted the expression of totipotency genes (for example, *TPRX1*, *ZSCAN4*, *ZSCAN5B*, *DUXA*, *DUXB* and *ZNF280A*), the levels of which were low in PSCs generated in different naive medium and particularly from EPSCs[11] (Extended Data Fig. 2b, c). Cells grown in e4CL preserved a normal karyotype (Extended Data Fig. 2d and Supplementary Fig. 1) but could not be passaged for extended culture, which may be related to the fact that totipotent cells lack self-renewal capacity[1,18]. Totipotency genes were also induced by culturing primed cells directly in e4CL for

7 days (Fig. 1a and Extended Data Fig. 2e). We termed this alternative approach direct e4CL to distinguish it from the stepwise, first 4CL and then e4CL (stepwise e4CL) conversion process.

To clarify whether the totipotency gene expression levels observed in the bulk RNA-seq and RT–qPCR analyses of stepwise e4CL-day 5 and direct e4CL-day 7 cells reflects a fraction of 8CLCs or low-level widespread expression of totipotency genes, we used a droplet-based method[19] to perform single-cell RNA-seq (scRNA-seq) spanning the entire conversion process. A total of 55,372 single cells passed quality control. We integrated them with a published SMART-seq2 scRNA-seq dataset of early human embryos[20] using Seurat[21] (Extended Data Fig. 3a). Uniform manifold approximation and projection (UMAP) representation showed that over 12 days in 4CL, primed PSCs gradually reverted to a state that resembled embryonic day 5 (E5; early blastocyst) cells. A few cells also clustered closer to E3 cells (8C embryo), which suggests there was a conversion to an 8CLC identity. This phenomenon was prominently increased when 4CL-day 12 naive PSCs were cultured in e4CL for 5 days. Culture of primed PSCs in direct e4CL produced cells close to E3 and E4 cells within 7 days, but the trajectory was different (Extended Data Fig. 3a). Primed-specific genes, such as *CD24* (ref. [22]), were gradually downregulated during the stepwise or direct e4CL conversion process (Extended Data Fig. 3b). A set of naive pluripotency genes, such as *DPPA3* and *KLF17*, were highly expressed in 4CL-day 12 naive PSCs and peaked in the putative 8CLC clustering near E3 cells, which is consistent with their expression pattern in the human embryo. Similarly, totipotency genes such as *TPRX1, DUXA* and *ZNF280A* were almost exclusively expressed in stepwise e4CL-day 5 or direct e4CL-day 7 cells (Extended Data Fig. 3c). Expression of KLF17 in 4CL-day 12 naive PSCs and TPRX1 in e4CL-day 5 or direct e4CL-day 7 cells was validated by immunostaining (Fig. 1b, top, and Extended Data Fig. 3d).

We focused on stepwise e4CL-day 5 cells to characterize the putative 8CLC population. Unsupervised clustering showed seven clusters, among which cluster 5 cells (11.9% of the entire population) displayed resemblance to the 8C embryo gene-expression pattern (Extended Data Fig. 4a–c). As in mouse 2C and human 8C embryos[20,23], key classical pluripotency transcription factors such as *OCT4* and *SOX2* were downregulated in cluster 5 cells (Extended Data Fig. 4b, c). Hereafter, we referred to cluster 5 cells as 8CLCs. Correlation analysis confirmed that 8CLCs in e4CL-day 5 cells closely resemble E3 cells (Extended Data Fig. 4d, e). Co-immunostaining verified the opposing pattern of OCT4 and SOX2 expression with TPRX1 in 8CLCs produced with stepwise e4CL (Fig. 1b, bottom, and Extended Data Fig. 4f). Direct e4CL-day 7 cells also contained a cluster (15.6%) that expressed 8C embryo genes, but the overall transcriptional match with E3 cells was less complete (Extended Data Fig. 4g, h). We further investigated the identity of 8CLCs in stepwise e4CL-day 5 cells by confirming the activation of early-embryo-enriched transposable elements (TEs) such as MLT2A1, MLT2A2 and LTR12C[24] (Extended Data Fig. 4i).

Droplet-based scRNA-seq covers a larger number of cells than SMART-seq2, but it has a 3′ bias rather than full-length gene coverage[25]; consequently, it is less sensitive. To test the fidelity with which 8CLCs generated with our culture medium reflect 8C embryo cells, we prepared a *TPRX1*–EGFP knock-in reporter cell line (Fig. 1c). We selected *TPRX1* because expression of this ETCHbox family transcription factor[4] is highly enriched in the 8C embryo (Extended Data Fig. 4c). We performed fluorescence-activated cell sorting (FACS) on *TPRX1*–GFP$^+$ cells (hereafter termed sorted 8CLCs) in e4CL-day 5 cells together with unsorted cells, 4CL-day 12 naive PSCs and primed PSCs, which were then subjected to SMART-seq2 analysis. We integrated the resulting 804 cells with two independent human early-embryo full-length scRNA-seq datasets[20,26]. A subset of e4CL cells and a substantial proportion of sorted 8CLCs clustered with E3 cells and the 8C embryo (Fig. 1d and Extended Data Fig. 5a), whereas some sorted 8CLCs stretched into the early E4 (morula) state. This was also confirmed using cell-to-cell correlation analysis (Extended Data Fig. 5b). These results are consistent with human ZGA starting at E3 but finishing by E4 (refs. [20,26]), which is reflected by the maintenance of lower levels of totipotency genes such as *TPRX1*, *ZSCAN4*, *DUXA* and *ZSCAN5B* until this stage in the human embryo (Extended Data Figs. 4c and 5c). Some 4CL-day 12 cells clustered together with E3 cells and the 8C embryo, which provides further evidence that our naive medium induces bona fide 8CLCs, albeit at low rate. Notably, sorted 8CLCs grown from e4CL medium expressed totipotency genes to the level of the 8C embryo, as well as high levels of naive pluripotency genes such as *DPPA3* and *KLF17* (Fig. 1e and Extended Data Fig. 5d). By contrast, classical (*OCT4* and *SOX2*) and primed (*CD24*) pluripotency genes were downregulated (Fig. 1e and Extended Data Fig. 5e). Sorted 8CLCs also displayed a pattern of TEs comparable to the 8C embryo and distinct from naive and primed PSCs (Fig. 1f, Extended Data Fig. 5f, g and Supplementary Table 2).

To monitor the dynamics of 8CLC generation, we sorted *TPRX1*$^+$ and *TPRX1*$^-$ cells from stepwise e4CL-day 5 cells and cultured them in 4CL and e4CL medium, respectively (Extended Data Fig. 5h). Around half of the sorted *TPRX1*$^+$ cells became *TPRX1*$^-$ within 24 h in 4CL medium and could be readily expanded in 4CL medium for prolonged cultures as naive PSCs. Similarly, sorted *TPRX1*$^-$ cells from e4CL-day 5 cultures could either be induced to *TPRX1*$^+$ cells (about 5%) over 24 h in e4CL medium or expanded in 4CL medium as naive PSCs. These results demonstrate that 8CLCs and naive PSCs are interconvertible, which is consistent with the properties of mouse 2CLCs[1].

Next, we examined the existence of 8CLCs in cultures produced using other naive media and in EPSCs. We analysed publicly available droplet-based scRNA-seq datasets for naive PSCs generated in PXGL[13], 5iLA[27], t2iLGö[28] or RSeT[28] medium and from EPSCs[11]. A small putative 8CLC population with variable and low levels of totipotency genes was detected in cultures grown in PXGL (1.7%) and 5iLA (0.7%). Notably, these percentages and the expression levels of totipotency genes are much lower than cultures grown using the stepwise e4CL-day 5 method (11.9% and substantially higher expression levels) (Extended Data Fig. 6a–c). t2iLGö and RSeT (a commercial medium based on NHSM[29]) cultures also contained a small fraction of putative 8CLCs (0.5% and 0.2% cells, respectively), but the expression of totipotency genes was more heterogenous. By contrast, we could not detect any putative 8CLC population in EPSCs. We then tested the effect of adding e4CL medium to naive PSCs induced by PXGL, 5iLA and HENSM[30] (an updated version of NHSM). There was a greatly enhanced expression of totipotency genes after 5 days, as assessed by RT–qPCR and immunostaining (Extended Data Fig. 7a–d). There was also an increased expression of totipotency genes using DZNep (50 nM) or TSA (20 nM) individually and more potently in combination, which provides evidence that these are the key e4CL components that promote the transition to 8CLCs. Similarly, DZNep (50 nM) and TSA (20 nM) synergistically increased the proportion (19.2%) of GFP$^+$ cells (2CLCs) and the expression of totipotency genes when added to mouse MuERV-L-LTR-GFP ES cells cultured in naive (serum+LIF) conditions[31] (Extended Data Fig. 7e–h).

DNA methylation is an important regulator of gene expression and undergoes dynamic changes in early development and during the primed-to-naive PSC conversion[32,33]. Moreover, substantial DNA demethylation in some naive PSC media is associated with increased risk of genome stability[34]. We studied whether DNA methylation at naive pluripotency and totipotency genes changes along the conversion to 8CLCs with e4CL and is comparable to the 8C embryo. First, we applied reduced representation bisulfite sequencing (RRBS) to primed PSCs, 4CL-day 12 naive PSCs, stepwise e4CL-day 5 cells and direct e4CL-day 7 cells. Notably, 4CL-day 12 naive PSCs showed overall lower DNA methylation levels than primed PSCs and a higher degree of resemblance at genomic imprinted loci with the ICM[32] than 5iLAF[17,35], t2iLGö[36] or tt2iLGö[37] naive PSCs (Extended Data Fig. 8a, b). Global DNA methylation was further decreased in stepwise e4CL-day 5 cells but did not reach the level of 5iLAF or t2iLGö cultures or ICM or the 8C embryo[32]. Direct e4CL-day 7 cells displayed similar DNA methylation levels to the 8C embryo (Extended Data Fig. 8a). Closer examination of naive

pluripotency loci (for example, *DPPA3*, *DNMT3L* and *DPPA5*) showed reduced DNA methylation in 4CL-day 12 naive PSCs, t2iLGö and 5iLAF cultures and stepwise e4CL-day 5 cells compared with primed PSCs (Extended Data Fig. 8c). Notably, as in the preimplantation ICM and the 8C embryo, totipotency loci (for example, *TPRX1*, *RFPL2* and *TRIM43*) also displayed reduced DNA methylation in 4CL-day 12 naive PSCs and in stepwise e4CL-day 5 cells and direct e4CL-day 7 cells compared with primed PSCs (Extended Data Fig. 8d). However, there was little change or even relative DNA hypermethylation in naive PSCs generated using other methods[29,35,36]. In addition, we performed whole-genome bisulfite sequencing (WGBS), which provides wider coverage than RRBS, with sorted 8CLCs, 4CL-day 12 naive PSCs and primed PSCs. This analysis showed that the global content of DNA methylation in sorted 8CLCs was lower than 4CL-day 12 naive or primed PSCs and comparable to the 8C embryo[32] (Extended Data Fig. 9a). CpG methylation at the transcription start site (TSS), gene bodies and specific genomic regions (for example, enhancers, CpG islands, long terminal repeats, short interspersed nuclear elements, long interspersed nuclear elements, intergenic and intragenic) was also lower in sorted 8CLCs than 4CL-day 12 naive and primed PSCs (Extended Data Fig. 9b, c). Analysis of naive pluripotency and totipotency genes (Extended Data Fig. 9d) validated the above observations with RRBS. We also noted that early-embryo-enriched TEs were more demethylated in 8CLCs than 4CL-day 12 naive and primed PSCs, which correlated with their expression (Extended Data Fig. 9e). Genome-wide comparison of the DNA methylome profiles confirmed that sorted 8CLCs, ICM and 8C embryo cells were globally hypomethylated and clustered closer to each other than primed PSCs and the postimplantation embryo (Extended Data Fig. 9f). Furthermore, we examined DNA methylation and gene expression for a set of well-known naive pluripotency and totipotency genes in mouse 2C embryo and 2CLCs compared with later developmental stages or naive ES cells, respectively (Extended Data Fig. 10a, b). In general, DNA methylation levels at these loci were lower in the 2C embryo[38] and 2CLCs[39]. Yet, although totipotency gene expression peaked in the 2C embryo[23] and 2CLCs[40], only some naive pluripotency genes, such as *Klf4* in the 2C embryo and *Dppa2* in 2CLCs, followed the same pattern.

Altogether, these results confirm the feasibility of a rapid and controllable methodology for producing 8CLCs that resemble the human 8C embryo from either primed or naive PSCs.

Chromatin landscape of 8CLCs

To assess whether 8CLCs grown in e4CL acquire the chromatin accessibility signature of the 8C embryo, we performed single-cell assay for transposase-accessible chromatin with high-throughput sequencing (scATAC-seq) of primed ES cells, 4CL-day 12 naive PSCs and stepwise e4CL-day 5 cells. Global visualization of scATAC-seq data by UMAP clearly separated these three cell states (Extended Data Fig. 11a). Chromatin around the TSS of primed PSC-enriched genes such as *ZIC2* was mostly open in primed PSCs, whereas naive pluripotency genes shared with 8CLCs such as *DPPA3* were more open in 4CL-day 12 cells and stepwise e4CL-day 5 cells (Extended Data Fig. 11b). By contrast, classical totipotency loci such as *ZSCAN5B*, *ZNF280A* and *ARGFX* were only open in a subset of stepwise e4CL-day 5 cells and a smaller fraction of 4CL-day 12 naive PSCs, which correlated with the closure of pluripotency loci such as *OCT4* and *SOX2* in the same population (Extended Data Fig. 11c, d).

To dissect the progression towards 8CLCs, we focused on stepwise e4CL-day 5 cells. Unsupervised clustering produced 12 different clusters, among which cluster 2 (5.2%) showed a chromatin accessibility pattern consistent with 8CLCs (Fig. 2a and Extended Data Fig. 11e, f). To study the trajectory of 8CLC conversion, we combined pseudotime analysis of chromatin openness around the TSS of totipotency loci and motif distribution for chromatin-bound factors. This analysis showed convergence at 8CLCs in cluster 2 (Fig. 2b). Analysis of chromatin openness along this route visibly illustrated the antagonistic behaviour of classical pluripotency loci and totipotency loci (Fig. 2c and Supplementary Table 3a). The latter included not only classical totipotency loci such as *TPRX1*, *ZSCAN4*, *ZNF280A*, *ARGFX* and *DPRX* but also lineage-specific factors such as *PITX2* (mesoderm and endoderm), *GSC* (mesendoderm) and *GATA6* (primitive endoderm) that are expressed in the human 8C embryo[16,41]. Similarly, motif analysis along the pseudotime trajectory showed enrichment in the open chromatin of potential binding sites for lineage regulators in addition to transcription factors shared between naive PSCs and 8CLCs (for example, TFAP2C) and classical totipotency transcription factors (Fig. 2d, e, Extended Data Fig. 11g and Supplementary Table 3b). Accordingly, we detected upregulation of *GATA6* and *PITX2* in 8CLCs compared with 4CL-day 12 naive PSCs in our droplet-based scRNA-seq dataset, and in RT–qPCR results of stepwise e4CL-day 5 cells compared with 4CL-day 12 naive PSCs, which matched the 8C embryo[26] (Extended Data Fig. 11h, i). However, there was negligible expression of *GATA2*, which is instead upregulated in human somatic cell reprogramming to primed or naive PSCs[28], and *GATA4*, of which the orthologue *Gata4* is enriched in the mouse 2C embryo instead of *Gata6* (ref. [23]) (Extended Data Fig. 11j). *Pitx2* was also enriched, albeit at low levels, in the mouse 2C embryo, but we did not detect upregulation of lineage-specific factors in mouse 2CLCs compared to naive ES cells. In addition, we observed that chromatin regions that contain potential DNA-binding motifs for classical pluripotency transcription factors such as OCT4 and SOX2 progressively became less accessible in cells approaching the 8CLC stage (Fig. 2e and Extended Data Fig. 11g).

Analysis of scATAC-seq genome browser tracks for selected classical (*OCT4*) and naive (*KLF17*) pluripotency genes, classical totipotency genes (*TPRX1*, *ZSCAN4* and *ZNF280A*) and the lineage-specific factor *GATA6* confirmed the above observations (Fig. 2f and Extended Data Fig. 11k). Moreover, the gain of chromatin accessibility at 8C-embryo-enriched genes in 8CLCs correlated with an increase in the active H3K27 acetylation (H3K27ac) mark in the 8C embryo compared to ICM cells[42]. Furthermore, there was a reduction in chromatin accessibility in the 8C embryo compared to ICM cells and in 8CLCs compared with 4CL-day 12 naive PSCs at the well-studied *OCT4* proximal enhancer[43]. However, chromatin accessibility in a putative *OCT4* intronic enhancer was substantially increased in the 8C embryo and 8CLCs compared to ICM cells[44] and 4CL-day 12 naive PSCs (Fig. 2f). Predicted chromatin interactions between potential regulatory elements based on co-accessibility of paired regions across individual cells were also different between 8CLCs, 4CL-day 12 naive and primed PSCs for classical (*OCT4*) and naive (*KLF17* and *TFAP2C*) pluripotency loci, as well as for classical totipotency loci (*TPRX1*, *ZSCAN4* and *ZNF280A*) and the lineage-specific factor *GATA6* (Extended Data Fig. 12a–d). These findings provide evidence of a distinctive mode of transcriptional regulation at both pluripotency and totipotency loci in 8CLCs compared to naive PSCs. Moreover, we identified motifs for both pluripotency and totipotency transcription factors at sites of gained chromatin accessibility in totipotency loci in the 8C embryo and 8CLCs (Fig. 2f and Extended Data Fig. 11k). Similarly, we observed motifs for totipotency transcription factors at classical pluripotency loci such as *OCT4*. These results suggest that a feedback mechanism controls the balance between the pluripotency and totipotency gene networks in naive PSCs and 8CLCs. In addition to genes, we examined chromatin accessibility for early-embryo-enriched TEs and found that 8CLCs have a similar open pattern to the 8C embryo. Specifically, chromatin opening at early-embryo-enriched ERV elements such as MLT2A1, MLT2A2 and LTR12C[24,45] occurred only in 8C embryo cells[16] and 8CLCs but not ICM cells[16] or 4CL-day 12 naive PSCs (Fig. 2g and Extended Data Fig. 12e). Besides, we analysed the chromatin landscape trajectory from primed and 4CL-day 12 naive PSCs to cluster 2 8CLCs, and observed that 4CL-day 12 naive PSCs sit in between primed and 8CLCs but closer to the latter (Extended Data Fig. 12f–h). The chromatin of naive pluripotency genes shared with 8CLCs (such as *DPPA3*) and totipotency genes (such as *ZNF280A*) became gradually more open as naive PSCs approach 8CLCs in this trajectory.

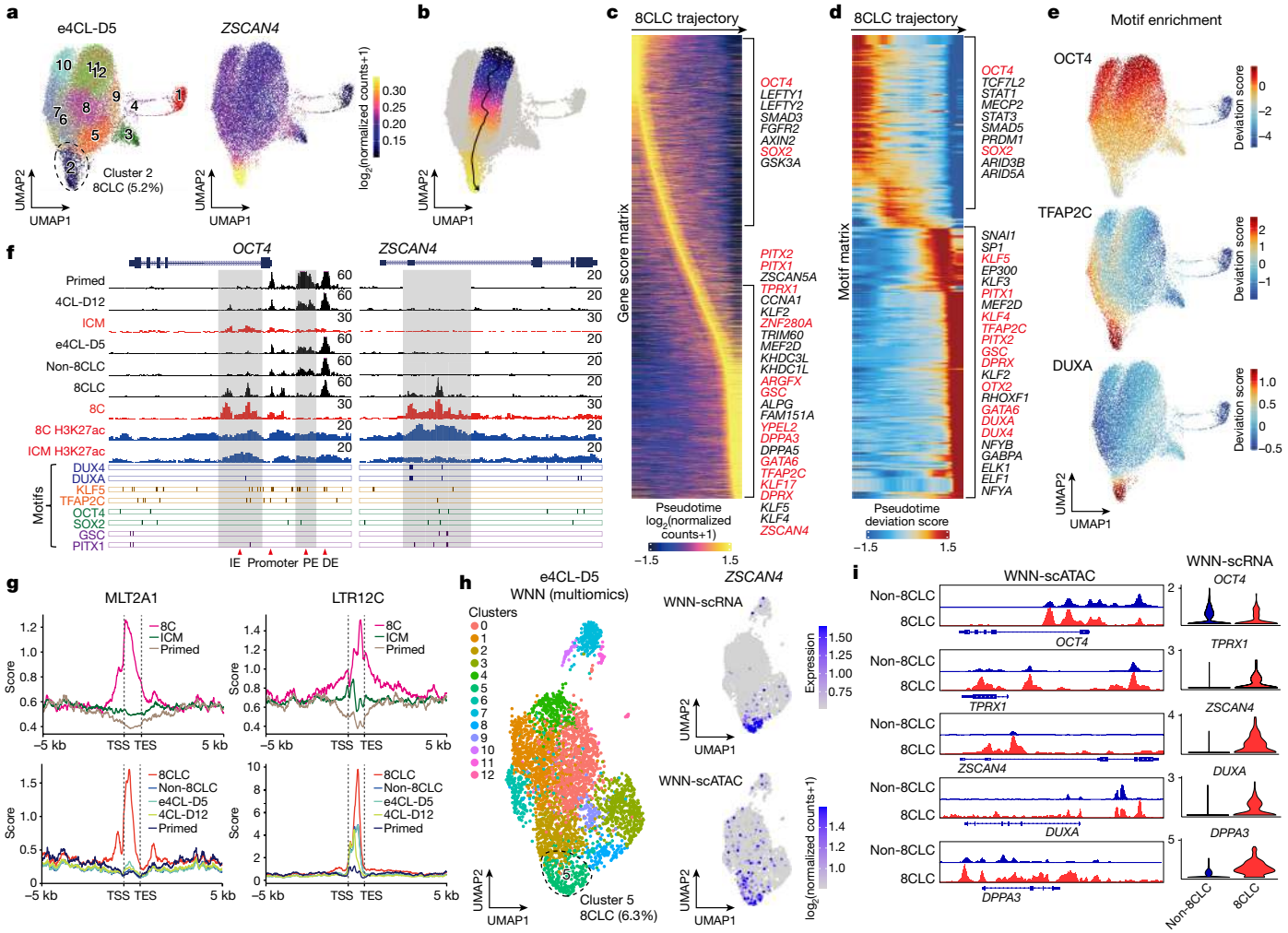


Fig. 2 | Chromatin landscape of 8CLCs. a, Left, UMAP depicting the Louvain clustering of scATAC-seq data from stepwise e4CL-D5 cells. Right, *ZSCAN4* gene score in cluster 2. H9 ES cells were used. **b**, Trajectory of 8CLC induction by combined pseudotime analysis of gene scores for 8C-embryo-enriched genes and DNA-binding motif distributions for transcription factors projected onto the UMAP of **a** using ArchR[61]. **c**, Heatmap showing the dynamics in gene score along the pseudotime trajectory of **b**. Representative genes are in red. **d**, Heatmap representing the DNA-binding motif deviation scores along the pseudotime trajectory of **b**. Representative genes are in red. **e**, DNA-binding motif deviation scores of representative pluripotency (classical, OCT4; naive, TFAP2C) and classical totipotency (DUXA) transcription factors projected onto the UMAP of **a**. **f**, Genome browser tracks showing chromatin accessibility, H3K27ac[42] and DNA-binding motifs for selected transcription factors at the *OCT4* and *ZSCAN4* loci in the indicated cell types. Our scATAC-seq data were analysed as pseudobulk and compared with human 8C embryo and ICM[16]. DE, distal enhancer; IE, intronic enhancer; PE, proximal enhancer. Reference datasets used in this study are summarized in Supplementary Table 8. **g**, Average normalized chromatin accessibility signals for the indicated early human embryo-enriched TEs in 8C embryo, ICM and primed ES cell bulk ATAC-seq[16] and pseudobulk of our scATAC-seq data. TES, transcription end site. **h**, Left, UMAP of the single-cell multiomics of stepwise e4CL-D5 cells analysed by WNN. Right, UMAP highlighting the expression (top) and chromatin accessibility (bottom) of the totipotency gene *ZSCAN4* in cluster 5. H9 ES cells were used. **i**, Genome browser tracks visualizing chromatin accessibility (left) and violin plots (right) showing the log-normalized expression for the indicated genes in 8CLCs (cluster 5) and non-8CLCs (other clusters) of **h**.

To determine the relationship between chromatin accessibility and gene expression in 8CLCs, we performed single-cell multiomics (combined scRNA-seq and scATAC-seq) of stepwise e4CL-day 5 cells. Annotation of the multiomics scRNA-seq data showed that 13.5% cells were 8CLCs (Extended Data Fig. 13a, b), which is consistent with our droplet-based scRNA-seq time-course analysis and the SMART-seq2 dataset. In the multiomics scATAC-seq data, 5.1% of cells displayed a pattern of chromatin accessibility consistent with an 8CLC identity (Extended Data Fig. 13c, d), which is similar to the percentage in the individual scATAC-seq data. Notably, most of the 8CLCs detected by multiomics scATAC-seq (79.7%) matched the 8CLCs detected by multiomics scRNA-seq, whereas 30% of the 8CLCs detected by multiomics scRNA-seq matched the multiomics scATAC-seq (Extended Data Fig. 13e). Furthermore, integrative analysis of scRNA-seq and scATAC-seq data using weighted nearest neighbour (WNN) showed that cells in cluster 5 (6.3%) have both a gene expression pattern and chromatin landscape compatible with 8CLCs (Fig. 2h, i and Extended Data Fig. 13f). These results imply that chromatin accessibility can predict 8CLC identity to a large extent and suggest that other mechanisms (for example, histone modifications and 3D genome configuration) may be important for inducing totipotency genes in 8CLCs.

We conclude that 8CLCs generated with stepwise e4CL display a chromatin configuration that mimics the 8C embryo.

Molecular roadmap of 8CLC conversion

Next, we performed a deeper analysis of our droplet-based scRNA-seq dataset to investigate the dynamics of gene expression during 8CLC

generation from primed and naive PSCs. Pseudotime analysis[46] of the stepwise e4CL conversion time-course identified five groups of highly variable genes (Extended Data Fig. 14a and Supplementary Table 4). There was good correlation between the dynamics of gene expression in all groups and chromatin accessibility at the corresponding loci (Extended Data Fig. 14b). We then performed gene ontology (GO) analysis (Extended Data Fig. 14c), which showed differential enrichment for multiple functional categories. Among these, the changes in metabolic regulators along the conversion process indicate that similar to mouse 2CLCs[7], human 8CLC generation might be amenable to modulation through metabolic interference. Notably, groups 1 and 5 were enriched in BMP–WNT and negative regulators of NOTCH–NF-κB signalling pathways, respectively. This is consistent with the observation that WNT activation has a negative effect on the generation of naive PSCs[36,47] and recent work showing that NOTCH and NF-κB inhibition facilitate it[30]. The expression pattern for many of these genes followed an inverse path in the early embryo[26] (Extended Data Fig. 14d, e). Pseudotime analysis of direct e4CL also showed five major gene groups, which included both shared and different regulators compared with the stepwise protocol (Extended Data Fig. 14f–h).

The study of gene regulatory networks (GRNs) and their main switches provides an important approach to uncover how specific cell states, such as primed and naive pluripotency or totipotency, are induced, maintained and dissolved. We therefore examined the GRNs of primed PSCs, 4CL-day 12 naive PSCs and sorted 8CLCs from e4CL-day 5 cells, taking advantage of our high-resolution scRNA-seq datasets generated with SMART-seq2. Co-expressed gene modules and differentially expressed hub genes were defined using single-cell weighted gene correlation network analysis (scWGCNA)[48] (Fig. 3a and Supplementary Table 5). This analysis revealed GRNs associated with each cell state (Fig. 3b and Extended Data Fig. 15a, b). The central players in the 8CLC GRN included totipotency genes such as *TPRX1*, *ZSCAN4* and *DUXA*. The 8CLC and naive GRNs contained shared central players such as *DPPA3* and *KLF17*, although their target genes were mostly different. GO terms related to chromatin and gene expression and the cell cycle were common to 8CLCs and naive PSCs, whereas meiotic recombination and protein ubiquitination were specifically enriched in 8CLCs. The primed PSC GRN was substantially different and was mainly enriched for early differentiation processes and signalling pathways. The specific function of many of the genes in all these GRNs is largely unknown, which reflects our limited understanding of human developmental stages in vivo and pluripotency and totipotency in vitro thus far.

