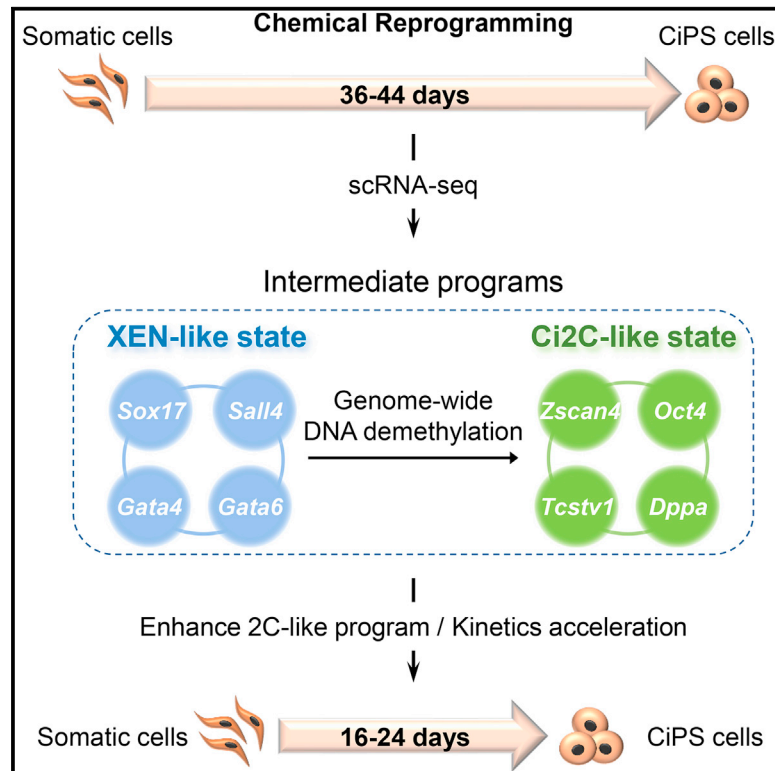# Single-Cell RNA-Seq Reveals Dynamic Early Embryonic-like Programs during Chemical Reprogramming

Graphical Abstract

Authors

Ting Zhao, Yao Fu, Jialiang Zhu, ..., Chao Tang, Cheng Li, Hongkui Deng

Correspondence

cheng_li@pku.edu.cn (C.L.), hongkui_deng@pku.edu.cn (H.D.)

In Brief

Single-cell RNA sequencing analysis of chemical reprogramming depicts its trajectory and highlights dynamic intermediate cellular programs resembling early embryonic signatures. Zhao et al. apply these insights to develop a faster reprogramming system.

Highlights

- Highly efficient chemical reprogramming enables single-cell transcriptomic analysis

- scRNA-seq reconstructs reprogramming trajectories and major molecular events

- An early embryonic-like gene network and epigenetic status drive CiPSC generation

- Chemical reprogramming is accelerated by enhancing the 2C-like program

CellPress

Single-Cell RNA-Seq Reveals Dynamic Early Embryonic-like Programs during Chemical Reprogramming

Ting Zhao,[1,2,10] Yao Fu,[1,10] Jialiang Zhu,[1,10] Yifang Liu,[4,10] Qian Zhang,[3,10] Zexuan Yi,[1,5,10] Shi Chen,[6] Zhonggang Jiao,[1] Xiaochan Xu,[7] Junquan Xu,[8] Shuguang Duo,[9] Yun Bai,[6] Chao Tang,[7] Cheng Li,[3,*] and Hongkui Deng[1,2,11,*]

[1]Department of Cell Biology, School of Basic Medical Sciences, Peking University Stem Cell Research Center, State Key Laboratory of Natural and Biomimetic Drugs, Peking University Health Science Center and the MOE Key Laboratory of Cell Proliferation and Differentiation, College of Life Sciences, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100191, China
[2]Shenzhen Stem Cell Engineering Laboratory, Key Laboratory of Chemical Genomics, Peking University Shenzhen Graduate School, Shenzhen 518055, China
[3]Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies and School of Life Sciences, Center for Statistical Science and Center for Bioinformatics, Peking University, Beijing 100871, China
[4]Joint Graduate Program of Peking-Tsinghua-NIBS, School of Life Sciences, Tsinghua University, Beijing 100084, China
[5]Joint Graduate Program of Peking-Tsinghua-NIBS, School of Life Sciences, Peking University, Beijing 100871, China
[6]Department of Cell Biology, School of Basic Medical Sciences, Peking University Stem Cell Research Center, State Key Laboratory of Natural and Biomimetic Drugs, Peking University Health Science Center, Beijing 100083, China
[7]Center for Quantitative Biology and Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China
[8]CapitalBio Technology Corporation, Beijing 102206, China
[9]Institute of Zoology, Chinese Academy Sciences, Beijing 100101, China
[10]These authors contributed equally
[11]Lead Contact
*Correspondence: cheng_li@pku.edu.cn (C.L.), hongkui_deng@pku.edu.cn (H.D.)
https://doi.org/10.1016/j.stem.2018.05.025

