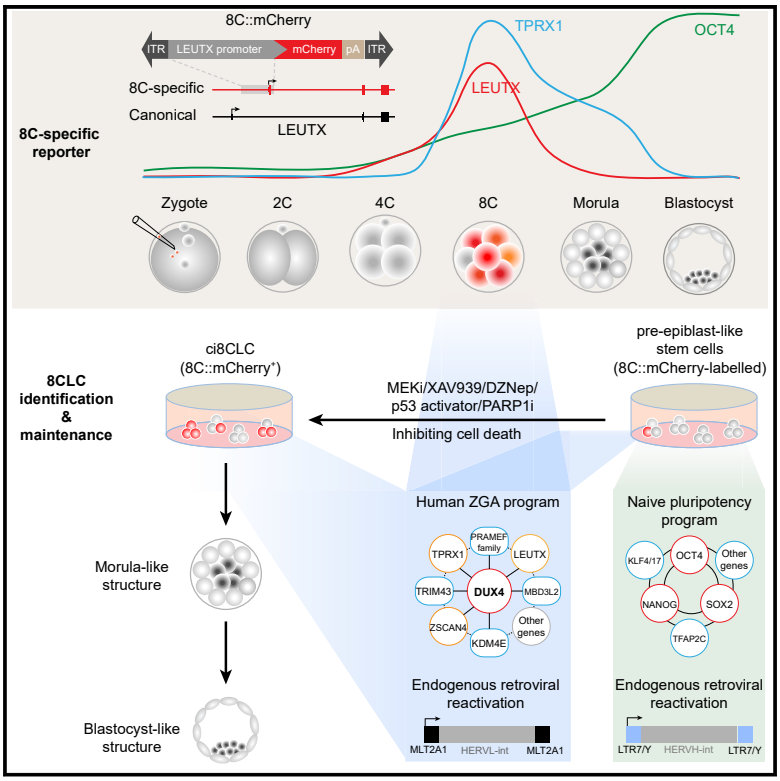
Cell Reports

Article

Recapitulating early human development with 8C-like cells

Graphical abstract



Highlights

- A human 8C-specific reporter is developed to isolate 8CLCs from human prEpiSC cultures
- A chemical-based culture condition increases and maintains the 8CLC population
- 8CLCs can self-organize to form blastocyst-like structures

Authors

Xiu Yu, Shiqi Liang, Manqi Chen, ..., Chuanbo Sun, Mingzhu Yang, Jichang Wang

Correspondence

bobo_1110@163.com (C.S.), yangmzh9@mail2.sysu.edu.cn (M.Y.), wangjch53@mail.sysu.edu.cn (J.W.)

In brief

By using one human eight-cell (8C) embryo-specific reporter, Yu et al. isolated 8C-like cells from human preimplantation epiblast-like stem cells and developed one chemical-based culture condition to increase and maintain the 8CLC population. 8CLCs capture the major zygotic genome activation program and can be used to recapitulate early human development.

Yu et al., 2022, Cell Reports *39*, 110994
June 21, 2022
https://doi.org/10.1016/j.celrep.2022.110994

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Recapitulating early human development with 8C-like cells

Xiu Yu,[1,2,7] Shiqi Liang,[1,2,7] Manqi Chen,[1,2,7] Hanwen Yu,[1,2] Ruiqi Li,[3] Yuliang Qu,[1,2] Xuhui Kong,[1,2] Ruirui Guo,[1,2] Rongyan Zheng,[1,2] Zsuzsanna Izsvák,[4] Chuanbo Sun,[5,*] Mingzhu Yang,[6,7,*] and Jichang Wang[1,2,8,*]

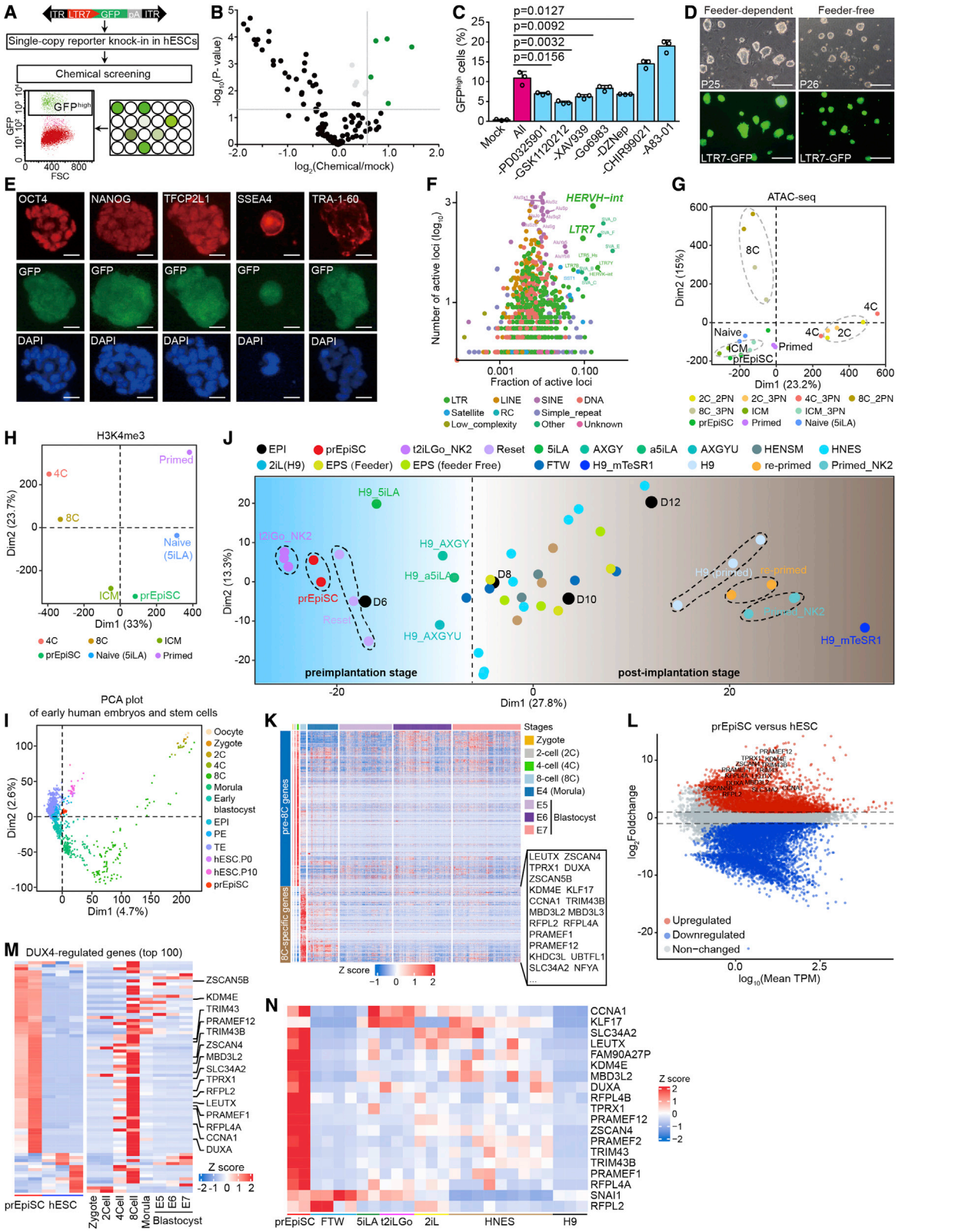
[1]Department of Histology and Embryology, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China
[2]Key Laboratory for Stem Cells and Tissue Engineering, Ministry of Education, Sun Yat-sen University, Guangzhou 510080, China
[3]Department of Obstetrics and Gynaecology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, China
[4]Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany
[5]Laboratory of Medical Systems Biology, Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou 510623, China
[6]State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou 510060, China
[7]These authors contributed equally
[8]Lead contact
*Correspondence: bobo_1110@163.com (C.S.), yangmzh9@mail2.sysu.edu.cn (M.Y.), wangjch53@mail.sysu.edu.cn (J.W.)
https://doi.org/10.1016/j.celrep.2022.110994

SUMMARY

In human embryos, major zygotic genome activation (ZGA) initiates at the eight-cell (8C) stage. Abnormal ZGA leads to developmental defects and even contributes to the failure of human blastocyst formation or implantation. An *in vitro* cell model mimicking human 8C blastomeres would be invaluable to understanding the mechanisms regulating key biological events during early human development. Using the non-canonical promoter of LEUTX that putatively regulates human ZGA, we developed an 8C::mCherry reporter, which specifically marks the 8C state, to isolate rare 8C-like cells (8CLCs) from human preimplantation epiblast-like stem cells. The 8CLCs express a panel of human ZGA genes and have a unique transcriptome resembling that of the human 8C embryo. Using the 8C::mCherry reporter, we further optimize the chemical-based culture condition to increase and maintain the 8CLC population. Functionally, 8CLCs can self-organize to form blastocyst-like structures. The discovery and maintenance of 8CLCs provide an opportunity to recapitulate early human development.

INTRODUCTION

After fertilization, the quiescent zygotic genome initiates bursts of transcription termed zygotic genome activation (ZGA) (Jukam et al., 2017; Lee et al., 2014; Ortega et al., 2018; Palfy et al., 2017; Schulz and Harrison, 2019). The major ZGA occurs at the late two-cell (L2C) stage in mice and eight-cell (8C) stage in humans, which is essential for subsequent lineage segregation (Jukam et al., 2017; Lee et al., 2014; Ortega et al., 2018; Palfy et al., 2017; Schulz and Harrison, 2019). The onset of ZGA depends on intricately coordinated mechanisms in which transcription factors and epigenetic regulation (e.g., DNA methylation, histone post-translational modification) play crucial roles (Ortega et al., 2018; Palfy et al., 2017; Schulz and Harrison, 2019). Advances in low-input multi-omics techniques have recently led to the discoveries of several mammalian genome activators. Intriguingly, DUX4/DUX family genes, which are specific to placental mammals, are critical drivers of ZGA in both mice and humans (De Iaco et al., 2017; Hendrickson et al., 2017; Whiddon et al., 2017). Recent studies also have shown that the orthologous genes have distinct functions in regulating ZGA in mammals. For instance, OCT4 knockdown results in downregulation of hundreds of ZGA genes in human embryos, implicating its essential roles in human ZGA (Fogarty et al., 2017; Gao et al., 2018), whereas OCT4 is not involved in ZGA in mice (Gao et al., 2018). On the other hand, misexpression of these ZGA regulators has serious consequences for embryonic development. OCT4 depletion at the pre-ZGA stage compromises human blastocyst formation but weakly affects mouse preimplantation development (Fogarty et al., 2017; Gao et al., 2018). DUX4 misexpression in human skeletal muscle cells, where it activates aberrant expression of ZGA genes, causes an untreatable form of muscular dystrophy called facioscapulohumeral dystrophy (Geng et al., 2012; Karpukhina et al., 2021; Lemmers et al., 2010). Besides transcription factors, proper epigenetic regulation is also critical for ZGA. Previous studies suggest that failed H3K9me3 reprogramming impedes ZGA in mammalian embryos generated by somatic cell nuclear transfer, compromising blastocyst formation (Chen et al., 2020; Chung et al., 2015; Deng et al., 2020; Matoba et al., 2014). Together, abnormal ZGA contributes to



(legend on next page)

substantial lethality of human embryos, and even may bring about long-term detrimental consequences on health. Therefore, understanding the regulatory mechanism of human ZGA is important for human reproduction and health. The studies on human embryos, however, are experimentally challenged and ethically restricted.

The derivation of pluripotent stem cells facilitates research about mammalian early development. A rare cell subpopulation is identified in mouse embryonic stem cell (mESC) culture, which is similar to the mouse late two-cell embryo, thus called 2C-like cells (2CLCs) (Macfarlan et al., 2012). The ZGA-like transcriptional program is reactivated in 2CLCs (Iturbide and Torres-Padilla, 2020; Macfarlan et al., 2012; Rodriguez-Terrones et al., 2018). Moreover, based on the 2CLC model, we and two other groups establish the mouse totipotent-like stem cell lines that faithfully recapitulate mouse two-cell embryos and have more robust developmental potency than mESCs (Shen et al., 2021; Xu et al., 2022; Yang et al., 2022). This unique cell type with 2C-like transcriptome holds promise to identify ZGA regulators and explore the regulatory mechanisms of early mouse development. However, to date the early blastomere-like cells have not been captured from pluripotent stem cell cultures in other species, including humans. Clearly, one suitable cell model like 2CLCs is crucial for dissecting the mechanisms that regulate key biological events (e.g., ZGA, lineage segregation) during early human development. Thus, it is worth exploring whether such early human blastomere-like cells can be cultured *in vitro*.

In this study, based on single-cell RNA-sequencing (scRNA-seq) analysis, we developed an 8C-specific mCherry reporter (8C::mCherry) and successfully captured the rare human 8C-like cells (8CLCs) from human preimplantation epiblast-like stem cells, in parallel with the other two latest reports (Mazid et al., 2022; Taubenschmid-Stowers et al., 2022). These 8CLCs express multiple signature genes and retrotransposons of human 8C embryos and exhibit a human ZGA-like transcriptional program. By chemical screening, we further optimized the culture condition to improve and maintain the 8CLC population in cultures. Finally, we evaluated the developmental potential of 8CLCs by the blastoid generation assay.

RESULTS

Derivation of preimplantation epiblast-like hPSCs expressing multiple 8C-specific genes

In mice, 2CLCs were discovered specifically from naive pluripotent stem cells (PSCs) representing the preimplantation epiblast (pre-EPI). We asked whether the PSCs representing the human pre-EPI might contain 8CLCs. To address this question, we first generated human pre-EPI-like stem cells (prEpiSCs) by chemical screening. Using the LTR7/HERVH-based, pluripotency reporter (pT2-LTR7-GFP) (Wang et al., 2014) (Figure 1A), we conducted chemical screening on human embryonic stem cells (hESCs) cultured in the N2B27 medium with a selected small molecule pool targeting multiple signaling pathways enriched in hESCs but underrepresented in the human pre-EPI (Figure S1A and STAR Methods). As interleukin (IL)6R rather than LIFR is expressed in human pre-EPI (Figure S1A), the cytokine IL6 (but not LIF) was added into the N2B27 medium. First, we screened the chemicals one by one, and selected those that could enhance the reporter GFP signal and increase the GFP^+ cell number in human PSC (hPSC) culture (Figure 1B). In the second round of screening, we identified the essential compounds to maintain the GFP^+ cells in hPSC culture. This strategy revealed that combined inhibition of MEK (GSK1120212), WNT (XAV939), and PKC (Go6983) signaling pathways as well as transiently inhibiting Src (A419259) and deposition of H3K27me3 marks (DZNep) were suitable for stable maintenance of GFP^+ cells in hPSC culture (Figures 1C and S1B). Of note, the bone morphogenetic protein (BMP) signaling pathway seems to be required for prEpiSC induction from hESCs, as BMP4 promoted, whereas the BMP inhibitor (K02288), blocked this induction process (Figure S1C).

We observed that once GFP^+ hPSCs were stabilized (indicated by bright domed-shape colonies without differentiation), DZNep and A419259 could be removed from the medium for the long-term culturing, while GFP^+ hPSC colonies were differentiated or died once GSK1120212, XAV939, and Go6983 were withdrawn from the medium (Figure S1D). With this strategy, we derived an hPSC line mimicking human pre-EPI, which was featured by HERVH hypertranscription and high expression of pre-EPI-specific genes (Figures 1E and 1F, S1E, and S1F), thus termed

Figure 1. Derivation and characterization of prEpiSCs

(A) The schematic for the screening strategy of culture conditions supporting prEpiSCs.

(B) The scatterplot showing effects of individual chemicals on LTR7-GFP^{high} hPSC percentage.

(C) Identification of chemicals essential for prEpiSC induction. Data are shown as means ± SD (n = 3 biological replicates). Statistical analyses are performed using unpaired, two-tailed t test.

(D) The images represent long-term maintenance of prEpiSCs in both feeder-dependent and independent conditions. Scale bar, 100 μm.

(E) Immunofluorescence analysis of pluripotent marker expression in prEpiSCs. Scale bar, 20 μm.

(F) The expression of transposable elements in prEpiSCs. The HERVH family is one of most abundant expressed retrotransposons.

(G) PCA of ATAC-seq data showing similarities in chromatin accessibilities between prEpiSCs and human ICM.

(H) PCA of genome-wide H3K4me3 signals in human preimplantation embryos and multiple stem cell lines (prEpiSCs, naive [5iLA] and primed hESCs).

(I) PCA of global expressed genes of prEpiSCs and human preimplantation embryos, detected by scRNA-seq.

(J) PCA of global expressed genes of prEpiSCs, multiple stem cell lines and human epiblast (EPI) cells.

(K) Definition of human 8C-specific genes by scRNA-seq of early human embryos.

(L) The plot showing the differentially expressed genes between prEpiSCs and hESCs. Multiple human 8C-specific genes are marked.

(M) The heatmap showing relative expression levels of top100 DUX4-regulated genes (Hendrickson et al., 2017) in hESCs, prEpiSCs, and early human embryos. Multiple human 8C-specific genes are marked.

(N) The heatmap showing relative expression levels of DUX4-regulated 8C-specific genes in prEpiSCs, naive (5iLA, t2iLGo, and 2iL), primed, and formative hESCs. See also Figure S1 and Table S1.

pre-EPI-like hPSCs (prEpiSC). However, unlike other naive hPSCs (e.g., 5iLA, t2iLGo hPSCs) (Pastor et al., 2016), most of prEpiSCs expressed SSEA4 instead of SSEA1 (Figure 1E), consistent with that SSEA4 is expressed in human inner cell mass (ICM) (Henderson et al., 2002). Unlike the primed hPSCs, prEpiSCs could be passaged as single cells and formed dome-shaped colonies under either feeder-dependent or feeder-free conditions (Figure 1D). Importantly, while other naive hPSCs usually exhibit quite abnormal karyotypes after 10 to 12 passages (Pastor et al., 2016), no karyotype aberration was observed in most of prEpiSCs even after 80 passages (normal/abnormal: 359/6) (Figure S1G). In addition, prEpiSCs could be readily and directly differentiated into three germ layers (Figure S1H).

To further characterize prEpiSCs, we performed comparative epigenome and transcriptome analyses between prEpiSCs and epiblast cells from human embryos as well as other hPSC lines. The assay for transposase-accessible chromatin using sequencing (ATAC-seq) analysis revealed that prEpiSCs displayed a similar chromatin accessibility landscape to human ICM and 5iLA naive hESCs (Figure 1G). However, the analysis for the genome-wide histone mark landscape revealed that the global H3K4me3 occupancy in prEpiSCs was comparable to that in ICM rather than naive and primed hPSCs (Figure 1H). In line with the chromatin state, the principal component analysis (PCA) showed that prEpiSCs transcriptomically resembled pre-implantation epiblast cells (day 6), but were distinctive from post-implantation epiblast cells (day 8–12) (Figures 1I and 1J) (Zhou et al., 2019).