To understand which regulators are relevant for generating 8CLCs, we compared the differentially expressed genes (DEGs) in 8CLCs (cluster 5 of stepwise e4CL-day 5 cells) with the other cells (non-8CLCs) at the same time point (Supplementary Table 6). The analysis showed 449 genes upregulated in 8CLCs, whereas few genes were downregulated (Fig. 3c). Among the upregulated genes, there were several of the predicted key regulators of the 8CLC GRN and the naive PSC and 8CLC GRNs, including *TPRX1*, *ZSCAN4*, *ZNF280A*, *DUXA* and *DPPA3*. We also observed that many upregulated genes in 8CLCs fell into GO categories related to protein synthesis and RNA metabolism (Extended Data Fig. 15c), which is consistent with the observation that these processes are enriched in the 8C embryo[41]. We focused on *DPPA3* and *TPRX1*, which are among the top five upregulated genes (four and five, respectively). DPPA3 controls passive DNA demethylation in mouse oocytes by inducing the cytoplasmic translocation of UHRF1, a partner of DNMT1. Consequently, oocytes lacking DPPA3 produce embryos with defective ZGA[3]. DPPA3 has also been implicated in the regulation of genomic imprinting in mouse somatic cell reprogramming[49], DNA methylation maintenance in naive mouse PSCs[50,51] and the efficiency of mouse somatic nuclear transfer[52]. In addition, *DPPA3* marks a population of mouse ES cells with ICM-like characteristics and, interestingly, its expression can be enhanced with TSA[53], a key component of 4CL and e4CL. *TPRX1* and other ETCHbox family members (*ARGFX* and *DPRX*), highly enriched in the human 8C embryo[20] and in 8CLCs (Extended Data Fig. 15d), were lost in mice during divergent evolution from primates[4]. Their function and targets are largely unknown.

Knocking out *DPPA3* blocked the generation of naive PSCs from primed PSCs grown in 4CL at day 12 and the acquisition of 8C embryo markers in direct e4CL at day 7 (Fig. 3d and Extended Data Fig. 15e, f). We also confirmed that *DPPA3* is necessary for the naive PSC-to-8CLC step by performing siRNA-mediated knockdown during the stepwise 4CL-day 12 to e4CL-day 5 conversion process (Fig. 3e). By contrast, knockout of *TPRX1* blocked the acquisition of 8C embryo markers in stepwise e4CL-day 5 and direct e4CL-day 7 cells, but did not significantly affect naive PSC generation in 4CL at day 12 (Fig. 3f and Extended Data Fig. 15g). We noted that *TPRX1* and *DPPA3* levels were reduced in cells depleted in *DPPA3* or *TPRX1*, respectively, in stepwise e4CL at day 5 and/or direct e4CL at day 7 (Fig. 3d–f). Moreover, downregulated genes in cells depleted in *DPPA3* (direct e4CL-day 7) or *TPRX1* (stepwise e4CL-day 5) were enriched for GO terms similar to those observed in 8CLCs compared with non-8CLCs (Extended Data Fig. 15c, h, i). Furthermore, as reported for oocytes[3], DPPA3 overexpression induced the nucleocytoplasmic translocation of UHRF1 in primed PSCs, naive 4CL-day 12 PSCs and stepwise e4CL-day 5 cells (Extended Data Fig. 16a), whereas UHRF1 protein and mRNA levels remained unchanged in *DPPA3*-depleted cells compared to controls (Extended Data Fig. 16b). Consistent with these observations, global DNA methylation levels were substantially higher in *DPPA3* knockout cells grown in 4CL-day 12 and more markedly in direct e4CL-day 7 than the wild-type control (Extended Data Fig. 16c). DNA methylation around the TSS of multiple naive pluripotency loci (*DPPA5* and *KHDC3L*) and totipotency loci (*TPRX1* and *TRIM43*) was higher in *DPPA3* knockout cells compared to wild-type cells (Fig. 3g and Extended Data Fig. 16d, e). There was also higher DNA methylation at intergenic regions in *DPPA3* knockout cells, which suggests that distal regulatory regions and potential sites of 3D chromatin interactions are involved (Extended Data Fig. 16f). Contrary to the human setting, suppression of *Dppa3* in mouse naive ES cells did not affect pluripotency gene levels or the induction of totipotency genes in 2CLCs (Extended Data Fig. 16g, h). We also investigated the function of TSA and DZNep on the activation of 8CLC regulators by eliminating them individually from stepwise or direct e4CL and performing bulk RNA-seq at day 5 and day 7, respectively (Extended Data Fig. 17a, b). Withdrawal of either TSA or DZNep impaired the activation of multiple regulators of the 8CLC GRN, including *ZSCAN4*, *DUXA* and *ZNF280A*, as well as *TPRX1* and *DPPA3*. Similarly, TSA was necessary for the reduction of classical pluripotency genes such as *ZIC2* and *OCT4*. Thus, the two inhibitors work synergistically to promote the 8CLC identity, but their function is not identical.

Overall, these findings indicate a substantial change in cell function and reorganization of the GRNs along the 8CLC conversion pathway. Further scrutiny of our dataset will be useful to identify additional regulators and their potential crosstalk with TPRX1 and DPPA3.

Developmental potency of 8CLCs

To comprehensively assess the differentiation potential of 8CLCs, we tested their capacity to generate trophoblast stem cells (TSCs) and sublineages, blastoids, interspecies chimeras and teratomas.

Human TSCs generated in vitro provide a unique experimental model to study early human placental development and related diseases in a dish. We produced TSCs with characteristic morphology, specific markers and gene expression pattern of human TSCs from sorted 8CLCs, stepwise e4CL-day 5 cells and 4CL-day 12 naive PSCs, as assessed by RT–qPCR, immunostaining, bisulfite PCR and bulk RNA-seq (Fig. 4a and Extended Data Fig. 18a–e). These TSCs could be further differentiated into syncytiotrophoblasts (SCTs) and extravillous trophoblasts (EVTs). In addition to RT–qPCR and immunostaining for specific markers, SCT identity was validated by measuring the secretion of human chorionic

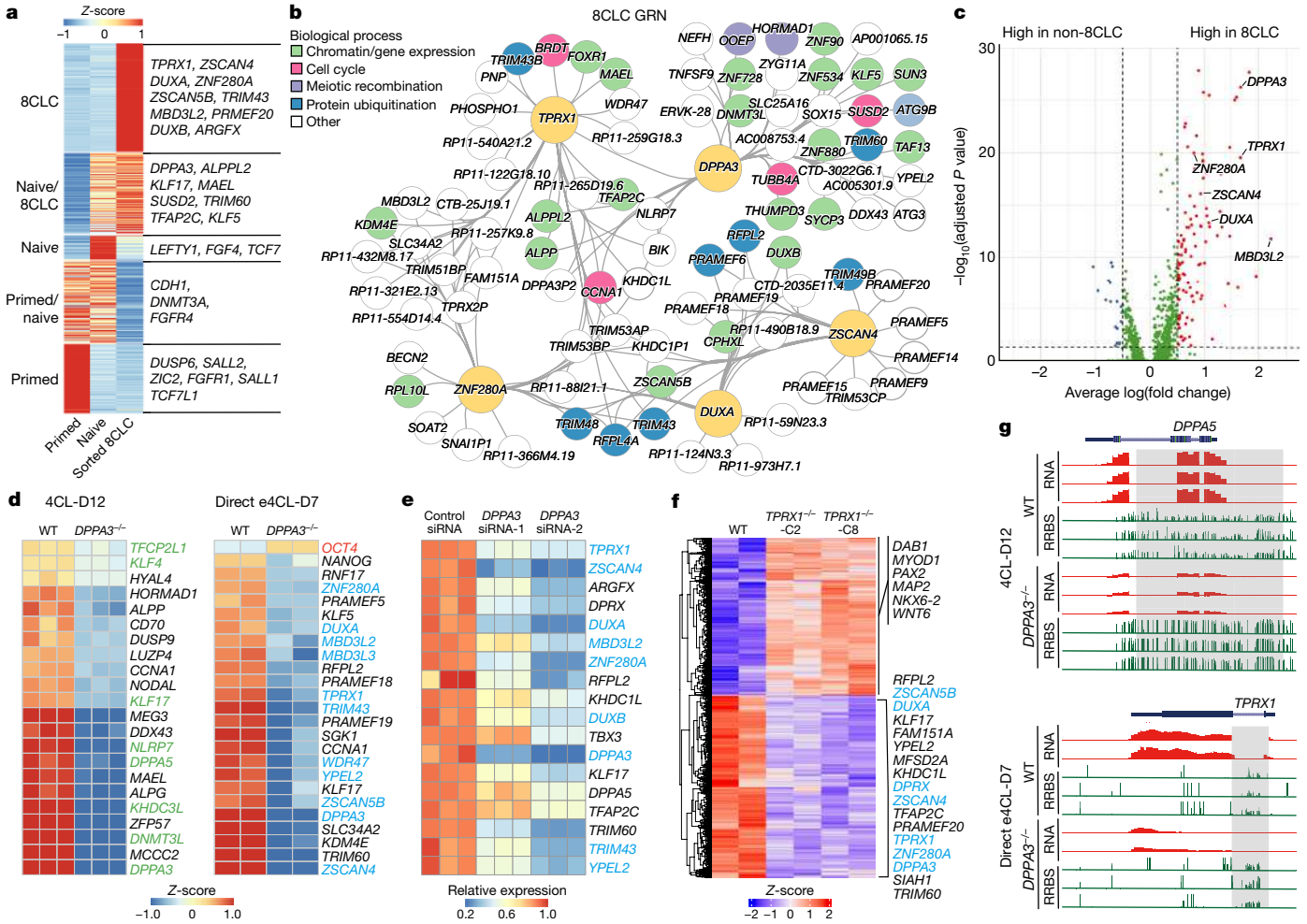


Fig. 3 | Molecular roadmap of 8CLC conversion. a, Heatmap showing cell-type-specific hub gene expression analysed by scWGCNA[56] using the dataset of Fig. 1d. **b**, Cytoscape GRN visualization of top 8CLC hub genes from **a**. **c**, Volcano plot showing DEGs between 8CLCs (cluster 5) and non-8CLCs (other clusters) in the dataset of Extended Data Fig. 3a. *P* values were calculated using Wald test and adjusted for multiple testing using Benjamini–Hochberg correction. **d**, Heatmaps showing downregulated genes in bulk RNA-seq of *DPPA3*-knockout H9 ES cells compared to wild-type (WT) after conversion from a primed state to 4CL-D12 naive PSCs (left; $n = 3$ technical replicates) or direct e4CL-D7 cells (right; $n = 2$ technical replicates). Stage-specific representative genes are highlighted (green, naive pluripotency; red, classical pluripotency; blue, totipotency). *DPPA3* knockout clone 1 was used. **e**, Heatmap showing downregulation of 8C-embryo-enriched genes in RT–qPCR of transient *DPPA3* knockdown compared to control in H9 ES cells converted from a 4CL-D12 naive PSC state to stepwise e4CL-D5 cells ($n = 3$ biological replicates). Representative 8C-embryo-enriched genes are in blue. **f**, Heatmap showing downregulation of 8C-embryo-enriched genes in bulk RNA-seq of *TPRX1*-knockout H9 ES cells compared to WT after conversion from a 4CL-D12 naive state to stepwise e4CL-D5 cells ($n = 2$ technical replicates). Representative 8C-embryo-enriched genes are in blue. *TPRX1* knockout clone 2 and clone 8 were used. **g**, Genome browser tracks showing DNA methylation measured by RRBS at the naive pluripotency *DPPA5* locus (top) and the totipotency *TPRX1* locus (bottom) in WT and *DPPA3*-knockout H9 ES cells converted from a primed state to 4CL-D12 naive PSC (top) or direct e4CL-D7 cells (bottom). Each bar represents a CpG and the height indicates the percentage of methylation. Bulk RNA-seq tracks are included.

gonadotropin (hCG) in the culture medium, whereas EVT identity was confirmed by RT–qPCR (Fig. 4b–d and Extended Data Fig. 18f, g). The ability of 4CL-day 12 naive PSCs to produce TSCs and sublineages is consistent with this property being maintained in human naive PSCs[13,54,55].

Blastocyst-like structures or blastoids generated in vitro represent a promising in vitro model of early human embryogenesis. When cultured inside AggreWells[5,27], sorted 8CLCs, stepwise e4CL-day 5 cells and 4CL-day 12 naive PSCs formed blastoids with variable efficiency (28–58%) and displayed a blastocoel-like cavity and distinctive inner and outer layers resembling the ICM and the trophectoderm, respectively. This spatial cellular organization was verified by co-immunostaining of the ICM marker OCT4 and the trophectoderm marker GATA2 (Fig. 4e, f and Extended Data Fig. 19a–c). For comprehensive characterization and to test the degree of similarity with human embryo lineages, we performed droplet-based scRNA-seq of 8CLC-derived blastoids. Integrated UMAP analysis of the resulting 1,112 cells showed clustering with epiblast, trophectoderm and hypoblast cells from the human blastocyst[5,20], which also maintained comparable frequencies (Fig. 4g, h and Extended Data Fig. 19d–f).

We evaluated the interspecies chimera competency of purified 8CLCs, 4CL-day 12 naive PSCs and primed PSCs labelled with DsRed by aggregating them with mouse 8C embryos[56] (Extended Data Fig. 20a). Substantial integration with cell distribution in both the ICM and the trophectoderm was observed in ex vivo cultured blastocysts at E3.5 for purified 8CLCs, stepwise e4CL-day 5 cells and 4CL-day 12 naive PSCs but not primed PSCs (Extended Data Fig. 20b, c). We then implanted chimeric blastocysts into pseudo-pregnant mice and allowed for further development until E10.5 in vivo. DsRed$^+$ cells were readily observed in embryonic (fetus) and extraembryonic (placenta and yolk sac) tissues of E10.5 embryos for purified 8CLCs, stepwise e4CL-day 5 cells and 4CL-day 12 naive PSCs (Extended Data Fig. 20d, e). Chimerism was validated by immunostaining for specific markers and mitochondrial

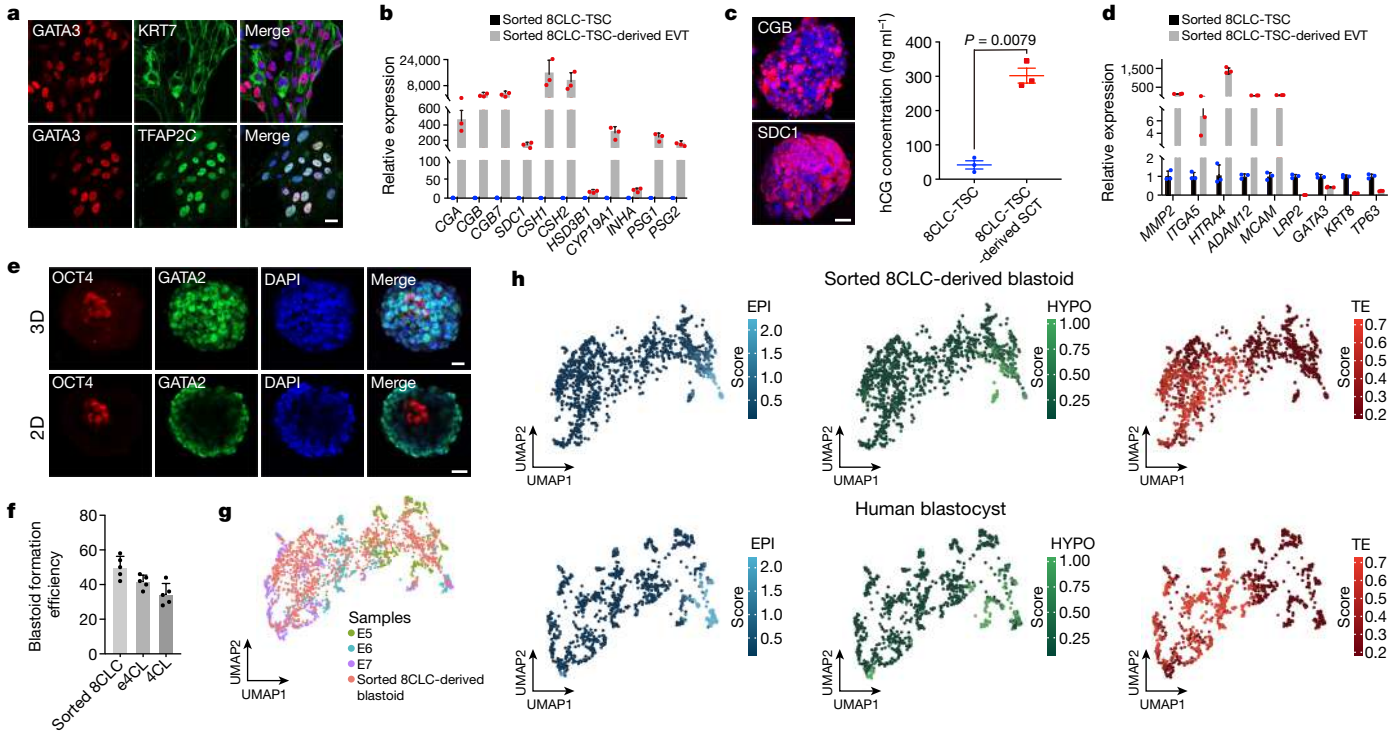


Fig. 4 | Generation of TSCs and blastoids from 8CLCs. a, Representative co-immunostaining images for GATA3 (red) and KRT7 or TFAP2C (green) in TSCs derived from sorted 8CLCs in stepwise e4CL-D5. Nuclei were counterstained with DAPI. Representative of four independent experiments. **b**, RT–qPCR for the indicated genes in SCTs generated from sorted 8CLC-derived TSCs. Data are presented as the mean ± s.d. of fold-change compared with sorted 8CLC-derived TSCs. $n = 3$ biological replicates. **c**, Left, representative immunostaining images for CGB and SDC1 in SCTs generated from sorted 8CLC-derived TSCs. Nuclei were counterstained with DAPI. Representative of three independent experiments. Right, ELISA assay detecting the concentration of hCG secreted from SCT culture medium. Data are presented as the mean ± s.d. $n = 3$ biological replicates. *P* value was calculated using two-tailed unpaired Student's *t*-test. **d**, RT–qPCR for the indicated genes in EVT generated from sorted 8CLC-derived TSC. Data are presented as the mean ± s.d. of fold-change compared with sorted 8CLC-derived TSCs. $n = 3$ biological replicates. **e**, Representative immunostaining images (top, combined sections; bottom, individual section) of a self-organized blastoid generated using sorted 8CLCs from stepwise e4CL-D5. OCT4 and GATA2 were used as ICM and trophectoderm markers, respectively. Nuclei were counterstained with DAPI (blue). Representative of three independent experiments. **f**, Quantification of blastoid formation efficiency using sorted 8CLCs, stepwise e4CL-D5 cells and 4CL-D12 naive ES cells. Data are presented as the mean ± s.d. $n = 5$ biological replicates. **g**, UMAP showing the integration of scRNA-seq data from sorted 8CLC-derived blastoids and E5–E7 blastocyst cells[20]. **h**, UMAP highlighting the average expression scores of gene signatures for different blastocyst lineages[5] in sorted 8CLC-derived blastoids (top) and human E5–E7 blastocysts[20] (bottom). EPI, epiblast; HYPO, hypoblast; TE, trophectoderm. Scale bars, 20 μm (**a**, **e**) or 50 μm (**c**).

PCR assay[57], which showed it constituted about 1% of the cells (Extended Data Fig. 20f–h). This value is comparable to the chimeric contribution of human PSCs bearing manipulation of cell survival pathways[57].

Teratoma formation is often used to evaluate the in vivo developmental potential of PSCs, especially human PSCs, and is also a good model for studying lineage specification[58]. Sorted 8CLCs, stepwise e4CL-day 5 cells and 4CL naive PSCs all formed teratomas at similar frequencies to primed PSCs, but their morphology was distinct (Extended Data Fig. 21a, b). Membranous and cystic teratomas were formed with sorted 8CLCs, stepwise e4CL-day 5 cells and 4CL naive PSCs compared to the more classical and solid consistency of primed PSC teratomas. Haematoxylin and eosin staining confirmed the presence of tissues derived from the three germ layers in all cases (Extended Data Fig. 21c). To dissect potential differences in cell types, we performed droplet-based scRNA-seq for these teratomas. A total of 227,598 cells passed quality control and were further analysed (Supplementary Table 7a). We integrated the data for all teratomas and annotated each of the cell clusters (Extended Data Fig. 21d) and identified a total of 24 cell types (Fig. 5a, b, Extended Data Fig. 21e and Supplementary Table 7b). Notably, cell composition in teratomas derived from sorted 8CLCs, stepwise e4CL-day 5 cells and 4CL naive PSCs was substantially different to that of teratomas derived from primed PSCs (Fig. 5b). Specifically, teratomas derived from sorted 8CLCs, stepwise e4CL-day 5 and 4CL naive PSCs contained extraembryonic trophoblast lineages and an enhanced number of other lineages, including haemogenic endothelium and immune cells (Fig. 5a–c, Extended Data Fig. 22a–e and Supplementary Table 7c). Closer inspection of the trophoblast cell population by subclustering and annotation revealed that sorted 8CLCs generated more differentiated villous cytotrophoblast sublineage than stepwise e4CL-day 5 cells or 4CL naive PSCs (Fig. 5d, e and Supplementary Table 7d). To our knowledge, substantial numbers of trophoblast cells have not been reported in teratomas formed with primed PSCs. Understanding the mechanisms of trophoblast differentiation in teratomas derived from sorted 8CLCs and 4CL naive PSCs may provide valuable insights into placenta formation in human embryonic development.

Collectively, our results show that 8CLCs have embryonic and extraembryonic developmental capacity, which makes them a valuable resource for investigating early human embryogenesis.

Discussion

Single-cell omics technologies are advancing our understanding of early human developmental processes, including ZGA[20,23,26,59], but the scarcity of embryos and relevant ethical considerations hinder progress. To overcome this caveat, we devised transgene-free, rapid and controllable culture conditions that generate 8CLCs from primed and naive PSCs that transcriptionally and epigenetically closely resemble the human 8C embryo. These 8CLCs can be purified with reporters such

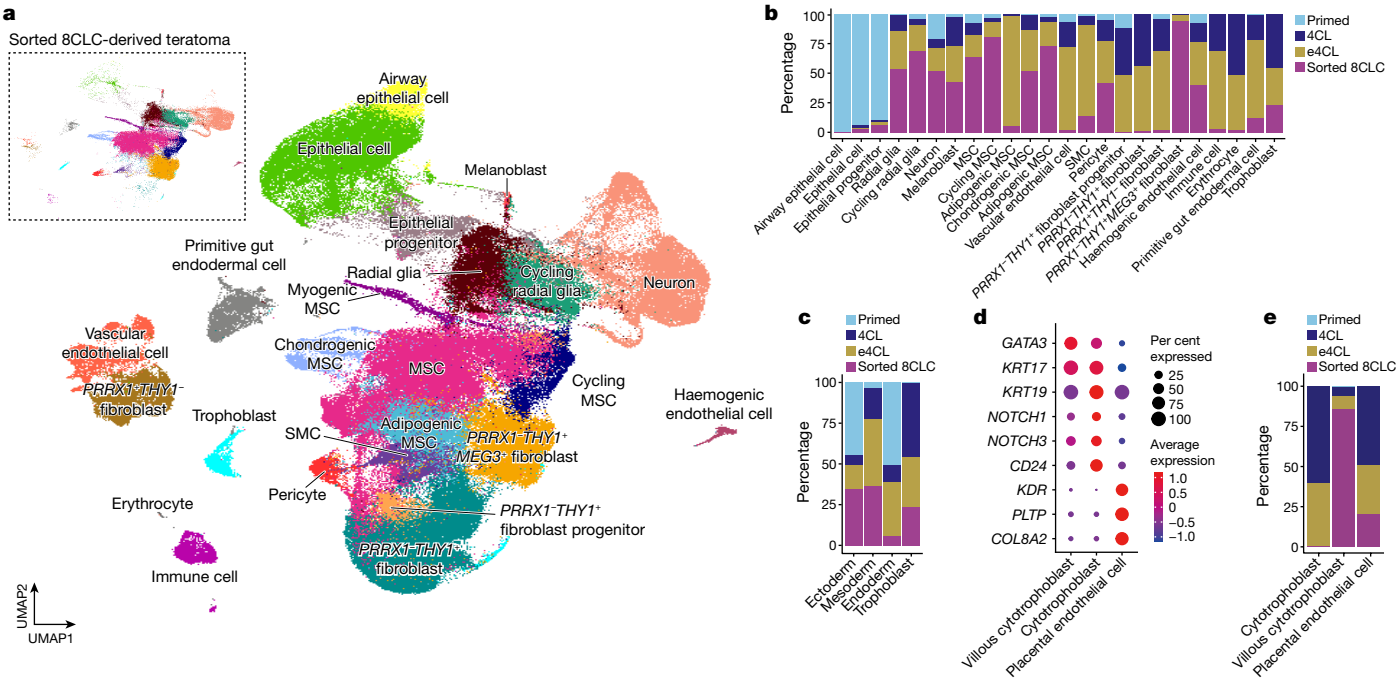


Fig. 5 | scRNA-seq of teratomas from 4CL naive PSCs and 8CLCs. a, UMAP of the identified cell types in scRNA-seq of teratomas derived from sorted 8CLCs, e4CL-D5 cells, 4CL naive ES cells and primed ES cells. H9 ES cells were used for producing all these teratomas. MSC, mesenchymal stem cell; SMC, smooth muscle cell. **b**, Bar plot showing the contribution of the different teratomas to the identified cell types. **c**, Bar plot showing the relative contribution of the different teratomas to each embryonic (ectoderm, mesoderm and endoderm) and extraembryonic (trophoblast) lineage. **d**, Bubble plot showing the frequency of expression and average expression of marker genes in cell subtypes of the extraembryonic trophoblast lineage. **e**, Bar plot showing the relative contribution of sorted 8CLCs, e4CL-D5 cells, 4CL naive ES cells and primed ES cell-derived teratomas to cell subtypes of the extraembryonic trophoblast lineage.

as the *TPRX1*–EGFP knock-in described here and used to systematically explore the molecular characteristics of the 8C human embryo. Naive PSCs generated using other methods[8,13,27,28] contained a small proportion of cells that displayed variable and generally low levels of totipotency genes. However, further study with high-resolution approaches, such as those used here, is necessary to determine the extent to which the rare putative 8CLCs produced using those methods are comparable to the 8C embryo, as well as their biological properties. Importantly, the addition of two key components of our medium, TSA and DZNep, potently and synergistically enhanced totipotency gene expression in other naive PSC cultures, which provides evidence that these inhibitors modulate critical pathways that induce the 8CLC state.

We also provide a molecular roadmap of the 8CLC conversion process and dissect the GRN controlling the 8CLC identity. This information could be useful to devise approaches that increase the efficiency of 8CLC generation, for example, through metabolic or signalling pathway perturbation. As proof of principle, we identified DPPA3 and TPRX1 as essential players in 8CLC generation. *TPRX1* was lost during divergent evolution in rodents[4], and it will be important to study whether other members of this family (for example, *ARGFX* and *DPRX*) that are absent in rodents are similarly required for 8CLC generation. Meanwhile, human DPPA3 only shares 35% identity and 53% similarity at the protein level with its mouse orthologue[60], which indicates that its function and/or regulation is not identical in the two species. In this regard, although combined TSA and DZNep also enhanced mouse 2CLC generation, DPPA3 is not necessary in this process. In addition to classical totipotency genes such as *TPRX1*, it will be informative to study the role of human 8C embryo and 8CLC-enriched lineage-specific factors such as *GATA6* or *PITX2*. These lineage-specific factors may participate in controlling the totipotent state or be present in preparation for blastocyst formation. Notably, the mouse 2C embryo is enriched in different lineage-specific factors compared to humans. This and other discrepancies provide a basis to study how mammalian embryogenesis is differentially controlled in a species-specific manner and suggest that evolutionary variation at the onset of development might be broader than anticipated.

We showed that 8CLCs can produce embryonic and extraembryonic lineages in vitro (TSCs and blastoids) and in vivo (teratomas). This property is shared with naive PSCs, which is consistent with the knowledge that the capacity to generate all blastocyst lineages extends beyond the time of ZGA[13]. However, given the transcriptional and epigenetic differences between the cell states, the differentiation processes are unlikely to be identical. Using cultured cells that more closely resemble the early human embryo is therefore potentially relevant to truly model the dynamics of developmental processes. This could facilitate the generation of optimal human blastoids and other embryonic structures, as well as functionally differentiated cells and tissues. Future studies using 8CLCs will explore these possibilities and expand our limited knowledge of human embryogenesis.

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/s41586-022-04625-0.

1. Macfarlan, T. S. et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).
2. De Iaco, A. et al. DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* **49**, 941–945 (2017).
3. Li, Y. et al. Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. *Nature* **564**, 136–140 (2018).
4. Maeso, I. et al. Evolutionary origin and functional divergence of totipotent cell homeobox genes in eutherian mammals. *BMC Biol.* **14**, 45 (2016).
5. Yanagida, A. et al. Naive stem cell blastocyst model captures human embryo lineage segregation. *Cell Stem Cell* **28**, 1014–1022.e4 (2021).