SUMMARY

Chemical reprogramming provides a powerful platform for exploring the molecular dynamics that lead to pluripotency. Although previous studies have uncovered an intermediate extraembryonic endoderm (XEN)-like state during this process, the molecular underpinnings of pluripotency acquisition remain largely undefined. Here, we profile 36,199 single-cell transcriptomes at multiple time points throughout a highly efficient chemical reprogramming system using RNA-sequencing and reconstruct their progression trajectories. Through identifying sequential molecular events, we reveal that the dynamic early embryonic-like programs are key aspects of successful reprogramming from XEN-like state to pluripotency, including the concomitant transcriptomic signatures of two-cell (2C) embryonic-like and early pluripotency programs and the epigenetic signature of notable genome-wide DNA demethylation. Moreover, via enhancing the 2C-like program by fine-tuning chemical treatment, the reprogramming process is remarkably accelerated. Collectively, our findings offer a high-resolution dissection of cell fate dynamics during chemical reprogramming and shed light on mechanistic insights into the nature of induced pluripotency.

INTRODUCTION

Induced pluripotency, which represents a remarkable reversion of developmental programs to embryonic transcriptome states, can be achieved by either somatic cell nuclear transfer (SCNT) (Gurdon, 1962) or transgene induction of defined transcription factors, originally *Oct4*, *Sox2*, *Klf4*, and *c-Myc* (OSKM) (Takahashi and Yamanaka, 2006). In 2013, our group reported a fundamental approach to induce mature cells into pluripotent stem cells (PSCs) using only small molecules, which is also known as chemical reprogramming (Hou et al., 2013). As a non-integrating and oncogene-free strategy, chemical reprogramming not only opens promising avenues for cell therapy and disease modeling but also provides an excellent platform for interrogating mechanisms of cell fate decisions and early developmental programs (Takahashi and Yamanaka, 2016; Xu et al., 2015a). Despite extensive efforts toward elucidating OSKM-induced reprogramming, the molecular principles underlying chemical reprogramming need to be further deciphered.

We previously demonstrated that, in contrast to OSKM-induced reprogramming, the first cornerstone of chemical reprogramming is the formation of plastic XEN-like cells marked by *Gata4*, *Gata6*, and *Sox17*, which resemble the extraembryonic endoderm (XEN) cells of the blastocyst at the preimplantation stage in terms of transcriptome, developmental ability, and reprogramming potential (Li et al., 2017; Ye et al., 2016; Zhao et al., 2015, 2016). Recently, our findings were also confirmed by other groups (Cao et al., 2018; Ping et al., 2018) using a similar small molecule cocktail that targets the same pathways identified by our group. While regarded as the most important part

of reprogramming, little was known about the molecular dynamics that led to the gradual establishment of complete pluripotency network. Due to the high heterogeneity of the cell population being reprogrammed, it was difficult to capture the small fraction of cells on the "right" route to chemically induced pluripotent stem cells (CiPSCs). Moreover, traditional bulk RNA profiling provided the average transcriptional information of the whole population, which prevented us from dissecting the minor populations that were progressing toward pluripotency.

Dissection of sophisticated biological processes at the single-cell resolution facilitates the capture of transcriptional snapshots of infrequent occurrences. Recently, the droplet-based single cell RNA sequencing (scRNA-seq) platform, which has unsurpassed advantages, was used as a powerful tool to analyze the genome-wide gene expressions of thousands of single cells in a single experiment (Zheng et al., 2017). We reasoned that high-throughput scRNA-seq analysis of a more efficient process would greatly help us to deconstruct the cellular heterogeneity during chemical reprogramming and infer the reprogramming trajectory across stages.

In this study, we employed high-throughput scRNA-seq to systematically analyze the molecular and cellular dynamics of a newly established chemical reprogramming system that generated CiPSCs with high efficiency. We characterized a series of transitional transcriptional waves throughout the reprogramming process and identified a unique early embryonic-like gene network and epigenetic status as the key molecular events that led to successful reprogramming. Finally, using the 2C-like program as an indicator of late transition, the chemical reprogramming process was notably accelerated.