Interestingly, prEpiSCs expressed a large number of 8C-specific genes (e.g., *LEUTX*, *TPRX1*, *ZSCAN4*, *DUXA*, *KLF17*, *MBD3L2*, *ZSCAN5B*) that were defined by the scRNA-seq data of early human embryos (Figures 1K, 1L, and Table S1) (Petropoulos et al., 2016; Yan et al., 2013). In particular, multiple DUX4-regulated genes were reactivated in prEpiSCs (Figure 1M) (Hendrickson et al., 2017). Notably, the DUX4-regulated ZGA genes (e.g., *ZSCAN4*, *LEUTX*, *DUXA*, *TPRX1*) were expressed in prEpiSCs at much higher levels than those in human naive and formative pluripotent cells (Figure 1N).

Identification of 8CLCs from prEpiSCs and PXGL hPSCs by scRNA-seq

According to the above results, we hypothesized that human 8CLCs might exist in the prEpiSC culture. To test this hypothesis, we performed scRNA-seq analysis on prEpiSCs. We identified one main cell cluster and two rare clusters from prEpiSCs (Figure 2A). The main cell cluster (cluster0) expressed multiple pluripotent genes at comparable levels to the epiblast from human blastocysts, suggesting the similarity between prEpiSCs and human pre-implantation epiblast. Intriguingly, one rare cell cluster (cluster2) expressed multiple 8C-specific genes (e.g., *ZSCAN4*, *LEUTX*, *KLF17*, *TPRX1*, *H3Y1*), while the other rare cell cluster (cluster1) expressed pluripotent genes at quite low levels (Figure 2B). To address the identities of these two rare cell clusters, we performed t-distributed stochastic neighbor embedding (tSNE) analysis by integrating these rare cell clusters into scRNA-seq datasets of early human embryos. Some cells from cluster2 were grouped with 8C embryos, thus named 8CLCs (Figure 2C). On the other hand, some cells from cluster1 were grouped with medium trophectoderm (TE), thus named TE-like cells (TELCs) (Figure 2C). The expression of multiple markers for the 8C embryo and TE was detected in 8CLCs and TELCs, respectively (Figures 2D and 2E). In line with the scRNA-seq analysis, immunofluorescence analysis revealed that LEUTX was heterogeneously expressed in prEpiSC colonies, while ZSCAN4 seemed to be homogeneously expressed (Figure 2F). Collectively, these results demonstrate that 8CLCs are present in prEpiSC culture.

To explore whether the existence of 8CLCs is dependent on culture conditions and pluripotent states, we analyzed the scRNA-seq data of naive and primed hPSCs cultured in different conditions (Kinoshita et al., 2021; Liu et al., 2020; Messmer et al., 2019). We first analyzed the scRNA-seq data (Sahakyan et al., 2017) of hESCs to interrogate any cell subset that expresses 8C-specific genes. The cell subpopulation expressing multiple 8C-specific genes was discovered in neither feeder-dependent nor feeder-free hESCs (Figures S2A–S2C). We also analyzed the human extended PSCs (hEPSCs) that were reported to be capable of differentiation into both embryonic and extraembryonic lineages (Liu et al., 2021a), but could not identify one subpopulation expressing 8C-specific genes (Figure S2C). Though the transcription of a few 8C-sepcific genes could be detected (Figures S3A and S3B), the unique 8C-like subpopulation could not be discovered from 5iLA and t2iLGo naive hPSCs. However, similar to prEpiSCs, 8CLCs were also identified from PXGL naive hPSCs (Figures S3C–S3G), in accordance with the latest studies (Mazid et al., 2022; Taubenschmid-Stowers et al., 2022). These data indicate that the existence of 8CLCs is dependent on the culture condition and pluripotent states of hPSCs.

Isolation of 8CLCs from prEpiSCs with an 8C-specific reporter

It is a challenge to characterize 8CLCs in detail and dissect the transition between 8CLCs and hPSCs, because of the quite low proportion of 8CLCs in prEpiSCs. Thus, we tried to develop an 8C-specific reporter to monitor and enrich the 8CLC population from the prEpiSC culture. By analyzing the scRNA-seq data of early human embryos (Petropoulos et al., 2016; Yan et al., 2013), we confirmed LEUTX as one of the most specific markers for human 8C embryos (Figures 3A, 3B, and S4A) (Jouhilahti et al., 2016). Thus, we next attempted to identify the regulatory element of LEUTX to develop an 8C-specific reporter. We analyzed the public ATAC-seq data of early human embryos (Wu et al., 2018) and determined the putative promoter and transcription start site (TSS) of 8C-specific isoform of LEUTX (Figure 3C). Finally, we developed the LEUTX-mCherry reporter in which mCherry was driven by the putative *LEUTX* promoter (Figure 3D).

Next, to test its specificity, we injected the LEUTX-mCherry plasmid into human zygotes, and then monitored the LEUTX-mCherry signals during early embryonic development. As an endogenous control, the continually activated CAG-GFP reporter was also injected into human zygotes, together with the LEUTX-mCherry reporter. While the GFP signal started to be detected at the 2C stage, and was maintained throughout the developmental stages, the mCherry signal started to be activated at the 4C stage, peaked at the 8C stage, and then got weak at the morula stage (Figure 3E), suggesting that the LEUTX-mCherry reporter activity during human preimplantation

CellPress
OPEN ACCESS

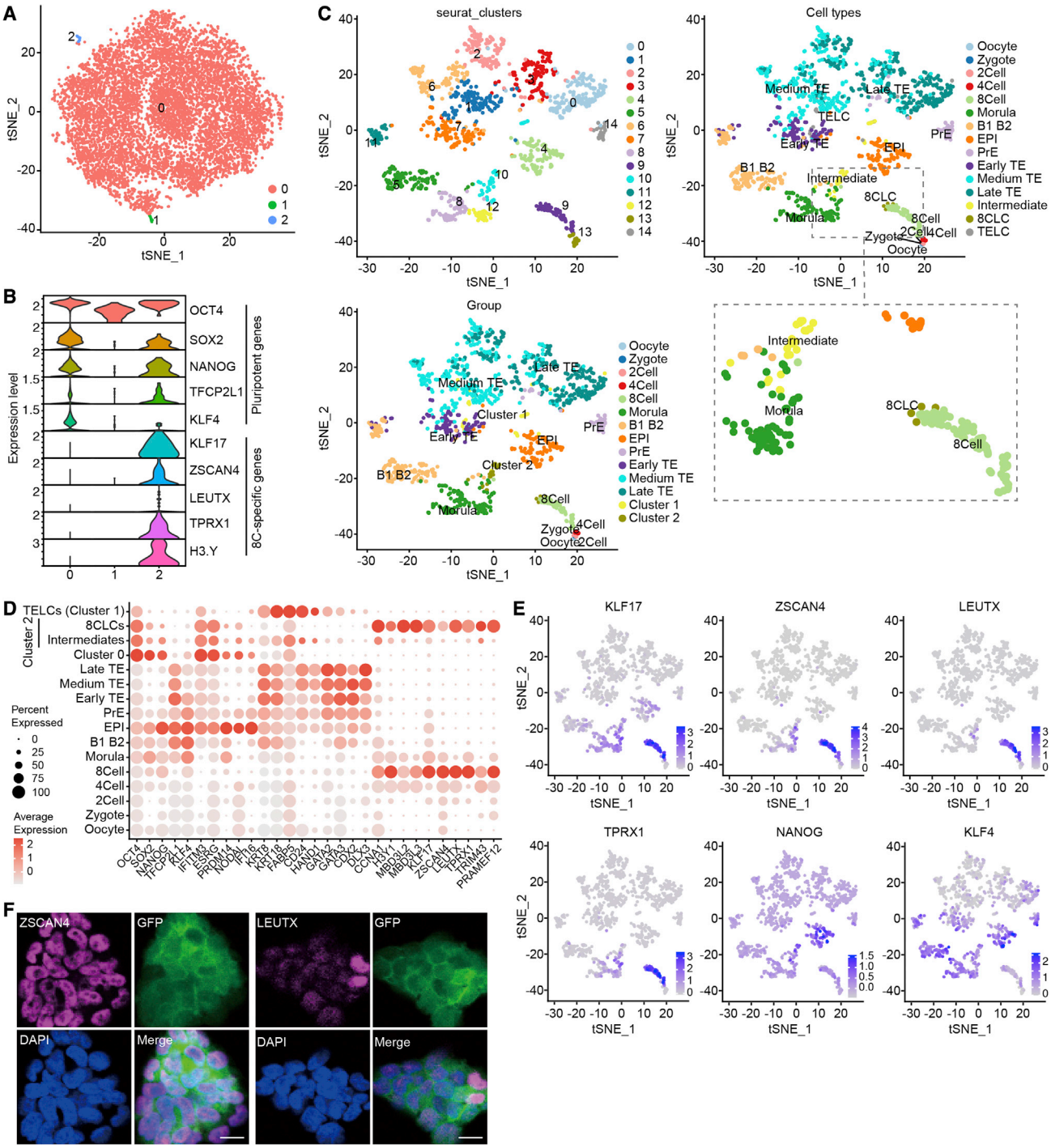


Figure 2. Identification of 8CLCs in human pre-epiblast stem cell (prEpiSC) culture

(A) tSNE analysis of scRNA-seq data of prEpiSCs. Three cell clusters were identified in prEpiSCs.

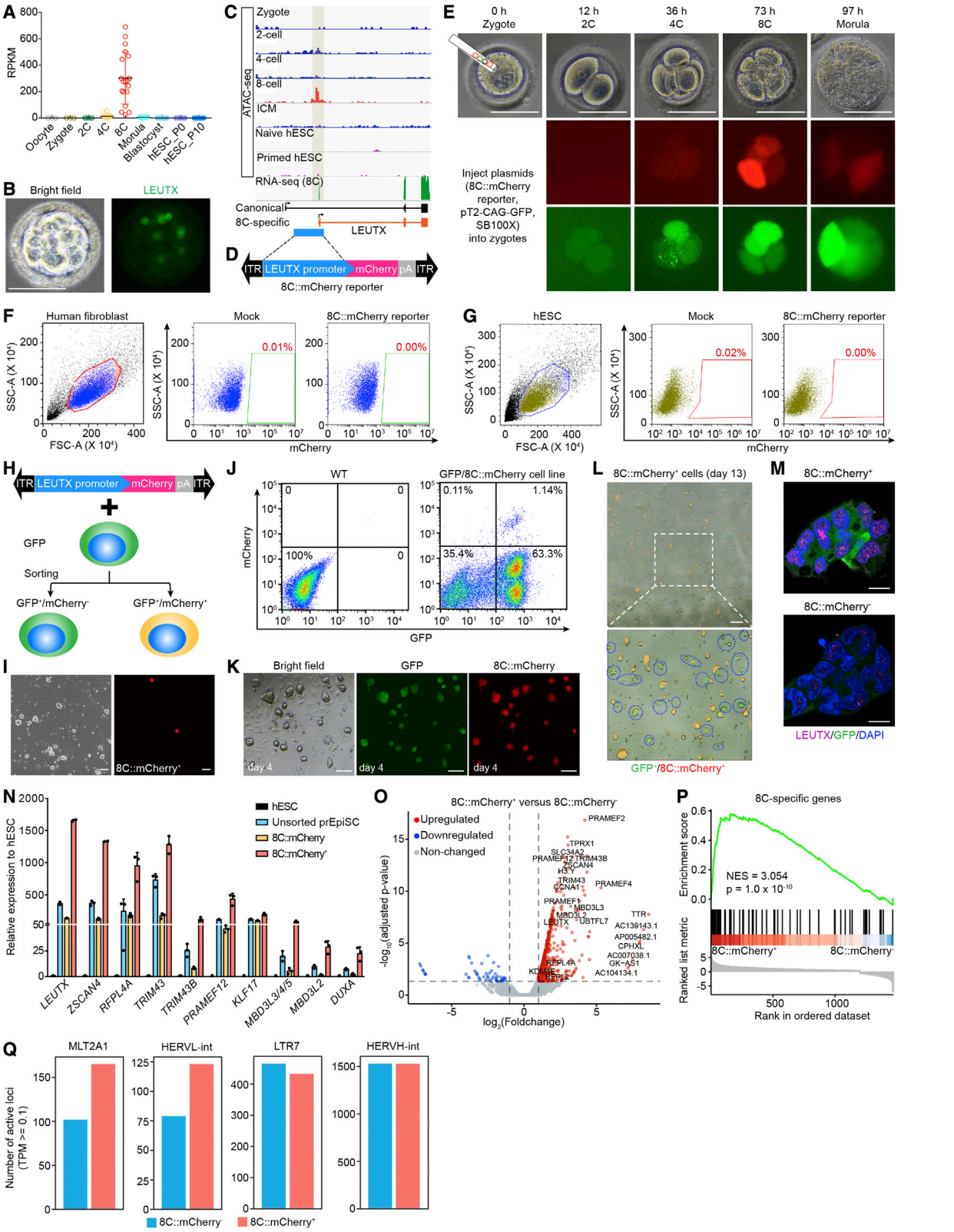
(B) Combined violin plots showing the relative expression levels of pluripotent and 8C-specific genes in three cell clusters in (A).

(C) Integrative tSNE analysis of scRNA-seq data of early human embryos and cell cluster1 and 2 in (A). Some cells from cell cluster2 are grouped into human 8C embryos while the ones from cell cluster1 are grouped into the medium TE.

(D) The frequency of expression and average expression levels for selected pluripotent, TE-specific, and 8C-specificgenes indicating individual cell types identified in prEpiSCs and early human embryos.

(E) Expression levels of multiple 8C-specific and pluripotent genes in human embryos and prEpiSCs.

(F) Immunofluorescence analysis of 8C-specific transcription factors (LEUTX and ZSCAN4) in prEpiSCs. Scale bar, 10 μm. See also Figures S2 and S3.



(legend on next page)

development can well recapitulate the dynamics of endogenous LEUTX transcription. Notably, the LEUTX-mCherry signal faithfully reflected the heterogeneity of early human blastomeres, in line with LEUTX heterogeneous expression in human 8C blastomeres (Figures 3A, 3B, and 3E). We also transfected the LEUTX-mCherry reporter into somatic cells (human fibroblasts) and hESCs, respectively. As expected, the LEUTX-mCherry reporter was activated in neither human somatic nor ESCs (Figures 3F and 3G). However, upon transient overexpression of 8C-specific genes in HEK293 cells, the LEUTX-mCherry was quickly reactivated (Figures S4B and S4C). Particularly, DUX4 overexpression reactivated the LEUTX-mCherry reporter in more than 30% of HEK293 cells (Figures S4B and S4C), indicating that the LEUTX-mCherry reporter is regulated by DUX4, one of the key regulators for ZGA (De Iaco et al., 2017; Hendrickson et al., 2017; Whiddon et al., 2017). Taken together, our LEUTX-mCherry reporter can specifically mark human 8C embryos, thus named 8C::mCherry reporter afterward.

Next, we generated one 8C::mCherry prEpiSC line by integrating the 8C::mCherry reporter into the prEpiSC genome (Figures 3H and 3I). About 1% of 8C::mCherry$^+$ cells (with strong signal) were detected from the prEpiSC culture by fluorescence-activated cell sorting (FACS) (Figure 3J). We enriched these 8C::mCherry$^+$ cells and re-seeded them on feeder cells for long-term culturing. These 8C::mCherry$^+$ cells grew as colonies from the single cell, retaining strong mCherry signal for four days at least (Figure 3K). The RT-qPCR results showed that these 8C::mCherry$^+$ cells expressed multiple 8C-specific genes at much higher levels than 8C::mCherry$^-$ prEpiSCs and hESCs, consistent with higher LEUTX protein level in 8C::mCherry$^+$ than that in 8C::mCherry$^-$ cells (Figures 3M and 3N). The comparative transcriptome analysis showed that the upregulated genes in 8C::mCherry$^+$ cells were enriched for the 8C-specific gene set, when compared with 8C::mCherry$^-$ cells (Figures 3O and 3P). As retrotransposons are dynamically transcribed during human preimplantation development (Goke et al., 2015; Wang et al., 2014), we then compared the transcription of retrotransposons between 8C::mCherry$^+$ and 8C::mCherry$^-$ cells. Consistent with previous studies showing that *HERVL* (*MLT2A1* and *HERVL-int*) is specifically expressed in human 8C embryos (Goke et al., 2015), *MLT2A1* and *HERVL-int* were expressed at higher levels in 8C::mCherry$^+$ than those in 8C::mCherry$^-$ cells, whereas the pluripotent-specific retrotransposon *HERVH* (*LTR7* and *HERVH-int*) (Wang et al., 2014) was expressed at comparable levels between 8C::mCherry$^+$ and 8C::mCherry$^-$ cells (Figure 3Q). These results indicate that the 8C::mCherry$^+$ 8CLCs have molecular features resembling the human 8C embryo.

However, the 8C::mCherry$^+$ cells could not be maintained in long-time culture, as the mCherry signal was lost with extension of the culturing time. By 2-week culturing, only 50% of the cells were mCherry$^+$ cells (Figure 3L), indicating that like 2CLCs, 8CLCs cannot be stably maintained under the pluripotent cell culture condition.