6. Manor, Y. S., Massarwa, R. & Hanna, J. H. Establishing the human naive pluripotent state. *Curr. Opin. Genet. Dev.* **34**, 35–45 (2015).
7. Rodriguez-Terrones, D. et al. A distinct metabolic state arises during the emergence of 2-cell-like cells. *EMBO Rep.* **21**, e48354 (2020).
8. Wang, Y. et al. Unique molecular events during reprogramming of human somatic cells to induced pluripotent stem cells (iPSCs) at naive state. *eLife* **7**, e29518 (2018).
9. Yang, Y. et al. Derivation of pluripotent stem cells with in vivo embryonic and extraembryonic potency. *Cell* **169**, 243–257.e25 (2017).
10. Yang, J. et al. Establishment of mouse expanded potential stem cells. *Nature* **550**, 393–397 (2017).
11. Gao, X. et al. Establishment of porcine and human expanded potential stem cells. *Nat. Cell Biol.* **21**, 687–699 (2019).
12. Posfai, E. et al. Evaluating totipotency using criteria of increasing stringency. *Nat. Cell Biol.* **23**, 49–60 (2021).
13. Guo, G. et al. Human naive epiblast cells possess unrestricted lineage potential. *Cell Stem Cell* **28**, 1040–1056.e6 (2021).
14. Esteban, M. A. et al. Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. *Cell Stem Cell* **6**, 71–79 (2010).
15. Miranda, T. B. et al. DZNep is a global histone methylation inhibitor that reactivates developmental genes not silenced by DNA methylation. *Mol. Cancer Ther.* **8**, 1579–1588 (2009).
16. Wu, J. et al. Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature* **557**, 256–260 (2018).
17. Theunissen, T. W. et al. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* **15**, 471–487 (2014).
18. Surani, M. A., Hayashi, K. & Hajkova, P. Genetic and epigenetic regulators of pluripotency. *Cell* **128**, 747–762 (2007).
19. Liu, C. et al. A portable and cost-effective microfluidic system for massively parallel single-cell transcriptome profiling. Preprint at https://doi.org/10.1101/818450 (2019).
20. Petropoulos, S. et al. Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* **165**, 1012–1026 (2016).
21. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
22. Shakiba, N. et al. CD24 tracks divergent pluripotent states in mouse and human cells. *Nat. Commun.* **6**, 7329 (2015).
23. Deng, Q., Ramskold, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196 (2014).
24. Goke, J. et al. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* **16**, 135–141 (2015).
25. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
26. Yan, L. et al. Single-cell RNA-seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1131–1139 (2013).
27. Yu, L. et al. Blastocyst-like structures generated from human pluripotent stem cells. *Nature* **591**, 620–626 (2021).
28. Liu, X. et al. Reprogramming roadmap reveals route to human induced trophoblast stem cells. *Nature* **586**, 101–107 (2020).
29. Gafni, O. et al. Derivation of novel human ground state naive pluripotent stem cells. *Nature* **504**, 282–286 (2013).
30. Bayerl, J. et al. Principles of signaling pathway modulation for enhancing human naive pluripotency induction. *Cell Stem Cell* **28**, 1549–1565.e12 (2021).
31. Zhang, M. et al. β-catenin safeguards the ground state of mouse pluripotency by strengthening the robustness of the transcriptional apparatus. *Sci. Adv.* **6**, eaba1593 (2020).
32. Guo, H. et al. The DNA methylation landscape of human early embryos. *Nature* **511**, 606–610 (2014).
33. Smith, Z. D. et al. DNA methylation dynamics of the human preimplantation embryo. *Nature* **511**, 611–615 (2014).
34. Pastor, W. A. et al. Naive human pluripotent cells feature a methylation landscape devoid of blastocyst or germline memory. *Cell Stem Cell* **18**, 323–329 (2016).
35. Di Stefano, B. et al. Reduced MEK inhibition preserves genomic stability in naive human embryonic stem cells. *Nat. Methods* **15**, 732–740 (2018).
36. Takashima, Y. et al. Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* **158**, 1254–1269 (2014).
37. Guo, G. et al. Epigenetic resetting of human pluripotency. *Development* **144**, 2748–2763 (2017).
38. Wang, Y. et al. Single-cell multiomics sequencing reveals the functional regulatory landscape of early embryos. *Nat. Commun.* **12**, 1247 (2021).
39. Eckersley-Maslin, M. A. et al. MERVL/Zscan4 network activation results in transient genome-wide DNA demethylation of mESCs. *Cell Rep.* **17**, 179–192 (2016).
40. Shen, H. et al. Mouse totipotent stem cells captured and maintained through spliceosomal repression. *Cell* **184**, 2843–2859.e20 (2021).
41. Stirparo, G. G. et al. Integrated analysis of single-cell embryo data yields a unified transcriptome signature for the human pre-implantation epiblast. *Development* **145**, dev158501 (2018).
42. Xia, W. et al. Resetting histone modifications during human parental-to-zygotic transition. *Science* **365**, 353–360 (2019).
43. Yeom, Y. I. et al. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881–894 (1996).
44. Pastor, W. A. et al. TFAP2C regulates transcription in human naive pluripotency by opening enhancers. *Nat. Cell Biol.* **20**, 553–564 (2018).
45. Liu, L. et al. An integrated chromatin accessibility and transcriptome landscape of human pre-implantation embryos. *Nat. Commun.* **10**, 364 (2019).
46. Macnair, W. & Claassen, M. psupertime: supervised pseudotime inference for single cell RNA-seq data with sequential labels. Preprint at https://doi.org/10.1101/622001 (2019).
47. Bredenkamp, N. et al. Wnt inhibition facilitates RNA-mediated reprogramming of human somatic cells to naive pluripotency. *Stem Cell Rep.* **13**, 1083–1098 (2019).
48. Morabito, S. et al. Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat. Genet.* **53**, 1143–1155 (2021).
49. Xu, X. et al. Dppa3 expression is critical for generation of fully reprogrammed iPS cells and maintenance of Dlk1–Dio3 imprinting. *Nat. Commun.* **6**, 6008 (2015).
50. Sang, H. et al. Dppa3 is critical for Lin28a-regulated ES cells naive-primed state conversion. *J. Mol. Cell. Biol.* **11**, 474–488 (2019).
51. Mulholland, C. B. et al. Recent evolution of a TET-controlled and DPPA3/STELLA-driven pathway of passive DNA demethylation in mammals. *Nat. Commun.* **11**, 5972 (2020).
52. Wang, L. et al. Overexpression of Stella improves the efficiency of nuclear transfer reprogramming. *J. Genet. Genomics* **44**, 363–366 (2017).
53. Hayashi, K., de Sousa Lopes, S. M. C., Tang, F. & Surani, M. A. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* **3**, 391–401 (2008).
54. Dong, C. et al. Derivation of trophoblast stem cells from naive human pluripotent stem cells. *eLife* **9**, e52504 (2020).
55. Castel, G. et al. Induction of human trophoblast stem cells from somatic cells and pluripotent stem cells. *Cell Rep.* **33**, 108419 (2020).
56. Velychko, S. et al. Excluding Oct4 from Yamanaka cocktail unleashes the developmental potential of iPSCs. *Cell Stem Cell* **25**, 737–753.e4 (2019).
57. Huang, K. et al. BMI1 enables interspecies chimerism with human pluripotent stem cells. *Nat. Commun.* **9**, 4649 (2018).
58. McDonald, D. et al. Defining the teratoma as a model for multi-lineage human development. *Cell* **183**, 1402–1419.e18 (2020).
59. Xue, Z. et al. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500**, 593–597 (2013).
60. Payer, B. et al. Stella is a maternal effect gene required for normal early development in mice. *Curr. Biol.* **13**, 2110–2117 (2003).
61. Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).

Methods

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

Animal study and ethics statement

All animal experiments were performed in accordance with the ethics guidelines of the Guangzhou Institutes of Biomedicine and Health. Human–mouse chimera and human blastoid experiments were approved and followed up by the Animal Care and Use Committee and Human Subject Research Ethics Committee under license numbers IACUC2016012 and GIBH-IRB2020-034, respectively, of the Guangzhou Institutes of Biomedicine and Health. These committees comprise experts in different fields (including scientists working on development and other disciplines, and non-scientists (doctors and lawyers)), and they evaluated the rationale of the experiment plan, origins and consent of human materials and the qualification of the investigators. Several of these experts followed the study until completion. Both experiments followed relevant international regulations, including the 2016 Guidelines for Stem Cell Research and Clinical Translation released by the International Society for Stem Cell Research (ISSCR). A specialized review process was activated in the human–mouse chimera experiments to assess whether the degree of functional integration was sufficiently high to raise concerns that the nature of the chimeric animal would be substantially changed. Human blastoid experiments were subjected to an embryo research oversight process. Several of these experts followed the study until completion. We have consent forms for all iPS cells generated in our laboratory. For the chimera experiments, we used the HN10-DsRed ES cell line, which were established by Hainan Medical University, China. We performed a *TPRX1*–EGFP knock-in into HN10-DsRed ES cells, as H9 ES cells cannot be used for interspecies chimeras. All human PSC lines were used anonymized.

Human primed PSC culture

Feeder cells derived from ICR mouse embryonic fibroblasts (mitomycin-treated) were maintained in DMEM (Corning, 10-013-CMR) supplemented with 10% FBS (Natocor, SFBE). Primed human ES cells, including H1, H9, WIBR3, HUES1, HUES7, HN10-DsRed[57], and primed human iPS cells, including UH10 (ref. [57]) (iPSC-1), Phoenix (iPSC-2), STiPS O-XX1 (iPSC-3) and DiPS 1016SevA (iPSC-4) were routinely cultured in mTeSR 1 medium (Stemcell Technologies, 85850) or TeSR-E8 medium (Stemcell Technologies, 05990) on plates coated with extracellular matrix (ECM; Geltrex (ThermoFisher, A1413302) or Matrigel (Corning, 354277)). Generally, primed PSCs were passaged every 4 days and medium was refreshed every day. For passaging, cells were washed with DPBS (Hyclone, SH30028.02) once and treated with 0.5 mM EDTA (Invitrogen, 15575020) for 5 min. Then, EDTA was removed, and cells were passaged as small clumps using a Pasteur pipette (Greiner bio-one, 612361). Primed human ES cells, including H9 and H1, were obtained from WiCell Research Institute, WIBR3 human ES cells were provided by J. Hanna, UH10 human iPS cells and HN10-DsRed human ES cells by G. Pan (Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, China), and Phoenix (iPSC-2) human iPS cells by U. Martin (Hannover Medical School, Germany). Primed human ES cells, including HUES1 and HUES7, and iPS cells, including DiPS 1016SevA (iPSC-1) and STiPS O-XX1 (iPSC-3), were obtained from the Howard Hughes Medical Institute. All cell lines were negative for mycoplasma.

Generation of human naive PSCs and 8CLCs

To generate 4CL naive PSCs, primed human PSCs were washed with DPBS once, dissociated into single cells with TrypLE and plated at a density of 1,000–1,500 cells cm^{-2} on feeders in mTeSR 1 medium supplemented with 10 μM Y-27632 (Axon, 1683). Twenty-four hours later, the medium was switched to 4CL medium composed of 1:1 mix of Neurobasal medium (Gibco, 21103049) and Advanced DMEM/F12 (Gibco, 12634028) supplemented with N2 (Gibco, 17502048) and B27 (Gibco, 17504044), sodium pyruvate (Corning, 25000CL), non-essential amino acids (Corning, 25025CL), GlutaMAX (Gibco, 35050061), penicillin–streptomycin (HyClone, SV30010), 10 nM DZNep (Selleck, S7120), 5 nM TSA (Vetec, V900931), 1 μM PD0325901 (Axon, 1408), 5 μM IWR-1 (Sigma, I0161), 20 ng ml^{-1} human LIF (Peprotech, 300-05), 20 ng ml^{-1} activin A (Peprotech, 120-14E), 50 μg ml^{-1} L-ascorbic acid (Sigma, A8960) and 0.2% (v/v) Geltrex or Matrigel. 4CL medium was refreshed every day, and cells were passaged as single cells (1:5 to 1:8) every 3–4 days; optionally, 5 μM Y-27632 was added in the medium for the first 24 h. To generate 8CLCs in stepwise e4CL, naive PSCs were dissociated into single cells and plated at 2,000–3,000 cells cm^{-2} on feeders in 4CL medium, replacing it with e4CL medium 24 h later. The composition of e4CL is the same as 4CL except for higher concentrations of DZNep (50 nM) and TSA (20 nM). e4CL medium was refreshed every day until day 5. To generate 8CLCs in direct e4CL, primed PSCs were dissociated into single cells and plated at 2,000–3,000 cells cm^{-2} on feeders in mTeSR 1 or TeSR-E8 medium supplemented with 10 μM Y-27632. Twenty-four hours later, medium was replaced with e4CL, which was refreshed every day. Cells were passaged as single cells at day 4 after conversion initiation. Naive PSCs and 8CLCs can also be generated on feeder-free ECM-coated plates. Cells in 4CL or e4CL were cultured in an incubator at 37 °C, 5% O_2 and 5% CO_2. To generate 8CLCs from other naive media (PXGL[47], 5iLA[17] and HENSM[30]), naive PSCs were induced using the respective protocols. Converted naive cells were dissociated into single cells and plated at 2,000–3,000 cells cm^{-2} on feeders in the respective naive medium. After 24 h, medium was switched to e4CL and refreshed every day until day 5.

Establishment of genetically engineered ES cells

Single guide RNA (sgRNA) inserts targeting critical *DPPA3* and *TPRX1* exons were cloned into lentiCRISPR-v2 plasmids (Addgene, 52961) for knockout generation. Primed H9 ES cells were transfected with sgRNA-coding plasmids using Lipofectamine 3000 (ThermoFisher, L3000015) following the manufacturer's instructions. For knock-in *TPRX1*–EGFP generation, sgRNA targeting the insertion site of the donor construct (containing homologous arms flanked by an EGFP–puromycin cassette) was cloned into lentiCRISPR-v2 plasmids. Primed H9 ES cells were electroporated using the donor construct and the sgRNA plasmids. All sgRNA sequences are listed in Supplementary Table 9. Forty-eight hours after transfection/electroporation, cells were treated with 3 μg ml^{-1} puromycin for 24–48 h for selection. The remaining cells were split into 96-wells using mTeSR 1 medium with 10 μM Y-27632. Y-27632 was withdrawn 24 h after splitting, and the medium was refreshed every day. Culture wells containing a single colony were selected and expanded, and DNA lysates of the derived cells were prepared by adding 20 μl lysis solution for blood (Sigma, L3289) and processed following the manufacturer's instructions. Amplicons covering the target regions (knockout and knock-in) were sequenced. For knockout, amplicons that had frameshift mutations were further confirmed by TA cloning and Sanger sequencing. All primers are listed in Supplementary Table 9.

Karyotype analysis

Karyotyping by G-banding was performed following standard protocols. In brief, cells were treated with 100 ng ml^{-1} Colcemid for 1 h for metaphase arrest. Arrested cells were trypsinized, pelleted and resuspended with 100 μl DPBS followed-by dropwise addition of 10 ml hypotonic solution (0.56 g KCl and 0.5 g sodium citrate in 200 ml H_2O). Cells were then fixed with a 3:1 mixture of methanol and acetic acid and dropped on slides. The slides were air dried and stained with Giemsa. A total of 20 randomly selected metaphase spreads were analysed for each sample.

Mouse ES cell culture and 2CLC generation

Mouse ES cells were cultured in serum+LIF medium (DMEM/high glucose containing 15% FBS (Biological Industries), GlutaMAX, penicillin–streptomycin, nonessential amino acids, sodium pyruvate, 2-mercaptoethanol and mouse LIF (1,000 U ml^{-1})) on feeders. To generate 2CLCs, cells were split onto 0.2% gelatin-precoated plates in serum+LIF medium. After 24 h, medium was switched to 2CLC induction medium (termed as 2CM) containing serum+LIF medium supplemented with DZNep (50 nM) and TSA (20 nM) and kept for 3 days. The MuERV-L-LTR-GFP ES cell reporter cell line was provided by M. Zhang (Guangzhou Laboratory, China).

Immunostaining, western blotting and FACS

For immunostaining, cells were fixed with 4% paraformaldehyde at room temperature (RT) for 10 min, then permeabilized with 0.1% Triton X-100 (Sigma, X100) for 30 min, and blocked with 3% BSA at RT for 1 h. Cells were incubated with primary antibodies at 4 °C overnight, followed by secondary antibodies at RT for 1 h. Antibodies were diluted in 3% BSA blocking buffer. DAPI (Sigma, D9542) was used for counterstaining the nuclei. Images were taken using a Zeiss LSM 710 NLO microscope. Western blot analysis was performed by isolating whole-cell protein extracts with conventional RIPA lysis buffer, and the extracts were quantified using a BCA kit (Thermo Fisher Scientific). Blots were incubated with specific antibodies. Uncropped scans of all western blots with molecular weight markers are shown in Supplementary Fig. 2. All antibodies used in this study are described in Supplementary Table 10. For flow cytometry, cells were trypsinized, washed twice with DPBS and filtered through a Falcon 40-μm cell strainer (Falcon, 352340). Cells were resuspended in 0.04% BSA in PBS. They were loaded into a BD LSRFortessa X-20 cell analyser. All results were processed using FlowJo (v.10.4) software. For cell sorting using *TPRX1*–EGFP reporter cells grown in stepwise e4CL-day 5, cells were trypsinized, resuspended in 0.04% BSA and loaded into a BD FACSAria III for sorting (BD FACSDiva 8.0.2). The FACS gating strategy is shown in Supplementary Fig. 3.

Derivation and differentiation of human TSCs

Six-well plates were coated with 5 μg ml^{-1} mouse collagen IV (Corning, 354233) or Geltrex following the manufacturer's instructions. 4CL-day 12 PSCs, stepwise e4CL-day 5 cells or sorted 8CLCs were seeded at a density of 50,000 cells per well in human TSC medium[62] with minor modifications. Modified TSC medium consisted of DMEM/F12 supplemented with 0.1 mM 2-mercaptoethanol, 0.2% FBS, penicillin–streptomycin, 0.3% BSA, N2, 1.5 μg ml^{-1} L-ascorbic acid, 50 ng ml^{-1} EGF, 2 μM CHIR99021, 0.5 μM A83-01, 1 μM SB431542, 0.8 mM valproic acid and 5 μM Y27632. Medium was refreshed every other day. For routine culture, cells reaching around 70% confluency were dissociated with TrypLE for 5–10 min at 37 °C and passaged to a new collagen IV-coated or Geltrex-coated plate at a 1:4–1:10 split ratio. TSCs at passages 5–20 were used for analysis. Differentiation into SCTs was done in suspension. In brief, TSCs were plated at 100,000 cells per well into AggreWell 800 wells (StemCell, 34811) using SCT medium (Advanced DMEM/F12 supplemented with GlutaMAX, 0.1 mM 2-mercaptoethanol, penicillin–streptomycin, N2, B27, 50 ng ml^{-1} human EGF, 2 μM CHIR99021, 0.8 mM valproic acid and 5 μM Y27632). Medium was refreshed every 2–3 days. At day 10, medium was refreshed, and 48 h later it was collected for measuring secreted hCG by ELISA (Abcam, ab100533). At the same time, aggregates of SCTs were collected for immunostaining. For EVT differentiation, TSCs were plated at 300,000 cells per well on Matrigel-coated 12-well plates using EVT medium (Advanced DMEM/F12 supplemented with GlutaMAX, 0.1 mM 2-mercaptoethanol, penicillin–streptomycin, N2, B27 and 5 μM Y27632). Medium was refreshed every day. After 11 days, cells were collected for detecting HLA-G by flow cytometry. Individual analysis of the methylation status of the *ELF5* promoter in TSCs was measured as previously described[63].

Self-organized blastoid formation

For blastoid formation, we used a previously described protocol[5] with minor modifications. In brief, dissociated single cells were washed and pelleted with DMEM/F12 with 0.1% BSA by centrifugation. The pellet was resuspended in N2B27 medium supplemented with PD0325901 (1.5 μM), A83-01 (1 μM) and Y-27632 (10 μM), and cells were seeded at a density of 25 cells per microwell into AggreWell 800 wells (StemCell, 34811) according to the manufacturer's instructions. After 2 days, the medium was switched to N2B27 medium supplemented only with A83-01 (0.5 μM) and maintained for another 3 days. At day 5, blastoids were collected for downstream analysis. Cultures were maintained in an incubator at 37 °C and 5% CO_2 for the entire process.

Teratoma formation

For teratoma formation, 1 million cells were prepared in suspension with a 200 μl prechilled 1:1 mixture of DMEM/F12 and Matrigel. The cell suspension was subcutaneously injected into 6–8-week-old male NOD-*scid IL2Rg*$^{-/-}$ mice. Teratomas were formed in 4–8 weeks. Mice were killed and tumours collected and either used directly for scRNA-seq or fixed with 4% paraformaldehyde and embedded in paraffin. Paraffin sections and haematoxylin and eosin staining were performed following standard protocols.

Interspecies chimera assay

Primed ES cell HN10 with DsRed inserted into the AAVS1 locus (HN10-DsRed)[57] converted by 4CL and e4CL were used. For the 8CLC chimera experiment, we knocked-in EGFP under the *TPRX1* locus in HN10-DsRed ES cells. GFP$^+$ cells were manually picked and purified at day 5 of stepwise e4CL induction. Chimeras were generated by aggregation of 4CL PSCs, e4CL cells or purified 8CLCs with BDF1 mouse 8C embryos as previously described[56]. The aggregated embryos were developed into early blastocysts in vitro, and phase contrast and fluorescence of embryos were imaged using a Nikon Eclipse Ti microscope. Chimeric embryos were then transferred into the uterus of pseudo-pregnant ICR mice and left to develop until E10.5. Conceptuses at E10.5 were collected. Embryos and placentas were fixed with 4% paraformaldehyde, dehydrated with 30% sucrose solution overnight, embedded with SDS Tissue-Tek OCT compound (Sakura Finetek, 4583) and cryosectioned at 10 μm. Sections were processed for immunostaining.

Quantitative mitochondria DNA PCR assay

DNA was extracted from mouse E10.5 fetus, placenta and yolk sac and used for quantitative PCR. Primer sequences are listed in Supplementary Table 9.

shRNA transduction, siRNA transfection, RNA isolation, RT–qPCR and RNA-seq library preparation

HEK293T cells were cultured in the same medium as mouse feeders. For shRNA experiments, shRNA inserts were cloned into pLKO.1. Lentiviral shRNA particles were generated with HEK293T cells using Lipofectamine 3000 (Invitrogen, L3000015) following the manufacturer's instructions. Virus supernatants were collected 48 h after transfection. siRNA oligonucleotides were purchased from GeneWiz. siRNA-mediated knockdown experiments were performed using Lipofectamine 3000 according to the manufacturer's instructions. All shRNA and siRNA target sequences are listed in Supplementary Table 9. Total RNA was prepared using RNAzol RT (MRC, RN190). RT–qPCR was performed using SYBR Green (Takara, RR420A) and an ABI7500 real-time PCR machine. Data were analysed based on the $2^{\Delta\Delta Ct}$ method. RT–qPCR primer sequences are listed in Supplementary Table 9. Bulk RNA-seq library preparation and sequencing were performed by Berry Genomics.

Library preparation for RRBS and low-input WGBS

For RRBS library preparation, 1 μg of genomic DNA was digested with 300 U of MspI (NEB, R0106) in 100 μl at 37 °C for 16 h. After purification, the digested products were blunt-ended, and then dA was added, followed by methylated-adapter ligation. To obtain DNA fractions of promoter and CGI regions from MspI-digested genomic DNA, 160–500 bp adapter-ligated fractions were excised from a 2% agarose gel. Bisulfite conversion was conducted using an EZ DNA Methylation-Gold kit (Zymo, D5008) following the manufacturer's instructions. The final libraries were generated by PCR amplification using JumpStart Taq DNA polymerase (Sigma, D9307), analysed using an Agilent 2100 Bioanalyzer (Agilent Technologies) and quantified by real-time PCR. Libraries were sequenced using an Illumina Xten platform at E-GENE. For low-input WGBS, genomic DNA was extracted using a QIAamp DNA Micro kit (Qiagen) and quantified using a Quant-iT dsDNA HS Assay kit with an Invitrogen Qubit fluorometer according to the manufacturer's protocols. Low-input DNA WGBS libraries were constructed as previously described[64]. In brief, bisulfite conversion was conducted using an EZ DNA Methylation-Gold kit (Zymo, D5008) following the manufacturer's instructions. The purified DNA was eluted in 10 mM Tris-HCl (pH 8.5) buffer. Post bisulfite extension was carried out with 0.4 mM dNTPs, 0.4 μM oligo 1 (Biotin-CTACACGACGCTCTTCCGATCTNN NNNNNNN) in a final volume of 24 μl. After incubation at 65 °C for 3 min and paused at 4 °C, 50 U of Klenow exo− (Enzymatics) was added and the samples were incubated at 4 °C for 5 min, +1 °C/15 s to 37 °C and kept at 37 °C for 30 min. After the first-strand synthesis, 40 U exonuclease I (NEB, M0568S) was added and incubated for 1 h at 37 °C followed by DNA purification using streptavidin beads. The second-strand synthesis was carried out with 0.8 mM dNTPs and 4 μM oligo 2 (TGCTGAACCGCTCTTCCGAT CTNNNNNNNNN). The final libraries were generated by PCR amplification using KAPA HiFi HotStart DNA polymerase (KAPA Biosystems). The resulting libraries were analysed using an Agilent 2100 Bioanalyzer (Agilent Technologies) and quantified by real-time PCR and were then sequenced in an Illumina Nova 6000 sequencer at E-GENE.

Library preparation for droplet-based scRNA-seq

scRNA-seq libraries were prepared using a DNBelab C Series Single-Cell Library Prep set (MGI, 1000021082)[19]. In brief, the single-cell suspensions were converted to barcoded scRNA-seq libraries through droplet encapsulation, emulsion breakage, mRNA captured bead collection, reverse transcription, cDNA amplification and purification. Indexed sequencing libraries were constructed according to the manufacturer's protocol. The sequencing library concentration was quantified using a Qubit ssDNA Assay kit (Thermo Fisher Scientific, Q10212). The resulting libraries were sequenced using a DIPSEQ T1 sequencer at China National GeneBank (CNGB).

Library preparation for scATAC-seq

scATAC-seq libraries were prepared using a DNBelab C Series Single-Cell ATAC Library Prep set (MGI, 1000021878)[65]. In brief, transposed single-nucleus suspensions were converted to barcoded scATAC-seq libraries through droplet encapsulation, pre-amplification, emulsion breakage, beads collection, DNA amplification and purification. Indexed sequencing libraries were constructed according to the manufacturer's protocol. Library concentrations were quantified using a Qubit ssDNA Assay kit and sequenced using a BGISEQ-500 sequencer at CNGB.

Library preparation for SMART-seq2

SMART-seq2 was performed according to a modified 96-well-plate-based protocol[25]. In brief, single cells were loaded into each well and subjected to cell lysis, reverse transcription and cDNA amplification. The resulting cDNA libraries were further labelled with unique barcodes and sequenced using a BGISEQ-500 sequencer at CNGB.

Single-cell multiomics library preparation

H9 ES cells cultured in stepwise e4CL-day 5 were trypsinized and passed through a 30-μm filter and centrifuged at 300*g* for 5 min at 4 °C. Libraries were prepared using a Chromium Next GEM Single Cell Multiome ATAC + Gene Expression kit (10X Genomics, PN-1000283) according to the manufacturer's instructions. They were sequenced using a BGISEQ-500 sequencer at CNGB with the following strategy: scRNA-seq read 1, 28 bp; scRNA-seq read 2, 100 bp; scATAC-seq read 1, 50 bp; scATAC-seq read 2, 66 bp.

Bulk RNA-seq data analysis

Reads were trimmed of adapters and for low-quality reads using Trim Galore (v.0.5.0) (https://github.com/FelixKrueger/TrimGalore) coupled with cutadapt (v.1.14) and FastQC (v.0.11.2) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) or fastp (v.0.21.0) (https://github.com/OpenGene/fastp) with −q 20 option. Gene counts were calculated using RSEM (v1.3.1)[66] using rsem-calculate-expression with --star, --estimate-rspd, --output-genome-bam and --paired-end options, STAR (v.2.7.3a)[67]. Reads were aligned to hg38 assembly indexed with v.81 Ensembl GTF. Pairwise differential expression, PCA plots and MA plots were generated using DESeq2 (v.1.24)[68] in R (v.3.6.0). BigWig files were generated from RSEM BAM output files using deeptools (v.3.1.3)[69] bamCoverage tool with and --normalizeUsing CPM options. When comparing with published data, counts were normalized using the withinLaneNormalization function in the EDASeq R package (v.2.18.0) before being processed in DESeq2.