RESULTS

Establishment of a Highly Efficient System for Chemical Reprogramming from XEN-like State to Pluripotency

To establish a more efficient system and a more homogeneous reprogrammed population that facilitate single-cell analysis, we attempted to adapt several conditions from embryonic stem cell (ESC) culture system for chemical induction. The previous XEN-like cell to CiPSC reprogramming system, which relied on serum and serum replacement in stage II (hereinafter referred to as the FBS/KSR-SII condition), generated up to ~500 CiPS colonies at the end of stage III from 100,000 re-plated cells from stage I (Figures 1A and S1A) (Zhao et al., 2015). We found that when re-plated cells were reprogrammed at low density on feeder cells with N2B27-based medium supplemented with LIF, vitamin C, and the same small molecules applied in FBS/KSR-SII condition (hereinafter referred to as the N2B27-SII condition, see Method Details) (Figures 1A and S1A), over 500 CiPS colonies were generated at the end of stage III from 1,000 re-plated cells (100-fold greater efficiency than that of FBS/KSR-SII condition) (Figures 1B and 1C).

Due to the presence of non-XEN-like cells in re-plated cells, we set to determine the accurate reprogramming potential of each XEN-like cell by single cell reprogramming with a XEN-like cell line. Single-cell sorting and reprogramming of this cell line in 96-well plates under FBS/KSR-SII condition yielded colonies that were unable to generate CiPSCs (Figure 1D), consistent with its low efficiency in inducing bulk reprogram-

ming (<1%) (Figure 1B). In contrast, up to 90% of the colonies induced from single XEN-like cells under N2B27-SII condition were capable of generating CiPSCs (Figures 1D and S1B). Characterization showed that the CiPSCs generated with N2B27-SII condition were fully reprogrammed (Figures S1C, S1D, and S7F). Collectively, these results demonstrated that the adaptations made based on ESC culture conditions substantially facilitated the access to pluripotency from XEN-like state, by which nearly all XEN-like cells gained the potential to be reprogrammed into CiPSCs, offering a tractable platform for further dissection.

High-Resolution Dissection of Chemical Reprogramming from Somatic Cells to Pluripotent Stem Cells using scRNA-Seq

Next, we analyzed the reprogramming trajectories by measuring single-cell gene expression profiles during chemical reprogramming with Chromium system (10x Genomics), a high-throughput and parallel droplet-based scRNA-seq platform (Zheng et al., 2017). Sequencing data were obtained from 36,199 individual cells at 12 time points throughout the reprogramming process (Figure 2A), with 4,378 median genes and 82,255 mean confidently mapped reads per cell (Figures S2A–S2D). We projected all single cells on t-distributed stochastic neighbor embedding (t-SNE) plot and identified 6 transcriptionally distinct clusters: MEFs, reprogrammed cells in stage I, XEN-like cells, further reprogrammed cells in stage II and III, CiPSCs, and Ci2C-like cells (discussed below) (Figures 2B, S2E, and S2F).

To further identify the cell population primed toward pluripotency at a higher resolution, we refined clustering analyses at each reprogramming stage. As reported previously (Zhao et al., 2015), XEN genes Sox17 and Sall4 were successively activated in stage I and reached maximal expression levels in XEN-like cells, in contrast with the gradual downregulation of fibroblast genes (Figures 2C and S2G). In stage II, the activation of pluripotency genes such as Oct4 and Dppa2 was concomitant with a gradual downregulation of XEN genes (Figures 2D and S2H). However, Sall4 remained at a high expression level throughout the induction from XEN-like cells to CiPSCs (Figures 2C–2E). At as early as stage III D3, CiPSCs were generated (Figures 2F and S2F). They continued forming and further growing and were present as large colonies at the end of stage III (Figure 1C), which constituted the major cell type (~77% of the whole population, Figures 2F and S2F). The majority of remaining cells at stage III remained a high expression level of XEN genes, such as Sox17 and Gata4 (Figures 2E and S2I).