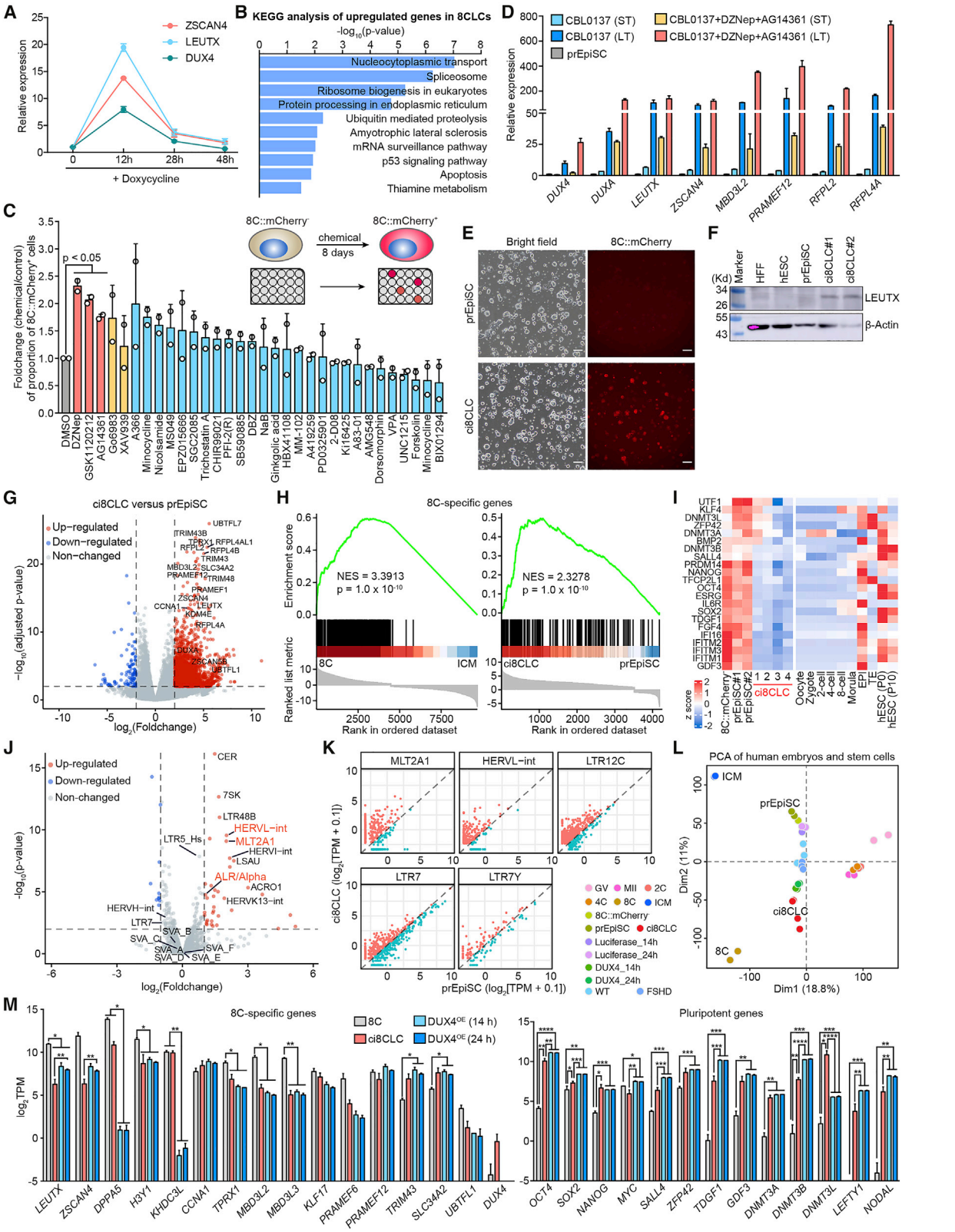
Increase and maintenance of 8CLC population with the optimized culture condition

We then explored whether DUX4 overexpression could improve the 8CLC population in cultures. In line with the previous report (Hendrickson et al., 2017), DUX4 overexpression increased the expression of 8C-specific ZGA genes (*LEUTX* and *ZSCAN4*) in prEpiSCs, but the increased expression level could not be sustained (Figure 4A). These data suggest that DUX4 alone is not sufficient to reactivate the 8C-specific transcriptional regulatory network.

Inspired by our previous study on 2CLCs (Yang et al., 2022), we expected that signaling pathway modulation and/or epigenetic regulation could increase the 8CLC population in prEpiSC culture. To identify potential targets, we performed the transcriptome comparison between human 8C embryos and the pluripotent ICM. Our analysis suggested that multiple signaling pathways (e.g., mTOR, Hippo, WNT, TGFβ, PKC, and MAPK) were downregulated in human 8C embryos, when compared with ICM (Figure S5A). Similarly, we also identified multiple downregulated epigenetic regulators that were enriched for DNA methylation, histone modification (H3K9, H3K27, and arginine methylation), p53 activity regulation, and so on (Figures S5B–S5D), consistent with previous reports revealing the lowest H3K27me3 enrichment at the 8C stage during human preimplantation development (Xia et al., 2019). In addition, compared with prEpiSCs, the Kyoto Encyclopedia of Genes and Genomes

Figure 3. Development of the human 8C-specific reporter for 8CLC enrichment from prEpiSCs

(A) The dynamic expression of LEUTX during human preimplantation development (Yan et al., 2013).
(B) Immunofluorescence analysis of LEUTX protein level in human 8C embryos. Scale bar, 100 μm.
(C) Identification of TSS and putative promoter of LEUTX in human 8C embryos by ATAC-seq and RNA-seq.
(D) Construction of 8C::mCherry reporter in which mCherry is driven by the putative LEUTX promoter.
(E) The activity and specificity of 8C::mCherry reporter in preimplantation embryos. Scale bar, 100 μm.
(F and G) The activity of 8C::mCherry reporter in human fibroblasts (F) and hESCs (G).
(H) The schematic illuminating generation of 8C::mCherry reporter cell line and isolation of 8C::mCherry$^+$ cells.
(I) The representative images showing rare 8C::mCherry$^+$ cells in prEpiSCs. Scale bar, 100 μm.
(J) Enrichment of 8C::mCherry$^+$ cells from prEpiSCs by FACS followed by long-term culturing.
(K) The representative images showing short-term culturing of 8C::mCherry$^+$ cells. Scale bar, 100 μm.
(L) The representative images showing long-term culturing of 8C::mCherry$^+$ cells. Scale bar, 100 μm.
(M) Immunofluorescence analysis of LEUTX protein levels in 8C::mCherry$^+$ and 8C::mCherry$^-$ cells. Scale bar, 10 μm.
(N) RT-qPCR analysis of key 8C-specific genes in hESCs, unsorted prEpiSCs, and 8C::mCherry$^+$ and 8C::mCherry$^-$ cells. Data are shown as means ± SD (n = 3 technical replicates from two independent experiments).
(O) The volcano plot showing the differentially expressed genes between 8C::mCherry$^+$ and 8C::mCherry$^-$ cells analyzed by edgeR. Red and blue dots indicate upregulated (foldchange > 2) and downregulated (foldchange < 0.5) genes with adjusted $p < 0.05$. Multiple 8C-specific genes are labeled.
(P) Gene set enrichment analysis (GSEA) of the 8C-specific gene set by comparing 8C::mCherry$^+$ with 8C::mCherry$^-$ cells.
(Q) The bar plot showing dynamic transcription of varied retrotransposons in 8C::mCherry$^+$ and 8C::mCherry$^-$ cells. See also Figure S4.



(legend on next page)

(KEGG) analysis revealed that the upregulated genes in 8CLCs were enriched for spliceosome, p53 signaling pathway, and apoptosis (Figure 4B), indicating that p53 activation may contribute to transition of prEpiSCs into the 8C-like state, in line with the recent finding that DUX4 is activated by p53 (Grow et al., 2021). Collectively, these analyses imply that multiple signaling pathways and epigenetic regulation may be involved in the transition between prEpiSCs and 8CLCs.

Based on the above analyses, we selected a panel of potent and highly selective small molecules targeting multiple signaling pathways as well as a set of epigenetic regulators (Key Resources Table). The screening identified DZNep (repressing H3K9me3 and H3K27me3), GSK1120212 (one MEK inhibitor), XAV939 (repressing the activity of WNT signaling pathway), AG14361 (one PARP1 inhibitor), and CBL0137 (one p53 activator) as potent compounds that promoted the conversion of prEpiSCs to 8CLCs (Figures 4C and 4D, and S5E–S5G). We found that dramatic cell death compromised 8CLC induction and maintenance, thus applied two additional chemical compounds (GSK872 and Ac-DEVD-CHO) to inhibit necroptosis and apoptosis, respectively. Finally, we developed one chemical-based culture condition to remarkably improve the 8CLC population in prEpiSCs and increase the expression of multiple ZGA genes (Figures 4D–4F). These converted 8CLCs were named chemical-induced 8CLCs (ci8CLCs). The ci8CLCs could be maintained in extended culture (2-3 weeks) at a relatively high proportion.

To characterize the ci8CLCs at the molecular level, we compared their transcriptome with those of prEpiSCs as well as human 8C embryos. The comparative transcriptome analysis showed that the upregulated genes in ci8CLCs were enriched for the 8C-specific gene set, when compared with prEpiSCs (Figures 4G and 4H), whereas the pluripotent genes were downregulated in ci8CLCs (Figure 4I). 8C-specific retrotransposons (*MLT2A1*, *HERVL-int*, and *LTR12C*) were expressed at higher levels in ci8CLCs than those in prEpiSCs (Figures 4J and 4K). Importantly, the PCA revealed that ci8CLCs were located closer to the human 8C embryo than DUX4-overexpressed ($DUX4^{OE}$) hESCs (Figure 4L). Multiple 8C-specific genes showed similar expression levels between ci8CLCs and $DUX4^{OE}$ hESCs, but many pluripotent genes expressed in ci8CLCs at lower levels than in $DUX4^{OE}$ hESCs (Figure 4M). Collectively, these data suggest that ci8CLCs have unique transcriptome profiles similar to human 8C embryos.

Blastocyst-like structures generated from 8CLCs

Finally, we tried to evaluate the developmental potency of 8CLCs. To mimic the developmental process of early human embryos, we injected 8CLCs into the empty mouse zona pellucida, which were cultured in the prEpiSC medium for one day and then switched to the blastoid induction medium (Figure 5A) (Kagawa et al., 2022). Ci8CLCs could self-organize to form the blastocyst-like structures (blastoids) in 6 days (Figure 5A). Intriguingly, from day 3, 8CLC aggregates exhibited clear accumulation of ZO1 (one tight junction protein) at the cell-cell junctions and a polarization of atypical protein kinase C (aPKC) toward the apical side of cells (Figure 5B), suggesting that 8CLC aggregates recapitulate blastomere compaction and polarization. In particular, late 8CLC aggregates showed the morula-like structures that were composed of inner ($OCT4^{+}$) and outer ($GATA3^{+}$) cells (Figure 5C). Then, we checked whether 8CLC-blastoids had ICM and TE lineages that are components of human blastocysts. The immunofluorescence analysis revealed that cells in the outer layer of 8CLC-blastoids expressed GATA3 (one marker for TE), while the inner cell population located at one side of the blastoid expressed OCT4 (one marker for ICM) (Figure 5D).

Similar to human blastocysts, 8CLC-blastoids could give rise to hESCs that expressed multiple pluripotent markers and could be differentiated into three germ layers *in vitro* (Figures 5E–5H). Human trophoblast stem cells (hTSCs) could also be derived from 8CLC-blastoids, expressing multiple hTSC markers (Figures 5I–5K).

In summary, the above results suggest that the development of 8CLCs into blastoids holds potential to recapitulate the key processes of human preimplantation development.

Figure 4. Chemical-induced 8CLCs

(A) RT-qPCR analysis of expression of LEUTX and ZSCAN4 in DUX4-overexpressed prEpiSCs at different time points. Data are shown as means ± SD (n = 3 technical replicates).

(B) Gene ontology analysis of upregulated genes in 8CLCs compared with prEpiSCs.

(C) Bar plot showing the effects of chemicals on the percentage of 8C::$mCherry^{+}$ cells. Data are shown as means ± SD (n = 2 independent experiments). Statistical analyses are performed using two-tailed, unpaired t test.

(D) RT-qPCR analysis of expression of multiple 8C-specific genes under chemical treatment as indicated. Data are shown as means ± SD (n = 3 technical replicates from two independent experiments). ST, short-term; LT, long-term.

(E) Representative images showing ci8CLCs derived from prEpiSCs. Scale bar, 100 μm.

(F) Western blot analysis showing upregulation of LEUTX protein level in ci8CLCs compared with human fibroblasts (HFF), hESCs, and prEpiSCs.

(G) Volcano plot showing the differentially expressed genes between ci8CLCs (n = 4) and prEpiSCs (n = 3) analyzed by DESeq2. Red and blue dots indicate upregulated (foldchange > 4) and downregulated (foldchange < 0.25) genes with adjusted $p < 0.05$. Multiple 8C-specific genes are marked.

(H) Gene set enrichment analysis (GSEA) of the 8C-specific gene set by comparison between human 8C embryos and ICM (left panel) and by comparing ci8CLCs with prEpiSCs (right panel).

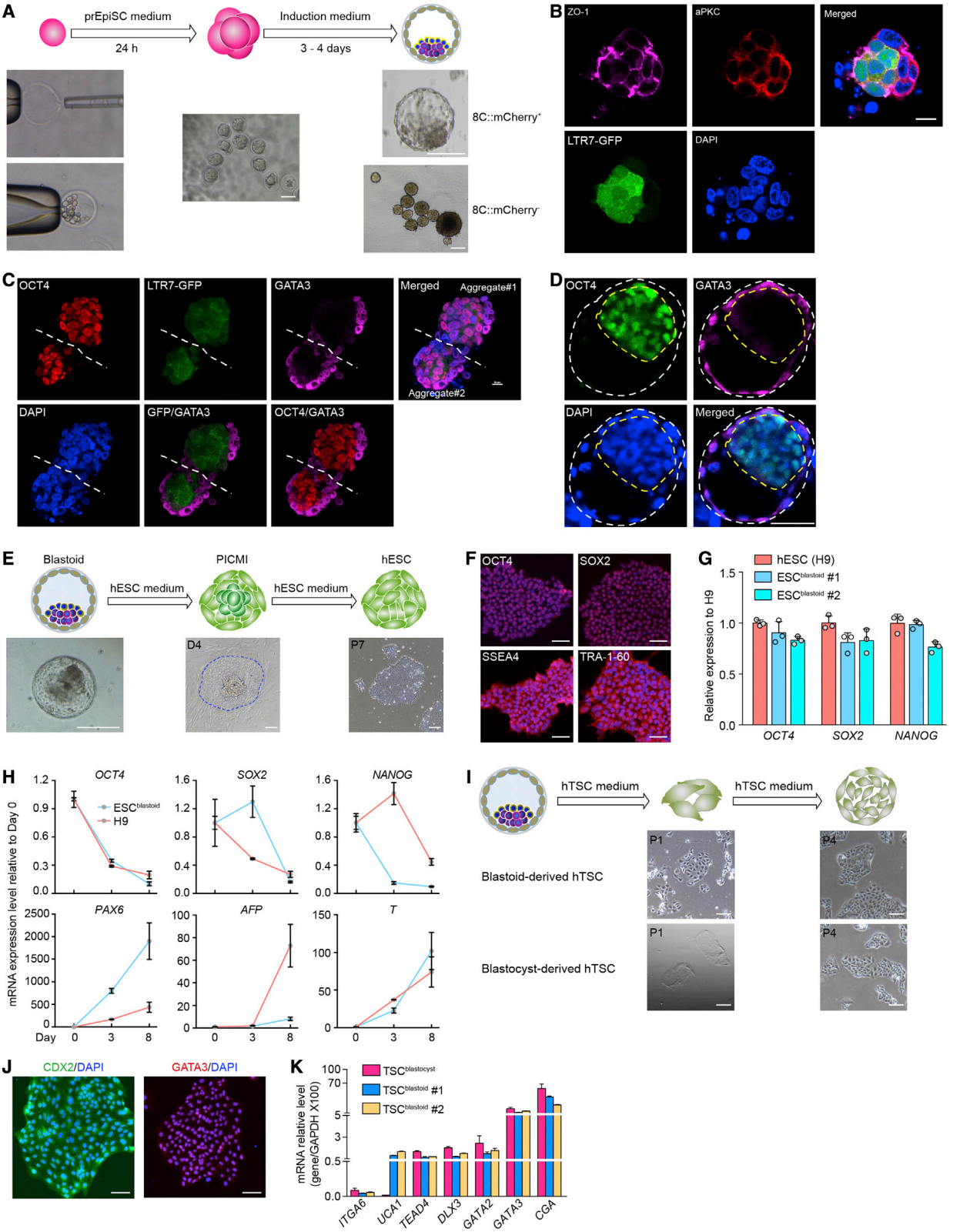
(I) Heatmap showing the relative expression levels of pluripotent genes in early human embryos, hESCs, prEpiSCs, and ci8CLCs.

(J) Volcano plot showing the differentially expressed repeat elements between ci8CLCs (n = 4) and prEpiSCs (n = 3) analyzed by DESeq2. Red and blue dots indicate upregulated (foldchange > 2) and downregulated (foldchange < 0.5) repeats with $p < 0.05$.

(K) Expression comparison of *HERVL* (*MLT2A1* and *HERVL-int*), *LTR12C*, and *LTR7(Y)* between ci8CLCs and prEpiSCs.

(L) PCA of global gene expression data from early human embryos, prEpiSCs, ci8CLCs, hESCs, $DUX4^{OE}$ hESCs, and FSHD muscle cells.

(M) Bar plots showing the expression levels of multiple 8C-specific (ZGA) and pluripotent genes in human 8C embryos (n = 2), ci8CLCs (n = 4), and $DUX4^{OE}$ hESCs (n = 2). Data are shown as means ± SD Statistical analyses are performed using two-tailed, unpaired t test with Welch correction. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$. See also Figure S5.



(legend on next page)

DISCUSSION

In this study, using scRNA-seq analysis and a human 8C-specific reporter, we identified rare 8CLCs in preimplantation epiblast-like stem cells and PXGL naive hPSCs, in parallel with the other two latest reports (Mazid et al., 2022; Taubenschmid-Stowers et al., 2022). The 8CLCs exhibit human 8C-specific molecular profiles and can self-organize to form blastoids through a transient morula-like stage. Given their unique molecular features and the potency of blastoid generation, the 8CLCs represent one of the best *in vitro* models to recapitulate early human development.

In contrast to the other two latest studies (Mazid et al., 2022; Taubenschmid-Stowers et al., 2022), we did not identify one 8C-like subpopulation from 5iLA and t2iLGo naive hPSCs, though some 8C genes were expressed with low levels in rare 5iLA and t2iLGo naive hPSCs. This inconsistency may be associated with the definition of 8C genes and different scRNA-seq datasets of 5iLA and t2iLGo naive hPSCs used in our and other studies. Another report (Wang et al., 2018) suggests that a few 8C-specific ZGA genes could be transiently detected during reprogramming of 5iLA naive hPSCs from primed hESCs, indicating that the expression of 8C genes may be also dependent on the passages of 5iLA naive hPSCs. Of note, based on the 8C gene set defined by the other two latest reports, which are also highly expressed in morula (e.g., *TPRX1*), the identified 8CLCs actually contain a large number of cells resembling morula (Mazid et al., 2022; Taubenschmid-Stowers et al., 2022), indicating that the proportion of 8CLCs in varied naive hPSC cultures is overestimated.

By small-scale chemical screening, we identified one chemical cocktail that can increase the 8CLC population and extend the culturing time. Consistent with one recent report (Mazid et al., 2022), we also showed that DZNep, one chemical repressing H3K27me3 and H3K9me3, promoted 8C gene expression, suggesting that epigenetic regulation may play crucial roles in transition between hPSCs and 8CLCs. Notably, the p53 activator CBL0137 significantly improved the expression of 8C genes, suggesting that the p53 signaling pathway is involved in the 8CLC transition from hPSCs. The p53 signaling pathway is activated as a response to DNA damage and replication stress that are closely associated with 2CLC transition from mESCs (Atashpaz et al., 2020; Grow et al., 2021; Nakatani et al., 2022). To address the role of the p53 signaling pathway in early blastomere-like cells, it needs to verify whether the DNA damage and replication stress contribute to ZGA during early mammalian development in future studies.

The discovery of 8CLCs suggests high plasticity of human pluripotent cells and demonstrates that the spontaneous transition between the pluripotent and early blastomere-like states is not restricted to mice, but may be a conserved feature in mammalian PSCs. Exploration of evolutionarily conserved mechanisms regulating the transition between the pluripotent and early blastomere-like states will not only deepen understanding of the regulatory circuitry for pluripotency, but also facilitate the studies on mammalian early embryonic development.