DNA methylation data analysis

Data were aligned to the hg38 genome using BS-Seeker2 (v.2.1.0)[70] with default options. CpG sites with read depth < 5 were filtered. The resultant wig file was converted to a bigwig file using wigToBigWig (v.4) and visualized in IGV (v.2.1.0). CGmapTools (v.0.1.2) mbin function was used to calculate the overall CpG methylation level across the entire genome. Specific loci were defined as overlapping 100-bp-tile regions at the 8C embryo, ICM and postimplantation embryo RRBS data. The CGmapTools mtr function was used to generate region-specific CpG methylation ratios.

scRNA-seq data analysis

Raw sequencing reads from DIPSEQ-T1 were filtered and demultiplexed using PISA (v.0.2) (https://github.com/shiquan/PISA). Reads were aligned to genome using STAR (v.2.7.4a) and sorted by sambamba (v.0.7.0). Cell versus gene unique molecule identifier (UMI) count matrices were generated by custom script. Count matrices were processed using the Seurat package (v.3.1.4)[21] in R (v.3.6.3). Data were also re-analysed from ArrayExpress: E-MTAB-3929 (ref. [20]) and GSE36552 (ref. [26]). Outlier cells were filtered out based on the distribution of the counts of RNA features and expression of mitochondrial genes. For our data, cells with fewer than 200 genes detected, greater than 12,000 genes detected or more than 10% mitochondrial reads were removed. For SMART-seq2 data, cells with fewer than 1,500,000 counts were filtered out from downstream analysis. For DNBelab C4 data of ES cells, cells with fewer than 200 genes detected, greater than 12,000 genes detected or more than 10% mitochondrial reads were filtered out. For DNBelab C4 data of teratomas, cells with fewer than 3,000 UMI detected or more than 10% mitochondrial reads were filtered out. Data were normalized, and variable features were identified. To integrate data from E-MTAB-3929 and GSE36552 and our data, the FindIntegrationAnchors function was used followed by the IntegrateData function. Merged data were scaled, PCA was run and then we performed UMAP using the first 20 principal components. Cells were clustered using Louvain at resolution 0.5, and plots were generated using DimPlot, DotPlot or custom stacked violin functions. scRNA-seq data of stepwise e4CL-day 5 or direct e4CL-day 7 cells were clustered in an unbiased manner, and the 8CLC cluster was defined by the following 8C embryo signature:

high levels of *ZSCAN4*, *ZSCAN5A*, *ZSCAN5B*, *TPRX1*, *ARGFX*, *DPRX*, *DUXA*, *MDB3L2*, *MBD3L3*, *SLC34A2*, *TUBA3D*, *DPPA3*, *KLF17*, *KDM4E*, *CCNA1*, *ZNF280A*, *RFPL2*, *MFSD2A* and *TRIM43*; and low levels of *OCT4* and *SOX2*. FindMarkersAll function was used to identify DEGs. For pseudotime analysis, the Seurat data object was converted to SingleCellExperiment class and then psupertime (v.0.2.6, https://www.biorxiv.org/content/10.1101/622001v1) was used to align the data along a temporal trajectory with the psupertime function using hvg_cutoff=0.01, bio_cutoff=0.01 and span=0.1 parameters. Individual genes were plotted using the plot_specific_genes_over_psupertime function. Next, the data were clustered using the psupertime_go_analysis function and the profiles were plotted using the plot_profiles_of_gene_clusters function.

Gene network analysis

scWGCNA (v.0.1.0) was used according to a previous report[48]. In brief, the top 20,000 genes from the three cell stages were used for downstream analysis. The same number of primed PSCs, 4CL PSCs and 8CLCs were used as input to construct the co-expression network. Gene modules were calculated with minModuleSize = 50. Modules related to each state were assigned by the enrichment of module eigengenes. Weighted gene correlation was calculated as Pearson's correlation to the power of 6 for each gene pair. Gene pairs with weighted correlation lower than 0.15 were filtered to define the target genes for each hub gene. Target genes for each hub gene were again filtered according to the DEGs of scRNA-seq data of the embryo[26] at the corresponding developmental stage. The resulting GRNs were visualized using Cytoscape (v.3.9.0).

scATAC-seq data analysis

Fragment files were loaded into ArchR (v.0.9.5)[61] using createArrowFiles with filterTSS=4, filterFrags=1000, addTileMat=T, addGeneScoreMat=T. Cells with a median fragment size of 9,429 were filtered. An ArchRProject was instantiated and then latent semantic indexing (LSI) reduction was used before clustering and calculating UMAP coordinates. For trajectory analysis, the addTrajectory function was used before plotting heatmaps over the pseudotime with plotTrajectoryHeatmap. Clusters of interest were selected as starting and end points. The greyed-out cells were deemed as unimportant by the software to generating the trajectory. The gene score is calculated by ArchR and is defined as the depth normalized summed accessibility across all tiles within the gene. To generate pseudobulk plots for certain groups, the getGroupBW function was used to make bigwig files. To calculate motif deviation scores the functions addGroupCoverages, addMotifAnnotations, addBgdPeaks and addDeviationsMatrix functions were used with default settings. UMAP plots were visualized using the plotEmbedding function. Chromatin peak co-accessibility was performed using the R package ArchR. To calculate co-accessibility between two peaks across single cells, we used addCoAccessibility function in ArchR packages with default options. Then, plotBrowserTrack function in ArchR was used to plot browser tracks of co-accessibility, corCutoff was set in different genes to filter the co-accessibility peaks with relative high correlation for visualization.

Single-cell multiomics data analysis

Single-cell multiomics data were first aligned to hg38 genome by Cell Ranger Arc (v.2.0.0) (https://support.10xgenomics.com/single-cell-multiome-atac-gex/software). The resulting data were analysed in Seurat[21] and Signac[71]. First, cells were filtered with the following criteria: UMI in scRNA-seq > 5,000, number of fragments in scATAC-seq > 5,000 and mitochondria genes higher than 15% in scRNA-seq. For scRNA-seq, gene expression was normalized by SCTransform and visualized by UMAP with dims = 1:15. Clusters were identified with the options "dims = 1:6, k.param = 25, resolution = 0.5". For scATAC-seq, peaks detected in more than 100 cells as the VaribaleFeature. Data were normalized using the RunTFIDF function and identified the top features using the FindTopFeatures with min.cutoff = 'q1'. The LSI space were built by the RunSVD function and visualized by the UMAP with dims = 2:50. Clusters were identified using the options "dims = 1:50, k.param = 20, resolution = 1.5. We also identified the clusters based on the joint embeded scRNA-seq and scATAC-seq of multiomics data. In brief, a WNN graph was constructed with the options "reduction.list = list("pca", "lsi"), dims.list = list(2:50, 2:50), k.nn = 10" using the FindMultiModalNeighbors function, the joint clusters were identified using the FindClusters with the options "resolution = 0.8" based on the WNN graph. The 8CLC cluster was defined as in the scRNA-seq analysis described above.

Statistics and reproducibility

Statistical analysis was done with GraphPad Prism (v.8.3.1). Data of bar charts are represented as the mean ± s.d. or s.e.m. *P* values were calculated using two-tailed unpaired Student's *t*-test. The number of replicates for each experiment is presented in the figure legends.

Reporting summary

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

Data availability

All data needed to evaluate the conclusions in the paper are included in the paper and/or the supplementary materials. Raw sequencing data have been deposited in the CNGB Nucleotide Sequence Archive under the accession number CNP0001454. Reference datasets from other works can be accessed either in the NCBI Gene Expression Omnibus (GEO) repository under their GSE numbers or in the database of the EMBL's European Bioinformatics Institute under their E-MTAB numbers: accession numbers of human embryo scRNA-seq data are E-MTAB-3929 and GSE36552; the accession number of human embryo ATAC-seq data is GSE101571; the accession number of human embryo histone CUT&RUN data is GSE124718; accession numbers of other naive PSC/EPSC media single-cell sequencing data are GSE150311 (RSeT and t2iLGö), GSE166422 (PXGL), GSE150578 (5iLA); accession numbers of other naive/EPSC media RNA-seq data are GSE52617 (NHSM), E-MTAB-2857 (t2iLGö), GSE59435 (5iLAF), E-MTAB-7254 (EPSC); accession numbers of DNA methylation data are GSE49828 (human early embryo), GSE52617 (NHSM), GSE60945 (t2iLGö-NK2), GSE111018 (5iLAF), GSE136715 (mouse early embryo) and GSE75751 (mouse ES cells and 2CLCs); the accession numbers of mouse early embryo scRNA-seq data is GSE45719; the accession number of mouse ES cells and 2CLCs is GSE168728; and the accession number of TSC RNA-seq data is GSE138762. The human reference genome is available at http://ftp.ensembl.org/pub/release-105/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna_rm.alt.fa.gz. All reference datasets with links are summarized in Supplementary Table 8. Processed datasets with links and descriptions are summarized in Supplementary Table 11. Processed bulk RNA-seq data can be accessed at https://figshare.com/s/8ee132ff1366fa89d35a, processed scRNA-seq data can be accessed at https://figshare.com/s/34110eebb58462a79dd5, processed scATAC-seq data can be accessed at https://figshare.com/s/760d3ff54f1214a50cc2, processed RRBS and WGBS data can be accessed at https://figshare.com/s/ff707bf8242f7b3ed8f5, processed teratoma and blastoid scRNA-seq data can be accessed at https://figshare.com/s/037b348b1da763fb41d0, processed SMART-seq2 data can be accessed at https://figshare.com/s/a1b03a1463865b8a56c8, processed single-cell multiomics data can be accessed at https://figshare.com/s/9c01c3b58d34b80de230. Source data are provided with this paper.

Code availability

The analysis pipelines for scRNA-seq data generated by DNBeLab C4 are implemented using an in-house workflow (https://github.com/MGI-tech-bioinformatics/DNBelab_C_Series_HT_scRNA-analysis-software).

62. Okae, H. et al. Derivation of human trophoblast stem cells. *Cell Stem Cell* **22**, 50–63.e6 (2018).
63. Lee, C. Q. et al. What is trophoblast? A combination of criteria define human first-trimester trophoblast. *Stem Cell Rep.* **6**, 257–272 (2016).
64. Smallwood, S. A. et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* **11**, 817–820 (2014).
65. Yu, Y. et al. Single-nucleus chromatin accessibility landscape reveals diversity in regulatory regions across distinct adult rat cortex. *Front. Mol. Neurosci.* **14**, 651355 (2021).
66. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
67. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
68. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
69. Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
70. Guo, W. et al. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics* **14**, 774 (2013).
71. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).

Acknowledgements We thank all members of our laboratories for their support; L. Lai and G. Pan (Guangzhou Institutes of Biomedicine and Health) for their helpful comments during the execution of this project; X. Quan (the Experimental Animal Center of Guangzhou Institutes of Biomedicine and Health) for technical help with the chimera experiments; staff at the CNGB for providing technical support; and staff at the instrument platforms of the Guangzhou Institutes of Biomedicine and Health and the Bioland Laboratory for their technical support. This work was supported by the National Key Research and Development Program of China (2018YFA0106903 and 2016YFA0100102 to M.A.E.), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA16030502 to M.A.E.), the National Natural Science Foundation of China (U20A2015 to M.A.E., 31900466 to L.L., 32150410348 to M.A.M., 32070861 to B.Q. and 31950410553 to C.W.), the Guangdong Provincial Key Laboratory of Genome Read and Write (2017B030301011 to X.X.), the Guangdong Basic and Applied Basic Research Foundation (2021A1515110180 to Y.L.), and the Guangzhou Science and Technology Foundation (2021000122 to W.L.). M.A.M. was funded by CAS-TWAS President's PhD Fellowship and CAS President's International Fellowship Initiative (PIFI) for special experts (2020FSB0002).

Author contributions M.A.E., M.A.M. and W.L. conceived the original idea and designed the experiments. M.A.E., M.A.M., W.L. and L.L. supervised the study. M.A.M., W.L., Z.L., C.L. and Y.L. conducted most of the experiments (with help from W.J., Y.J., H.L., L.F., Y. Yang, D.P.I., J. Lai, P.G., Y. Yuan, Q.D., Y.W., Y. Liu, J.W. and G.W.). C.W., Y. Lai, L.W., J. Li, W.J., X.W. and J.A. performed bioinformatics analysis. M.A.E., M.A.M. and W.L. interpreted the data. F.G., S.Z., B.Q., G.W., P.H.M. and X.X. provided relevant advice regarding data interpretation and manuscript preparation. M.A.E. provided most of the financial support and L.L. provided essential materials and infrastructural support for the single-cell technologies. M.A.M., W.L. and M.A.E. wrote the manuscript with input from all authors. All authors approved the final version of the manuscript.

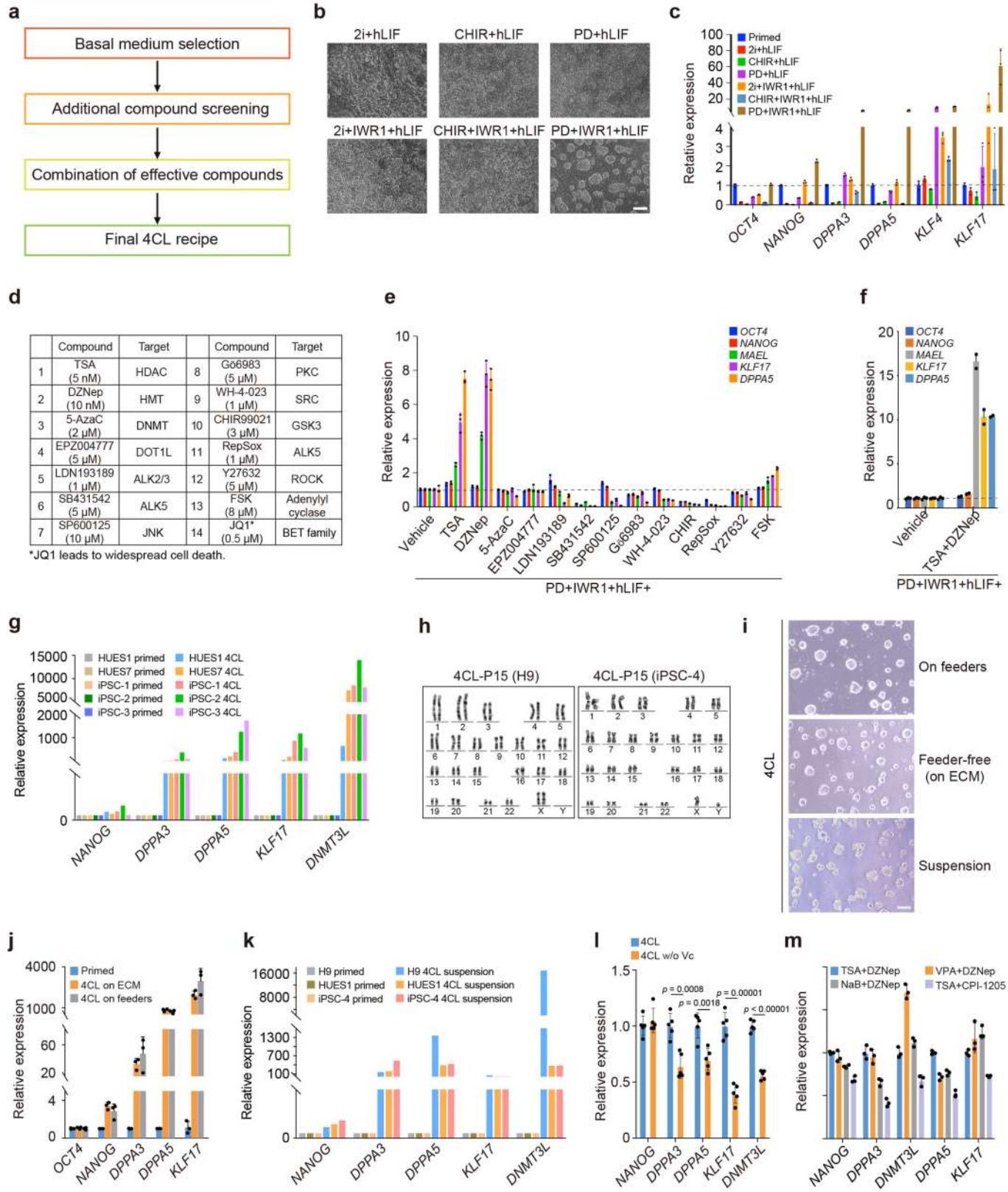
Competing interests The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-04625-0.

Correspondence and requests for materials should be addressed to Md. Abdul Mazid, Longqi Liu, Wenjuan Li or Miguel A. Esteban.

Peer review information *Nature* thanks Janet Rossant and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

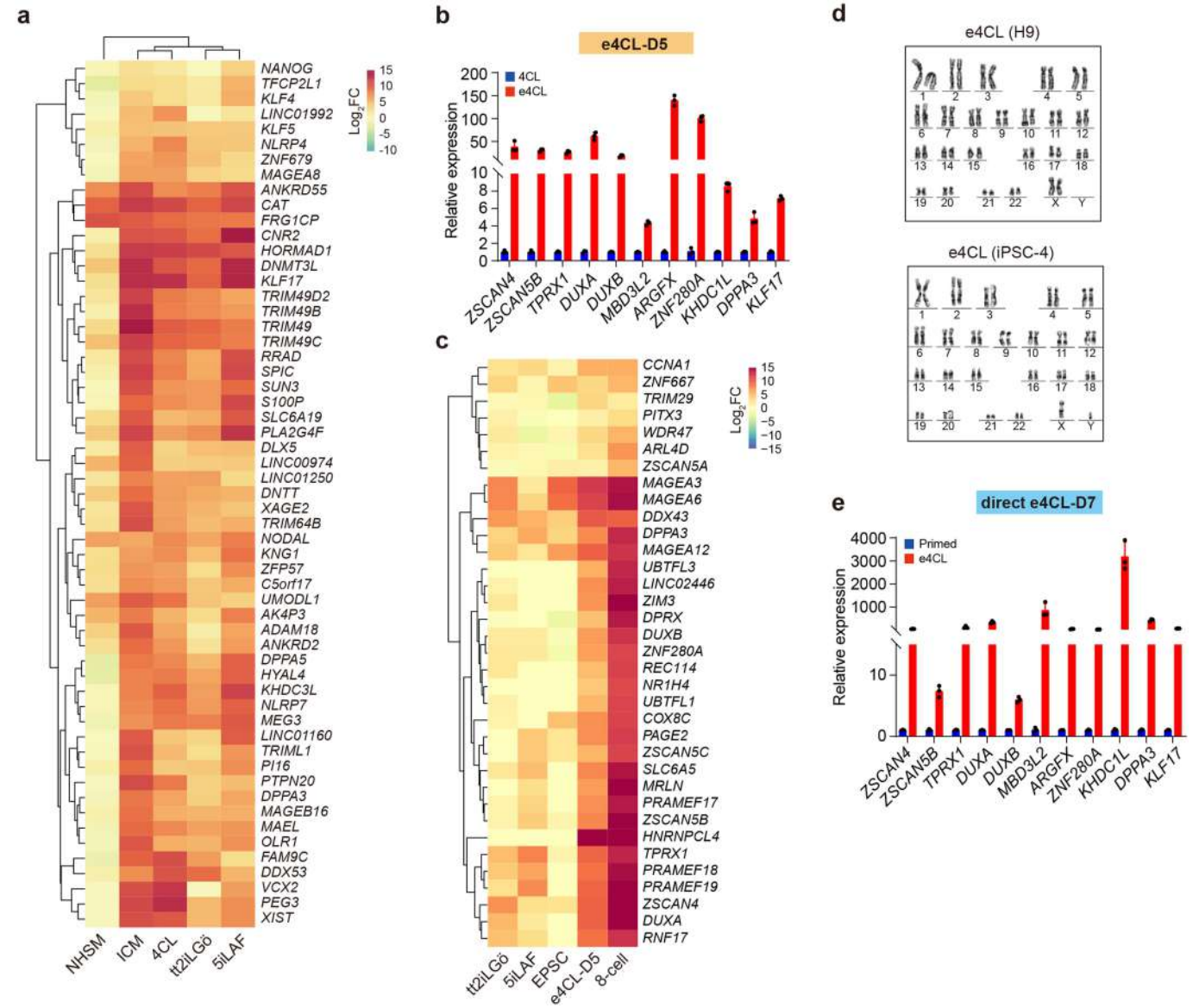


	Compound	Target		Compound	Target
1	TSA (5 nM)	HDAC	8	Gö6983 (5 μM)	PKC
2	DZNep (10 nM)	HMT	9	WH-4-023 (1 μM)	SRC
3	5-AzaC (2 μM)	DNMT	10	CHIR99021 (3 μM)	GSK3
4	EPZ004777 (5 μM)	DOT1L	11	RepSox (1 μM)	ALK5
5	LDN193189 (1 μM)	ALK2/3	12	Y27632 (5 μM)	ROCK
6	SB431542 (5 μM)	ALK5	13	FSK (8 μM)	Adenylyl cyclase
7	SP600125 (10 μM)	JNK	14	JQ1* (0.5 μM)	BET family

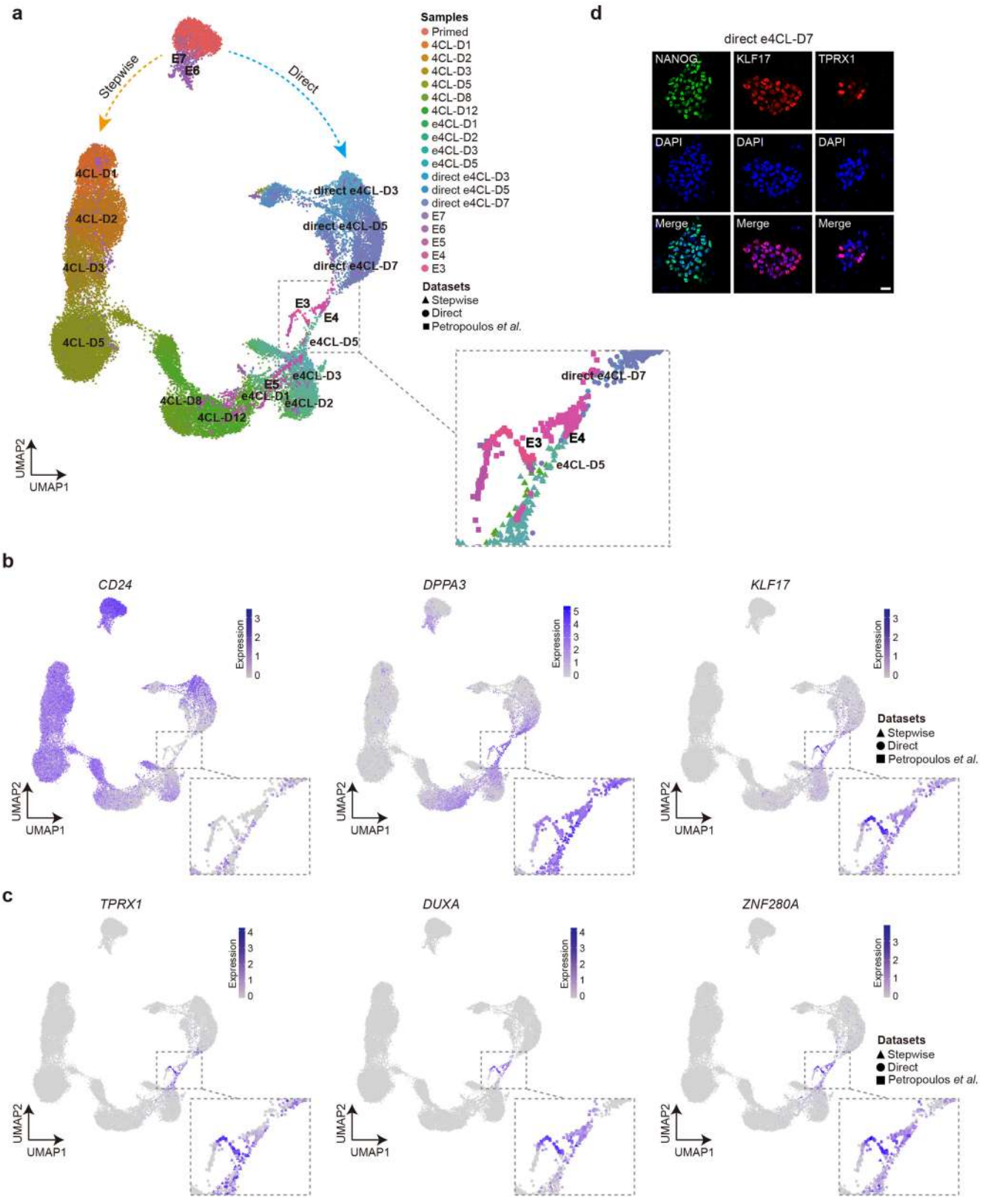
*JQ1 leads to widespread cell death.

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

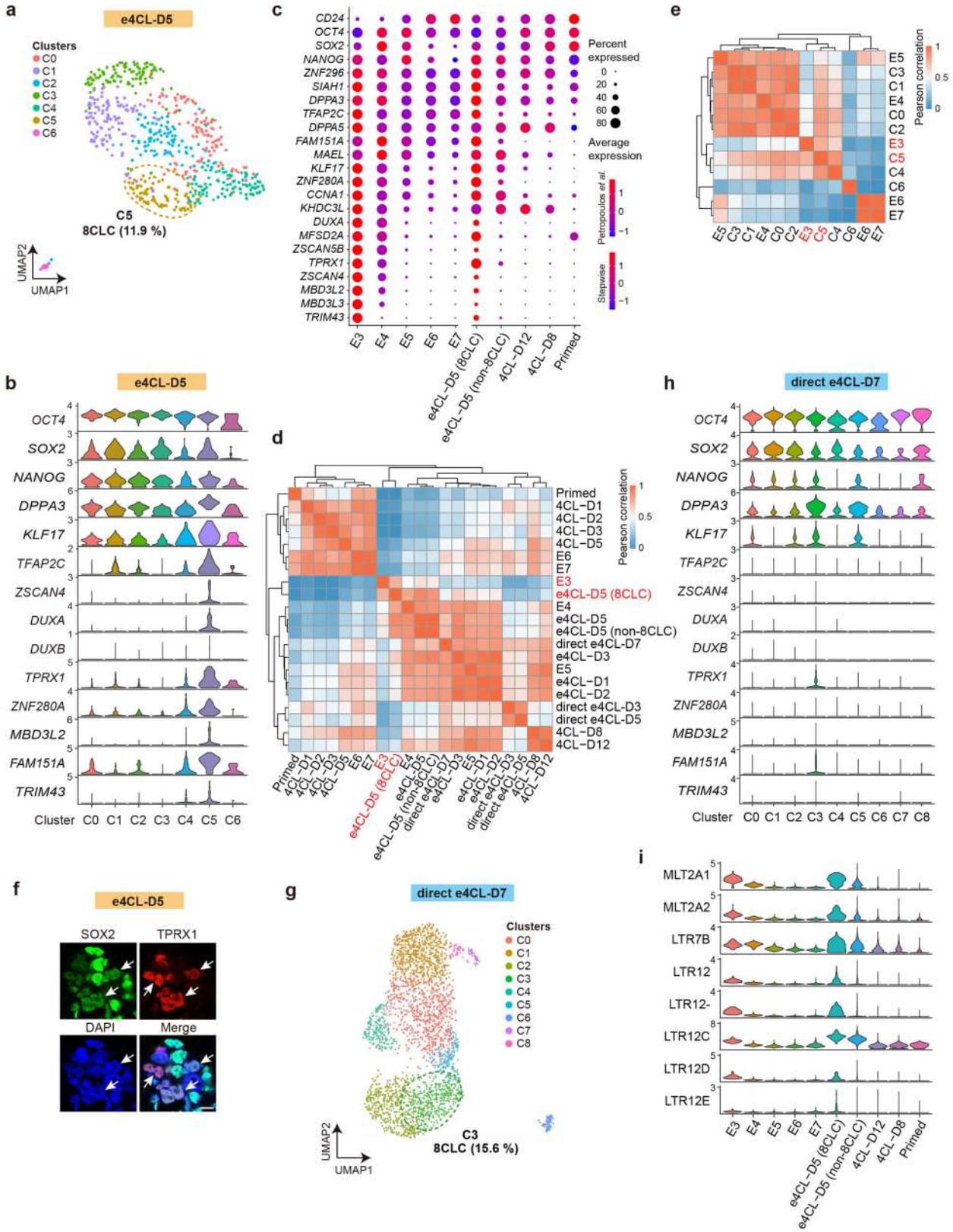
Extended Data Fig. 1 | Medium optimization for generating human naïve PSC. a. Schematic of the compound screening workflow. **b**. Representative phase contrast images of primed H9 ESC cultured in different basal recipes for three passages. hLIF, human LIF; CHIR, CHIR99021; PD, PD0325901; 2i, PD+CHIR. Scale, 40 μm. Representative of three independent experiments. **c**. RT-qPCR for the indicated pluripotency genes in primed H9 ESC cultured in different basal recipes for three passages. Data are the mean values ± standard error of the mean (SEM) of the fold-change compared to primed ESC. n = 3 biological replicates. **d**. Table of tested compounds and their known targets. **e and f**. RT-qPCR for the indicated pluripotency genes in primed H9 ESC cultured in different modified basal recipes for three passages. Data are presented as mean values ± SD of fold-change compared to primed ESC in basal medium with PD, IWR1 and human LIF. n = 3 biological replicates. **g**. RT-qPCR for the indicated pluripotency genes in human primed HUES1 and HUES7 ESC, and human primed iPSC-1, iPSC-2 and iPSC-3 clones cultured in 4CL for three passages (day 12). **h**. Representative images of G-banding karyotype of primed H9 ESC and primed iPSC-4 clone cultured in 4CL for 15 passages. Twenty metaphases were counted for each. All 20 metaphase spreads for H9 ESC are shown in Supplementary Figure 1. **i**. Representative phase contrast images of primed H9 ESC cultured in 4CL on feeders, ECM-coated surface (feeder-free), or in suspension, for three passages (day 12). Scale, 40 μm. Representative of three independent experiments. **j**. RT-qPCR for the indicated pluripotency genes in H9 ESC cultured in 4CL on either feeders or ECM-coated surface for three passages (day 12). Data are presented as mean values ± SD of fold-change compared to primed ESC. n = 3 biological replicates. **k**. RT-qPCR for the indicated pluripotency genes in primed H9 ESC, HUES1 ESC and iPSC-4 clone converted by 4CL in suspension for three passages (day 12). **l**. RT-qPCR for the indicated pluripotency genes in primed H9 ESC cultured in 4CL with or without (w/o) Vc for three passages (day 12). Data are presented as mean values ± SD of fold-change compared to 4CL with Vc. n = 3 biological replicates. *P* value was calculated using two-tailed unpaired Student's *t*-test. **m**. RT-qPCR for the indicated genes in H9 ESC cultured in 4CL with different TSA or DZNep substitutes for three passages (day 12). Data are presented as mean values ± SD of fold-change compared to 4CL with TSA and DZNep. n = 3 biological replicates.