Interestingly, a unique cell cluster was identified in stages II and III, accounting for ~4% of the whole population around the end of stage II (Figures 2F and S2F). This cluster was marked by the expression of genes including Zscan4 family members and Tcstv1 albeit at varied levels (Figures 2C–2E, S2H, and S2I), all of which were reported to be restricted to express in the 2C embryos at the preimplantation developmental stage (Deng et al., 2014; Falco et al., 2007; Tang et al., 2011) or in a minor cell subpopulation in ESC culture termed as 2C-like (ES-2C-like) cells (Eckersley-Maslin et al., 2016; Falco et al., 2007; Macfarlan et al., 2012; Zhang et al., 2016b). Accordingly, we designated these 2C gene-expressing cells in chemical reprogramming as chemically induced two-cell embryonic-like cells
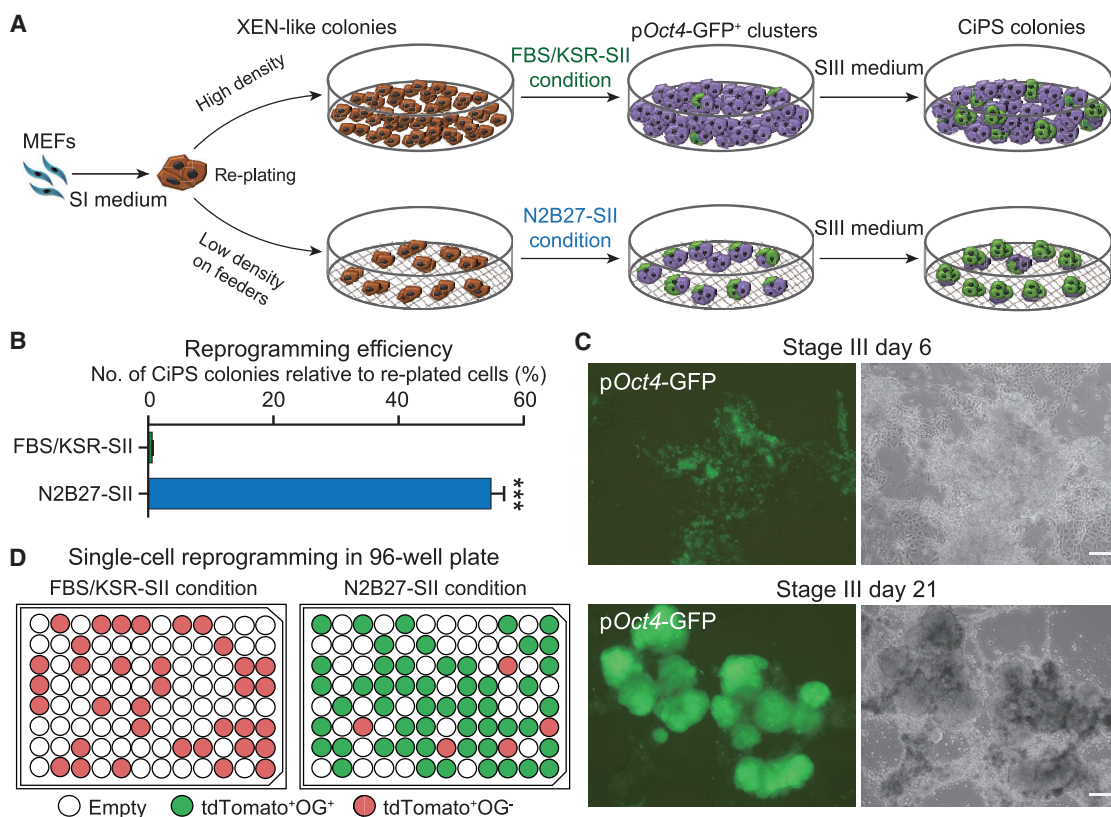
Figure 1. Establishment of a Highly Efficient System for Chemical Reprogramming from XEN-like State to Pluripotency

(A) Schematic diagram of reprogramming system with FBS/KSR-SII condition (Zhao et al., 2015) or N2B27-SII condition (inducing XEN-like cells at low density on feeders in N2B27-based medium) (this study). Green indicates pOct4-GFP+ (OG+) cells.

(B) Reprogramming efficiency induced with FBS/KSR-SII and N2B27-SII condition (n = 3).

(C) Typical CiPS colonies induced under N2B27-SII condition at stage III D6 and D21. Scale bar, 200 μm.

(D) Comparison of single-cell reprogramming of individual tdTomato-labeled XEN-like cells with OG reporter under FBS/KSR-SII and N2B27-SII condition (2 replicates of 96-well plate for each condition). CiPSC conversion efficiency was 91.1% ± 0.2% for N2B27-SII condition. Red, wells with colonies but no CiPSCs. Green, wells with CiPSCs. White, wells without colonies.

The data are presented as the mean ± SD. Significance was assessed compared with the controls using a one-tailed Student's t test. ***p < 0.001.

See also Figure S1.

(Ci2C-like cells) and referred to their top 100 upregulated genes as the 2C-like program (Figures 2D, 2E, and 6B; Table S1).

Reconstruction of Chemical Reprogramming Trajectory in a Pseudotime Manner

The gene expression profiling data of all the single cells allowed us to deconstruct the population heterogeneity and reconstruct the reprogramming trajectory. We ordered cells in a pseudo-temporal manner using Monocle 2, an algorithm for the lineage reconstruction of biological processes based on transcriptional similarity (Qiu et al., 2017).