Limitations of the study

We showed that small molecular compounds targeting epigenetic regulators could increase the 8CLC population, but the mechanisms behind this are unknown. We found that *DNMT3B* was highly expressed in 8CLCs (Figure 4M), indicating a distinct DNA methylation state from human 8C embryos. In addition, many pluripotent genes were also expressed in 8CLCs with higher levels than those in human 8C embryos (Figure 4M). Comprehensive analyses of epigenetic dynamics (e.g., DNA methylation, histone modification) over 8CLC transition may facilitate culture condition optimization to downregulate pluripotent genes and derive an alternative state much closer to the 8C embryo.

To better understand the developmental potency of 8CLCs, the comparative analysis between 8CLC-balstoids and those generated from hPSCs is also required (Fan et al., 2021; Kagawa et al., 2022; Liu et al., 2021b; Sozen et al., 2021; Yanagida et al., 2021; Yu et al., 2021a). In particular, a transient stage like morula can be captured during blastoid generation from 8CLCs. Such morula-like structures need to be characterized in detail in future studies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS

Figure 5. Blastoid generation from 8CLCs

(A) The protocol of blastoid generation from 8CLCs and the representative image showing typical morphology of blastoids similar to the human blastocyst. Scale bar, 100 μm.

(B) Immunofluorescence analysis of compaction (indicated by ZO-1) and polarization (indicated by aPKC) of cells in 8CLC aggregates. Scale bar, 10 μm.

(C) Immunofluorescence analysis of OCT4 and GATA3 protein levels in 8CLC aggregates. Scale bar, 10 μm.

(D) Immunofluorescence staining of 8CLC-derived blastoids for OCT4 and GATA3. Scale bar, 50 μm.

(E) The hESC derivation from 8CLC-derived blastoids. Scale bar, 100 μm.

(F) Immunofluorescence analysis of pluripotent markers (OCT4, SOX2, SSEA4, and TRA-1-60) in 8CLC-blastoid-derived hESCs. Scale bar, 50 μm.

(G) RT-qPCR analysis of pluripotent gene expression in 8CLC-blastoid-derived hESCs. Data are shown as means ± SD (n = 3 technical replicates).

(H) RT-qPCR analysis of expression of differentiation markers for three germ layers in embryoid bodies generated from 8CLC-blastoid-derived hESCs. Data are shown as means ± SD (n = 3 technical replicates).

(I) The hTSC derivation from 8CLC-derived blastoids. Scale bar, 100 μm.

(J) Immunofluorescence analysis of hTSC markers (CDX2 and GATA3) in 8CLC-blastoid-derived hTSCs. Scale bar, 50 μm.

(K) RT-qPCR analysis of expression of hTSC marker genes in 8CLC-blastoid-derived hTSCs. Data are shown as means ± SD (n = 3 technical replicates).

- Human embryos
- Cell culture
- Derivation and culture of prEpiSCs
- Human ci8CLC culture

- METHOD DETAILS
  - Construct cloning
  - Microinjection of 8C::mCherry and pT2-CAG-GFP reporters into human zygotes
  - Generation of blastoids from 8CLCs
  - Derivation of hESCs from 8CLC-blastoids
  - Derivation of hTSCs from 8CLC-blastoids
  - *In vitro* differentiation assay
  - Flow cytometry analysis
  - RT-qPCR
  - Western blot
  - Immunofluorescence
  - RNA-seq
  - Smart-seq2 scRNA-seq of prEpiSCs
  - Single-cell transcriptome profiling of prEpiSCs by 10X Genomics
  - ATAC-seq
  - ChIP-seq
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Processing of bulk RNA-seq and scRNA-seq data
  - Differential expression analysis
  - Downstream analysis of scRNA-seq data
  - Definition of human 8C-specific genes
  - Processing of ChIP-seq and ATAC-seq data
  - Principal component analysis
  - Statistical analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2022.110994.

ACKNOWLEDGMENTS

We thank Jinhe Guo for his assistance in plasmid construction. We thank Junhua Liu for her assistance in chemical screening. This study was supported by grants from the National Key R&D Program of China (2018YFA0107003), the National Natural Science Foundation of China (31871444), the Fundamental Research Funds for Central Universities (19ykzd35), and the Guangdong Innovative and Entrepreneurial Team Program (2016ZT06S029) to J.W.; M.Y. is supported by the Innovative Development Program of Sun Yat-sen University for Outstanding PhD Students (19ykyjs62).

AUTHOR CONTRIBUTIONS

C.S. and J.W. conceived ideas and supervised the project; M.Y. and J.W. designed the experiments; C.S. and J.W. generated LTR7-GFP hESCs under supervision by Z.I; M.Y. established 8C::mCherry cell lines, performed chemical screening, and generated blastoids with assistance from X.Y., S.L., X.K., R.G., and R.Z; R.L. performed embryo injection; X.Y. performed RNA-seq experiments; M.C. and H.Y. performed bioinformatics analyses; Y.Q., C.S., and J.W. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 24, 2022
Revised: April 21, 2022
Accepted: May 31, 2022
Published: June 21, 2022

REFERENCES

Atashpaz, S., Samadi Shams, S., Gonzalez, J.M., Sebestyén, E., Arghavanifard, N., Gnocchi, A., Albers, E., Minardi, S., Faga, G., Soffientini, P., et al. (2020). ATR expands embryonic stem cell fate potential in response to replication stress. Elife *9*, e54756. https://doi.org/10.7554/elife.54756.

Bayerl, J., Ayyash, M., Shani, T., Manor, Y.S., Gafni, O., Massarwa, R., Kalma, Y., Aguilera-Castrejon, A., Zerbib, M., Amir, H., et al. (2021). Principles of signaling pathway modulation for enhancing human naive pluripotency induction. Cell Stem Cell *28*, 1549–1565. https://doi.org/10.1016/j.stem.2021.04.001.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218. https://doi.org/10.1038/nmeth.2688.

Chen, M., Zhu, Q., Li, C., Kou, X., Zhao, Y., Li, Y., Xu, R., Yang, L., Yang, L., Gu, L., et al. (2020). Chromatin architecture reorganization in murine somatic cell nuclear transfer embryos. Nat. Commun. *11*, 1813. https://doi.org/10.1038/s41467-020-15607-z.

Chung, Y.G., Matoba, S., Liu, Y., Eum, J.H., Lu, F., Jiang, W., Lee, J.E., Sepilian, V., Cha, K.Y., Lee, D.R., and Zhang, Y. (2015). Histone demethylase expression enhances human somatic cell nuclear transfer efficiency and promotes derivation of pluripotent stem cells. Cell Stem Cell *17*, 758–766. https://doi.org/10.1016/j.stem.2015.10.001.

Collier, A.J., Panula, S.P., Schell, J.P., Chovanec, P., Plaza Reyes, A., Petropoulos, S., Corcoran, A.E., Walker, R., Douagi, I., Lanner, F., and Rugg-Gunn, P.J. (2017). Comprehensive cell surface protein profiling identifies specific markers of human naive and primed pluripotent states. Cell Stem Cell *20*, 874–890.e7. https://doi.org/10.1016/j.stem.2017.02.014.

De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nat. Genet. *49*, 941–945. https://doi.org/10.1038/ng.3858.

Deng, M., Liu, Z., Chen, B., Wan, Y., Yang, H., Zhang, Y., Cai, Y., Zhou, J., and Wang, F. (2020). Aberrant DNA and histone methylation during zygotic genome activation in goat cloned embryos. Theriogenology *148*, 27–36. https://doi.org/10.1016/j.theriogenology.2020.02.036.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047–3048. https://doi.org/10.1093/bioinformatics/btw354.

Fan, Y., Min, Z., Alsolami, S., Ma, Z., Zhang, E., Chen, W., Zhong, K., Pei, W., Kang, X., Zhang, P., et al. (2021). Generation of human blastocyst-like structures from pluripotent stem cells. Cell Discov. *7*, 81. https://doi.org/10.1038/s41421-021-00316-8.

Fogarty, N.M.E., McCarthy, A., Snijders, K.E., Powell, B.E., Kubikova, N., Blakeley, P., Lea, R., Elder, K., Wamaitha, S.E., Kim, D., et al. (2017). Genome editing reveals a role for OCT4 in human embryogenesis. Nature *550*, 67–73. https://doi.org/10.1038/nature24033.

Gao, L., Wu, K., Liu, Z., Yao, X., Yuan, S., Tao, W., Yi, L., Yu, G., Hou, Z., Fan, D., et al. (2018). Chromatin accessibility landscape in human early embryos and its association with evolution. Cell *173*, 248–259.e15. https://doi.org/10.1016/j.cell.2018.02.028.

Geng, L.N., Yao, Z., Snider, L., Fong, A.P., Cech, J.N., Young, J.M., van der Maarel, S.M., Ruzzo, W.L., Gentleman, R.C., Tawil, R., and Tapscott, S. (2012). DUX4 activates germline genes, retroelements, and immune

mediators: implications for facioscapulohumeral dystrophy. Dev. Cell *22*, 38–51. https://doi.org/10.1016/j.devcel.2011.11.013.

Göke, J., Lu, X., Chan, Y.S., Ng, H.H., Ly, L.H., Sachs, F., and Szczerbinska, I. (2015). Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. Cell Stem Cell *16*, 135–141. https://doi.org/10.1016/j.stem.2015.01.005.

Grow, E.J., Weaver, B.D., Smith, C.M., Guo, J., Stein, P., Shadle, S.C., Hendrickson, P.G., Johnson, N.E., Butterfield, R.J., Menafra, R., et al. (2021). p53 convergently activates Dux/DUX4 in embryonic stem cells and in facioscapulohumeral muscular dystrophy cell models. Nat. Genet. *53*, 1207–1220. https://doi.org/10.1038/s41588-021-00893-0.

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849. https://doi.org/10.1093/bioinformatics/btw313.

Guo, G., von Meyenn, F., Santos, F., Chen, Y., Reik, W., Bertone, P., Smith, A., and Nichols, J. (2016). Naive pluripotent stem cells derived directly from isolated cells of the human inner cell mass. Stem Cell Rep. *6*, 437–446. https://doi.org/10.1016/j.stemcr.2016.02.005.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

Henderson, J.K., Draper, J.S., Baillie, H.S., Fishel, S., Thomson, J.A., Moore, H., and Andrews, P.W. (2002). Preimplantation human embryos and embryonic stem cells show comparable expression of stage-specific embryonic antigens. Stem Cells *20*, 329–337. https://doi.org/10.1634/stemcells.20-4-329.

Hendrickson, P.G., Doráis, J.A., Grow, E.J., Whiddon, J.L., Lim, J.W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., et al. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat. Genet. *72*, 483–484. https://doi.org/10.1097/ogx.0000000000000471.

Hu, Z., Li, H., Jiang, H., Ren, Y., Yu, X., Qiu, J., Stablewski, A.B., Zhang, B., Buck, M.J., and Feng, J. (2020). Transient inhibition of mTOR in human pluripotent stem cells enables robust formation of mouse-human chimeric embryos. Sci. Adv. *6*, eaaz0298. https://doi.org/10.1126/sciadv.aaz0298.

Iturbide, A., and Torres-Padilla, M.E. (2020). A cell in hand is worth two in the embryo: recent advances in 2-cell like cell reprogramming. Curr. Opin. Genet. Dev. *64*, 26–30. https://doi.org/10.1016/j.gde.2020.05.038.

Jagannathan, S., Shadle, S.C., Resnick, R., Snider, L., Tawil, R.N., van der Maarel, S.M., Bradley, R.K., and Tapscott, S.J. (2016). Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells. Hum. Mol. Genet. *25*, 4419–4431. https://doi.org/10.1093/hmg/ddw271.

Jin, Y., Tam, O.H., Paniagua, E., and Hammell, M. (2015). TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics *31*, 3593–3599. https://doi.org/10.1093/bioinformatics/btv422.

Jouhilahti, E.M., Madissoon, E., Vesterlund, L., Töhönen, V., Krjutškov, K., Plaza Reyes, A., Petropoulos, S., Månsson, R., Mansson, R., Linnarsson, S., et al. (2016). The human PRD-like homeobox gene LEUTX has a central role in embryo genome activation. Development *143*, 3459–3469. https://doi.org/10.1242/dev.134510.

Jukam, D., Shariati, S.A.M., and Skotheim, J.M. (2017). Zygotic genome activation in vertebrates. Dev. Cell *42*, 316–332. https://doi.org/10.1016/j.devcel.2017.07.026.

Kagawa, H., Javali, A., Khoei, H.H., Sommer, T.M., Sestini, G., Novatchkova, M., Scholte Op Reimer, Y., Castel, G., Bruneau, A., Maenhoudt, N., et al. (2022). Human blastoids model blastocyst development and implantation. Nature *601*, 600–605. https://doi.org/10.1038/s41586-021-04267-8.

Karpukhina, A., Tiukacheva, E., Dib, C., and Vassetzky, Y.S. (2021). Control of DUX4 expression in facioscapulohumeral muscular dystrophy and cancer. Trends Mol. Med. *27*, 588–601. https://doi.org/10.1016/j.molmed.2021.03.008.

Khan, S.A., Park, K.M., Fischer, L.A., Dong, C., Lungjangwa, T., Jimenez, M., Casalena, D., Chew, B., Dietmann, S., Auld, D.S., et al. (2021). Probing the signaling requirements for naive human pluripotency by high-throughput chemical screening. Cell Rep. *35*, 109233. https://doi.org/10.1016/j.celrep.2021.109233.

Kinoshita, M., Barber, M., Mansfield, W., Cui, Y., Spindlow, D., Stirparo, G.G., Dietmann, S., Nichols, J., and Smith, A. (2021). Capture of mouse and human stem cells with features of formative pluripotency. Cell Stem Cell *28*, 2180. https://doi.org/10.1016/j.stem.2021.11.002.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

Lee, M.T., Bonneau, A.R., and Giraldez, A.J. (2014). Zygotic genome activation during the maternal-to-zygotic transition. Annu. Rev. Cell. Dev. Biol. *30*, 581–613. https://doi.org/10.1146/annurev-cellbio-100913-013027.

Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics *28*, 882–883. https://doi.org/10.1093/bioinformatics/bts034.

Lemmers, R.J.L.F., van der Vliet, P.J., Klooster, R., Sacconi, S., Camaño, P., Dauwerse, J.G., Snider, L., Straasheijm, K.R., Jan van Ommen, G., Padberg, G.W., et al. (2010). A unifying genetic model for facioscapulohumeral muscular dystrophy. Science *329*, 1650–1653. https://doi.org/10.1126/science.1189044.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

Liu, X., Ouyang, J.F., Rossello, F.J., Tan, J.P., Davidson, K.C., Valdes, D.S., Schröder, J., Sun, Y.B.Y., Chen, J., Knaupp, A.S., et al. (2020). Reprogramming roadmap reveals route to human induced trophoblast stem cells. Nature *586*, 101–107. https://doi.org/10.1038/s41586-020-2734-6.

Liu, B., Chen, S., Xu, Y., Lyu, Y., Wang, J., Du, Y., Sun, Y., Liu, H., Zhou, H., Lai, W., et al. (2021a). Chemically defined and xeno-free culture condition for human extended pluripotent stem cells. Nat. Commun. *12*, 3017. https://doi.org/10.1038/s41467-021-23320-8.

Liu, X., Tan, J.P., Schröder, J., Aberkane, A., Ouyang, J.F., Mohenska, M., Lim, S.M., Sun, Y.B.Y., Chen, J., Sun, G., et al. (2021b). Modelling human blastocysts by reprogramming fibroblasts into iBlastoids. Nature *591*, 627–632. https://doi.org/10.1038/s41586-021-03372-y.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome. Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature *487*, 57–63. https://doi.org/10.1038/nature11244.

Mátés, L., Chuah, M.K.L., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D.P., Schmitt, A., Becker, K., Matrai, J., et al. (2009). Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. Nat. Genet. *41*, 753–761. https://doi.org/10.1038/ng.343.

Matoba, S., Liu, Y., Lu, F., Iwabuchi, K.A., Shen, L., Inoue, A., and Zhang, Y. (2014). Embryonic development following somatic cell nuclear transfer impeded by persisting histone methylation. Cell *159*, 884–895. https://doi.org/10.1016/j.cell.2014.09.055.

Mazid, M.A., Ward, C., Luo, Z., Liu, C., Li, Y., Lai, Y., Wu, L., Li, J., Jia, W., Jiang, Y., et al. (2022). Rolling back human pluripotent stem cells to an eight-cell embryo-like stage. Nature *605*, 315–324. https://doi.org/10.1038/s41586-022-04625-0.

McCarthy, D.J., Campbell, K.R., Lun, A.T.L., and Wills, Q.F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics *33*, 1179–1186. https://doi.org/10.1093/bioinformatics/btw777.

Messmer, T., von Meyenn, F., Savino, A., Santos, F., Mohammed, H., Lun, A.T.L., Marioni, J.C., and Reik, W. (2019). Transcriptional heterogeneity in naive and primed human pluripotent stem cells at single-cell resolution. Cell Rep. *26*, 815–824.e4. https://doi.org/10.1016/j.celrep.2018.12.099.

Nakatani, T., Lin, J., Ji, F., Ettinger, A., Pontabry, J., Tokoro, M., Altamirano-Pacheco, L., Fiorentino, J., Mahammadov, E., Hatano, Y., et al. (2022). DNA replication fork speed underlies cell fate changes and promotes reprogramming. Nat. Genet. *54*, 318–327. https://doi.org/10.1038/s41588-022-01023-0.

Okae, H., Toh, H., Sato, T., Hiura, H., Takahashi, S., Shirane, K., Kabayama, Y., Suyama, M., Sasaki, H., and Arima, T. (2018). Derivation of human trophoblast stem cells. Cell Stem Cell *22*, 50–63. https://doi.org/10.1016/j.stem.2017.11.004.

Ortega, N.M., Winblad, N., Plaza Reyes, A., and Lanner, F. (2018). Functional genetics of early human development. Curr. Opin. Genet. Dev. *52*, 1–6. https://doi.org/10.1016/j.gde.2018.04.005.

Pálfy, M., Joseph, S.R., and Vastenhouw, N.L. (2017). The timing of zygotic genome activation. Curr. Opin. Genet. Dev. *43*, 53–60. https://doi.org/10.1016/j.gde.2016.12.001.

Pastor, W.A., Chen, D., Liu, W., Kim, R., Sahakyan, A., Lukianchikov, A., Plath, K., Jacobsen, S.E., and Clark, A.T. (2016). Naive human pluripotent cells feature a methylation landscape devoid of blastocyst or germline memory. Cell Stem Cell *18*, 323–329. https://doi.org/10.1016/j.stem.2016.01.019.

Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., and Lanner, F. (2016). Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. Cell *165*, 1012–1026. https://doi.org/10.1016/j.cell.2016.03.023.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181. https://doi.org/10.1038/nprot.2014.006.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165. https://doi.org/10.1093/nar/gkw257.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. https://doi.org/10.1093/bioinformatics/btp616.

Rodriguez-Terrones, D., Gaume, X., Ishiuchi, T., Weiss, A., Kopp, A., Kruse, K., Penning, A., Vaquerizas, J.M., Brino, L., and Torres-Padilla, M.E. (2018). A molecular roadmap for the emergence of early-embryonic-like cells in culture. Nat. Genet. *50*, 106–119. https://doi.org/10.1038/s41588-017-0016-5.

Rojas, L.A., Valentine, E., Accorsi, A., Maglio, J., Shen, N., Robertson, A., Kazmirski, S., Rahl, P., Tawil, R., Cadavid, D., et al. (2020). p38α regulates expression of DUX4 in a model of facioscapulohumeral muscular dystrophy. J. Pharmacol. Exp. Ther. *374*, 489–498. https://doi.org/10.1124/jpet.119.264689.

Sahakyan, A., Kim, R., Chronis, C., Sabri, S., Bonora, G., Theunissen, T.W., Kuoy, E., Langerman, J., Clark, A.T., Jaenisch, R., and Plath, K. (2017). Human naive pluripotent stem cells model X chromosome dampening and X inactivation. Cell Stem Cell *20*, 87–101. https://doi.org/10.1016/j.stem.2016.10.006.

Schulz, K.N., and Harrison, M.M. (2019). Mechanisms regulating zygotic genome activation. Nat. Rev. Genet. *20*, 221–234. https://doi.org/10.1038/s41576-018-0087-x.

Shen, H., Yang, M., Li, S., Zhang, J., Peng, B., Wang, C., Chang, Z., Ong, J., and Du, P. (2021). Mouse totipotent stem cells captured and maintained through spliceosomal repression. Cell *184*, 2843–2859.e20. https://doi.org/10.1016/j.cell.2021.04.020.

Sozen, B., Jorgensen, V., Weatherbee, B.A.T., Chen, S., Zhu, M., and Zernicka-Goetz, M. (2021). Reconstructing aspects of human embryogenesis with pluripotent stem cells. Nat. Commun. *12*, 5550. https://doi.org/10.1038/s41467-021-25853-4.

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., et al. (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell *158*, 1254–1269. https://doi.org/10.1016/j.cell.2014.08.029.

Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. Bioinformatics *31*, 2032–2034. https://doi.org/10.1093/bioinformatics/btv098.

Taubenschmid-Stowers, J., Rostovskaya, M., Santos, F., Ljung, S., Argelaguet, R., Krueger, F., Nichols, J., and Reik, W. (2022). 8C-like cells capture the human zygotic genome activation program in vitro. Cell Stem Cell *29*, 449–459.e6. https://doi.org/10.1016/j.stem.2022.01.014.

Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., et al. (2014). Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell *15*, 524–526. https://doi.org/10.1016/j.stem.2014.09.003.

Wang, J., Xie, G., Singh, M., Ghanbarian, A.T., Rasko, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N.V., et al. (2014). Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature *516*, 405–409. https://doi.org/10.1038/nature13804.

Wang, Y., Zhao, C., Hou, Z., Yang, Y., Bi, Y., Wang, H., Zhang, Y., and Gao, S. (2018). Unique molecular events during reprogramming of human somatic cells to induced pluripotent stem cells (iPSCs) at naïve state. Elife *7*, e29518. https://doi.org/10.7554/elife.29518.

Whiddon, J.L., Langford, A.T., Wong, C.J., Zhong, J.W., and Tapscott, S.J. (2017). Conservation and innovation in the DUX4-family gene network. Nat. Genet. *49*, 935–940. https://doi.org/10.1038/ng.3846.

Wu, J., Xu, J., Liu, B., Yao, G., Wang, P., Lin, Z., Huang, B., Wang, X., Li, T., Shi, S., et al. (2018). Chromatin analysis in human early development reveals epigenetic transition during ZGA. Nature *557*, 256–260. https://doi.org/10.1038/s41586-018-0080-8.

Xia, W., Xu, J., Yu, G., Yao, G., Xu, K., Ma, X., Zhang, N., Liu, B., Li, T., Lin, Z., et al. (2019). Resetting histone modifications during human parental-to-zygotic transition. Science *365*, 353–360. https://doi.org/10.1126/science.aaw5118.

Xu, Y., Zhao, J., Ren, Y., Wang, X., Lyu, Y., Xie, B., Sun, Y., Yuan, X., Liu, H., Yang, W., et al. (2022). Derivation of totipotent-like stem cells with blastocyst-like structure forming potential. Cell Res. *32*, 513–529. https://doi.org/10.1038/s41422-022-00668-0.

Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1131–1139. https://doi.org/10.1038/nsmb.2660.

Yanagida, A., Spindlow, D., Nichols, J., Dattani, A., Smith, A., and Guo, G. (2021). Naive stem cell blastocyst model captures human embryo lineage segregation. Cell Stem Cell *28*, 1016–1022.e4. https://doi.org/10.1016/j.stem.2021.04.031.

Yang, M., Yu, H., Yu, X., Liang, S., Hu, Y., Luo, Y., Izsvák, Z., Sun, C., and Wang, J. (2022). Chemical-induced chromatin remodeling reprograms mouse ESCs to totipotent-like stem cells. Cell Stem Cell *29*, 400–418.e13. https://doi.org/10.1016/j.stem.2022.01.010.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287. https://doi.org/10.1089/omi.2011.0118.

Yu, L., Wei, Y., Duan, J., Schmitz, D.A., Sakurai, M., Wang, L., Wang, K., Zhao, S., Hon, G.C., and Wu, J. (2021a). Blastocyst-like structures generated from human pluripotent stem cells. Nature *591*, 620–626. https://doi.org/10.1038/s41586-021-03356-y.

Yu, L., Wei, Y., Sun, H.X., Mahdi, A.K., Pinzon Arteaga, C.A., Sakurai, M., Schmitz, D.A., Zheng, C., Ballard, E.D., Li, J., et al. (2021b). Derivation of intermediate pluripotent stem cells amenable to primordial germ cell specification. Cell Stem Cell *28*, 550–567.e12. https://doi.org/10.1016/j.stem.2020.11.003.

Zhou, F., Wang, R., Yuan, P., Ren, Y., Mao, Y., Li, R., Lian, Y., Li, J., Wen, L., Yan, L., et al. (2019). Reconstituting the transcriptome and DNA methylome landscapes of human implantation. Nature *572*, 660–664. https://doi.org/10.1038/s41586-019-1500-0.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-OCT4	Abcam	Cat#ab181557; RRID: AB_2687916
anti-SOX2	R&D	Cat#AF2018; RRID: AB_355110
anti-TFCP2L1	R&D	Cat#AF5726; RRID: AB_2202564
anti-SSEA4	Applied Stemcell	Cat#ASA-0150; RRID: AB_2827682
anti-TRA-1-60	Applied Stemcell	Cat#ASA-0160; RRID: AB_2827667
anti-CDX2	Abcam	Cat#ab76541, RRID: AB_1523334
anti-GATA3	Invitrogen	Cat#MA1-028; RRID: AB_2536713
anti-ZO1	Thermo Fisher Scientific	Cat#339100; RRID: AB_2533147
anti-aPKC	Abcam	Cat#ab4124; RRID: AB_2168401
anti-LEUTX	Thermo Fisher Scientific	Cat#PA5-59595; RRID: AB_2643351
anti-ZSCAN4	Novus Biologicals	Cat#NBP1-77120; RRID: AB_11005275
anti-β-Actin	Beyotime	Cat#AF0003; RRID: AB_2893353
anti-H3K4me3	Millipore	Cat#07-473; RRID: AB_1977252
Bacterial and virus strains		
DH5α Chemically Competent Cells	Tsingke	Cat#TSC-C14
Biological samples		
Human embryos	This study	N/A
Chemicals and recombinant proteins		
G-MOPS	Vitrolife	Cat#10130
G-1 PLUS medium	Vitrolife	Cat#10128
G-2 PLUS medium	Vitrolife	Cat#10132
Essential 8™ Medium	Thermo Fisher Scientific	Cat#A1517001
Matrigel hESC-qualified matrix	Corning	Cat#354277
Neurobasal Medium	Thermo Fisher Scientific	Cat#21103049
DMEM/F12	Thermo Fisher Scientific	Cat#11320033
Fetal bovine serum (FBS)	ExCell Bio	Cat#FSP500
BSA Fraction V (BSA)	Thermo Fisher Scientific	Cat#15260037
2-Mercaptoethanol	Thermo Fisher Scientific	Cat#21985023
N2 Supplement	Thermo Fisher Scientific	Cat#17502048
B27 Supplement	Thermo Fisher Scientific	Cat#12587010
ITS-X	Thermo Fisher Scientific	Cat#51500056
Primocin	InvivoGen	Cat#Ant-pm-2
L-Glutamine	Thermo Fisher Scientific	Cat#25030081
MEM Non-Essential Amino Acids Solution	Thermo Fisher Scientific	Cat#11140050
L-Ascorbic acid	Sigma	Cat#A4544
Accutase	Thermo Fisher Scientific	Cat#A11105-01
TrypLE™ Express Enzyme	Thermo Fisher Scientific	Cat#12604013
Recombinant Human LIF	PeproTech	Cat#AF-300-05
Recombinant Human IL-6	PeproTech	Cat#AF-200-06
Recombinant Human sIL-6 Receptor α	PeproTech	Cat#200-06RC
Recombinant Human EGF	PeproTech	Cat#AF-100-15

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
2-D08	Selleckchem	Cat#S8696
A366	Tocris	Cat#5163
A83-01	Tocris	Cat#2939
Ac-DEVD-CHO	Selleckchem	Cat#S7901
AG14361	Selleckchem	Cat#S2178
A419259	Tocris	Cat#3914
AM580	Tocris	Cat#0760
AMG548	Tocris	Cat#3920
AMI-5	Millipore	Cat#539211
AS8351	Tocris	Cat#6044
AZD2461	Tocris	Cat#6060
BIX01294	Tocris	Cat#3364
CGP77675	Cayman Chemical	Cat#21089
Chaetocin	Tocris	Cat#4504
CHIR99021	Selleckchem	Cat#S1263
CP-673451	Selleckchem	Cat#S1536
Crenolanib	Selleckchem	Cat#S2730
CTPB	Cayman Chemical	Cat#19570
DBZ	Tocris	Cat#4489
Decitabine	Selleckchem	Cat#S1200
Dimethindene	Tocris	Cat#1425
Dorsomorphin	Sigma	Cat#P5499
DZNep	Cayman Chemical	Cat#13828
EPZ015666	Tocris	Cat#6516
EZM2302	ProbeChem	Cat#PC-61030
Forskolin	Tocris	Cat#1099
GDC0994	Selleckchem	Cat#S7554
Ginkgolic acid	Tocris	Cat#6326
Go6983	Tocris	Cat#2285
GSK1120212	Selleckchem	Cat#S2673
GSK872	Tocris	Cat#6492
GSK-LSD1 dihydrochloride	Tocris	Cat#5361
HBX41108	Tocris	Cat#4285
IM-12	Enzo Life Sciences	Cat#BML-WN102-0005
JIB04	Tocris	Cat#4972
K02288	Tocris	Cat#4986
Ki16425	Selleckchem	Cat#S1315
LL507	Tocris	Cat#5365
LY2874455	Selleckchem	Cat#S7057
MM-102	Tocris	Cat#5307
Minocycline, Hydrochloride	Santa Cruz	Cat#sc-203339
Sodium butyrate (NaB)	Sigma	Cat#B5887
Nicolsamide	Tocris	Cat#4079
MS049 oxalate salt	Tocris	Cat#5685
NSC636819	Tocris	Cat#5287
OSI-744	Selleckchem	Cat#S1023
PD0325901	Selleckchem	Cat#S1036
PFI-2 HCl	Selleckchem	Cat#S7294

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Pladienolide B	Tocris	Cat#6070
Ponatinib	Selleckchem	Cat#S1490
PRT4165	Tocris	Cat#5047
RG-108	Selleckchem	Cat#S2821
SANT-1	Tocris	Cat#1974
SB431542	Selleckchem	Cat#S1067
SB590885	Tocris	Cat#2650
SGC0946	Cayman Chemical	Cat#13967
SGC2085	Selleckchem	Cat#S8340
SGC707	Tocris	Cat#5367
TCS-JNK-6o	Tocris	Cat#3222
Torin-1	Tocris	Cat#4247
Tranylcypromine	Enzo Life Sciences	Cat#BML-EI217
TSA	Selleckchem	Cat#S1045
TTNPB	Tocris	Cat#0761
UNC1215	Selleckchem	Cat#S7088
Verteporfin	Tocris	Cat#5305
Valproic acid (VPA)	Selleckchem	Cat#S3944
WH-4-023	Selleckchem	Cat#S7565
XAV939	Tocris	Cat#3748
Y-27632 2HCl	Selleckchem	Cat#S1049
Dynabeads™ Protein A	Thermo Fisher Scientific	Cat#10001D
Critical commercial assays		
HiScript III 1st Strand cDNA Synthesis Kit (+gDNA wiper)	Vazyme	Cat#R312
HiScript III RT SuperMix for qPCR (+gDNA wiper)	Vazyme	Cat#R323-01
ChamQ Universal SYBR qPCR Master Mix	Vazyme	Cat#Q711-02
Neon™ Transfection System 100 μL Kit	Thermo Fisher Scientific	MPK10096
VAHTS mRNA-seq V3 Library Prep Kit for Illumina	Vazyme	Cat#NR611
TruePrep DNA Library Prep Kit V2 for Illumina	Vazyme	Cat#TD501
TruePrep DNA Library Prep Kit V2 for Illumina	Vazyme	Cat#TD503
Chromium Next GEM Single Cell 3′ Kit v3.1	10x Genomics	Cat#1000268
Chromium Next GEM Chip G Single Cell Kit	10x Genomics	Cat#1000120
Deposited data		
prEpiSC: bulk RNA-seq and scRNA-seq	This study	GSE204801
8C::mCherry^{+} cell: RNA-seq	This study	GSE204801
8C::mCherry^{-} cell: RNA-seq	This study	GSE204801
ci8CLC: RNA-seq	This study	GSE204801
prEpiSC: ATAC-seq	This study	GSE204801
prEpiSC: H3K4me3 ChIP-seq	This study	GSE204801
Experimental models: Cell lines		
hESC (the H9 line)	WiCell	N/A
prEpiSC	This study	N/A
8C::mCherry prEpiSC	This study	N/A
ci8CLC	This study	N/A
Oligonucleotides		
Primers for RT-qPCR	See Table S2	N/A
Recombinant DNA		

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pT2-LEUTX-mCherry (8C::mCherry)	This study	N/A
pCW57.1-DUX4-WT	(Jagannathan et al., 2016)	Addgene#99282; RRID: Addgene_99,282
PB-TetON-MiniCMV-DUX4-PGK-puro-T2A-rtTA	This study	N/A
PB-TetON-MiniCMV-LEUTX-PGK-puro-T2A-rtTA	This study	N/A
PB-TetON-MiniCMV-MBD3L2-PGK-puro-T2A-rtTA	This study	N/A
PB-TetON-MiniCMV-ZSCAN4-PGK-puro-T2A-rtTA	This study	N/A
PB-TetON-MiniCMV-UBTFL1-PGK-puro-T2A-rtTA	This study	N/A
PB-TetON-MiniCMV-KDM4E-PGK-puro-T2A-rtTA	This study	N/A
PB-TetON-MiniCMV-KLF17-PGK-puro-T2A-rtTA	This study	N/A
PB-TetON-MiniCMV-ARGFX-PGK-puro-T2A-rtTA	This study	N/A
pT2-EF1a-FLAG-Biotin-KHDC3L-PGK-puro	This study	N/A
pT2-EF1a-FLAG-Biotin-FAM46B-PGK-puro	This study	N/A
SB100X	(Mates et al., 2009)	Izsvák lab
Software and algorithms		
FastQC (v0.11.9)	http://www.bioinformatics.babraham.ac.uk/projects	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
MultiQC (v1.9)	(Ewels et al., 2016)	https://multiqc.info/
TrimGalore (v0.6.5)	http://www.bioinformatics.babraham.ac.uk/projects	http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
STAR (v2.7.4a)	(Dobin et al., 2013)	https://github.com/alexdobin/STAR
featureCounts (v2.0.0)	(Liao et al., 2014)	http://subread.sourceforge.net/
TEtranscripts (v2.2.1)	(Jin et al., 2015)	https://github.com/mhammell-laboratory/TEtranscripts
Seurat (v4.0.5)	(Hao et al., 2021)	https://satijalab.org/seurat/
DESeq2 (v1.28.1)	(Love et al., 2014)	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
edgeR (v3.34.1)	(Robinson et al., 2010)	https://bioconductor.org/packages/release/bioc/html/edgeR.html
Scater (v1.20.1)	(McCarthy et al., 2017)	https://bioconductor.org/packages/release/bioc/html/scater.html
clusterProfiler (v3.16.1)	(Yu et al., 2012)	http://www.bioconductor.org/packages/release/bioc/html/clusterProfiler.html
ComplexHeatmap (v2.7.4)	(Gu et al., 2016)	https://github.com/jokergoo/ComplexHeatmap
FactoMineR (v2.4)	N/A	http://factominer.free.fr
Sambamba (v0.8.2)	(Tarasov et al., 2015)	https://lomereiter.github.io/sambamba/index.html
bowtie2 (v2.4.5)	(Langmead and Salzberg, 2012)	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
deepTools (v3.5.1)	(Ramirez et al., 2016)	https://deeptools.readthedocs.io/en/develop/
sva (v3.40.0)	(Leek et al., 2012)	http://www.bioconductor.org/packages/release/bioc/html/sva.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Cell Ranger (6.0.2)	10X Genomics	https://github.com/10XGenomics/cellranger
Other		
Human embryos: scRNA-seq	(Yan et al., 2013)	GSE36552
Human embryos: scRNA-seq	(Petropoulos et al., 2016)	E-MTAB-3929
Human epiblast (EPI): scRNA-seq	(Zhou et al., 2019)	GSE109555
Human embryos: H3K4me3 CUT&RUN	(Xia et al., 2019)	GSE124718
Human embryos: RNA-seq and ATAC-seq	(Wu et al., 2018)	GSE101571
Naive and primed hESC: ATAC-seq	(Wu et al., 2018)	GSE101571
Naive and primed hESCs: H3K4me3 ChIP-seq	(Theunissen et al., 2014)	GSE59434
Naive hiPSC (5iLA): scRNA-seq	(Liu et al., 2020)	GSE147564
Naive hESC (t2iGo): scRNA-seq	(Messmer et al., 2019)	E-MTAB-6819
Naive hESC (PXGL): scRNA-seq	(Kinoshita et al., 2021)	GSE156589
Naive hESC (Reset): bulk RNA-seq	(Takashima et al., 2014)	E-MTAB-2857
Naive (5iLA, a5iLA, AXGY, and AXGYU) and primed hESC_H9: bulk RNA-seq	(Khan et al., 2021)	GSE153212
Naive hESC_H9 (HENSM): bulk RNA-seq	(Bayerl et al., 2021)	GSE150772
Naive hESC (HNES): bulk RNA-seq	(Guo et al., 2016)	E-MTAB-4461
Naive (t2iLGo_NK2) and primed (Primed_NK2) hESCs: bulk RNA-seq	(Collier et al., 2017)	GSE93241
Formative hiPSC (FTW): bulk RNA-seq	(Yu et al., 2021b)	GSE135989
Naive (2iL) and primed hESC_H9: bulk RNA-seq	(Hu et al., 2020)	GSE87452
hEPSC and hESC (mTeSR1): scRNA-seq	(Liu et al., 2021a)	GSE156920
hEPSC: bulk RNA-seq	(Liu et al., 2021a)	GSE147839
hESC: scRNA-seq	(Sahakyan et al., 2017)	GSE87237
hESC_DUX4^{OE}: RNA-seq	(Hendrickson et al., 2017)	GSE95516
FSHD: RNA-seq	(Rojas et al., 2020)	GSE153301

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jichang Wang (wangjch53@mail.sysu.edu.cn).