Extended Data Fig. 2 | Validation of 4CL and e4CL media. a. Heatmap showing the expression of preimplantation ICM-enriched genes in human naïve PSC cultured in NHSM[29], tt2iLGö (this study), 5iLAF[17] and 4CL (passage 5), and human ICM cells[16] compared with primed PSC. All reference datasets used in this study are summarized in Supplementary Table 8. H9 ESC were used to generate our dataset. The full list of DEG is included in Supplementary Table 1. **b**. RT-qPCR validation of totipotency genes in primed H9 ESC cultured in stepwise e4CL (day 5). Data are presented as mean values ± SD of fold-change compared to 4CL-day 12 naïve ESC. n = 3 biological replicates. **c**. Heatmap showing the expression of totipotency genes in naïve ESC cultured in NHSM, tt2iLGö, 5iLAF or stepwise e4CL (day 5), EPSC[11], sorted 8CLC from e4CL-day 5 cells, and human 8C-embryo cells[16] compared with primed PSC. H9 ESC were used to generate our dataset. **d**. Representative images of G-banding karyotype of primed H9 ESC and iPSC-4 cultured in stepwise e4CL (day 5). Twenty metaphases were counted for each. All 20 metaphase spreads for H9 ESC are shown in Supplementary Figure 1. **e**. RT-qPCR validation of totipotency genes in primed H9 ESC cultured in direct e4CL (day 7). Data are presented as mean values ± SD of fold-change compared to primed ESC. n = 3 biological replicates.

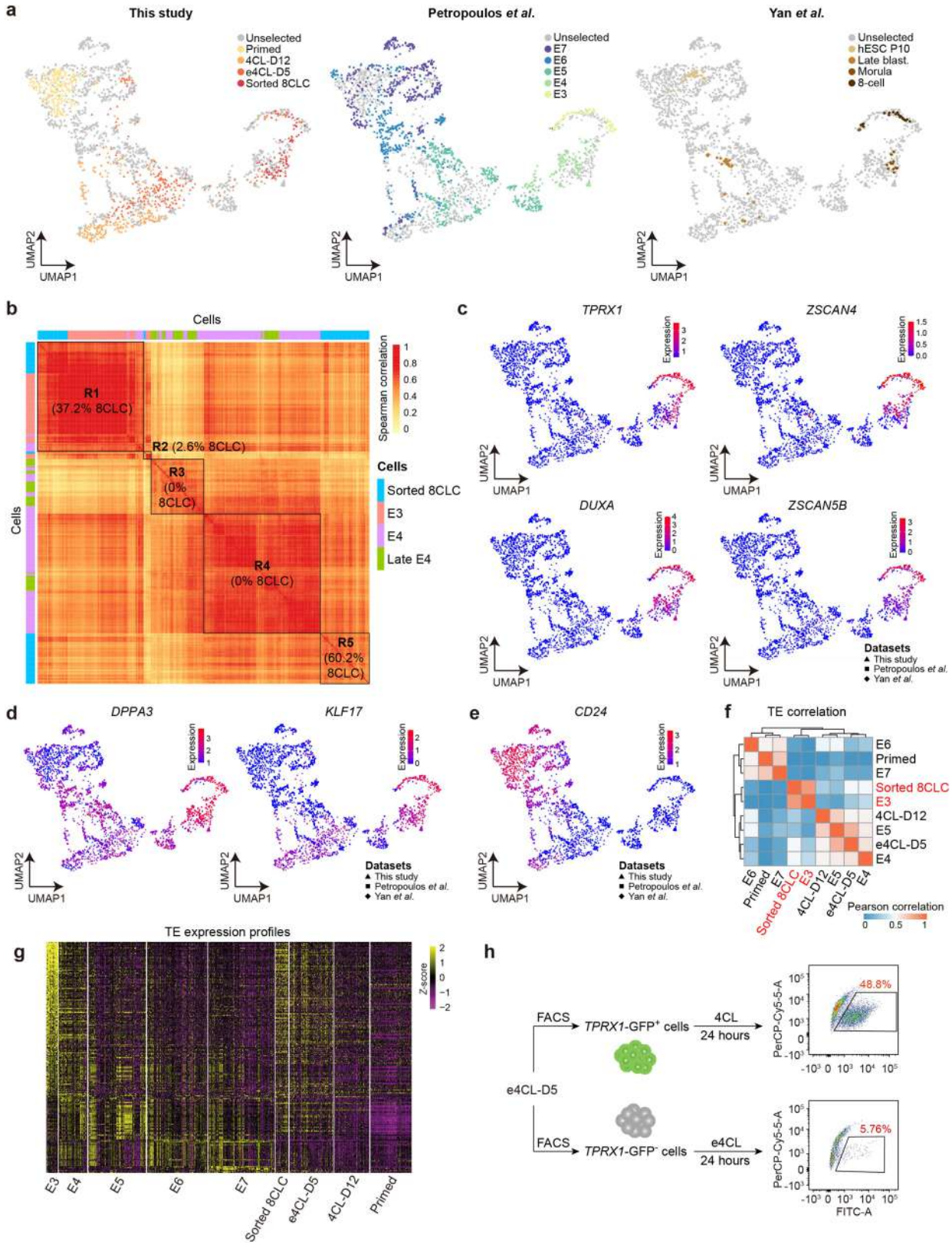


Extended Data Fig. 3 | scRNA-seq (droplet-based) analysis of 8CLC generated from human PSC. a. UMAP comparing the developmental rolling back from human E7 to E3 embryonic stages[20] in stepwise or direct e4CL induction scRNA-seq time courses. All reference datasets used in this study are summarized in Supplementary Table 8. H9 ESC were used to generate our dataset. **b and c**. UMAP visualization based on **panel a** highlighting the expression of the primed-enriched gene *CD24* and the shared naïve pluripotency/8CLC genes *DPPA3* and *KLF17* (**b**), and several totipotency genes (*TPRX1*, *DUXA* and *ZNF280A*) (**c**) as human ESC transition from a primed state to a naïve and then 8CLC state. **d**. Representative immunostaining images for TPRX1, KLF17 and NANOG of primed H9 ESC converted by direct e4CL (day 7). Scale, 20 µm. Representative of three independent experiments.



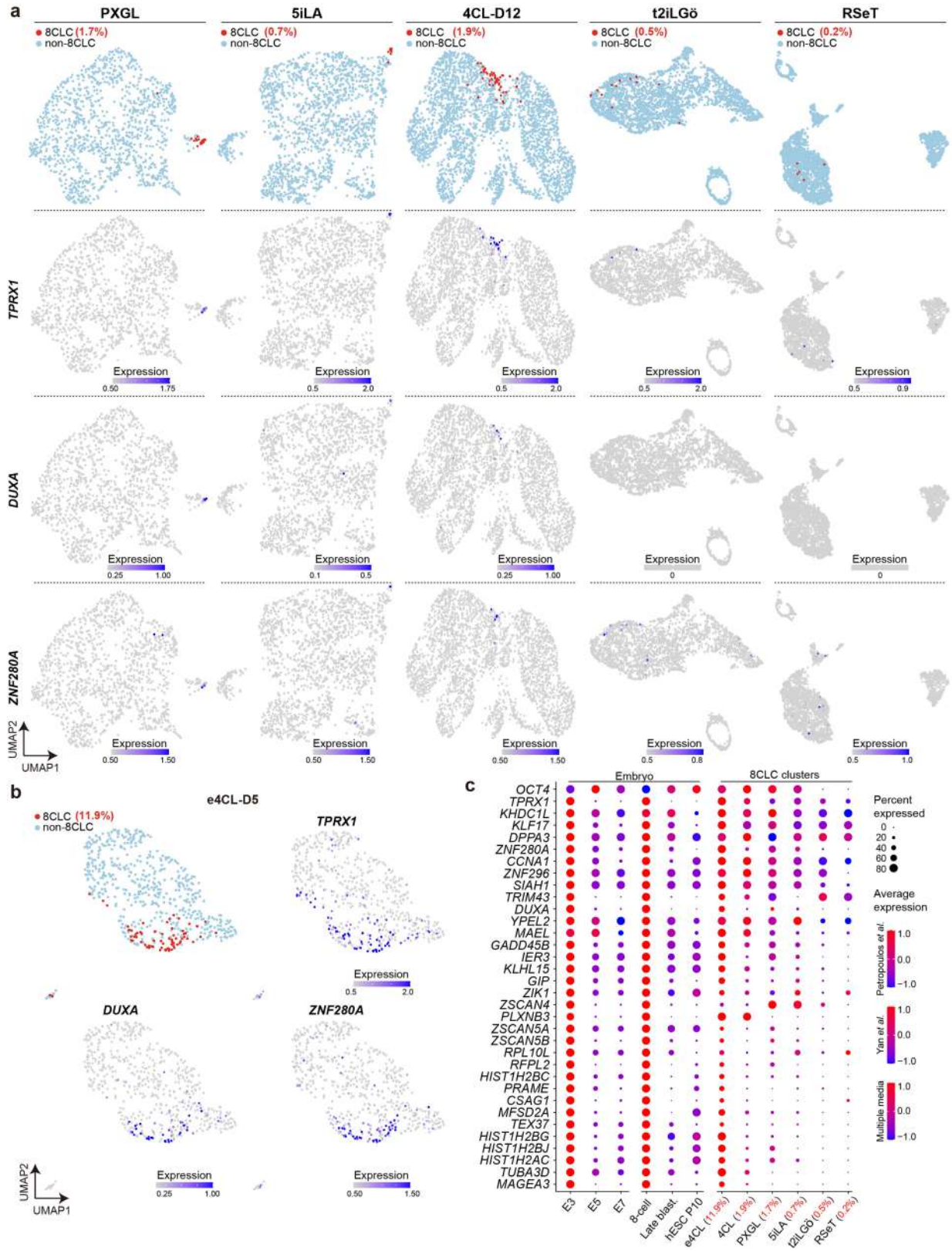
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Extended analysis of scRNA-seq (droplet-based) analysis of human 8CLC generated from PSC. a. UMAP visualization of stepwise e4CL-day 5 cells shows seven clusters. The encircled cluster 5 (8CLC) comprises 11.9% of the whole population. **b**. Violin plot showing the log normalized expression of representative pluripotency and totipotency genes for each of the seven clusters (n = 111, 109, 99, 97, 92, 70 and 12 cells for clusters 0, 1, 2, 3, 4, 5 and 6, respectively) of panel **a**. **c**. Bubble plot representing the frequency of expression and average expression of representative pluripotency and totipotency genes in early human embryonic stages[31] and primed ESC untreated or converted by 4CL (days 8 [passage 2] and 12 [passage 3]) and e4CL (day 5 C5 [8CLC] and non-8CLC [all other clusters summed]). D, day. All reference datasets used in this study are summarized in Supplementary Table 8. **d**. Pseudobulk correlation analysis of the *in vitro* time course data with the embryo data based on the top 2,000 most variable genes. The average linkage hierarchical clustering of the Pearson correlation is shown. **e**. Pseudobulk correlation analysis of e4CL-D5 clusters (in Extended Data Fig. 4a) with the embryo data based on top 2,000 most variable genes. The average linkage hierarchical clustering of the Pearson correlation is shown. **f**. Representative co-immunostaining images of stepwise e4CL-day 5 cells showing mutual exclusivity between SOX2 and TPRX1 (see arrows). Nuclei were counterstained with DAPI. Scale, 20 μm. Representative of three independent experiments. **g**. UMAP visualization of primed ESC converted by direct e4CL at day 7 shows nine clusters. The encircled cluster 3 (C3) cells resemble 8CLC and comprise 15.6% of the whole population. **h**. Violin plot showing the log normalized expression of representative pluripotency and totipotency genes for each of the nine clusters (n = 738 cells, 621 cells, 464 cells, 432 cells, 205 cells, 165 cells, 72 cells, 38 cells, and 36 cells for clusters 0, 1, 2, 3, 4, 5, 6, 7, and 8, respectively) of panel **g**. **i**. Violin plot showing the log normalized expression pattern of representative early embryo-enriched TE in early human embryonic stages and primed ESC untreated or converted by 4CL (days 8 and 12) or stepwise e4CL (day 5; 8CLC and non-8CLC).

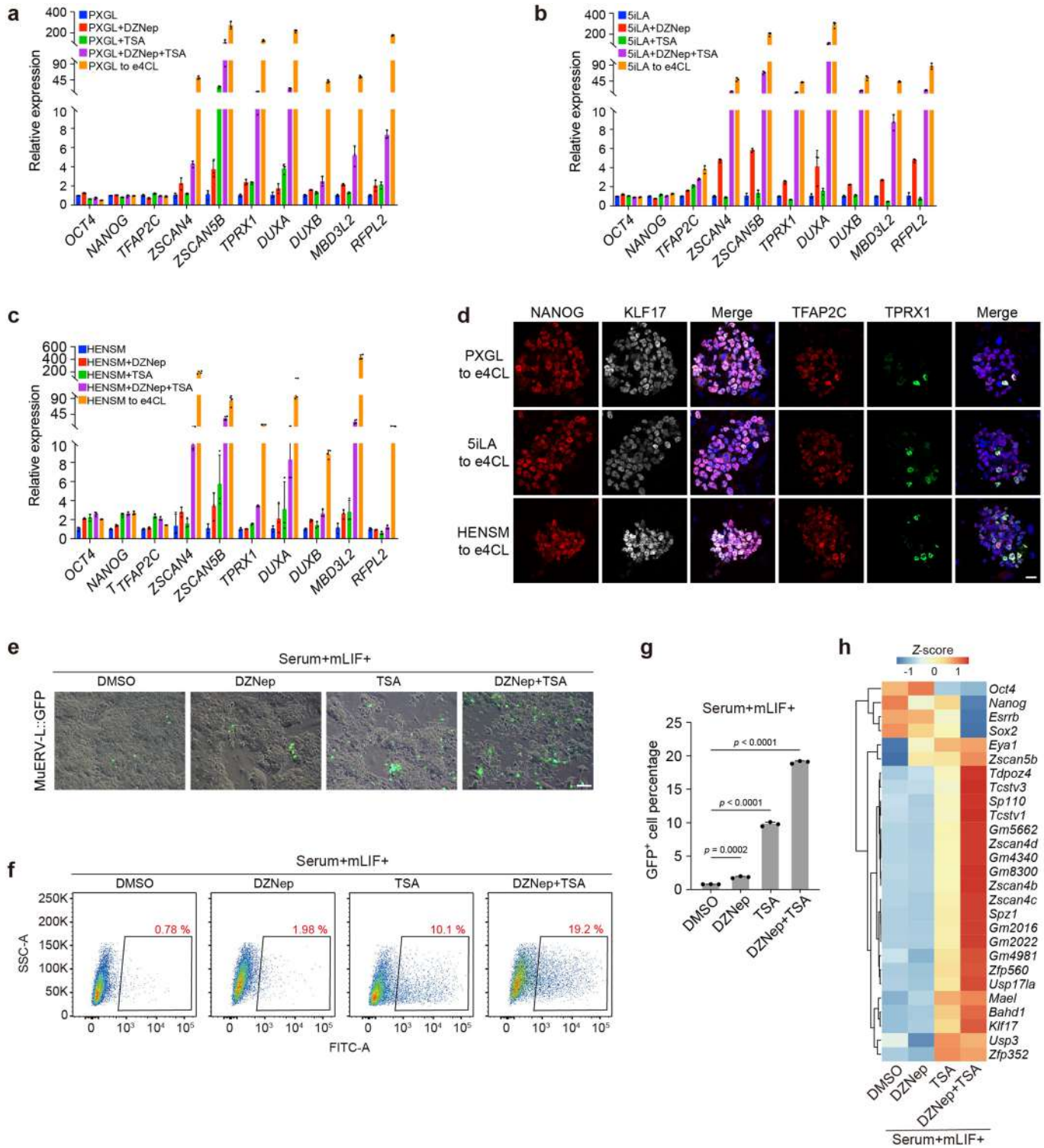


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Extended analysis of scRNAseq data generated by SMART-seq2. a. UMAP visualization based on Fig. 1d highlighting the cells in this study (left panel), in Petropoulos *et al*.[20] (middle panel) and in Yan *et al*.[26] (right panel). P, passage; blast., blastocyst. All reference datasets used in this study are summarized in Supplementary Table 8. **b**. Hierarchical clustering diagram of sorted 8CLC (78 cells) and embryo E3/E4 cells showing the cell-to-cell similarity on the basis of Spearman correlation coefficients. Five closely correlated regions are defined: R1 and R2 include all E3 cells, 10 E4 cells and 31 sorted 8CLC (39.8%). R3 and R4 include E4, all late E4 cells and no sorted 8CLC. R5 includes 47 sorted 8CLC (60.2%) which share similarity with both E3 and E4. **c**–**e**. UMAP visualization based on Fig. 1d highlighting the expression of several totipotency genes (*TPRX1*, *ZSCAN4*, *DUXA* and *ZSCAN5B*) (**c**), the shared naïve pluripotency/8CLC genes *DPPA3* and *KLF17* (**d**) and the primed-enriched gene *CD24* (**e**) as human ESC transition from a primed state to an 8CLC state. **f**. Pseudobulk correlation analysis of the *in vitro* culture data with the embryo data based on all the TE. The average linkage hierarchical clustering of the Pearson correlation is shown. **g**. Heatmap showing the embryonic stage-specific TE expression profiles in human E3 to E7, sorted 8CLC, 4CL-day 12 naïve ESC and primed ESC. The full list of stage-specific TE are included in Supplementary Table 2. **h**. GFP^+ sorted 8CLC from e4CL-day 5 cells were cultured in 4CL for 24 h and subjected to FACS analysis. 48.8% remained GFP^+ and 51.2% became GFP^- (exited from the 8CLC state) (upper panels). GFP^- cells were sorted from stepwise e4CL-day 5 and were cultured in e4CL for 24 h and subjected to FACS analysis. 5.76% became GFP^+ (entered the 8CLC state) (lower panels). D, day.

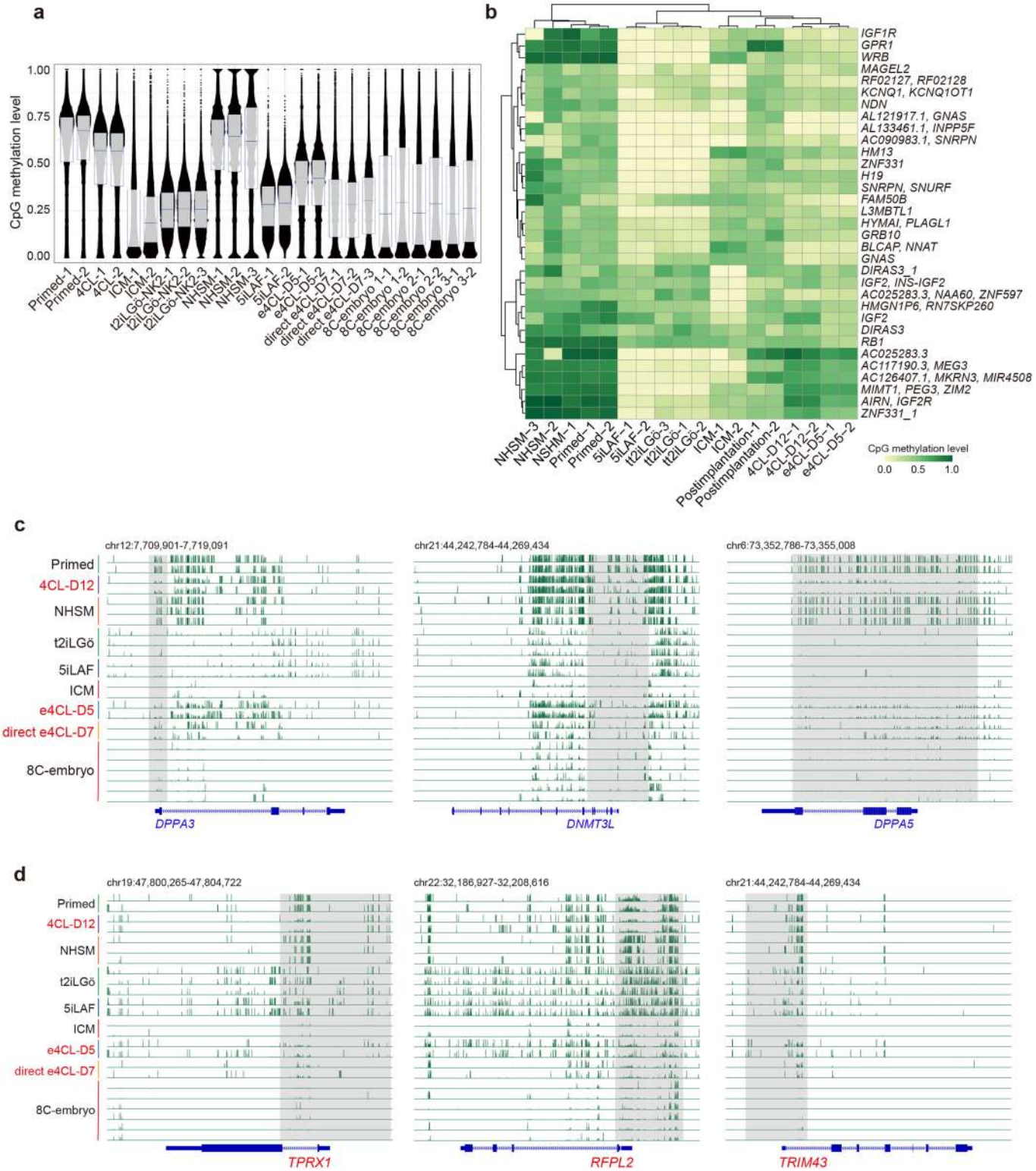


Extended Data Fig. 6 | Identification of 8CLC in scRNA-seq from other naïve PSC recipes. a. UMAP visualization highlighting putative 8CLC and non-8CLC and the expression of the totipotency genes *TPRX1*, *DUXA* and *ZNF280A* for the indicated naïve media[13,27,28]. Annotation was based on the averaged expression scores of the gene signature used for defining 8CLC in this study. All reference datasets used in this study are summarized in Supplementary Table 8. **b**. UMAP visualization highlighting putative 8CLC and non-8CLC and the expression of the totipotency genes *TPRX1*, *DUXA* and *ZNF280A* in stepwise e4CL-day 5 cells. **c**. Bubble plot representing the frequency of expression and average expression of representative pluripotency and totipotency genes in early human embryonic stages[20,26] and the annotated putative 8CLC in the indicated culture conditions. OCT4 reads have been removed from reprogramming t2iLGö and RSeT datasets due to overlap with the signal of the overexpressed factor.

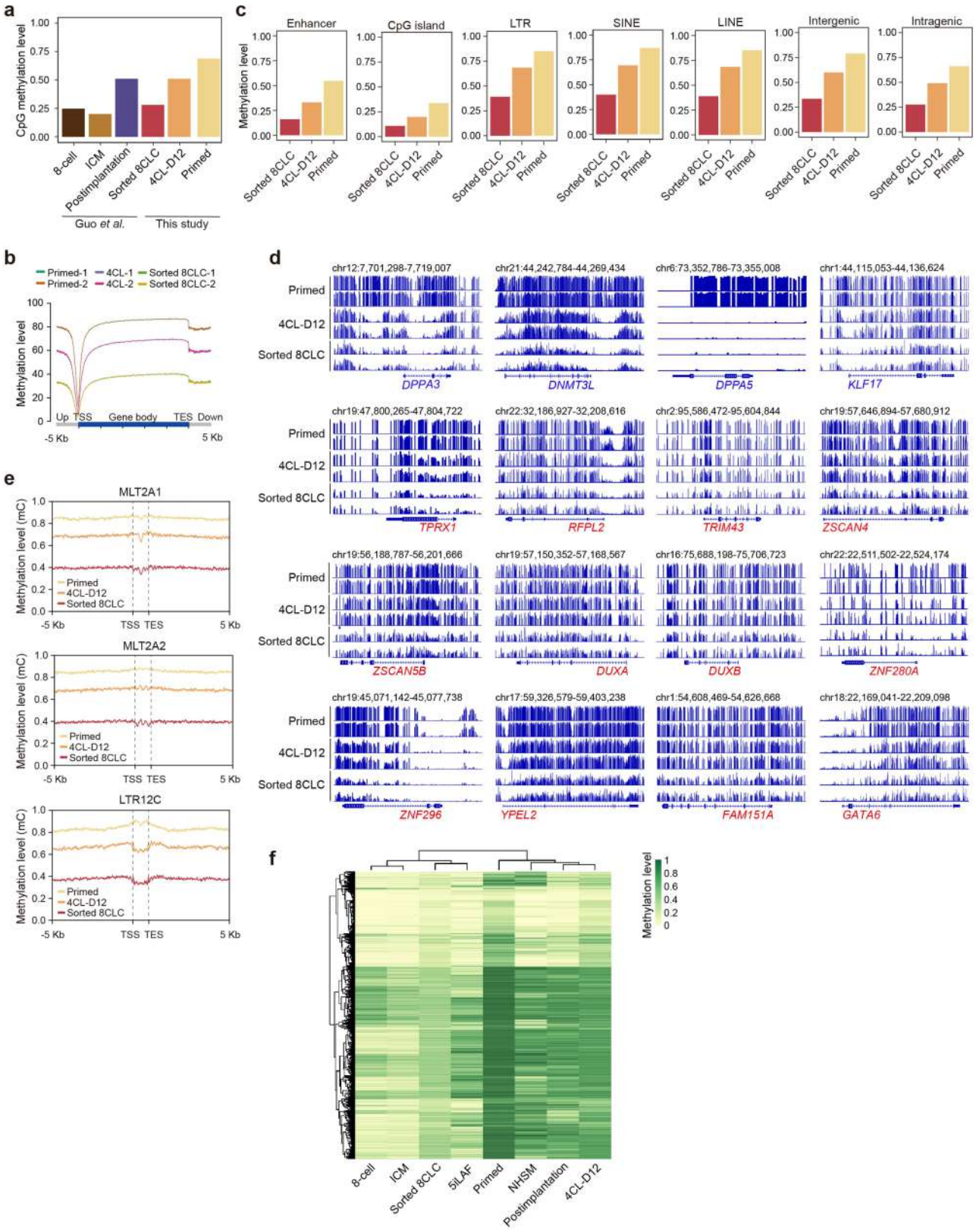


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Generation of 8CLC from other human naïve PSC and 2CLC from mouse naïve PSC. a. RT-qPCR for the indicated pluripotency and totipotency genes in primed H9 ESC cultured in PXGL[47], PXGL+DZNep, PXGL+TSA and PXGL+DZNep+TSA, and PXGL cells cultured in e4CL (PXGL to e4CL) for five days. Data are presented as mean values ± SD of the fold-change compared to cells in PXGL medium. n = 3 biological replicates. **b**. RT-qPCR for the indicated pluripotency and totipotency genes in primed H9 ESC cultured in 5iLA[17], 5iLA+DZNep, 5iLA+TSA, and 5iLA+DZNep+TSA, and 5iLA cells cultured in e4CL (5iLA to e4CL) for five days. Data are presented as mean values ± SD of the fold-change compared to cells in 5iLA medium. n = 3 biological replicates. **c**. RT-qPCR for the indicated pluripotency and totipotency genes in primed H9 ESC cultured in HENSM[30], HENSM+DZNep, HENSM+TSA, HENSM+DZNep+TSA and HENSM cells cultured in e4CL (HENSM to e4CL) for five days. Data are presented as mean values ± SD of the fold-change compared to cells in HENSM medium. n = 3 biological replicates. **d**. Representative immunofluorescence images for pluripotency (NANOG, KLF17 and TFAP2C) and totipotency (TPRX1) markers in PXGL, 5iLA or HENSM-derived naïve H9 ESC cultured in e4CL for five days. Scale, 20 µm. Representative of three independent experiments. **e**. Representative fluorescent (green) images of *MuERV-L-LTR-GFP* reporter mouse ESC cultured in the indicated conditions for three days. Scale, 40 µm. Representative of three independent experiments. m, mouse. **f**. Mouse *MuERV-L-LTR-GFP* reporter ESC cultured in the indicated conditions for three days and subjected to FACS analysis to determine the percentage of GFP^+ cells. m, mouse. **g**. Bar plot indicating the percentage of *MuERV-L-LTR-GFP*$^+$ cells generated in the indicated conditions measured by FACS. Data are presented as mean values ± SD. n = 3 biological replicates. *P* value was calculated using two-tailed unpaired Student's *t*-test. **h**. Heatmap showing the expression of pluripotency and 2C-embryo enriched genes in *MuERV-L-LTR-GFP* reporter mouse ESC cultured in the indicated conditions for three days. m, mouse.