We found that MEFs progressed toward the XEN-like state during stage I induction before bifurcation (pre-branch) (Figures 3A and S3A), which confirmed our previous findings that CiPSCs were exclusively generated from XEN-like cells (Zhao et al., 2015). However, shortly after XEN-like cell formation, cells bifurcated into two diverse branches, representing two major cell lineages in the late reprogramming stage (Figure 3A). The cells in one terminal expressed pluripotency genes (Figure 3B) and shared a highly similar global expression profile

with ESCs (Figure S3B), representing successfully reprogrammed CiPSCs in the successful branch. Meanwhile, the cells in another terminal maintained high expression levels of XEN genes, representing unsuccessfully reprogrammed cells with a remaining XEN-like program in the failed branch (Figures 3B, 3C, 3D, and S3D). In addition to gene expression, we analyzed the gene regulatory relationships in terminal cells using SCENIC, an algorithm for identifying cell state based on the reconstructed gene regulatory network (GRN) activities (Aibar et al., 2017). Two regulon (a group of co-expressed genes regulated by a transcription factor) clusters were clearly identified to represent successful (pluripotency regulons) and failed (XEN regulons) terminal cells, respectively (Figure S3C), confirming the results of gene expression analysis.

To gain insights into the gene expression dynamics along the trajectory, we analyzed the expression changes of 1,518 top differentially expressed genes (DEGs) and observed five major categories of transcriptional gene clusters in characterized patterns (Figures 3C–3E, 4A, and S3D; Table S2). We next analyzed the successful trajectory (blue lines). Genes in cluster I (fibroblast
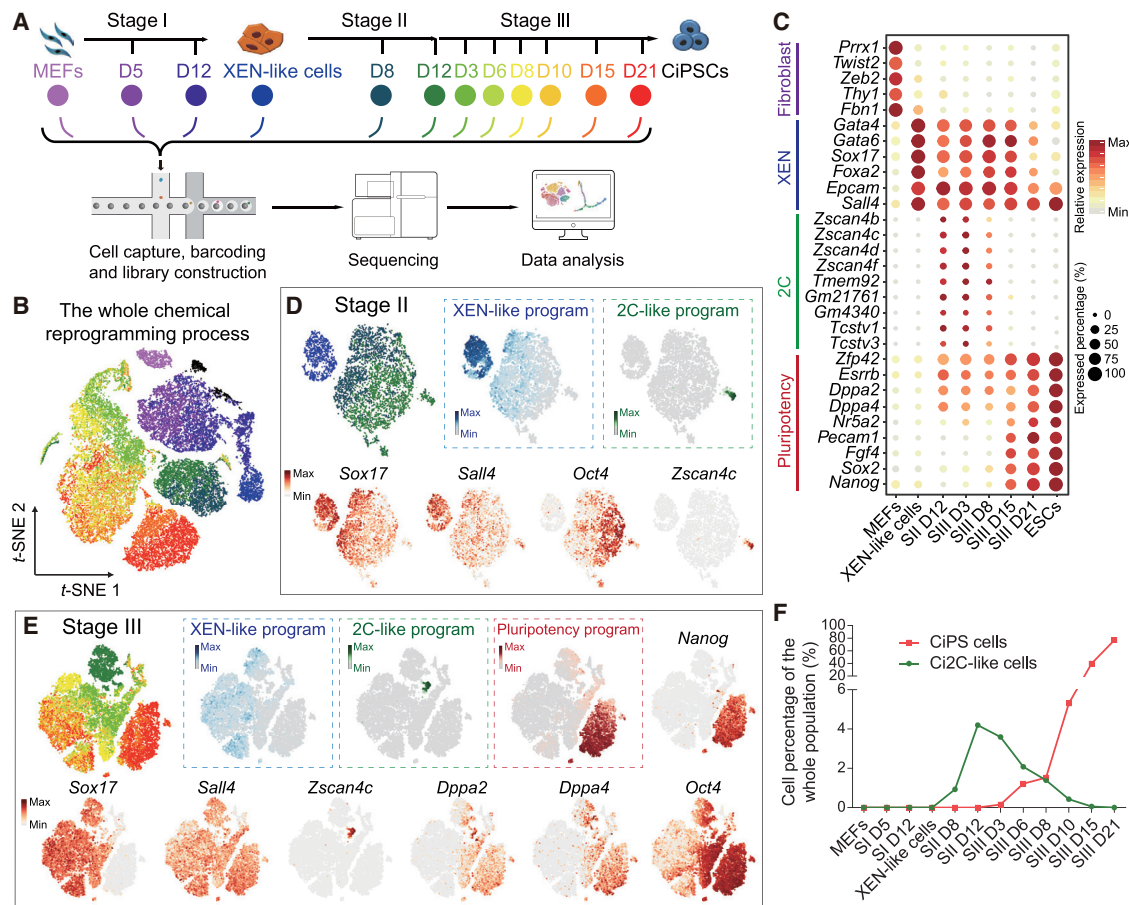
Figure 2. High-Resolution Dissection of Chemical Reprogramming from Somatic Cells to Pluripotent Stem Cells using scRNA-Seq
(A) Schematic diagram of scRNA-seq analysis strategy during chemical reprogramming. Overall, 963, 3,824, 4,739, 1,068, 2,367, 2,763, 3,729, 2,879, 2,366, 3,530, 3,377, and 4,594 cells were analyzed for MEFs, SI D5, SI D12, XEN-like cells, SII D8, SII D12, SIII D3, SIII D6, SIII D8, SIII D10, SIII D15, and SIII D21 (indicated by different colors), respectively. Cells were re-plated on feeders at stage I D12. The flow chart of the scRNA-seq analysis was adopted from 10x Genomics.
(B) t-SNE projection of all 36,199 individual cells during the whole reprogramming process, colored by indicated time points. Feeder cells are presented as black dots.
(C) Expression of genes in 4 different categories at the indicated time points.
(D) t-SNE projection of cells and typical gene expressions in stage II.
(E) t-SNE projection of cells and typical gene expressions in stage III.
The XEN-like and pluripotency program is the average expression of top 100 upregulated genes in XEN-like or CiPS cells relative to all cells, respectively.
(F) Percentages of CiPS cells and Ci2C-like cells at different time points.
See also Figure S2.