Materials availability

All stable reagents, plasmids, and cell lines generated in this study are available from the lead contact.

Data and code availability

- The raw data of RNA-seq have been deposited in NCBI's GeneExpression Omnibus (GEO) with the accession number GEO: GSE204801.
- This paper does not report original code. The scripts of bioinformatics analysis are put in Github (https://github.com/ChenManqi2/ci8CLC_scripts).
- Any additional information required to re-analyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human embryos

The regulatory framework about the use of human embryos for this study was based on the Human Biomedical Research Ethics Guidelines (set by National Health Commission of the People's Republic of China on 1 December 2016), the 2021 Guidelines for Stem Cell Research and Clinical Translation (issued by the International Society for Stem Cell Research, ISSCR) and the Human Embryonic Stem Cell Research Ethics Guidelines (set by China National Center for Biotechnology Development on 24 December 2003). All the experiments in this study are in compliance with these relevant ethical regulations.

The aim and protocols for the use of human three-pronuclear (3PN) embryos in the study have been reviewed and approved by the Ethical Committee of Sun Yat-sen Memorial Hospital, Sun Yat-sen University (SYSEC-KY-KS-2021-095). The embryos were donated by patients undergoing Assisted Reproductive Technology (ART) treatment at the Center for Reproductive Medicine, Sun Yat-sen Memorial Hospital, who have signed informed consent form for the use of their embryos or immature oocytes and leftover sperms for research. This study was performed in compliance with all relevant ethical regulations.

Cell culture

Human embryonic stem cells (hESCs) (the H9 line provided by WiCell) were cultured with the Essential 8 hESC medium in Matrigel-coated plates at 37°C, 5% CO_2, 5% O_2 conditions. The medium was changed daily. Human ESCs were treated with 0.48 mM EDTA and then passaged onto new Matrigel-coated plates every 4-5 days.

HEK293 cells (provided by ATCC) were cultured in the DMEM medium containing 10% fetal bovine serum and Penicillin/Streptomycin.

The cells are routinely determined for mycoplasma contamination. All of the cells used for experiments are not infected with mycoplasma.

Derivation and culture of prEpiSCs

The LTR7-GFP hESC_H9 cell line used in our previous report (Wang et al., 2014) was employed for chemical screening. The basal culture medium (500 mL) is prepared as follows: 240 mL DMEM/F12, 240 mL Neurobasal medium, 2 mM L-glutamine, 1X non-essential amino acids, 1% N2 supplement, 2% B27 supplement, 100 μg/mL Vitamin C, 50 ng/mL Bovine Albumin Fraction V, 20 ng/mL IL6, 20 ng/mL sIL-6R, 0.1 mM β-mercaptoethanol, Primocin. LTR7-GFP hESCs were passaged onto feeder cells and cultured in Essential 8 hESC medium for around 2-3 days until the hESC colonies grew large enough (Diameter: 200-500 μm). Then the medium was switched to the induction medium (basal culture medium plus 5 ng/mL BMP4 and chemical cocktail [0.05 μM DZNep, 0.05 or 0.1 μM GSK1120212, 0.2 μM A419259, 2.5 μM XAV939, 1 μM Go6983]). The hESCs were passaged and cultured in the GIX medium (basal culture medium plus chemical cocktail [0.2 μM GSK1120212, 2.5 μM XAV939, 1 μM Go6983]) from day 5. On day 5–10 the GFP^+, dome-shaped colonies exist. The medium was changed every other day. The dome-shaped colonies were dissociated with Accutase into single cells, and then passaged onto new Matrigel- or feeder-coated plates every 3 - 4 days.

To establish the 8C:mCherry prEpiSC line, the pT2-LEUTX-mCherry vector was integrated into the prEpiSC genome using SB100X transgenesis (Mates et al., 2009), according to our previous protocol (Yang et al., 2022).

Human ci8CLC culture

The basal culture medium (500 mL) is prepared as follows: 240 mL DMEM/F12, 240 mL Neurobasal medium, 2 mM L-glutamine, 1X non-essential amino acids, 1% N2 supplement, 2% B27 supplement, 100 μg/ml L-ascorbic acid, 50 ng/mL Bovine Albumin Fraction V, 20 ng/mL IL6, 20 ng/mL sIL-6R, 5 ng/mL BMP4, 0.1 mM β-mercaptoethanol, Primocin. Human prEpiSCs were cultured with the basal culture medium plus the chemical cocktail (0.2 μM GSK1120212, 2.5 μM XAV939, 1 μM Go6983, 0.05 μM DZNep, 0.25 μM CBL0137, 0.5 μM AG14361, 1 μM GSK872, and 0.25 μM Ac-DEVD-CHO).

METHOD DETAILS

Construct cloning

The TetON-DUX4-PGK-puro-T2A-rtTA fragment was cut from pCW57.1-DUX4-WT plasmid that was a gift from Stephen Tapscott (Addgene plasmid # 99,282; http://n2t.net/addgene:99282; RRID:Addgene_99,282) and then inserted into the empty *Piggybac* vector. The DUX4 fragment was then removed from the *Piggybac* vector to generate the *PB-TRE-PGK-puro-rtTA* vector. The CDS (coding sequence) of ARGFX, LEUTX, ZSCAN4, KDM4E, KLF17, MBD3L2 or UBTFL1 was amplified using prEpiSC cDNA as the template, and then cloned into the *PB-TRE-PGK-puro-rtTA* vector, to generate a Doxycycline-inducible expressing vector. The CDS of FAM46B or KHDC3L was amplified and then cloned into the pEF1a-Flagbio(FLBIO)-puro vector (gift from Prof. Junjun Ding).

Microinjection of 8C::mCherry and pT2-CAG-GFP reporters into human zygotes

Three-pronuclear (3PN) zygotes were collected 16-18h after the fertilization *in vitro*. G-MOPS medium was used as *in vitro* operating medium. The ICSI injection needle (TPC, Australia) with a diameter of 5 μm was used to inject the 8C::mCherry or pT2-CAG-GFP plasmid (200 ng/μL) into 3PN zygotes. The injected zygotes are cultured in G-1 PLUS medium at 37°C in 6% CO_2, 5% O_2 and 89% N_2 for three days to the 8-cell embryo stage, and then switched to G-2 PLUS medium until the morula stage.

Generation of blastoids from 8CLCs

20-25 8CLCs were injected into the empty mouse zona pellucida, and then cultured in the modified G-2 PLUS medium containing 10 ng/mL human LIF, 5 ng/mL BMP4, 1.5 μM PD0325901, 0.5 μM LPA, and 1 μM A83-01.

Derivation of hESCs from 8CLC-blastoids

The 8CLC-blastoids were seeded on the feeders and cultured with the Essential 8 hESC medium in the four-well plate in 5% CO_2 and 5% O_2. Half of the medium was changed every other day. About one week later, the colonies pre-treated with Y-27632 (10 μM) overnight were dissociated into single cells with Accutase and seeded in the Matrigel-coated plate in the Essential 8 hESC medium containing 10 μM Y-27632. From the second day after passaging, the Essential 8 hESC medium was changed routinely every day.

Derivation of hTSCs from 8CLC-blastoids

Human trophoblast stem cells (hTSCs) were derived from 8CLC-blastoids according to the method from a previous report (Okae et al., 2018). Briefly, The 8CLC-blastoids used for hTSC derivation were cultured with the G-2 PLUS medium in the four-well plate that was pre-treated with collagen IV for at least 30 min, then switched to the hTSC medium and cultured in 5% CO_2 and 5% O_2. Half of the medium was changed every other day until the clone was formed. About one week later, the clones were dissociated into single cells with TrypLE and seeded in the Collagen IV-coated plate in the hTSC medium. The hTSC medium was prepared as the previous report (Okae et al., 2018): DMEM/F12 supplemented with 0.2% FBS, 0.1 mM 2-mercaptoethanol, 1% ITS-X, 50 μg/ml L-ascorbic acid, 0.3% BSA, 50 ng/mL hEGF, 2 μM CHIR99021, 0.5 μM A83-01, 1 μM SB431542, 0.8 mM VPA, 5 μM Y27632 and Primocin.

In vitro differentiation assay

To spontaneously differentiate prEpiSCs and 8CLC-blastoid-derived hESCs to embryoid bodies (EBs), the cells from one well were dissociated with Accutase for 3-5 minutes into single cells. The single hESCs were transferred into one 10-cm low-attached dish, and cultured in the differentiation medium (KnockOut DMEM, 20% KnockOut serum replacement, 2 mM L-glutamine, 1% nonessential amino acids, 0.1 mM 2-mercaptoethanol and primocin). The medium was changed every other day. The EBs were cultured for eight days followed by collection for RT-qPCR.

Flow cytometry analysis

8C::mCherry prEpiSCs and ci8CLCs were dissociated with Accutase. After the dissociation, cells were washed once and analyzed in PBS supplemented with 0.1% BSA using the CytoFLEX S flow cytometer (Beckman).

RT-qPCR

Total RNA was extracted from cells using the Trizol kit (Thermo Fisher Scientific, Cat#15596018). 0.5 μg purified RNA was used for reverse transcription (RT) with the HiScript III first Strand cDNA Synthesis Kit (+gDNA wiper) (Vazyme, Cat#R312). Quantitative RT–PCR (RT–qPCR) was performed using the ChamQ Universal SYBR qPCR Master Mix (Vazyme, Cat#Q711) on the LightCycle480 II (Roche). Data were normalized to *H3F3A* and *GAPDH* expression using the $2^{-\Delta\Delta CT}$ method. Error bars represent the standard deviation (SD) of samples carried out in triplicates.

Western blot

Human fibroblasts, ESCs, prEpiSCs, and ci8CLCs cultured on 6-well plate were washed twice with ice-cold DPBS and then lysed on ice in extraction buffer containing 50 mM Tris-base (pH 7.4), 100 mM NaCl, 1% NP-40, 10 mM EDTA, 20 mM NaF, 1 mM PMSF, 3 mM Na_3VO4, and protease inhibitors. Protein samples were separated by 4–20% SurePAGE precast gel (Genscript, cat#M00656) and then transferred to nitrocellulose membranes (Millipore, catalog. no. HATF00010). After blocking with 5% BSA in Tris-buffered saline containing 0.1% Tween 20 (TBST) for one hour at room temperature, transferred membranes were incubated overnight at 4°C with different primary antibodies against β-Actin (1:2000 dilution) or LEUTX (1:1000 dilution). After horseradish peroxidase-conjugated secondary antibody incubation, bands were visualized using a ChemiDoc Touch Imaging System (Bio-Rad).

Immunofluorescence

Cells were grown on Matrigel/feeder-coated glass coverslips (Nest biotechnology, cat#801008). The cells were fixed with fresh 4% paraformaldehyde/phosphate buffer for 10 minutes at room temperature, washed three times with DPBS, and permeabilized in 0.2% Triton/DPBS for 30 minutes. Cells were blocked with 5% BSA/DPBS and incubated with primary antibodies diluted in 1% BSA/DPBS overnight at 4°C. Cells were then washed three times with DPBS, incubated with secondary antibodies for one hour at room temperature, washed with DPBS, mounted by Fluoroshield with DAPI (Sigma, catalog. no. F6057) and imaged with a confocal microscope (Nikon C2). The following antibodies were used at the indicated dilutions: rabbit anti-OCT4 (1:200), goat anti-SOX2 (1:200), goat anti-TFCP2L1 (1:100), anti-SSEA4 (1:50), anti-TRA-1-60 (1:50), rabbit anti-ZSCAN4 (1:200), rabbit anti-LEUTX (1:200), mouse anti-GATA3 (1:200), rabbit anti-CDX2 (1:200).

RNA-seq

Total RNA was extracted from hESCs, prEpiSCs, 8C::mCherry$^-$, 8C::mCherry$^+$ and ci8CLCs using the Trizol kit (Thermo Fisher Scientific, catalog. no. 15596018) following the manufacturer's instructions. After extraction, PolyA(+) RNA was isolated and RNA-seq library was constructed following the VAHTS mRNA-seq V3 Library Prep Kit protocol. Sequencing was performed on the Illumina NovaSeq platform.

Smart-seq2 scRNA-seq of prEpiSCs

Single prEpiSCs were individually picked into 2.5 μL RLT plus buffer (Qiagen, catalog. no. 1053393). RNA was captured on streptomycin C1 beads conjugated with Biotinylated Oligo-dT30VN primer. The volume of reverse transcription was adjusted to 10 μL. cDNA was generated according to the smart-seq2 protocol (Picelli et al., 2014). Sequencing libraries were generated and indexed from 1ng of cDNA per sample using the TruePrep DNA Library Prep Kit V2 for Illumina (Vazyme, catalog. no. TD503) according to manufacturer's instructions. Samples were pooled and sequenced on the Illumina Hiseq X Ten platform and 150-bp paired-end reads were generated.

Single-cell transcriptome profiling of prEpiSCs by 10X Genomics

Human prEpiSCs were dissociated into single cells using Accutase. The cells were suspended in 0.04% BSA/PBS and centrifuged for five minutes at 300 g. The cells were resuspended, and the cell suspension was filtered using Falcon 40-μM Cell Strainers (Corning). GFP^+ cells were sorted from the filtered cells by FACS and centrifuged for five minutes at 300 g. The cells were diluted to a concentration of 700–1200 cells/μL (viability $\geq$85%) as determined using the Countess II Automated Cell Counter (Thermo Fisher Scientific) and then subjected to single-cell RNA sequencing (10x Genomics, Chromium Single Cell 3′ Reagent Kits, v3). Single-cell libraries were generated following the manual instructions of 10x Genomics (Chromium Single Cell 3′ Reagent Kits, v3). The libraries were sequenced on the NovaSeq platform (Illumina) to generate 150-bp paired-end reads.

ATAC-seq

ATAC-seq was performed as previously described (Buenrostro et al., 2013). First, a total of 50,000 cells were resuspended in 50 μL lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% (v/v) NP-40). The nuclei were collected by centrifugation for 10 minutes at 500g at 4°C and then mixed with the transposition reaction mix (10 μL TTBL, 5 μL TTE Mix V50 and nuclease-free H_2O up to 50 μL in total) of TruePrep DNA Library Prep Kit V2 for Illumina for 10 minutes at 55°C. DNA samples were amplified by PCR according to the manufacturer's instructions. PCR products were mixed with 1.2 volume of DNA clean beads to clean up. Library integrity was evaluated by the Agilent bioanalyzer. The libraries were sequenced on an Illumina Hiseq X Ten platform and 150-bp paired-end reads were generated.

ChIP-seq

A total of 5×10^6 prEpiSCs were fixed in freshly made 1% formalin/prEpiSC medium (v/v) for 10 minutes followed by quenching with glycine (125 mM) for five minutes at room temperature. Fixed cells were washed twice in ice-cold PBS, re-suspended in lysis buffer 1 (50 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5%NP-40, 0.25% Triton X-100 and protease inhibitors) and incubated on ice for 10 minutes. Cells were then pelleted by centrifugation and re-suspended in lysis buffer 2 (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and protease inhibitors) followed by incubation on ice for 10 minutes. Cells were then pelleted and re-suspended in 0.25 mL of freshly prepared lysis buffer 3 (10 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-deoxycholate, 0.5%N-lauroylsarcosine and protease inhibitors) and sonicated to 100 - 500 bp fragments using a Bioruptor (Diagenode). Anti-H3K4me3 antibody was incubated with preblocked Dynabeads Protein A at 4°C with gentle agitation on a rotator for four hours. Antibody-bead complexes were added into the sonicated chromatin solution and rotated for overnight at 4°C. Chromatin-bead immunoprecipitates were collected by the magnetic device (Thermo Fisher Scientific) and washed sequentially with high-salt wash buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 2 mM EDTA, 0.1% SDS, and 1% Triton X-100), LiCl wash buffer (10 mM Tris-HCl, pH 8.0, 250 mM LiCl, 1 mM EDTA, and 1% Nonidet P-40), and then with Tris-EDTA-NaCl buffer (1 mM EDTA, 10 mM Tris-HCl, pH 8.0, and 50 mM NaCl). Immunoprecipitated DNA was eluted by incubating the beads with 100 μL elution buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, and 1% SDS), shaking at 1300 rpm for one hour at room temperature. For reverse crosslinking, samples were treated with proteinase K and incubated at 65°C for four hours. The DNA was purified using phenol-chloroform extraction and suspended in 40 μL TE buffer. The purified DNA was subjected to library preparation with the Accel-NGS 2S Plus DNA Library kit (Swift Biosciences, Cat#21096) following the manufacturer's instructions. The libraries were sequenced on an Illumina Hiseq platform and 150-bp paired-end reads were generated.

QUANTIFICATION AND STATISTICAL ANALYSIS

Processing of bulk RNA-seq and scRNA-seq data

For bulk RNA-seq data, we first examined the quality of raw sequencing reads by FastQC (v0.11.9), whose results were parsed with MultiQC (v1.9) (Ewels et al., 2016). In order to trim adapter sequences and remove poor quality reads (quality <20 or length <30), we utilized TrimGalore (v0.6.6) for reads filtering. Pre-processed reads were mapped to human reference genome GRCh38/hg38 using STAR (release 2.7.5b) (Dobin et al., 2013) with parameters "–outSAMmultNmax 1 –outFilterMultimapNmax 10" for gene expression analysis and "–outFilterMultimapNmax 1000 –outSAMmultNmax 1 –outFilterMismatchNmax 3 –outMultimapperOrder Random –winAnchorMultimapNmax 1000 –alignEndsType EndToEnd –alignIntronMax 1 –alignMatesGapMax 350" for repeat expression analysis. Raw count tables of genes and repeats were computed using featureCounts package (release 2.0.0) (Liao et al., 2014) with default parameters, only considering uniquely mapped reads. Quantification of repeat expression on the subfamily level was

conducted by featureCounts with parameter ''-M''. The gene and repeat expression levels were normalized by transforming raw counts into Transcripts Per Million (TPM). Those repeat loci with TPM $\geq$ 0.1 were considered as active loci.