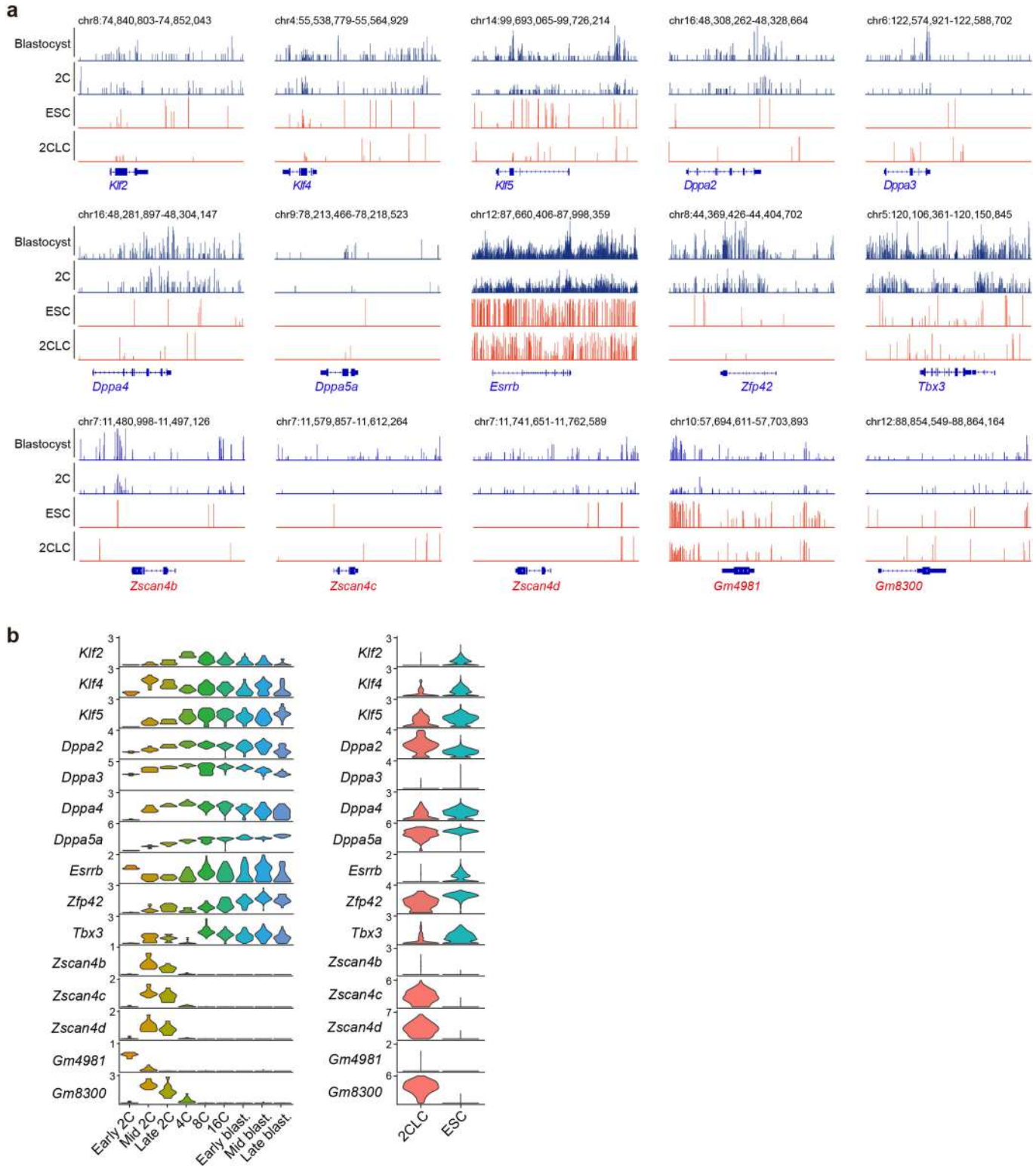


Extended Data Fig. 8 | DNA methylation (measured by RRBS) status of human PSC converted by 4CL and e4CL. a. Violin plot showing global CpG methylation levels measured by RRBS of human PSC cultured in primed conditions (n = 2 technical replicates), 4CL (day 12) (n = 2 technical replicates), 5iLAF[35] (n = 2 technical replicates), t2iLGö-NK2[36] (n = 3 technical replicates), NHSM[29] (n = 3 technical replicates), stepwise e4CL (day 5) (n = 2 technical replicates) and direct e4CL (day 7) (n = 3 technical replicates), human 8C-embryo (n = 3 biological replicates with 2 technical replicates each) and ICM[32] (n = 2 technical replicates). All reference datasets used in this study are summarized in Supplementary Table 8. Our dataset was generated using H9 ESC. For boxplots, the blue central line is the median, the boxes indicate the upper and lower quartiles. **b**. Heatmap showing CpG methylation levels at a panel of imprinting control regions of human PSC cultured in primed conditions, 4CL (day 12), 5iLAF[17], tt2iLGö[37], NHSM[29] and stepwise e4CL (day 5), ICM and postimplantation embryo[32]. **c and d**. Genome browser tracks showing CpG methylation levels at the indicated naïve pluripotency (blue) (**c**) and totipotency (red) loci (**d**) of PSC cultured in primed conditions, 4CL (day 12), 5iLAF, t2iLGö, NHSM, stepwise e4CL (day 5) and direct e4CL (day 7), human 8C-embryo and ICM. Each bar represents a single CpG and the height indicates the percentage of methylation.

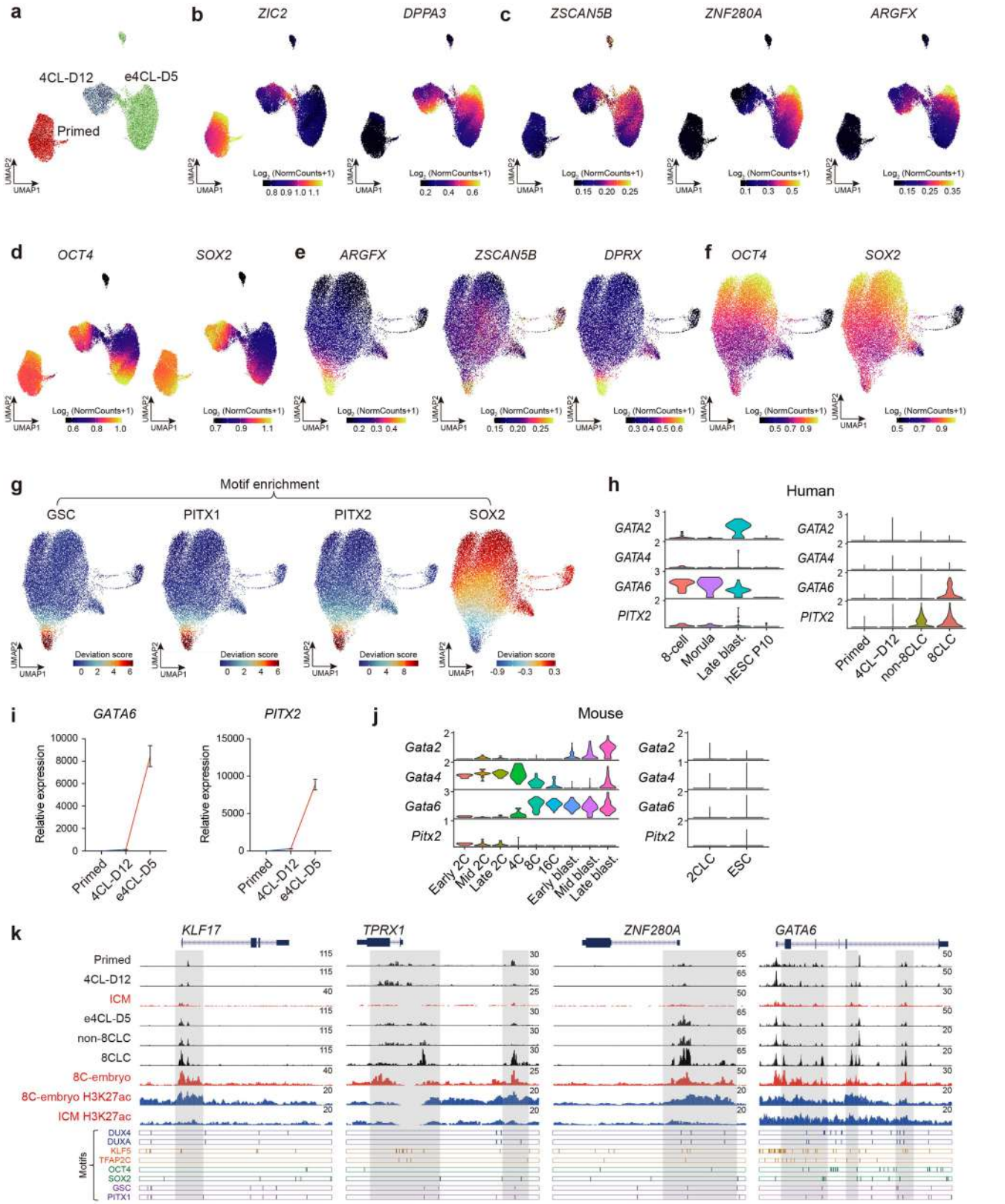


Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | DNA methylation (measured by WGBS) status of primed human PSC converted by 4CL and sorted 8CLC. a. Bar plot comparing global CpG methylation levels in the human 8C-embryo, ICM and postimplantation embryo[32], and primed ESC, 4CL-day 12 naïve ESC and sorted 8CLC from e4CL-day 5. All reference datasets used in this study are summarized in Supplementary Table 8. Our dataset was generated using H9 ESC. **b**. Comparative DNA methylation level at the TSS, gene body and TES in primed ESC, 4CL-day 12 naïve ESC and sorted 8CLC. **c**. Comparative DNA methylation level in different genomic regions in ESC cultured in primed, 4CL-day 12 naïve ESC and sorted 8CLC. **d**. Genome browser tracks visualization of DNA methylation levels measured by WGBS at the indicated naïve pluripotency (blue) and totipotency (red) loci of primed ESC, 4CL-day 12 naïve ESC and sorted 8CLC. Each bar represents a single CpG and the height indicates the percentage of methylation. **e**. Comparative DNA methylation level at human early embryo-enriched TE (MLT2A1, MLT2A2 and LRT12C) in primed ESC, 4CL-day 12 naïve ESC and sorted 8CLC. **f**. Genome-wide comparative DNA methylation profiles in human 8C-embryo, ICM, postimplantation embryo, primed ESC, 4CL-day 12 naïve ESC, sorted 8CLC from e4CL-D5, 5iLAF[35] and NHSM[29]. Data from different biological replicates were pooled and binned into consecutive 10-kb tiles. Only tiles covered by all the datasets were shown.

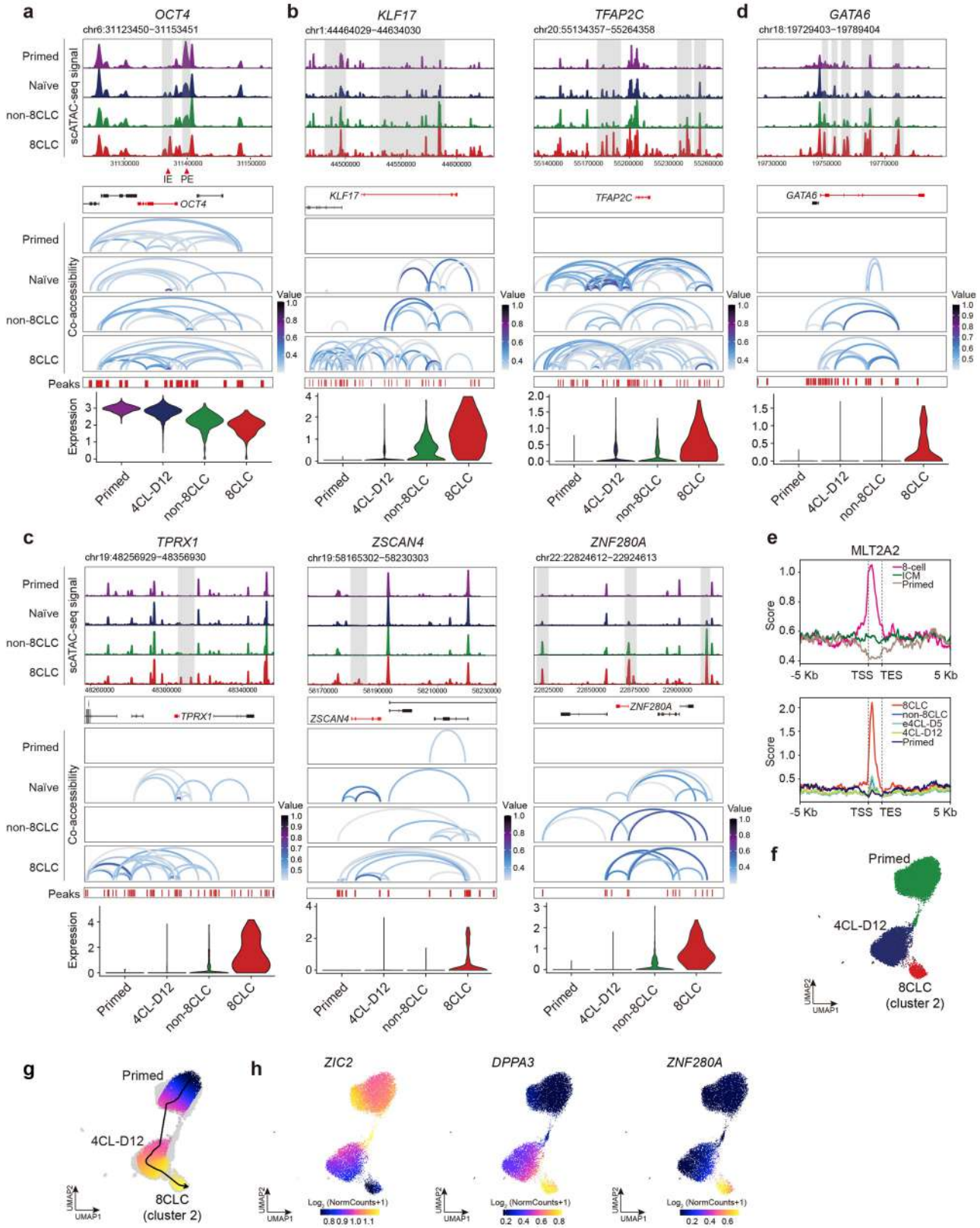


Extended Data Fig. 10 | DNA methylation and gene expression analysis in mouse embryo and 2CLC. a. Genome browser tracks visualization of DNA methylation levels measured by RRBS at the indicated naïve pluripotency (blue) and totipotency (red) loci of mouse embryonic stages[38], ESC and 2CLC[39]. Each bar represents a single CpG and the height indicates the percentage of methylation. All reference datasets used in this study are summarized in Supplementary Table 8. **b.** Violin plots showing the log normalized expression of representative naïve pluripotency and totipotency genes in mouse early embryonic stages[23] (left panel), ESC and 2CLC[40] (right panel).



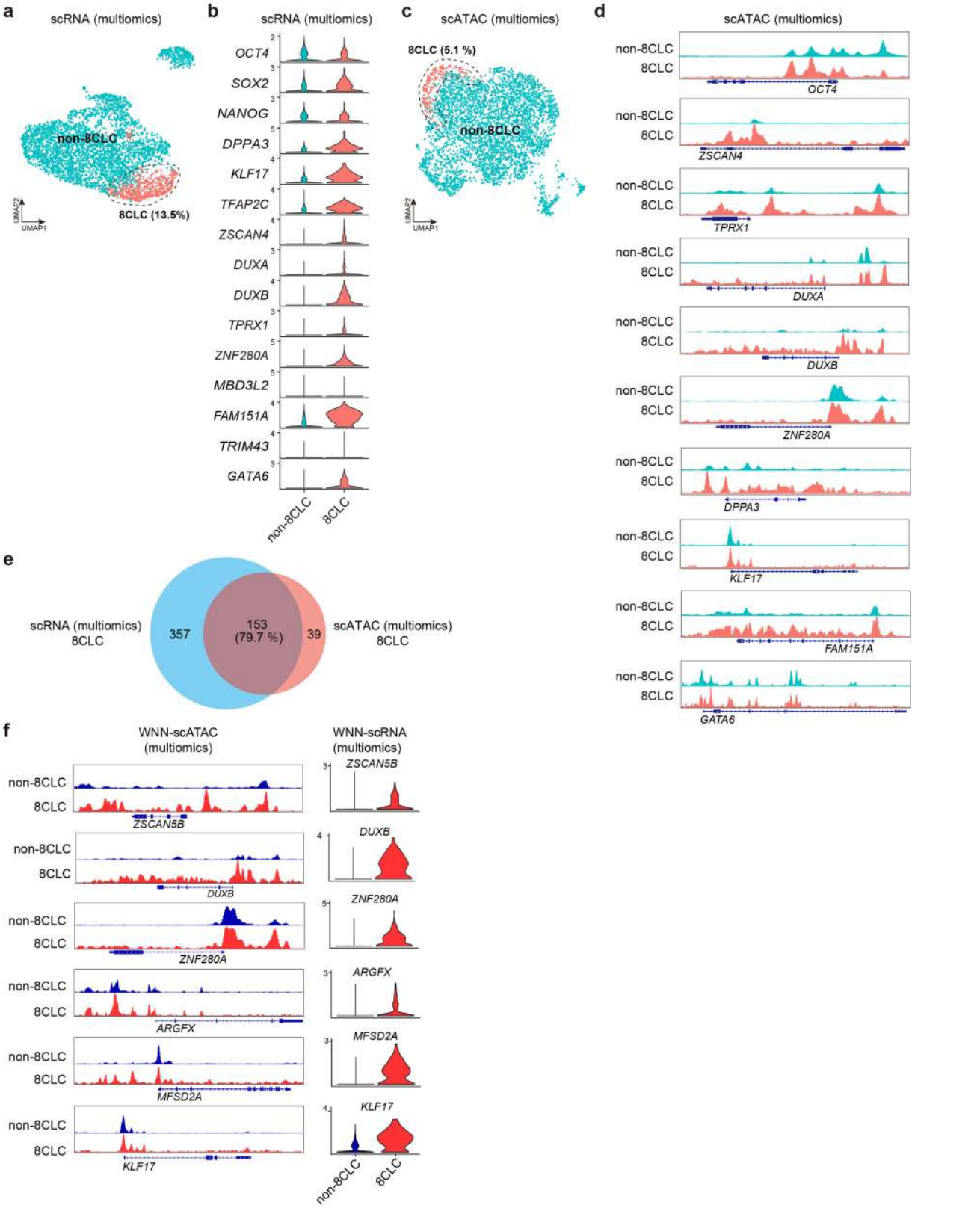
Extended Data Fig. 11 | See next page for caption.

Extended Data Fig. 11 | Extended analysis of 8CLC chromatin landscape. **a**. UMAP visualization of gene score for all genes in the scATAC-seq of primed ESC untreated (red) or converted by 4CL (day 12; blue) or stepwise e4CL (day 5; green). **b and c**. UMAP visualization based on panel **a** highlighting the gene score for primed (*ZIC2*), shared naïve pluripotency/8CLC (*DPPA3*) (**b**) and totipotency (**c**) genes projected onto each individual cell in the scATAC-seq of primed ESC untreated or converted by 4CL (day 12) and stepwise e4CL (day 5). **d**. UMAP visualization based on panel **a** highlighting the gene score for classical pluripotency genes (*OCT4* and *SOX2*) projected onto each individual cell in the scATAC-seq of primed ESC untreated or converted by 4CL (day 12) and stepwise e4CL (day 5). **e and f**. UMAP visualization showing the gene score for totipotency (**e**) and classical pluripotency (**f**) genes projected onto the individual cells of stepwise e4CL-day 5 of Fig. 2a. **g**. UMAP visualization showing DNA binding motif deviation scores of the lineage specifiers (GSC, PITX1 and PITX2) and classical pluripotency (SOX2) gene projected onto the UMAP of Fig. 2a. **h**. Violin plot showing the log normalized expression of the lineage specifiers *GATA2*, *GATA4*, *GATA6* and *PITX2* in human embryonic stages and human ESC passage 10[26] (left panel), and in primed ESC untreated or converted by 4CL (day 12) and e4CL (day 5 C5 [8CLC] and nnon-8CLC [all other clusters summed]) (right panel). Data were taken from the droplet-based scRNA-seq time course. All reference datasets used in this study are summarized in Supplementary Table 8. hESC, human ESC; P, passage. **i**. RT-qPCR showing upregulation of lineage specifiers (*GATA6* and *PITX2*) in stepwise e4CL day 5 cells compared to primed ESC and 4CL-day 12 naïve ESC. H9 ESC were used to generate these data. Data are presented as mean values ± SD. n = 3 biological replicates. D, day. **j**. Violin plot showing the log normalized expression of indicated lineage specifiers in early mouse embryonic stages[23] (left panel) and mouse 2CLC compared to naïve ESC[39] (right panel). blast., blastocyst. **k**. Genome browser tracks showing chromatin accessibility, H3K27ac level[42] and transcription factor DNA binding motif location at the naïve pluripotency *KLF17* locus, multiple totipotency (*TPRX1* and *ZNF280A*) loci and lineage specifier (*GATA6*) locus in the indicated cell types.

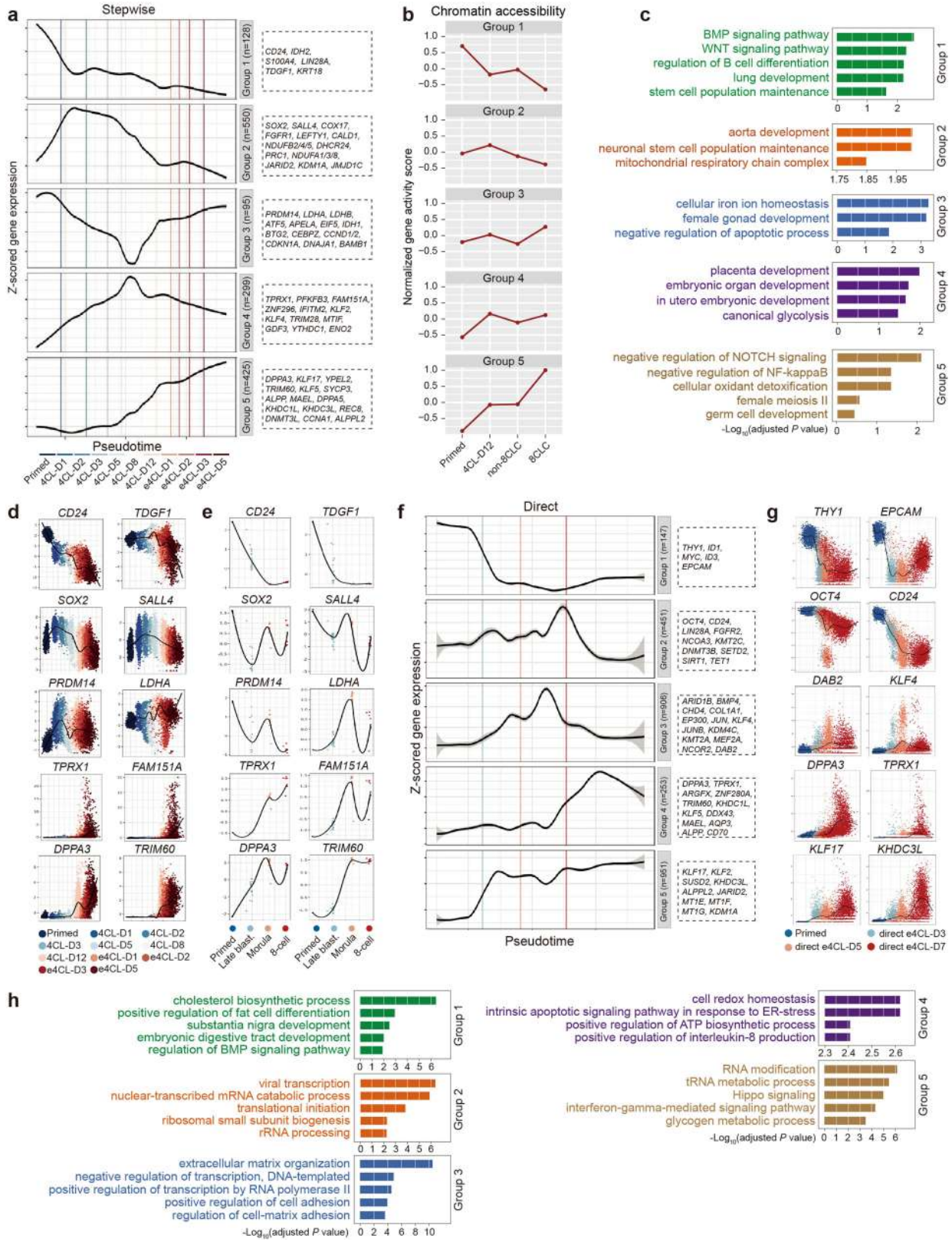


Extended Data Fig. 12 | See next page for caption.

Extended Data Fig. 12 | Cell type-specific chromatin accessibility and interactions. a–**d**. Genome browser tracks visualizing cell type-enriched (primed ESC, 4CL-day 12 naïve ESC, and non-8CLC and 8CLC in cluster 5 of stepwise e4CL-day 5 cells) chromatin accessibility (upper panel) and co-accessibility loops (middle panel) for the indicated classical pluripotency (**a**), naïve pluripotency (**b**) and classical totipotency loci (**c**), and the 8C-enriched lineage specifier *GATA6* locus (**d**). Color intensities of the loops represent the significance level of peak co-accessibility. A violin plot showing the log normalized expression level in the same cell types is displayed in the bottom panel for each locus. **e**. Average normalized chromatin accessibility signal for the early human embryo-enriched TE MLT2A2 in early human embryo bulk ATAC-seq[16] and pseudobulk of our scATAC-seq data. All reference datasets used in this study are summarized in Supplementary Table 8. **f**. UMAP visualization of gene score of all genes in the scATAC-seq of primed H9 ESC untreated (green) or converted by 4CL (day 12; blue) and 8CLC (red) detected in Fig. 2a. **g**. Trajectory based on gene score and DNA binding motif distribution along the stepwise process of 8CLC induction from primed ESC to 8CLC passing through 4CL-day 12 naïve ESC projected onto the UMAP of panel **f**. **h**. UMAP visualization based on panel **g** highlighting the gene score for primed (*ZIC2*) and shared naïve pluripotency/8CLC (*DPPA3*) and totipotency (*ZNF280A*) genes projected onto each individual cell of primed ESC untreated or converted by 4CL (day 12) and 8CLC detected in Fig. 2a.

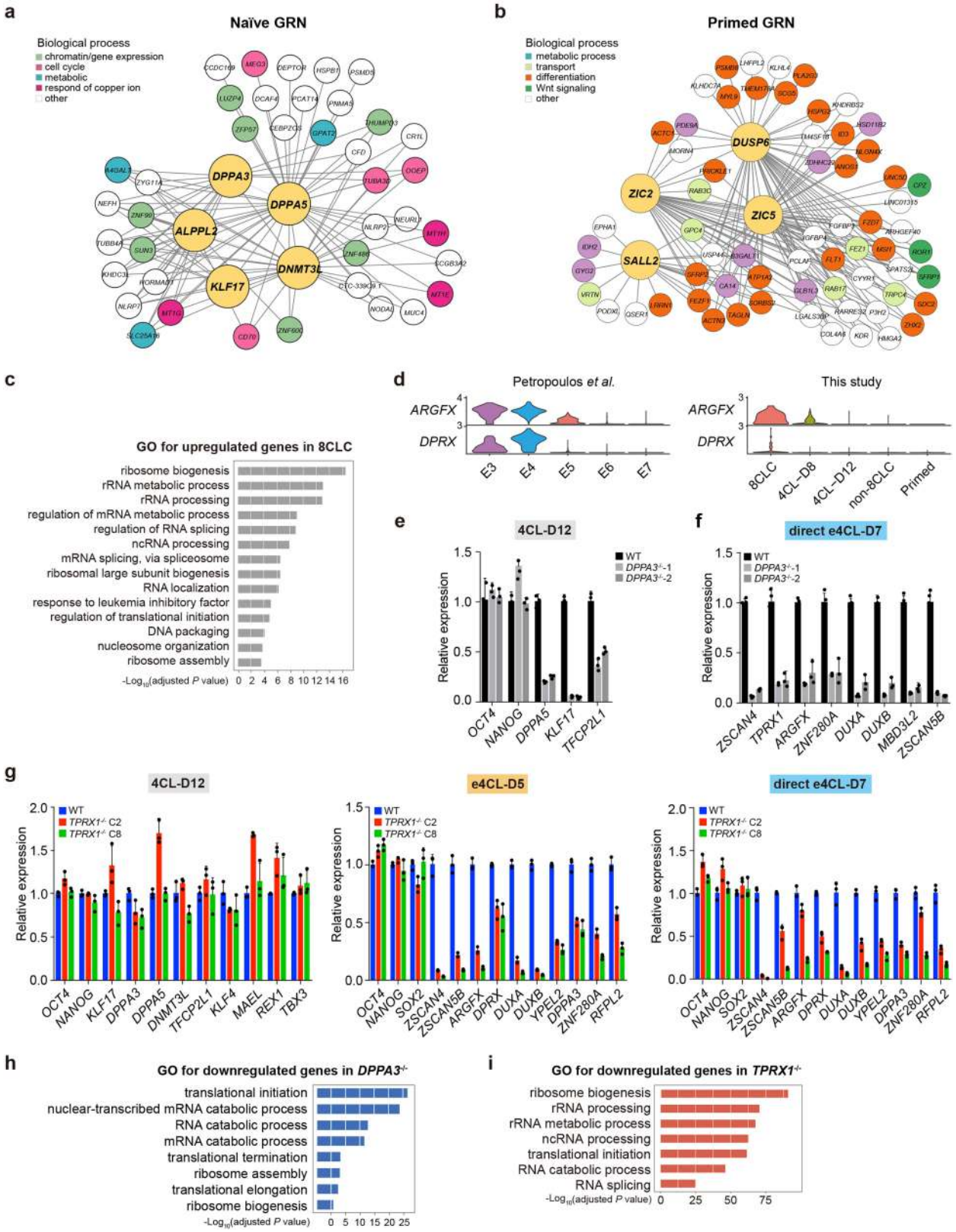


Extended Data Fig. 13 | Extended multiomics analysis of 8CLC chromatin landscape. a. UMAP visualization of scRNA-seq in the multiomics analysis of stepwise e4CL-day 5 cells showing annotated 8CLC (13.5%) and non-8CLC clusters. **b**. Violin plot showing the log normalized expression of representative pluripotency and totipotency genes for 8CLC and non-8CLC of panel **a**. **c**. UMAP visualization of scATAC-seq in the multiomics analysis of stepwise e4CL-day 5 cells showing annotated 8CLC (5.1%) and non-8CLC clusters. **d**. Genome browser tracks visualization of chromatin accessibility for the indicated pluripotency and totipotency loci in 8CLC and non-8CLC of panel **c**.**e**. Venn diagram showing the overlap between 8CLC detected by scATAC-seq and 8CLC detected by scRNA-seq. **f**. Genome browser tracks visualization of chromatin accessibility and violin plot showing the gene expression for the indicated genes in 8CLC (cluster 5) and non-8CLC (all other clusters summed) of Fig. 2h.