genes) were gradually downregulated from the beginning of reprogramming and were largely involved in the regulation of biological processes such as "collagen fibril organization" (e.g., *Thy1*, *Twist1*, and *Col6a1*) (Figures 3C–3E and S3D). Subsequently, cluster II and III genes were transiently upregulated but finally downregulated, representing two temporary transcriptional waves (i.e., XEN genes and 2C genes) (Figures 3C–3E and S3D). Concurrent with cluster III activation, cluster IV genes (early pluripotency genes) were upregulated and maintained at high expression levels until the final stage (e.g., *Dppa* family, *Klf2*, and *Oct4*) (Figures 3C–3E and S3D). Finally, cluster V genes (late pluripotency genes) were activated at the end of reprogramming with predominant involvement in the GO term "stem cell population maintenance" (e.g., *Sox2* and *Nanog*) (Figures 3C–3E and S3D).

In addition to the scRNA-seq approach, we further confirmed the sequential activation of major molecular events by bulk RNA profiling (Figure S3E). Therefore, these data depicted the trajectory of chemical reprogramming and revealed the ordered activation of transcriptional waves throughout this process.

2C-like Program Activation, Pluripotency Program Upregulation, and XEN-like Program Downregulation Represent Late Molecular Series in Successful Reprogramming

Because the reprogramming trajectory bifurcated into two branches at the late stages, we tried to elucidate the molecular dynamics that distinguished two branches. Four major gene clusters were identified accounting for the distinctions, including
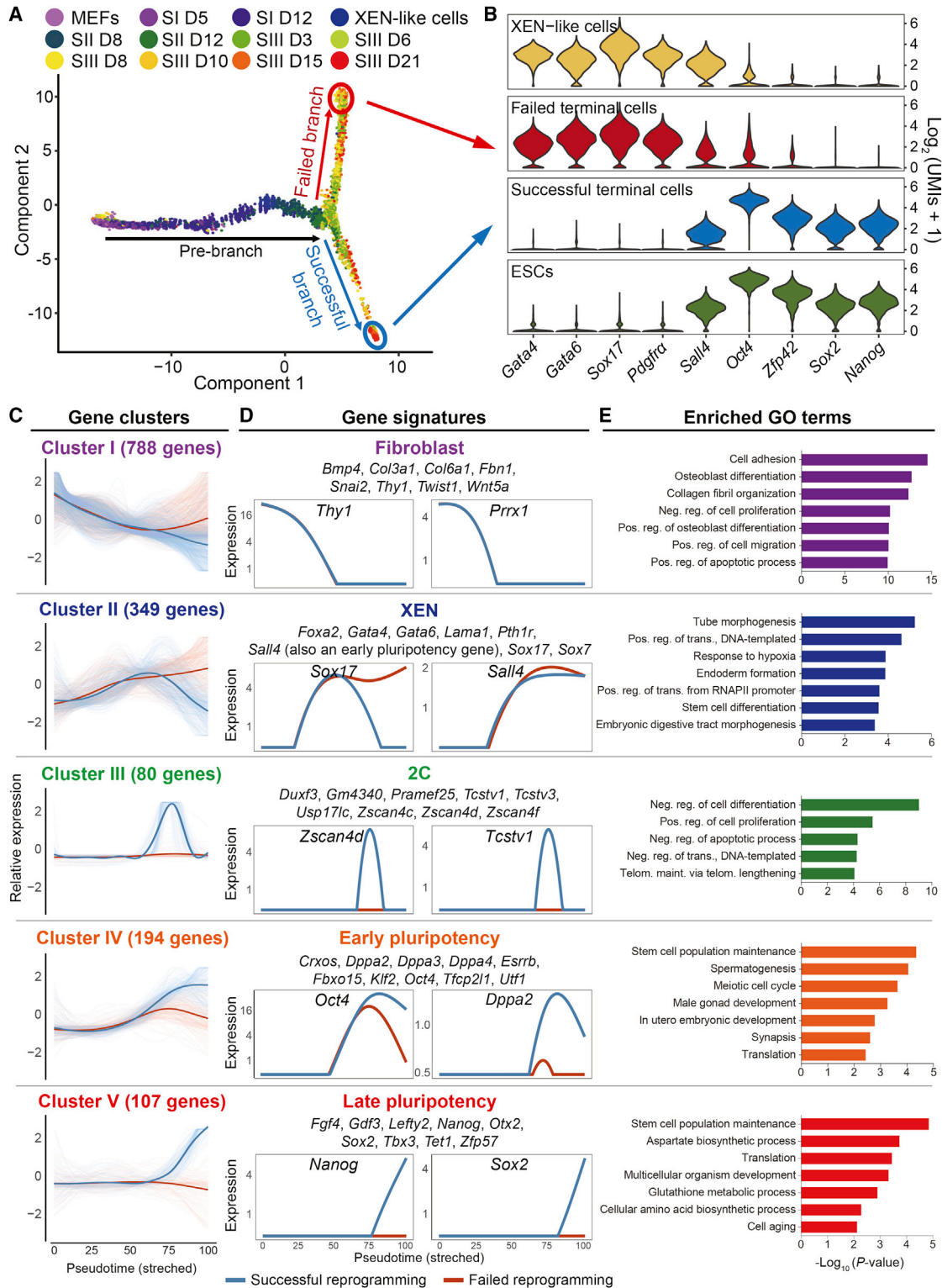
Figure 3. Reconstruction of Chemical Reprogramming Trajectory in a Pseudotime Manner

(A) Trajectory reconstruction of all single cells throughout chemical reprogramming reveals three branches: pre-branch (before bifurcation), successful branch, and failed branch (after bifurcation). Blue and red circles indicate cells at the terminus of the successful and failed branches, respectively.

(B) Violin plot displaying the expression of representative XEN and pluripotency genes in indicated cell types.

(C) The expression dynamics of 1,518 top DEGs were cataloged into 5 major clusters in a pseudotime manner shown as blue lines (successful reprogramming) and red lines (failed reprogramming). Thick lines indicate the average gene expression patterns in each cluster.