For 10X Genomics scRNA-seq data, raw sequencing reads were aligned to the GRCh38 human genome and quantified using the toolkit Cell Ranger (v.6.0.2).

Differential expression analysis

R packages edgeR (v3.34.1) (Robinson et al., 2010) and DESeq2 (v1.32.0) (Love et al., 2014) were applied for differential expression analyses of bulk RNA-seq data across different groups (presented in Figures 3O, 4G, and 4J, respectively). Genes or repeats with absolute $\log_2$(Foldchange) value $\geq$ 1 or 2 and adjust p value (p value for repeats) < 0.05 were selected as significant differentially expressed genes/repeats. Functional enrichment analyses and gene set enrichment analyses were performed by R package clusterProfiler (v4.0.5) (Yu et al., 2012).

Downstream analysis of scRNA-seq data

The downstream analysis of scRNA-seq data was performed with Seurat package (v4.0.5) (Hao et al., 2021). First, we filtered cells with low quality according to the number of detected genes, total number of reads and percentages of mitochondrial genes using 'isOutlier' function in R package Scater (v1.20.1) (McCarthy et al., 2017). Gene count values were normalized and scaled by ''LogNormalize'' and ''ScaleData'' function. Principle component analysis was based on top 2000 highly variable genes defined by 'FindVariableGenes' function. Taking background noise into account, we only used the significant principal components selected by function ''ElbowPlot'' for further analyses. Cell clustering was then conducted through the ''FindClusters'' and ''RunTSNE'' function. For cell type annotation, maker genes of each cell cluster were identified using ''FindAllMarkers'' function. In order to correct batch effect, the human embryo scRNA-seq datasets (E-MTAB-3929 and GSE36552) were integrated with naive hESC scRNA-seq dataset by functions ''FindIntegrationAnchors'' and ''IntegrateData'' using default parameters. Plots were generated by ''DimPlot'', ''FeaturePlot'', ''DotPlot'' and ''DoHeatmap'', separately.

Definition of human 8C-specific genes

The public human embryo scRNA-seq datasets (E-MTAB-3929 and GSE36552) were re-analyzed for definition of 8C-specific genes. We first performed differential gene expression analyses between 8C embryos and other development stages (from zygote to E7 blastocyst) by R package edgeR (Robinson et al., 2010), identifying a set of genes with a minimal two-fold significantly higher expression level than those in all other stages. Then genes with average TPM lower than 5 in 8C embryos were filtered. *Z* score of TPM values was calculated for visualizing 8C-specific genes in heatmap by R package ComplexHeatmap (v2.8.0) (Gu et al., 2016).

Processing of ChIP-seq and ATAC-seq data

After quality control with FastQC and MultiQC, the adapters were trimmed and reads with poor quality were filtered using TrimGalore. Then the duplicate reads were removed with package Sambamba (v0.8.2) (Tarasov et al., 2015). Cleaned reads were aligned to human genome GRCh38/hg38 using bowtie2 (v2.4.5) (Langmead and Salzberg, 2012), reporting the alignments with best mapping quality. Using the deepTools toolkits (v3.5.1) (Ramirez et al., 2016), we calculated genome-wide histone modification and chromatin accessibility signal for downstream analyses.

Principal component analysis

Principal component analyses (PCA) of cell lines and embryo datasets were processed using R package FactoMineR (v2.4). Before PCA, we transformed the gene raw counts and genome-wide ATAC-seq or H3K4me3 signals into TPM (RNA-seq) and RPKM (ATAC-seq and H3K4me3), respectively. Those genes or genomic regions with low TPM/RPKM in all samples were filtered out. Batch effects between datasets were corrected via 'ComBat' function in R package sva (v3.40.0) (Leek et al., 2012).

Statistical analysis

The statistical results were presented as the means and error bars represented standard derivation (SD). Statistical significance was determined by unpaired student t test as indicated in figure legends. The p values below 0.05 were considered statistically significant. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Cell Reports, Volume *39*

Supplemental information

Recapitulating early human development with 8C-like cells

Xiu Yu, Shiqi Liang, Manqi Chen, Hanwen Yu, Ruiqi Li, Yuliang Qu, Xuhui Kong, Ruirui Guo, Rongyan Zheng, Zsuzsanna Izsvák, Chuanbo Sun, Mingzhu Yang, and Jichang Wang

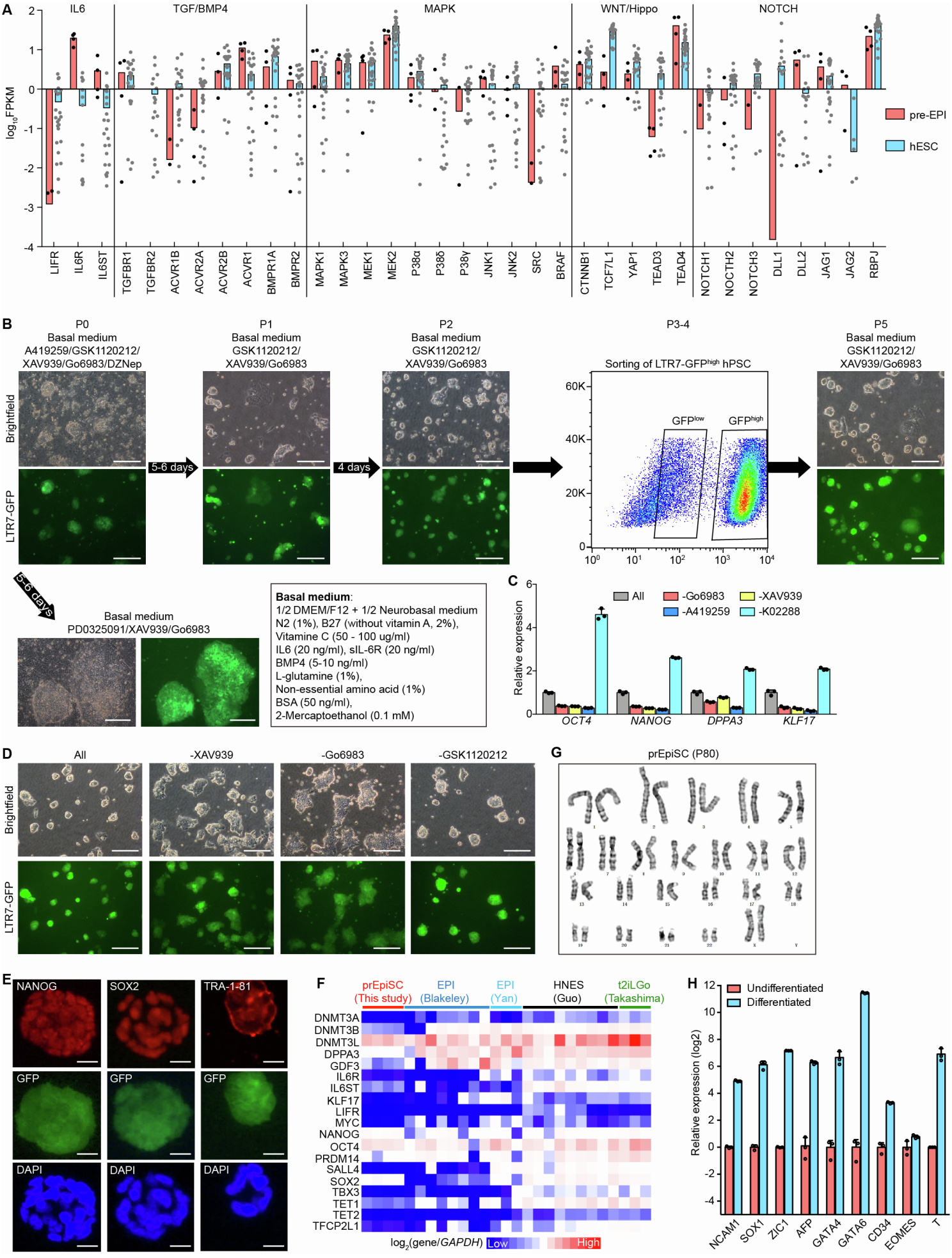

A
IL6
TGF/BMP4
MAPK
WNT/Hippo
NOTCH
log10FPKM
pre-EPI
hESC
LIFR
IL6R
IL6ST
TGFBR1
TGFBR2
ACVR1B
ACVR2A
ACVR2B
ACVR1
BMPR1A
BMPR2
MAPK1
MAPK3
MEK1
MEK2
P38α
P38δ
P38γ
JNK1
JNK2
SRC
BRAF
CTNNB1
TCF7L1
YAP1
TEAD3
TEAD4
NOTCH1
NOCTH2
NOTCH3
DLL1
DLL2
JAG1
JAG2
RBPJ
B
P0
Basal medium
A419259/GSK1120212/
XAV939/Go6983/DZNep
P1
Basal medium
GSK1120212/
XAV939/Go6983
P2
Basal medium
GSK1120212/
XAV939/Go6983
P3-4
Sorting of LTR7-GFPhigh hPSC
P5
Basal medium
GSK1120212/
XAV939/Go6983
Brightfield
LTR7-GFP
5-6 days
4 days
GFPlow
GFPhigh
5-6 days
Basal medium
PD0325091/XAV939/Go6983
Basal medium:
1/2 DMEM/F12 + 1/2 Neurobasal medium
N2 (1%), B27 (without vitamin A, 2%),
Vitamine C (50 - 100 ug/ml)
IL6 (20 ng/ml), sIL-6R (20 ng/ml)
BMP4 (5-10 ng/ml)
L-glutamine (1%),
Non-essential amino acid (1%)
BSA (50 ng/ml),
2-Mercaptoethanol (0.1 mM)
C
All
-Go6983
-XAV939
-A419259
-K02288
Relative expression
OCT4
NANOG
DPPA3
KLF17
D
All
-XAV939
-Go6983
-GSK1120212
Brightfield
LTR7-GFP
G
prEpiSC (P80)
E
NANOG
SOX2
TRA-1-81
GFP
DAPI
F
prEpiSC
(This study)
EPI
(Blakeley)
EPI
(Yan)
HNES
(Guo)
t2iLGo
(Takashima)
DNMT3A
DNMT3B
DNMT3L
DPPA3
GDF3
IL6R
IL6ST
KLF17
LIFR
MYC
NANOG
OCT4
PRDM14
SALL4
SOX2
TBX3
TET1
TET2
TFCP2L1
log2(gene/GAPDH)
Low
High
H
Undifferentiated
Differentiated
Relative expression (log2)
NCAM1
SOX1
ZIC1
AFP
GATA4
GATA6
CD34
EOMES
T

Figure S1. Derivation and characterization of prEpiSCs, related to Figure 1

(A) Comparison of multiple genes enriched for IL6, TGF/BMP4, MAPK, WNT/Hippo, and NOTCH signaling pathways between human epiblast and ESCs (Yan et al., 2013).

(B) The procedure of derivation of prEpiSCs from primed hESCs. Scale bar, 100 μm.

(C) The expression levels of pluripotent genes during prEpiSCs induction under the indicated treatment. Data are shown as means ± S.D. of technical replicates from two independent experiments.

(D) The representative images showing essential chemicals for prEpiSC maintenance. Scale bar, 100 μm.

(E) The immunofluorescence analysis showing expression of NANOG, SOX2 and TRA-1-81 in prEpiSCs. Scale bar, 20 μm.

(F) The heatmap showing the expression levels (normalized to *GAPDH*) of pluripotent genes in prEpiSCs, naïve hPSCs (HNES and t2iLGo) and human epiblast (EPI).

(G) Karyotyping analysis for prEpiSCs.

(H) The lineage marker expression levels showing the *in vitro* differentiation potential of prEpiSCs. Data are shown as means ± S.D. (n=3 biological replicates).

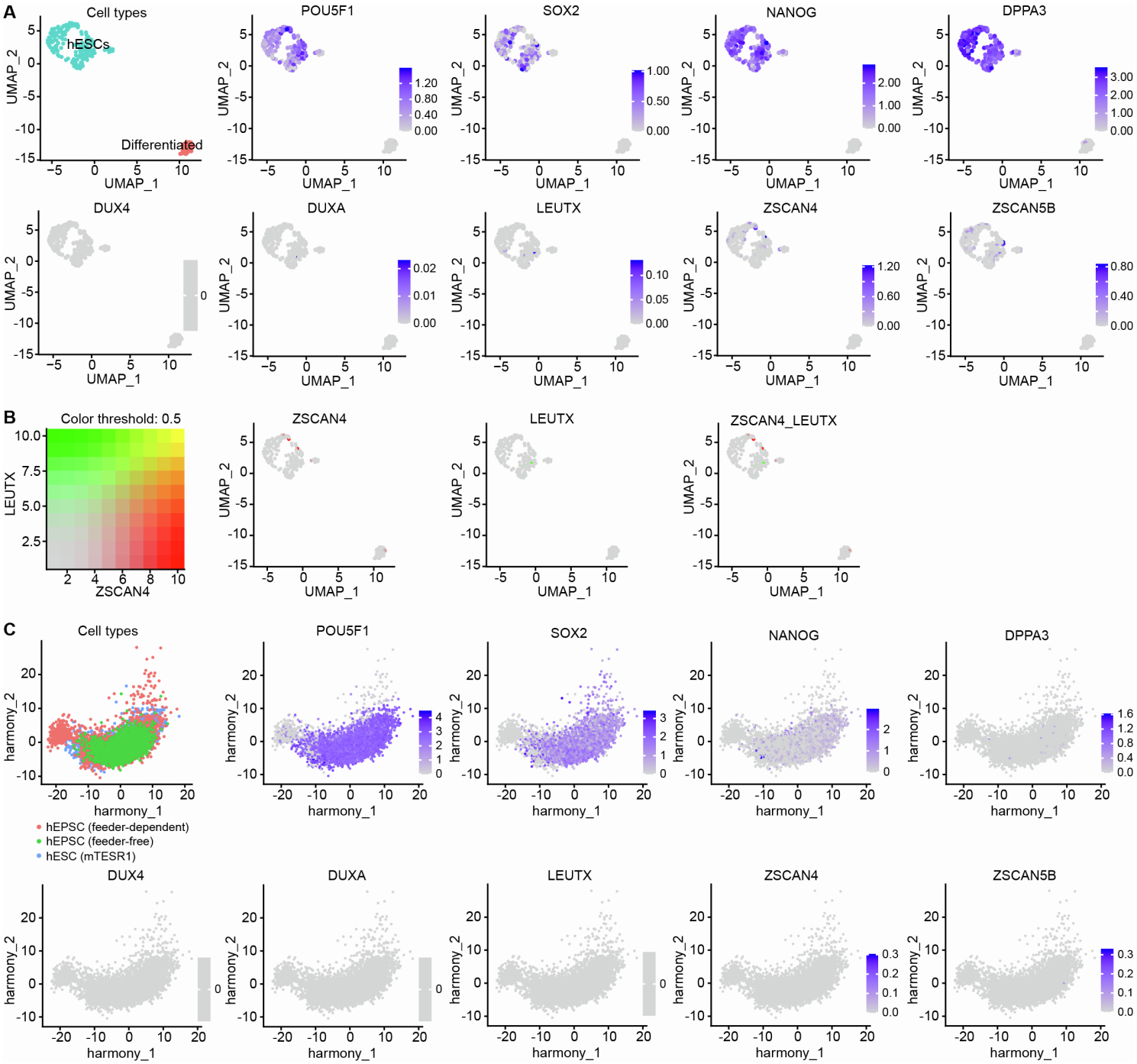


Figure S2. Expression of pluripotent and 8C-specific genes in human primed and extended pluripotent stem cells, related to Figure 2

(A) Expression levels of pluripotent and 8C-specific genes in human primed pluripotent stem cells, detected by scRNA-seq.

(B) Analysis for co-expression of *LEUTX* and *ZSCAN4* in human primed pluripotent stem cells.

(C) Expression of pluripotent and 8C-specific genes in human extended pluripotent stem cells, detected by scRNA-seq.

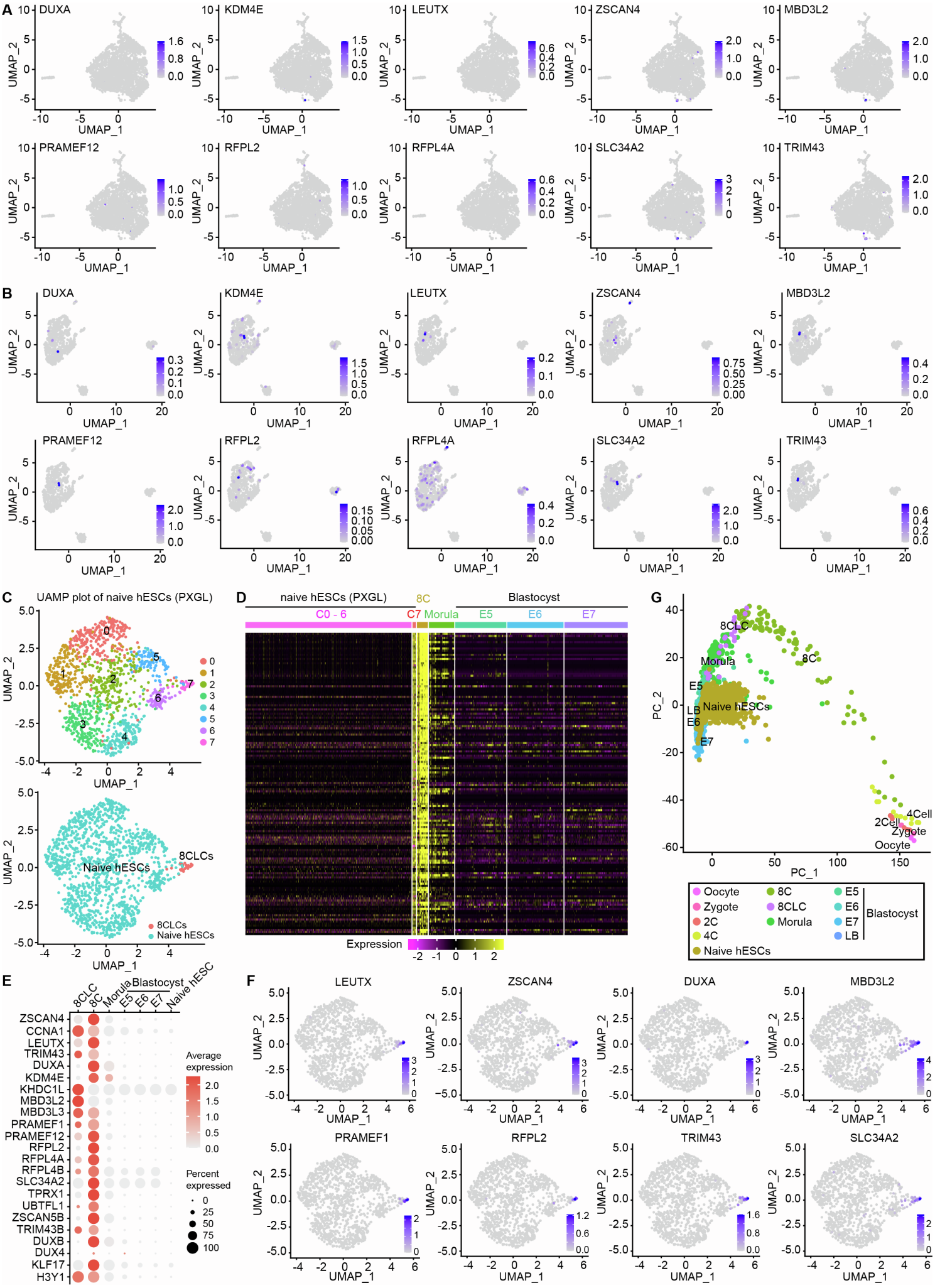

A
DUXA
KDM4E
LEUTX
ZSCAN4
MBD3L2
PRAMEF12
RFPL2
RFPL4A
SLC34A2
TRIM43
UMAP_1
UMAP_2
B
DUXA
KDM4E
LEUTX
ZSCAN4
MBD3L2
PRAMEF12
RFPL2
RFPL4A
SLC34A2
TRIM43
UMAP_1
UMAP_2
C
UAMP plot of naive hESCs (PXGL)
UMAP_1
UMAP_2
Naive hESCs
8CLCs
D
naive hESCs (PXGL)
8C
Blastocyst
C0 - 6
C7 Morula
E5
E6
E7
Expression
G
8CLC
8C
Morula
E5
Naive hESCs
LB
E6
E7
2Cell
4Cell
Zygote
Oocyte
PC_1
PC_2
Oocyte
Zygote
2C
4C
Naive hESCs
8C
8CLC
Morula
E5
E6
E7
LB
Blastocyst
E
8CLC
8C
Morula
E5
E6
E7
Naive hESC
Blastocyst
ZSCAN4
CCNA1
LEUTX
TRIM43
DUXA
KDM4E
KHDC1L
MBD3L2
MBD3L3
PRAMEF1
PRAMEF12
RFPL2
RFPL4A
RFPL4B
SLC34A2
TPRX1
UBTFL1
ZSCAN5B
TRIM43B
DUXB
DUX4
KLF17
H3Y1
Average expression
Percent expressed
F
LEUTX
ZSCAN4
DUXA
MBD3L2
PRAMEF1
RFPL2
TRIM43
SLC34A2
UMAP_1
UMAP_2

Figure S3. Identification of 8CLCs in human naïve pluripotent stem cells, related to Figure 2

(A) Expression levels of multiple 8C-specific genes in human naive pluripotent stem cells cultured in the t2iLGo condition.