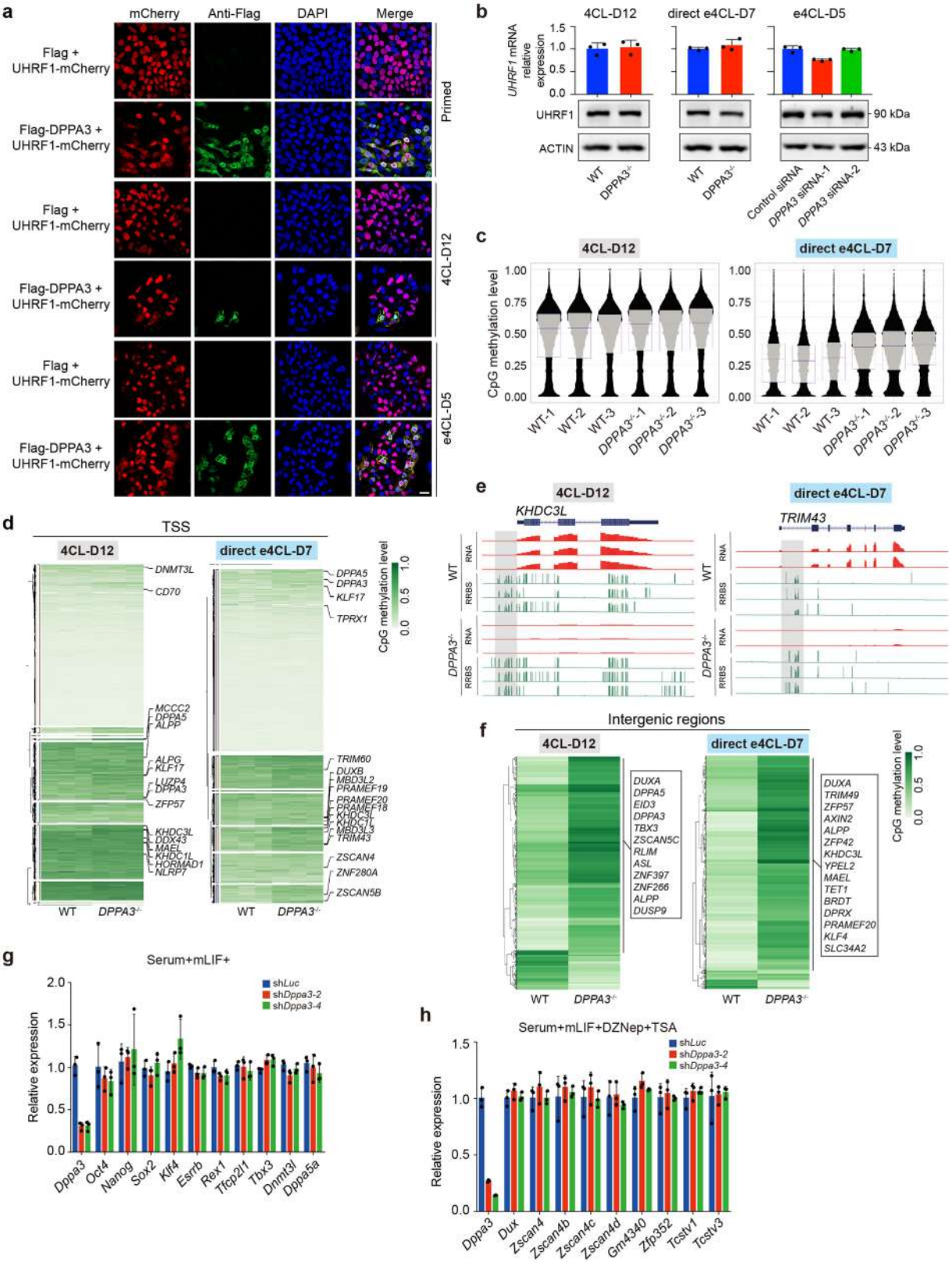
Extended Data Fig. 14 | See next page for caption.

Extended Data Fig. 14 | Dynamics of gene expression during the 8CLC conversion. a. Hierarchical clustering of 1,497 variable genes among the indicated scRNA-seq (droplet based) samples along the pseudotime of stepwise e4CL induction of 8CLC from primed ESC identifies five gene groups. $k = 5$. **b**. Line plots showing the mean standardized gene score of the indicated scATAC-seq samples for each gene group of panel **a**. 8CLC (cluster 2) and non-8CLC (all other clusters summed) are based on Fig. 2a. The red line represents the median values for the group center. **c**. Enriched GO terms for biological processes in each of the five groups of panel **a** relative to all other gene groups. *P* value was calculated using hypergeometric test and adjusted for multiple testing using Benjamini-Hochberg correction. **d**. Pseudotime showing the expression pattern of the indicated genes for each cluster of panel **a** during the stepwise e4CL induction of 8CLC from primed ESC. **e**. Pseudotime showing the expression pattern of the same genes of panel **d** during early human embryonic development[26]. blast., blastocyst. All reference datasets used in this study are summarized in Supplementary Table 8. **f**. Hierarchical clustering of variable genes among the indicated scRNA-seq samples along the pseudotime of direct e4CL induction of 8CLC from primed ESC identifies five groups. $k = 5$. Each transcript expression is shown as a gray line; the black line represents the mean expression. **g**. Pseudotime showing the expression pattern of the indicated genes for each cluster of panel **f** during the direct e4CL induction of 8CLC from primed ESC. **h**. Enriched GO terms for biological processes in each of the five groups of panel **f** relative to all other gene clusters. *P* value was calculated using hypergeometric test and adjusted for multiple testing using Benjamini-Hochberg correction.



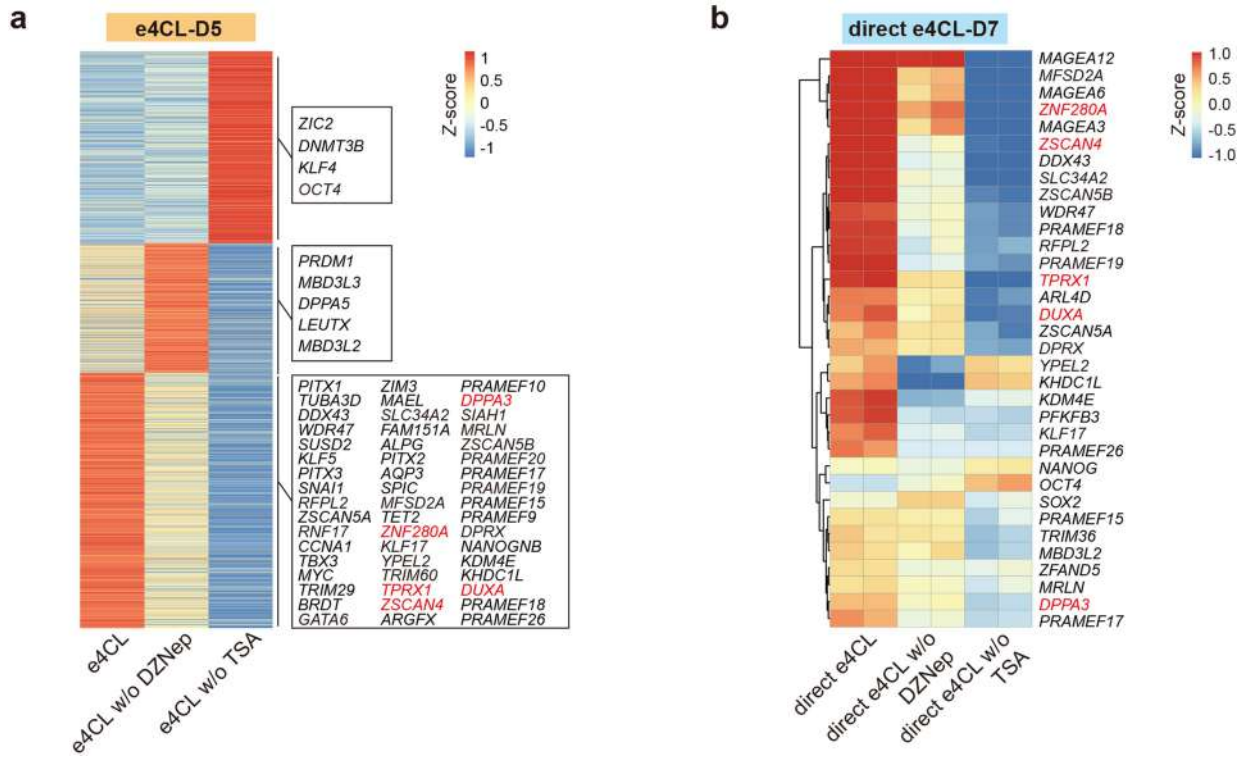
Extended Data Fig. 15 | See next page for caption.

Extended Data Fig. 15 | Networks and regulators controlling the 8CLC conversion. a and b. Cytoscape network visualization of top hub genes (highlighted in yellow) from Fig. 3a and their targets with interconnections in 4CL-day 12 naïve ESC (**a**) and primed ESC (**b**). **c**. Enriched GO terms for biological processes in DEG of 8CLC in stepwise e4CL-day 5 cells of Fig. 3c compared to non-8CLC. *P* value was calculated using hypergeometric test and adjusted for multiple testing using Benjamini-Hochberg correction. **d**. Violin plot showing the log normalized expression pattern of 8C-embryo enriched ETCHbox family members in early human embryonic stages[20] and primed ESC untreated or converted by 4CL (days 8 and 12) and non-8CLC and 8CLC of stepwise e4CL (day 5) of Extended Data Fig. 4a. D, day. All reference datasets used in this study are summarized in Supplementary Table 8. **e and f**. RT-qPCR showing the expression of the indicated genes in *DPPA3* knockout (*DPPA3*$^{-/-}$) ESC compared to wild-type after conversion from a primed state to 4CL-day 12 naïve PSC (**e**) or direct e4CL-day 7 cells (**f**). Data are presented as mean values ± SD. Two independent *DPPA3* knockout clones were included. n = 3 biological replicates. **g**. RT-qPCR showing expression of the indicated genes in *TPRX1* knockout (*TPRX1*$^{-/-}$) compared to wild-type after conversion of ESC from a primed state to 4CL-day 12 PSC (left pane), stepwise e4CL-day 5 cells (middle panel) or direct e4CL-day 7 cells (right panel). Data are presented as mean values ± SD. Two independent *TPRX1* knockout clones were included. n = 3 biological replicates. **h**. Enriched GO terms for downregulated genes in *DPPA3* knockout (*DPPA3*$^{-/-}$) ESC compared to wild type in direct e4CL-day 7. *P* value was calculated using hypergeometric test and adjusted for multiple testing using Benjamini-Hochberg correction. **i**. Enriched GO terms for downregulated genes in TPRX1 knockout (TPRX1-/-) ESC compared to wild type in stepwise e4CL-day 5. P value was calculated using hypergeometric test and adjusted for multiple testing using Benjamini-Hochberg correction.

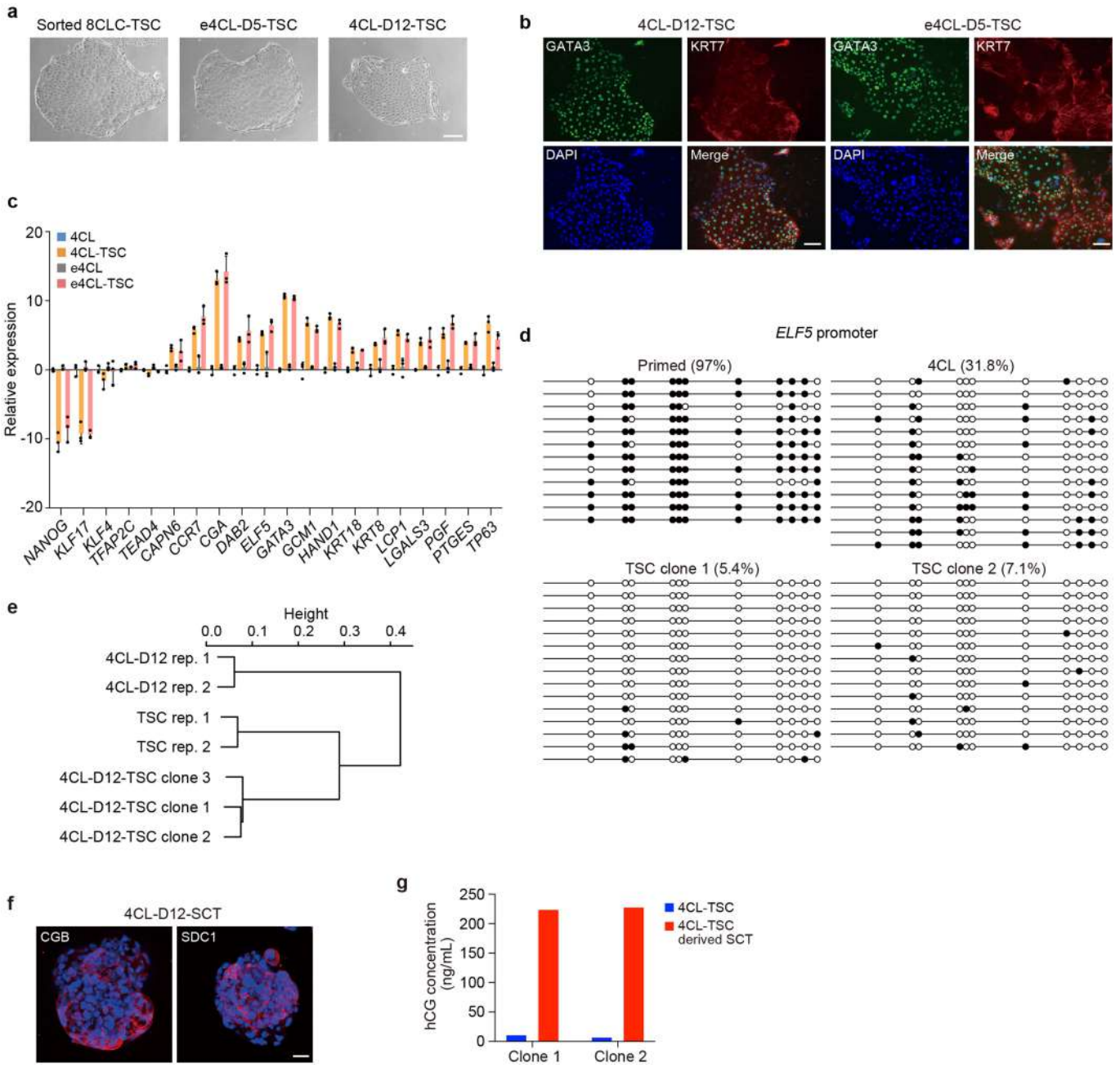


Extended Data Fig. 16 | See next page for caption.

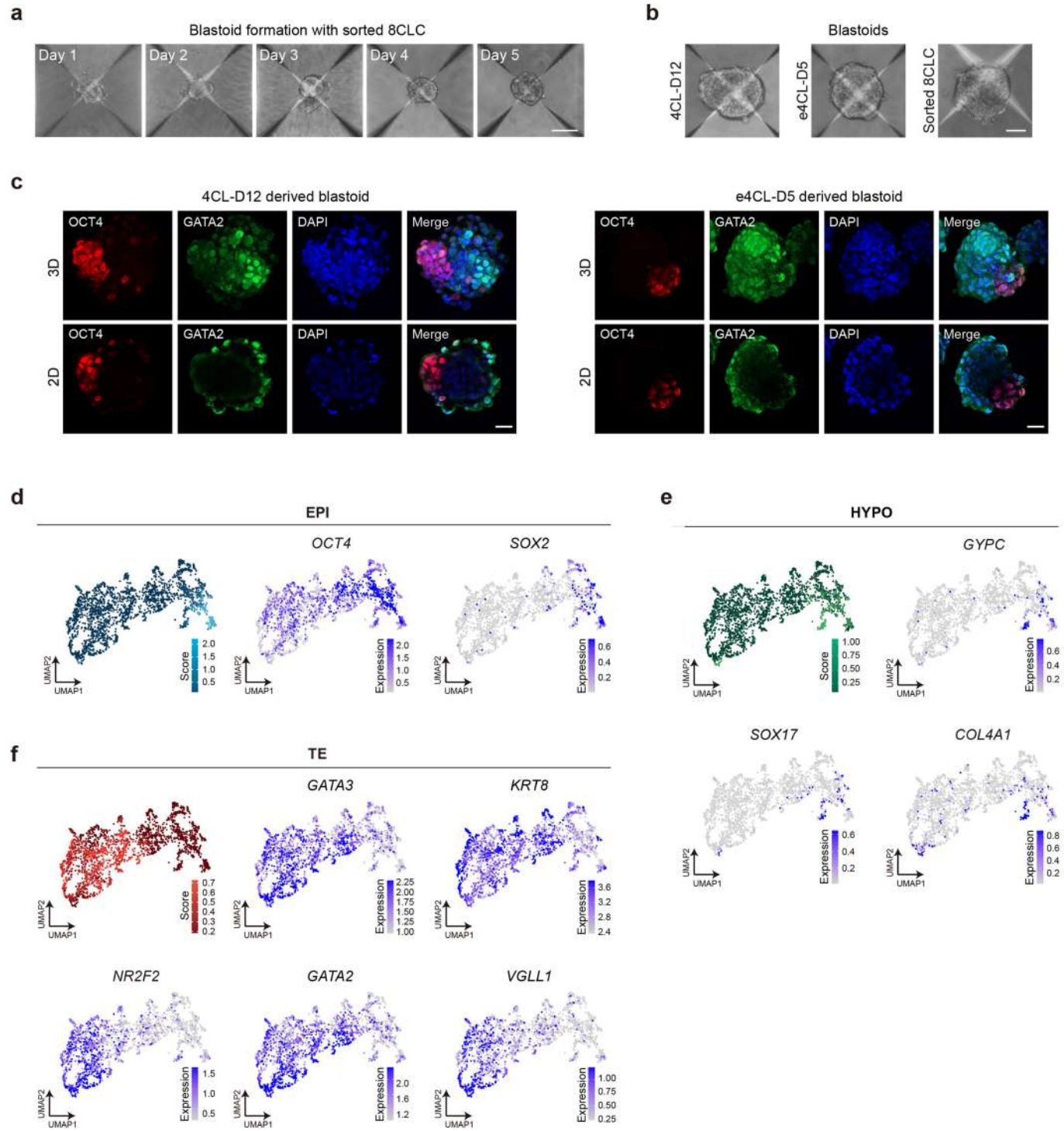
Extended Data Fig. 16 | Extended analysis of 8CLC induction regulators. **a**. Representative immunostaining images demonstrating the nucleocytoplasmic translocation of UHRF1 (UHRF1-mCherry) after DPPA3 overexpression in H9 ESC cultured in primed condition, 4CL (day 12) or e4CL (day 5). Nuclei were counterstained with DAPI. Scale, 20 µm. Representative of three independent experiments. **b**. RT-qPCR (upper panels, data are presented as mean values ± SD, n = 3 biological replicates) and western blot (lower panels, representative of three independent experiments.) analysis of UHRF1 mRNA and protein levels, respectively, in 4CL-day 12 naïve ESC (left), direct e4CL-day 7 cells (middle) and stepwise e4CL-day 5 cells (right) for the indicated samples. D, day. **c**. Violin plot showing global CpG DNA methylation levels assessed by RRBS in knockout *DPPA3* ESC compared to wild-type after conversion from a primed state to 4CL-day 12 naïve PSC or direct e4CL-day 7 cells. *DPPA3* knockout clone 1 was used. n = 3 technical replicates. The blue central line is the median, the boxes indicate the upper and lower quartiles. **d**. Unbiased clustering of differentially methylated CpG within 3 kb of the TSS in wild-type and *DPPA3* knockout (*DPPA3*$^{-/-}$) ESC converted from a primed state to 4CL-day 12 PSC (left panel; n = 3 technical replicates) or direct e4CL-day 7 cells (right panel; n = 3 technical replicates). *DPPA3* knockout clone 1 was used. **e**. Genome browser tracks showing CpG DNA methylation levels at the naïve pluripotency *KHDC3L* locus (left panel) and the totipotency locus *TRIM43* (right panel) in wild-type and *DPPA3* knockout (*DPPA3*$^{-/-}$) ESC converted from a primed state to 4CL-day 12 PSC. Each bar represents a single CpG and the height indicates the percentage of methylation. Bulk RNA-seq tracks showing *KHDC3L* and *TRIM43* expression are also included. **f**. Unbiased clustering of differentially methylated CpG in the intergenic regions in wild-type and *DPPA3* knockout (*DPPA3*$^{-/-}$) ESC converted from a primed state to 4CL-day 12 PSC (left panel; n = 3 technical replicates) or direct e4CL-day 7 cells (right panel; n = 3 technical replicates). Representative genes closest to the differentially methylated intergenic regions are shown. *DPPA3* knockout clone 1 was used. **g**. RT-qPCR showing expression of the indicated genes after shRNA lentivirus-mediated *Dppa3* knockdown (sh*Dppa3*-2 and sh*Dppa3*-4) compared to control (sh*Luc*) in mouse ESC cultured in serum+mLIF. Data are presented as mean values ± SD. n = 3 biological replicates. m, mouse. **h**. RT-qPCR showing the expression of the indicated genes in shRNA lentivirus-mediated knockdown of *Dppa3* (sh*Dppa3*-2 and sh*Dppa3*-4) compared to control (sh*Luc*) in mouse 2CLC converted in serum+mLIF with combined TSA and DZNep. Data are presented as mean values ± SD. n = 3 biological replicates. m, mouse.



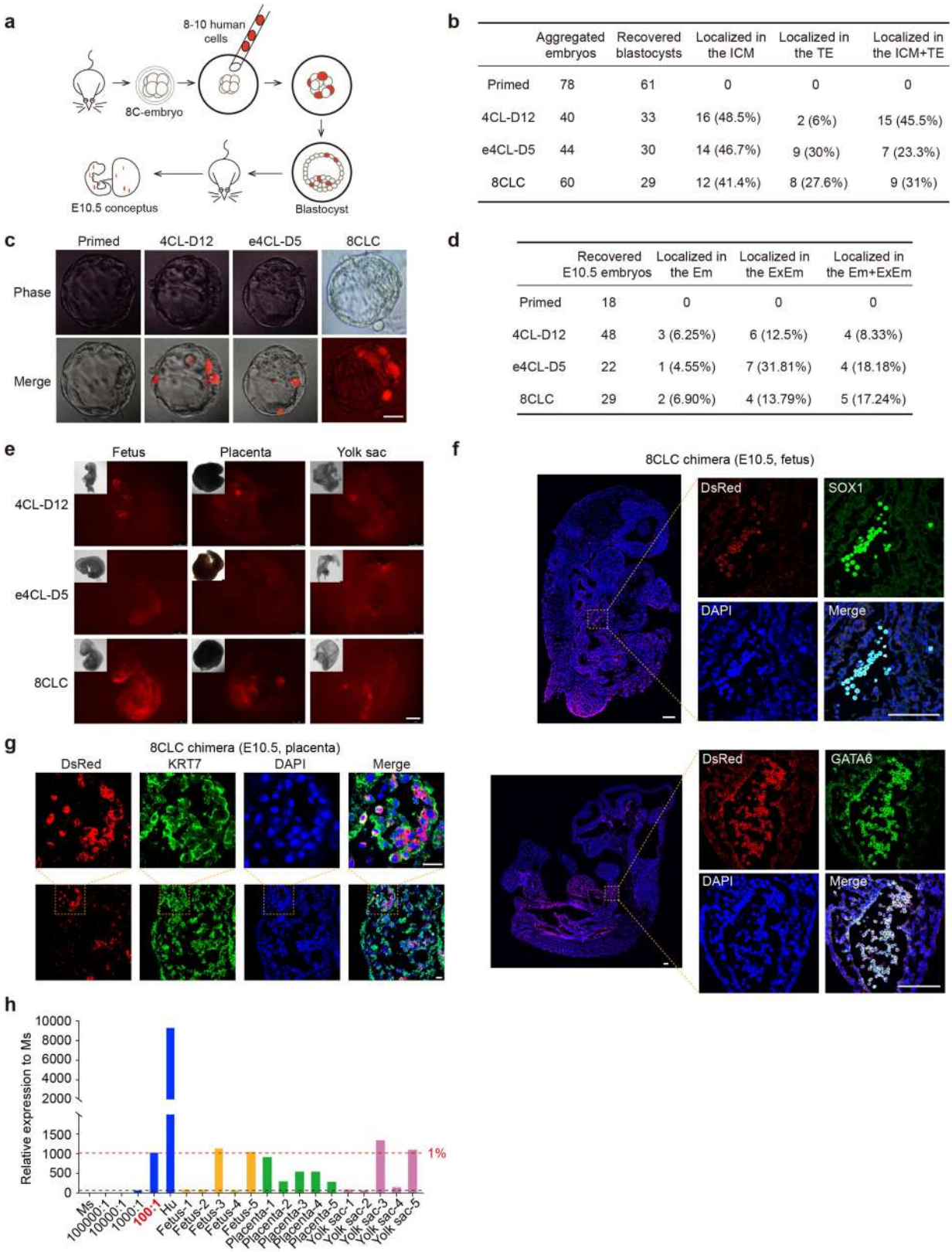
Extended Data Fig. 17 | Synergistic effect of TSA and DZNep on 8CLC induction. a. Heatmaps showing the global gene expression differences in H9 ESC cultured in stepwise e4CL at day 5 with or without (w/o) either DZNep or TSA. Example genes are shown for each cluster. n = 2 technical replicates. 8CLC network hub genes are highlighted in red. **b**. Heatmaps showing the expression of totipotency genes in human PSC cultured in direct e4CL at day 7 with or without DZNep or TSA. n = 2 technical replicates. 8CLC network hub genes are highlighted in red.



Extended Data Fig. 18 | Extraembryonic differentiation potency of primed PSC converted by 4CL and e4CL and sorted 8CLC. a. Representative phase contrast images showing TSC derived from sorted 8CLC in e4CL-day 5, stepwise e4CL-day 5 cells or 4CL-day 12 naïve PSC. H9 ESC were used for these experiments. Scale, 40 µm. D, day. Representative of four independent experiments. **b**. Representative immunostaining images for GATA3 (green) and KRT7 (red) in TSC derived from primed ESC converted by 4CL (day 12) or stepwise e4CL (day 5). Nuclei were counterstained with DAPI. Scale, 50 µm. D, day. Representative of three independent experiments. **c**. RT-qPCR for the indicated genes in TSC derived from primed ESC converted by 4CL (day 12) or stepwise e4CL (day 5). Data are presented as mean values ± SD of fold-change compared to PSC in 4CL or stepwise e4CL. n = 3 biological replicates. **d**. DNA methylation plots for the *ELF5* promoter in primed ESC untreated or converted by 4CL (day 12) and 4CL-day 12 naïve ESC-derived TSC. Percentages are the proportion of methylated (closed circles) to non-methylated (open circles) CpG sites. A representative experiment is shown. **e**. Hierarchical clustering of the bulk RNA-seq of primed ESC converted by 4CL (day 12), 4CL-day 12 naïve ESC-derived TSC and a primary human TSC dataset[54]. All reference datasets used in this study are summarized in Supplementary Table 8. Our dataset was generated using H9 ESC. **f**. Representative immunostaining images for CGB and SDC1 in SCT differentiated from 4CL-day 12 naïve PSC-derived TSC. Nuclei were counterstained with DAPI. Scale, 50 µm. Representative of three independent experiments. **g**. ELISA assay detecting the concentration of hCG secreted from 4CL-day 12 naïve ESC-derived TSC and SCT differentiated from 4CL-day 12 naïve ESC-derived TSC. Representative of two independent experiments.