(legend continued on next page)

the inability to downregulate cluster II genes and the failure to up-regulate cluster III/IV/V genes in the failed reprogramming trajectory (red line) (Figures 3C–3E). Coincident with this result, XEN, 2C, and pluripotency genes were identified as top DEGs between the two branches after bifurcation at late stage (Figures 3B and S4A; Table S3).

To investigate the sequential changes of all DEGs, we analyzed their expression patterns by dividing cells after bifurcation into two classes: earlier stage (class A) and later stage (class B) (Figure 4A). In comparison with the failed ones, successful branch cells in class A expressed higher levels of genes enriched for the GO terms "positive regulation of cell proliferation," "negative regulation of differentiation," and "stem cell maintenance," including almost all 2C genes and some early pluripotency genes, such as the *Dppa* family (Figures 4A–4C). Downregulated genes were strongly enriched for mesendoderm development regulators such as XEN genes (Figures 4A, 4B, and 4D). Together, these unique gene expression patterns characterized the earliest transcriptomic events indicating CiPSC generation. Subsequently, the cells in class B of successful branch acquired pluripotency by further silencing XEN gene expressions and activating the late pluripotency genes (Figures 4A and S4B–S4D).

Next, we investigated the different gene regulatory relationships in late reprogramming stage. Consistent with the gene expression analysis (Figure 2E), *Nanog* and *Sox17* regulons were predominantly presented in CiPSCs and failed reprogrammed cells, respectively, distinguishing the pronounced heterogeneity in stage III (Figure 4E). Apart from CiPSCs, the *Oct4* regulon was also present in the intermediate cells expressing 2C genes and early pluripotency genes *Dppa* family (Figures 2E, 4E, and S2I), suggesting that *Oct4* may act in concert with these genes to facilitate pluripotency acquisition. In addition, we projected cells in stage III on *t*-SNE plot based on different regulon activities (Figure S4E) and obtained results similar to those based on gene expression profiling (Figure 4E), further supporting our findings.