(B) Expression levels of multiple 8C-specific genes in human naive pluripotent stem cells cultured in the 5iLA condition.

(C) UMAP plots of scRNA-seq data of human naive pluripotent stem cells cultured in the PXGL condition.

(D) Expression levels of 8C-specific genes in human embryos and naïve pluripotent stem cells cultured in the PXGL condition.

(E) The frequency of expression and average expression levels for selected 8C-specific genes in human embryos and 8CLCs discovered from PXGL-cultured naïve pluripotent stem cells.

(F) The UMAP plot showing expression of selected 8C-specific genes in single PXGL-cultured naïve pluripotent stem cell.

(G) Principle component analysis of transcriptome of early human embryos and 8CLCs at the single-cell level.

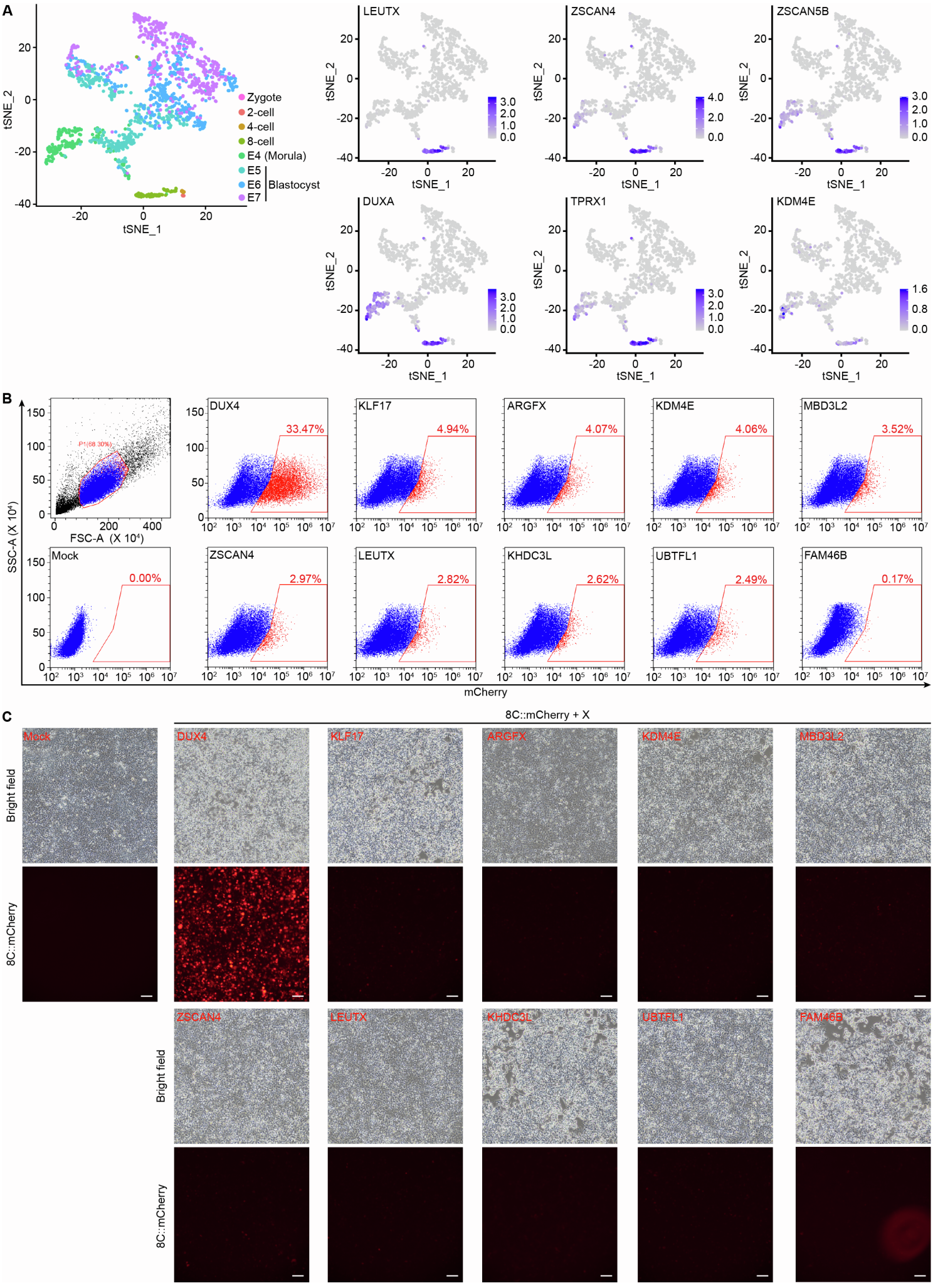

A
tSNE_2
tSNE_1
Zygote
2-cell
4-cell
8-cell
E4 (Morula)
E5
E6
E7
Blastocyst
LEUTX
ZSCAN4
ZSCAN5B
DUXA
TPRX1
KDM4E
B
SSC-A (X 10⁴)
FSC-A (X 10⁴)
P1(68.30%)
DUX4
33.47%
KLF17
4.94%
ARGFX
4.07%
KDM4E
4.06%
MBD3L2
3.52%
Mock
0.00%
ZSCAN4
2.97%
LEUTX
2.82%
KHDC3L
2.62%
UBTFL1
2.49%
FAM46B
0.17%
mCherry
C
8C::mCherry + X
Bright field
8C::mCherry
Mock
DUX4
KLF17
ARGFX
KDM4E
MBD3L2
ZSCAN4
LEUTX
KHDC3L
UBTFL1
FAM46B

Figure S4. Development of the 8C::mCherry reporter, related to Figure 3

(A) Expression levels of selected 8C-specific transcription factors and epigenetic modifiers in early human embryos.

(B) FACS analysis of the percentage of 8C::mCherry^{+} HEK293 cells under overexpression of individual 8C-specific gene.

(C) The representative images showing differential activation of the 8C::mCherry reporter by 8C-specific transcription factors and epigenetic modifiers. Scale bar, 100 μm.


A
B
Stage
MTOR
HIPPO
JAK-STAT
PKC
PKA
TGF
NOTCH
WNT
MAPK & AMPK
MAPKAP1
MTOR
RAC1
PRR5
RICTOR
MLST8
RPS6KB1
CDC42
RPTOR
LPAR2
TEAD4
LPAR1
TEAD1
YAP1
TEAD3
TEAD2
LIFR
STAT3
IL6ST
IL6R
STAT2
JAK1
STAT1
PRKCD
PRKCI
PRKACA
PRKAR2B
PRKAR2A
PRKAR1A
PRKAR1B
TGFBR3
SMAD5
ACVR1
BMPR1A
ID2
ID1
ID3
ACVR2A
TGFBR1
ACVR2B
BMPR2
SMAD1
ACVR1B
SMAD2
SMAD3
RBPJ
DLL3
JAG1
JAG2
FZD7
LDLR
PORCN
TCF7L1
FZD4
WNT9A
WNT3
APC
LRP6
FZD3
TCF7
AXIN2
FZD5
TCF7L2
GSK3B
CTNNB1
MAP2K5
MAP2K3
FGFR4
ATF4
FOS
JUNB
MAP2K2
MAP2K7
SRF
RAC1
MKNK2
JUND
MAPK9
MAP2K4
MAPK14
TP53
ARAF
RAF1
MAP3K8
GRB2
RPS6KA3
RRAS
MAPK1
KRAS
MAP3K1
PRKAA1
NRAS
MAP4K3
MAP2K1
PAK2
MAP3K3
MAPK8
ATF2
ELK4
MAPK7
MAP4K1
PAK1
MAP3K2
FGFR3
MAPK3
MAP3K11
FGFR2
FGFR1
MAP4K4
ELK1
MAPK13
MKNK1
ACTR5
SIRT3
KDM4D
KDM5C
KAT8
MBTD1
JMJD8
SUV39H2
RBBP7
RBBP4
ELP3
INO80D
KAT7
HAT1
CBX5
SMYD3
SETD4
HINFP
ING4
SIRT4
SMYD4
CHD1
INO80
SIRT1
ING2
YY1
ING1
KDM5B
TFP1
AURKC
INO80C
KDM4A
KAT5
MTA2
KMT2E
KDM5A
JARID2
MBD4
AURKA
KDM3A
DNMT1
KMT2C
CHAF1A
ING3
BRPF1
KDM7A
TAF1
BRD7
MTA3
NCOA1
SETDB2
MGMT
SETD1B
EP300
BRD2
SHMT1
KDM6B
KMT2B
MECP2
KDM1B
BAZ2A
PHF8
SMARCC2
PRDM2
SMARCD1
SETMAR
CHAF1B
NFRKB
SETD6
KDM4C
KAT6B
KDM6A
NCOA3
PRMT2
BRD4
SETD5
RING1
KAT6A
KDM2A
ASH1L
CREBBP
KMT2D
DOT1L
SETD1A
NSD1
L3MBTL3
HDAC2
SMARCA5
SETD2
CARM1
KAT2B
INO80B
JMJD4
EHMT2
CHRAC1
WDR5
EED
EZH2
PRMT3
SETD3
UCHL5
ING5
HDAC3
NAT10
SMYD5
MBD1
SIRT6
SMYD2
PRMT6
L3MBTL2
AURKB
HDAC1
PRMT5
RUVBL1
ACTL6A
JMJD6
PRMT1
SIRT2
JMJD1C
GTF3C4
PRMT9
PHF2
HDAC6
HDAC5
HSPBAP1
SETDB1
SIRT5
BRD3
DNMT3A
PRMT7
RUVBL2
DIP2AN
ACTR8
MCRS1
SUV39H1
KDM3B
KDM2B
KAT2A
CDYL
TET2
TET1
BRD9
SIRT7
SMARCD2
MBD3
CHD4
KDM1A
MBD2
CHD3
ARID1A
SMARCC1
SMARCA4
DNMT3B
Stage
Zygote
2C
4C
8C
Morula
E5
E6
E7
Blastocyst
Z score
2
1
0
−1
−2

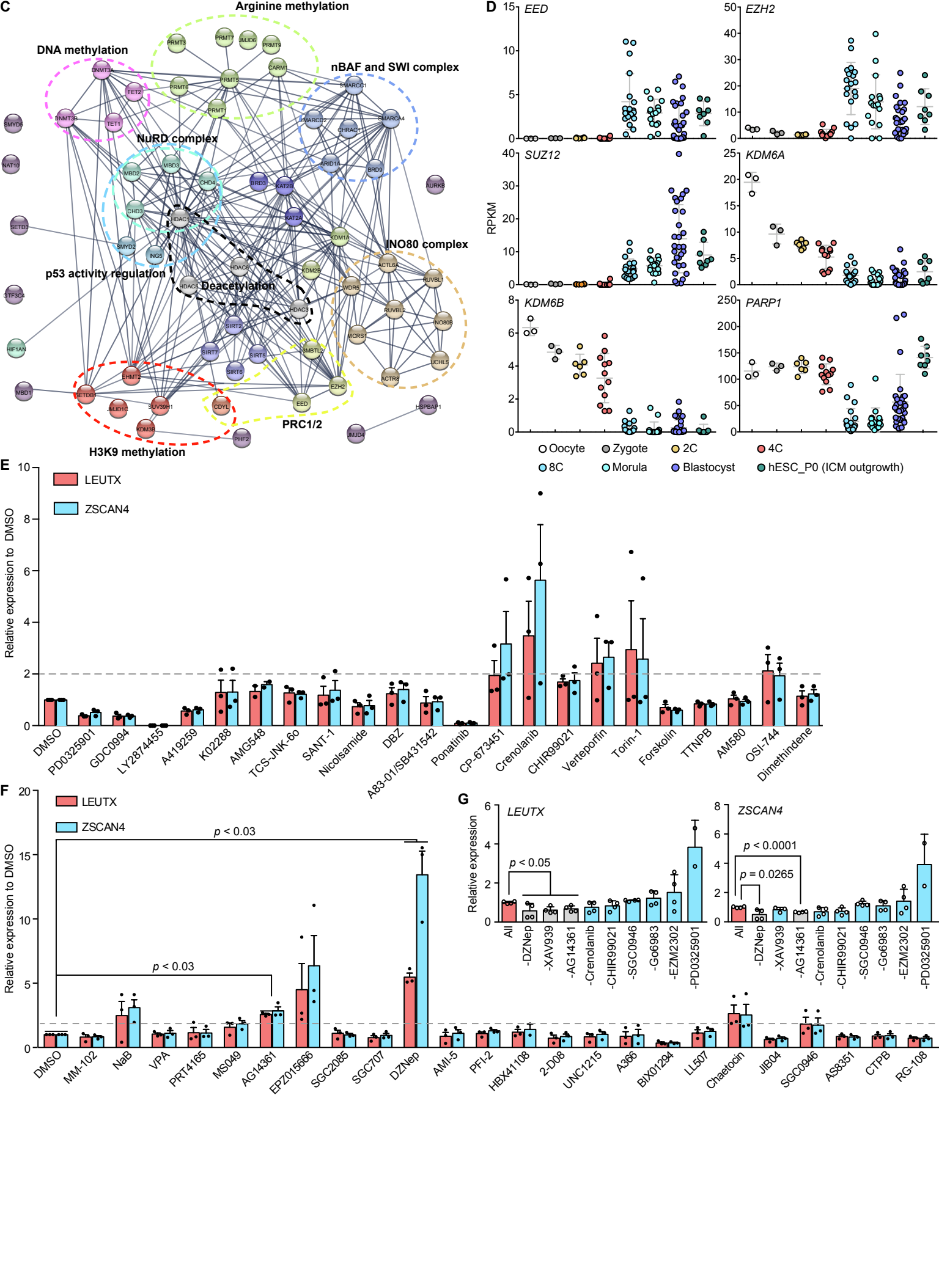

C
Arginine methylation
DNA methylation
nBAF and SWI complex
NuRD complex
INO80 complex
p53 activity regulation
Deacetylation
PRC1/2
H3K9 methylation
D
EED
EZH2
SUZ12
KDM6A
KDM6B
PARP1
RPKM
Oocyte
Zygote
2C
4C
8C
Morula
Blastocyst
hESC_P0 (ICM outgrowth)
E
LEUTX
ZSCAN4
Relative expression to DMSO
DMSO
PD0325901
GDC0994
LY2874455
A419259
K02288
AMG548
TCS-JNK-6o
SANT-1
Nicolsamide
DBZ
A83-01/SB431542
Ponatinib
CP-673451
Crenolanib
CHIR99021
Verteporfin
Torin-1
Forskolin
TTNPB
AM580
OSI-744
Dimethindene
F
LEUTX
ZSCAN4
Relative expression to DMSO
p < 0.03
p < 0.03
DMSO
MM-102
NaB
VPA
PRT4165
MS049
AG14361
EPZ015666
SGC2085
SGC707
DZNep
AMI-5
PFI-2
HBX41108
2-D08
UNC1215
A366
BIX01294
LL507
Chaetocin
JIB04
SGC0946
AS8351
CTPB
RG-108
G
LEUTX
ZSCAN4
Relative expression
p < 0.05
p < 0.0001
p = 0.0265
All
-DZNep
-XAV939
-AG14361
-Crenolanib
-CHIR99021
-SGC0946
-Go6983
-EZM2302
-PD0325901

Figure S5. Optimization of culture conditions for maintenance of 8CLCs, related to Figure 4

(A) Heatmap showing the relative expression of genes associated with multiple signaling pathways in early human embryos (Petropoulos et al., 2016; Yan et al., 2013).

(B) Heatmap showing the relative expression of multiple epigenetic regulators in early human embryos (Petropoulos et al., 2016; Yan et al., 2013).

(C) The interaction of downregulated epigenetic regulators in human 8C embryos compared to blastocysts.

(D) The expression of PRC2-associated genes, demethylases of H3K27, and PARP1 in early human embryos (Yan et al., 2013).

(E) and (F) RT-qPCR analysis of expression of *LEUTX* and *ZSCAN4* in prEpiSCs under chemical treatment indicated. Data are shown as means ± S.D. (n = 3 independent experiments).

(G) RT-qPCR analysis of expression of *LEUTX* and *ZSCAN4* in prEpiSCs under withdrawal of indicated chemical. Data are shown as means ± S.D. (n = 2 - 4 independent experiments).