Extended Data Fig. 19 | Blastoid formation capacity of 4CL naïve PSC, e4CL-day 5 cells and sorted 8CLC. a. Representative phase contrast images of sorted 8CLC-derived blastoids from day 1 to day 5. n = 5 biological replicates. Scale, 100 µm. **b**. Representative phase contrast images showing self-organized blastoids from sorted 8CLC, stepwise e4CL-day 5 cells and 4CL-day 12 naïve ESC. H9 ESC were used in all cases. D, day. Representative of five independent experiments. Scale, 50 µm. **c**. Representative immunofluorescence images of self-organized blastoids from 4CL-day 12 naïve ESC (left panel) and stepwise e4CL-day 5 cells (right panel). OCT4 and GATA2 were used as ICM and trophectoderm markers, respectively. Nuclei were counterstained with DAPI (blue). Scale, 20 µm. H9 ESC were used for these experiments. D, day. Representative of three independent experiments. **d**. UMAP visualization based on Fig. 4g highlighting the averaged expression scores of gene signatures and examples for the epiblast lineage in sorted 8CLC-derived blastoids and human E5–E7 blastocysts[20]. EPI, epiblast. All reference datasets used in this study are summarized in Supplementary Table 8. **e**. UMAP visualization based on Fig. 4g highlighting the averaged expression scores of gene signatures and examples for hypoblast lineage in sorted 8CLC-derived blastoids and human E5–E7 blastocysts. HYPO, hypoblast. **f**. UMAP visualization based on Fig. 4g highlighting the averaged expression scores of gene signatures and examples for the trophectoderm lineage in sorted 8CLC-derived blastoids and human E5–E7 blastocysts. TE, trophectoderm.

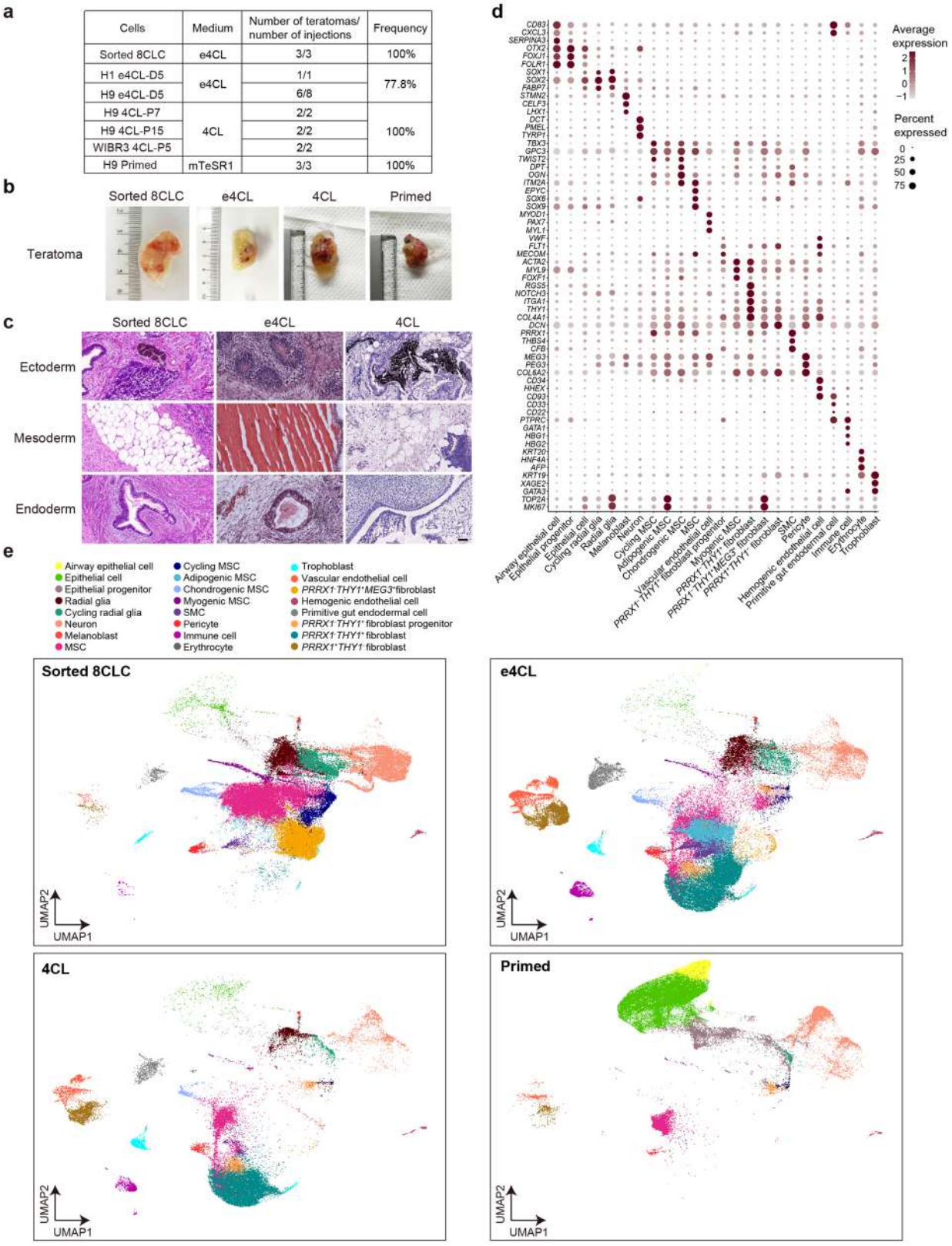


	Aggregated embryos	Recovered blastocysts	Localized in the ICM	Localized in the TE	Localized in the ICM+TE
Primed	78	61	0	0	0
4CL-D12	40	33	16 (48.5%)	2 (6%)	15 (45.5%)
e4CL-D5	44	30	14 (46.7%)	9 (30%)	7 (23.3%)
8CLC	60	29	12 (41.4%)	8 (27.6%)	9 (31%)

	Recovered E10.5 embryos	Localized in the Em	Localized in the ExEm	Localized in the Em+ExEm
Primed	18	0	0	0
4CL-D12	48	3 (6.25%)	6 (12.5%)	4 (8.33%)
e4CL-D5	22	1 (4.55%)	7 (31.81%)	4 (18.18%)
8CLC	29	2 (6.90%)	4 (13.79%)	5 (17.24%)

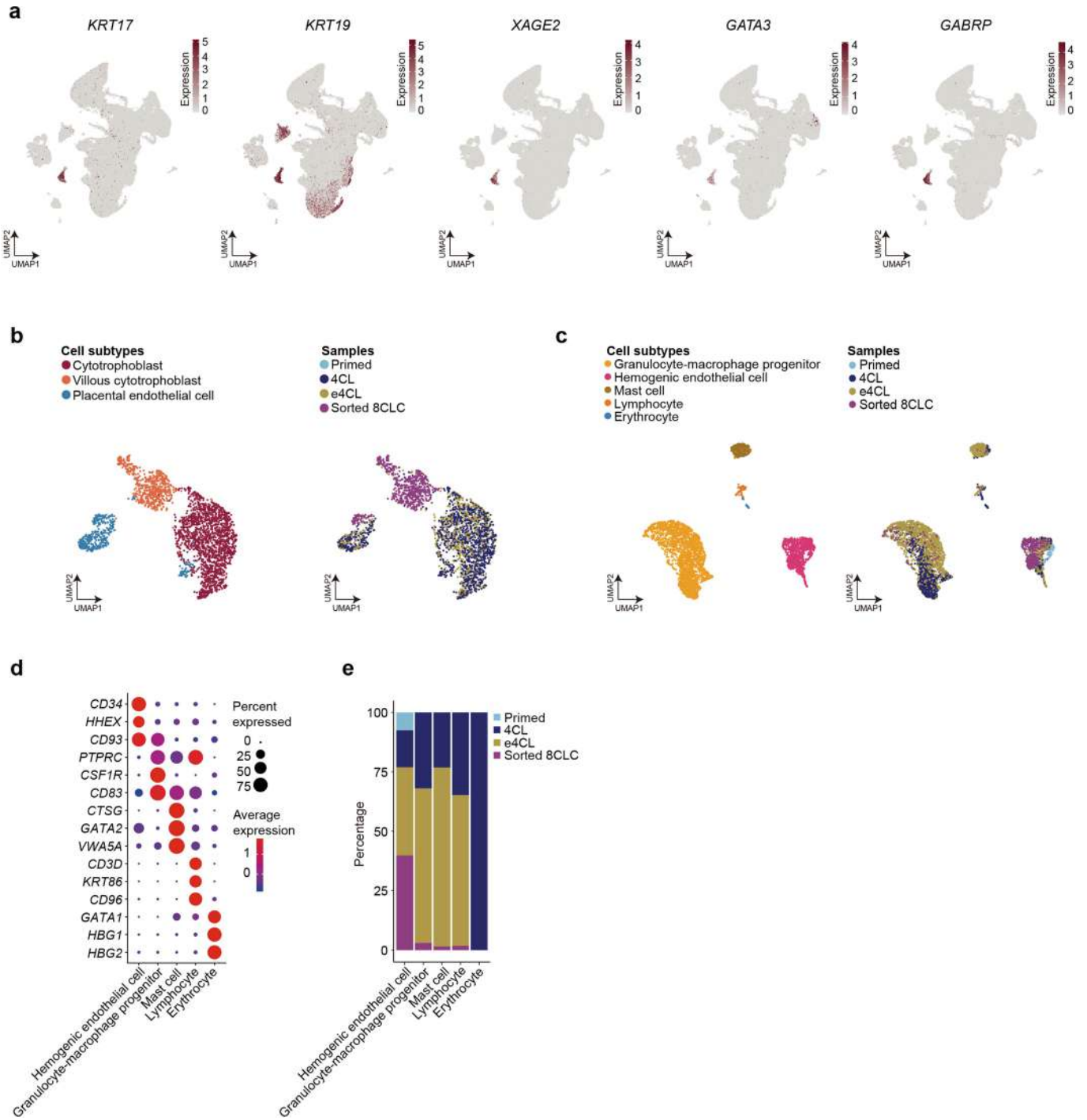
Extended Data Fig. 20 | See next page for caption.

Extended Data Fig. 20 | Interspecies chimeras using 4CL and e4CL and purified 8CLC. a. Schematic showing the workflow of the interspecies chimera assay by aggregation method[56]. Eight to ten cells for each condition (primed PSC, 4CL-day 12 naïve PSC and purified 8CLC derived from *TPRX1*-EGFP KI HN10 ESC labelled with DsRed [HN10-DsRed][57]) were aggregated with 8C-embryos of BDF1 mice. The aggregated embryos were cultured *in vitro* for 24 h to reach the blastocyst stage. Blastocysts were then implanted into ICR pseudopregnant mice and allowed to develop until E10.5. **b**. Summary of chimera assay results at blastocyst stage. TE, trophectoderm; D, day. **c**. Representative phase contrast (upper panel), fluorescence (DsRed, middle panel) and merged (lower panel) images showing the integration of DsRed$^+$ cells from the indicated injected cell types into mouse blastocysts. Scale, 20 µm. Representative of three independent experiments. **d**. Summary of chimera assay results at E10.5 mouse embryos. Em, embryonic lineage; ExEm, extraembryonic lineage. D, day. **e**. Representative fluorescence (DsRed) images showing cell integration into fetus, placenta, and yolk sac at E10.5 mouse embryos. Scale, 1 mm. Representative of three independent experiments. **f**. Representative images showing integration of DsRed$^+$ cells into mouse E10.5 embryos. Anti-DsRed antibodies were used together with anti-SOX1 (upper panel) or anti-GATA6 (lower panel) antibodies. Nuclei were counterstained with DAPI. Scale, 100 µm. Representative of three independent experiments. **g**. Representative immunofluorescence images showing integration of DsRed$^+$ cells into mouse E10.5 placenta. Anti-DsRed antibodies were used together with anti-KRT7 antibodies. Nuclei were counterstained with DAPI. Scale, 20 µm. Representative of three independent experiments. **h**. Quantitative PCR testing the presence of human mitochondrial DNA in fetus, placenta, and yolk sac of mouse E10.5 embryos. Chimeric embryos (E10.5) containing 8CLC were used for this assay. Ms, mouse; Hu, human.



Extended Data Fig. 21 | See next page for caption.

Extended Data Fig. 21 | Teratoma formation by 4CL naïve PSC, e4CL-day 5 cells and sorted 8CLC. a. Summary of teratoma assays using ESC primed and converted by 4CL or stepwise e4CL (day 5) and sorted 8CLC. D, day; P, passage. **b**. Images of teratomas derived from primed ESC converted by 4CL (passage 15, representative of two independent experiments) or stepwise e4CL (day 5, representative of eight independent experiments) and sorted 8CLC (representative of three independent experiments). H9 ESC were used for this assay. D, day; P, passage. **c**. Hematoxylin and eosin staining of teratomas derived from primed ESC converted by 4CL (passage 15, representative of two independent experiments) or stepwise e4CL (day 5, representative of eight independent experiments) and sorted 8CLC (representative of three independent experiments). Representative images of tissues corresponding to the three germ layers are shown. Scale, 50 µm. H9 ESC were used for this assay. D, day; P, passage. **d**. Bubble plot representing the frequency of expression and average expression of the marker set used to identify the cell types for each lineage. **e**. UMAP visualization based on Fig. 5a highlighting the identified cell types in the scRNA-seq datasets generated from primed ESC, 4CL naïve PSC, stepwise e4CL-day 5 cells and sorted 8CLC.



Extended Data Fig. 22 | Extended analysis of teratoma formation. a. UMAP visualization based on Fig. 5a highlighting the distribution and the expression of related markers in trophoblast cells from the indicated teratomas. **b**. UMAP visualization showing the annotated sub-clusters of extraembryonic lineages (left panel) and the relative contribution of teratomas from sorted 8CLC, e4CL-day 5 cells, 4CL naïve ESC and primed ESC (right panel). **c**. UMAP visualization showing the annotated sub-clusters of immune cells (left panel) and the contribution of primed PSC, 4CL naïve ESC, stepwise e4CL-day 5 cells and sorted 8CLC-derived teratoma cells (right panel). **d**. Bubble plot representing the frequency of expression and average expression of marker genes in different immune cell subtypes. **e**. Bar plot showing the contribution of sorted 8CLC, stepwise e4CL-day 5, 4CL naïve and primed PSC-derived teratomas to different immune cell subtypes.

Corresponding author(s): Mazid Md. Abdul, Longqi Liu, Wenjuan Li, Miguel A. Esteban

Last updated by author(s): 2022/03/08

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	Zeiss LSM 710 NLO for fluorescence microscope imaging, Nikon Eclipse Ti microscope for fluorescence imaging of embryos, BD LSRFortessa X-20 cell analyzer for flow cytometry, ABI7500 real-time PCR machine for RT-qPCR, Illumina Xten platform for RRBS, DIPSEQ T1 sequencer for droplet-based scRNA-seq and SMART-seq2. BGISEQ-500 sequencer for scATAC-seq and single-cell multiomics sequencing. Illumina Nova 6000 sequencer for WGBS.
Data analysis	Data were analyzed using: GraphPad Prism (v8.3.1), FlowJo (V10.4), Trim Galore (v0.5.0), cutadapt (v1.14), FastQC (v0.11.2), fastp (v0.21.0), RSEM (v1.3.1), STAR (v2.7.3a, v2.7.4a), DESeq2 (v1.24), deeptools (v3.1.3), EDASeq R package (v2.18.0), BS-Seeker2 (v2.1.0), wigToBigWig (v4), CGmapTools (v0.1.2), PISA (v0.2), sambamba (v0.7.0), Seurat package (v3.1.4), psupertime (v0.2.6), IGV (v2.1.0), Cell Ranger Arc (2.0.0), Signac and ArchR (v0.9.5), Cytoscape (v3.9.0), and BD FACSDiva 8.0.2 for FACS.

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

All data needed to evaluate the conclusions in the paper are included in the paper and/or the Supplementary Materials. Raw sequencing data have been deposited

in the CNGB Nucleotide Sequence Archive under the accession number CNP0001454 (https://db.cngb.org/search/project/CNP0001454/). Reference datasets from other works can be accessed either in NCBI Gene Expression Omnibus (GEO) repository under their GSE numbers or in the database of the EMBL's European Bioinformatics Institute under their E-MTAB numbers: accession numbers of human embryo scRNA-seq data are E-MTAB-3929 and GSE36552; the accession number of human embryo ATAC-seq data is GSE101571; the accession number of human embryo histone CUT&RUN data is GSE124718; accession numbers of other naïve/EPSC media single-cell sequencing data are GSE150311 (RSeT and t2iLGö), GSE166422 (PXGL), GSE150578 (5iLA); accession numbers of other naïve/EPSC media RNA-seq data are GSE52617 (NHSM), E-MTAB-2857 (t2iLGö), GSE59435 (5iLAF), E-MTAB-7254 (EPSC); accession numbers of DNA methylation data are GSE49828 (human embryo), GSE52617 (NHSM), GSE60945 (t2iLGö-NK2), GSE111018 (5iLAF), GSE136715 (mouse early embryo) and GSE75751 (mouse ESC and 2CLC); the accession numbers of mouse early embryo scRNA-seq data is GSE45719; the accession number of mouse ESC and 2CLC is GSE168728; and the accession number of TSC RNA-seq data is GSE138762. Human reference genome is available at http://ftp.ensembl.org/pub/release-105/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna_rm.alt.fa.gz. All reference datasets with links are summarized in Supplementary Table 8. Processed datasets with links and descriptions are summarized in Supplementary Table 11. Processed bulk RNA-seq data can be accessed at https://figshare.com/s/8ee132ff1366fa89d35a, processed scRNA-seq data can be accessed at https://figshare.com/s/34110eebb58462a79dd5, processed scATAC-seq data can be accessed at https://figshare.com/s/760d3ff54f1214a50cc2, processed RRBS and WGBS data can be accessed at https://figshare.com/s/ff707bf8242f7b3ed8f5, processed teratoma and blastoid scRNA-seq data can be accessed at https://figshare.com/s/037b348b1da763fb41d0, processed SMART-seq2 data can be accessed at https://figshare.com/s/a1b03a1463865b8a56c8, processed single-cell multiomics data can be accessed at https://figshare.com/s/9c01c3b58d34b80de230. 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	The sample size was estimated from the preliminary experiments or from our previously published studies. No statistical method was applied to predetermine sample size. The RT-qPCR, FACS and ELISA were conducted with at least three independent experiments except otherwise stated. Single-cell sequencing were performed independently with scRNA-seq, scATAC-seq, multiomics scRNA-seq, and SMART-seq2, and their results are consistent. Bulk RNA-seq, RRBS and WGBS were conducted with two or three independent samples. Animal or embryo experiments have been performed at least three times independently. Mice or embryo number used in each experiment was determined by data consistency in control and experiment group, resources available and ethical reduction of animal use. Please refer to figure legends and methods for details. All replications were successful.
Data exclusions	No data were excluded.
Replication	Each experiment was reproduced in at least 3 biological replicates if not it is otherwise stated. Please refer to figure legends and methods for details. All replications were successful.
Randomization	Mice were randomly allocated into experimental groups. No randomization methods were utilized for other experiments as all samples are randomly collected from the population, sample collections were performed with controls, all replications were successful and no data were excluded.
Blinding	No blinding was used. Blinding was not used because all analysis was performed on all relevant samples in reproducible ways so no bias could have been removed by blinding. Besides, all the experiments have been performed by at least two independent researchers.

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
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Antibodies

Antibodies used

Antibodies used:
Details of all antibodies used in this study were provided in Supplementary Table 10 with catalog number, and commercial sources supplied.

For Immunostaining:
Antibody, Company, Catalogue Number, Dilution
TPRX1 antibody, Novus Biologicals, Cat# NBP1-92524, 1:100
KLF17 antibody, Sigma-Aldrich, Cat# HPA024629, 1:200
OCT-3/4 antibody, Santa Cruz Biotechnology , Cat# sc-5279, 1:200
SOX2 antibody, R&D Systems, Cat# AF2018, 1:50
GATA-3 antibody, R&D Systems, Cat# AF2605, 1:200
Cytokeratin 7 (KRT7) antibody, Dako, Cat# M7018 , 1:200
Human chorionic gonadotropin antibody, Dako, Cat# IS50830-2, 1:50
SDC1 antibody, Abcam, Cat# ab39969, 1:200
GATA-6 antibody, Cell Signaling Technology, Cat# 5851, 1:200
SOX1 antibody, R&D Systems, Cat# AF3369, 1:100
NANOG antibody, Abcam, Cat# ab21624, 1:200
Flag M2 mouse monoclonal antibody, clone M2, Sigma-Aldrich, Cat# F1804, 1:200
Living Colors® DsRed antibody, Takara Bio, Cat# 632496, 1:200
Human AP-2 gamma antibody R&D Systems Cat# AF5059, RRID:AB_2255891
GATA2 antibody Thermo Fisher Scientific Cat# PA1-100, RRID:AB_2539871
Alexa Fluor 488-AffifiniPure Donkey Anti-Mouse IgG, Jackson ImmunoResearch, Cat# 715-545-150, 1:500
Alexa Fluor 488-AffifiniPure Donkey Anti-Rabbit IgG, Jackson ImmunoResearch, Cat# 711-545-152, 1:500
Alexa Fluor 488-AffifiniPure Donkey Anti-Goat IgG, Jackson ImmunoResearch Cat# 705-545-147, 1:500
Cy3-AffifiniPure Donkey Anti-Rabbit IgG, Jackson ImmunoResearch, Cat# 711-165-152, 1:500
Cy3-AffifiniPure Donkey Anti-Mouse IgG, Jackson ImmunoResearch, Cat# 715-545-150, 1:500
Cy3-AffifiniPure Donkey Anti-Goat IgG, Jackson ImmunoResearch, Cat# 705-165-147, 1:500

For Western Blot:
Antibody, Company, Catalogue Number, Dilution
UHRF1 antibody, GeneTex, Cat# GTX113963, 1:1000
ACTIN antibody, Sigma-Aldrich, Cat# A2066, 1:2000
Goat anti-Rabbit IgG (H+L) Secondary Antibody, HRP, Thermo Fisher Scientific, Cat# 32460, 1:5000

Validation

Antibodies obtained from the commercial source were validated by the suppliers, detailed validation analysis relevant literatures are provided on the company website for the products used in this study. Some antibodies were validated in a previously published study as indicated in methods or relevant literature was cited.
1. TPRX1 (NBP1-92524) https://www.novusbio.com/products/tprx1-antibody_nbp1-92524
2. KLF17 (HPA024629) https://www.sigmaaldrich.cn/CN/zh/product/sigma/hpa024629
3. OCT-3/4 (sc-5279) https://www.scbt.com/p/oct-3-4-antibody-c-10
4. SOX2(AF2018)https://www.rndsystems.com/cn/products/human-mouse-rat-sox2-antibody_af2018
5. GATA-3 (AF2605) https://www.rndsystems.com/cn/products/human-gata-3-antibody_af2605
6. KRT7 (M7018) https://www.labome.com/product/Dako/M7018.html
7. hCG (IS50830-2) https://www.agilent.com/en/dako-products
8. SDC1 (ab39969) https://www.abcam.cn/scd1-antibody-ab39969.html
9. GATA-6 (5851) https://www.cellsignal.cn/products/primary-antibodies/gata-6-d61e4-xp-rabbit-mab/5851
10. SOX1 (AF3369) https://www.rndsystems.com/cn/products/human-mouse-rat-sox1-antibody_af3369
11. NANOG (ab21624) https://www.abcam.com/nanog-antibody-ab21624.html
12. FLAG (F1804) https://www.sigmaaldrich.com/IN/en/product/sigma/f1804
13. DsRed (632496) https://www.labome.com/product/Takara-Bio-Clontech/632496.html
14. AP-2 gamma (AF5059) https://www.rndsystems.com/cn/products/human-ap-2gamma-antibody_af5059
15. GATA2 (PA1-100) https://www.thermofisher.cn/cn/zh/antibody/product/GATA2-Antibody-Polyclonal/PA1-100
16. Alexa Fluor 488-AffifiniPure Donkey Anti-Mouse IgG (715-545-150) https://www.jacksonimmuno.com/catalog/products/715-545-150
17. Alexa Fluor 488-AffifiniPure Donkey Anti-Rabbit IgG (711-545-152) https://www.jacksonimmuno.com/catalog/products/715-545-152
18. Alexa Fluor 488-AffifiniPure Donkey Anti-Goat IgG (705-545-147) https://www.jacksonimmuno.com/catalog/products/715-545-147
19. Cy3-AffifiniPure Donkey Anti-Rabbit IgG (711-165-152) https://www.jacksonimmuno.com/catalog/products/711-165-152
20. Cy3-AffifiniPure Donkey Anti-Mouse IgG (715-545-150) https://www.jacksonimmuno.com/catalog/products/715-545-150
21. Cy3-AffifiniPure Donkey Anti-Goat IgG (705-165-147) https://www.jacksonimmuno.com/catalog/products/705-165-147
22. UHRF1 (GTX113963) https://www.genetex.com/Product/Detail/UHRF1-antibody/GTX113963
23. Actin (A2066) https://www.sigmaaldrich.com/IN/en/product/sigma/a2066
24. Goat anti-Rabbit IgG (H+L) Secondary Antibody, HRP (32460) https://www.thermofisher.cn/cn/zh/antibody/product/Goat-anti-Rabbit-IgG-H-L-Secondary-Antibody-Polyclonal/32460

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

WIBR3 human ESC were kindly provided by Dr. Jacob Hanna (Weizmann Institute of Science, Israel), UH10 (iPSC-4) human

Cell line source(s)	iPSC and HN10-DsRed human ESC by Dr. Guangjin Pan (Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, China), Phoenix (iPSC-2) human iPSC by Dr. Ulrich Martin (Hannover Medical School, Germany). Primed human ESC including HUES1 and HUES7 and iPSC including DiPS 1016SevA (iPSC-1) and STiPS O-XX1 (iPSC-3) were purchased from Howard Hughes Medical Institute (USA) and H1 and H9 human ESC from WiCell Research Institute (USA). HEK293T were purchased from ATCC. MuERV-L-LTR-GFP ESC reporter cell line was kindly provided by Dr. Man Zhang (Guangzhou Laboratory, Guangzhou, China).
Authentication	HEK293T cells were validated by ATCC. Primed human ESC including HUES1 and HUES7, and iPSC including DiPS 1016SevA (iPSC-1) and STiPS O-XX1 (iPSC-3) were validated by Howard Hughes Medical Institute (USA). H1 ESC was validated by WiCell Research Institute. WIBR3 was validated by Dr. Jacob Hanna. Phoenix iPSC were validated by Dr. Ulrich Martin. MuERV-L-LTR-GFP mESC was authenticated by Dr. Man Zhang. We authenticated H9 ESC, HN10 ESC and UH10 iPSC by STR analysis.
Mycoplasma contamination	Yes, they all were monthly tested and were found devoid of mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	To our knowledge, no cell lines used in this study are listed in the database of commonly misidentified cell lines maintained by ICLAC.

Animals and other organisms

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

Laboratory animals	6 to 8-week-old male NOD-scid-IL2Rg−/− mouse (Mus musculus); female BDF1 mouse; female ICR mice. Mice were housed in a temperature (20-26 °C)-, humidity (40-70)- and light-controlled (12 hour light/dark cycle) specific pathogen-free (SPF) animal facility, in individually ventilated cages.
Wild animals	No wild animals were used.
Field-collected samples	No field-collected samples were used.
Ethics oversight	All animal experiments were performed in accordance with the ethical guidelines of the Guangzhou Institutes of Biomedicine and Health. Human-mouse chimera and human blastoid experiments were approved and followed up by the Animal Care and Use Committee and Human Subject Research Ethics Committee under license numbers IACUC2016012 and GIBH-IRB2020-034, respectively, of the Guangzhou Institutes of Biomedicine and Health. These committees are constituted of experts in different fields (including scientists working on development and other disciplines, and non-scientists [doctors and lawyers]), and they evaluated the rationale of the experiment plan, origins and consents of human materials, and the qualification of the investigators. Several of these experts followed the study until completion. Both experiments followed relevant international regulations including the 2016 Guidelines for Stem Cell Research and Clinical Translation released by the International Society for Stem Cell Research (ISSCR). A specialized review process was activated in the human-mouse chimera experiments to assess whether the degree of functional integration was high enough to raise concerns that the nature of the chimeric animal would be substantially changed. Human blastoid experiments were subjected to an embryo research oversight process. Several of these experts followed the study until completion.

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	For flow cytometry, cells were trypsinized, washed twice with DPBS and filtered through a Falcon® 40 µm cell strainer (Falcon, 352340). Cells were resuspended with 0.04% BSA in PBS. They were loaded into a BD LSRFortessa X-20 cell analyzer. For cell sorting using TPRX1-EGFP reporter cells in stepwise e4CL-day 5, cells were trypsinized, resuspended with 0.04% BSA and loaded into a BD FACSAria III for sorting.
Instrument	BD LSRFortessa X-20 cell analyzer was used for FACS analysis and BD FACSAria III for cell sorting.
Software	Flow cytometry data were collected with BD FACSDiva™ software, and were analyzed by Flowjo (v10.4).
Cell population abundance	Distinct cell populations were determined using controls.

Gating strategy	Cell debris were excluded using a FSC-A vs SSC-A gate, aggregates were excluded using a FSC-H vs FSC-A gate and the boundary between negative and positive was defined by the isotype control.

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