In addition to bioinformatics analyses (successful versus failed branch), we intended to experimentally identify the intermediate transcriptional signatures of reprogramming. The N2B27-SII condition, which yields 100-fold greater CiPSC generation efficiency than FBS/KSR-SII condition (Figure 1B), significantly stimulated the expression of genes that highly overlapped the upregulated genes in class A of successful branches at the end of stage II (Figures 4F and S4F; Table S4). These included 2C genes (e.g., *Zscan4d*, *Tcstv1*) and early pluripotency genes (e.g., *Oct4*, *Dppa* family), confirming their important roles in successful trajectory. We further compared two groups of cells with great differential reprogramming efficiency: OG⁺PDGFRA^low/⁻PECAM1⁻ versus unsorted cells (Figures S5I–S5K; Table S5) and optimal VPA treatment versus control (Figures 7C–7H; Table S6), obtaining the similar panel of upregulated genes. Moreover, we compared the commonly upregulated (>4-fold) genes in four groups of cells and surprisingly

found that all of them (n = 15) belonged to the 2C-like program with no exception (Figure 4G; Table S1), further suggesting a strong positive correlation between reprogramming potential and 2C gene expression.

Finally, we constructed the gene network based on pairwise correlation of the top DEG expressions in progressive cell fate transitions from XEN-like state to pluripotency (Figure 4H). Three subnetworks were revealed chronologically, in which the XEN subnetwork connected with 2C genes and early pluripotency genes connected with the pluripotency subnetwork. The sequential switch of transcriptional circuits highlighted the intermediate subnetwork as the bridge linking XEN-like state to pluripotency, which consisted of several 2C genes (*Zscan4c*, *Zscan4f*, *Tcstv1*, *Tcstv3*, *Lmx1a*, and *Sp110*) and early pluripotency genes *(Dppa2*, *Dppa3*, *Dppa4*, *Klf2*, and *Gm13154)* (Figure 4H). Due to the similar expression patterns observed between *Oct4* regulon and these intermediate genes (Figures 2E, 4E, and S2I), we examined the co-expression of representative genes *Zscan4c*, *Dppa2* with *Oct4*. At the end of stage II, 3% of cells were identified to be triple positive for these three genes, 68% of which were Ci2C-like cells (Figure 4I), suggesting the predominant enrichment of the co-expressions in Ci2C-like cells. On the other hand, 52% of Ci2C-like cells were found to co-express *Zscan4c*, *Dppa2*, and *Oct4* (Figures S4G and S4H).

Together, these findings depicted the sequential molecular dynamics that lead to successful reprogramming and identified the key intermediate subnetwork that bridges XEN-like state to pluripotency. Importantly, all of the intermediate subnetwork genes including 2C genes (express at the 2C stage) and early pluripotency genes (express from the 2C stage to blastocysts) exhibited expression patterns at late chemical reprogramming stages that were similar to early embryogenesis from the 2C stage to blastocyst (Figure 4J), implying the critical role of remarkable early embryonic transcriptional signature in acquiring chemically induced pluripotency.

Transcriptomic and Epigenetic Signatures of Early Embryos Mediate Late Reprogramming Process

To verify the function of intermediate subnetwork as well as its relationship with *Oct4* in late reprogramming stage, we performed knockdown experiments of the representative genes during reprogramming from XEN-like cells to CiPSCs (Figure S5A). CiPSC generation was significantly impaired by the knockdowns of either 2C gene *Zscan4* or early pluripotency genes *Oct4*, *Dppa2*, and *Dppa4* (Figure 5A). Moreover, interference with *Oct4*, *Dppa2*, or *Dppa4* severely disrupted the expressions of early pluripotency and 2C genes and consequently resulted in failure activation of the late pluripotency gene *Nanog* (Figure 5B). Although the disturbance of *Zscan4* or *Tcstv1* by knockdown greatly decreased CiPSC generation (Figure 5A), they led to a relatively less reduction of early pluripotency gene expressions at the end of stage II (Figure S5B). This was possibly because of the relatively small proportion of *Zscan4*⁺ cells within the whole population and the expression of these early

(D) Gene signatures and expression dynamics of representative genes in each gene cluster.
(E) Gene ontology analyses of each gene cluster. Pos. reg., positive regulation; Neg. reg., negative regulation; Trans., transcription; Telom. maint., telomere maintenance.
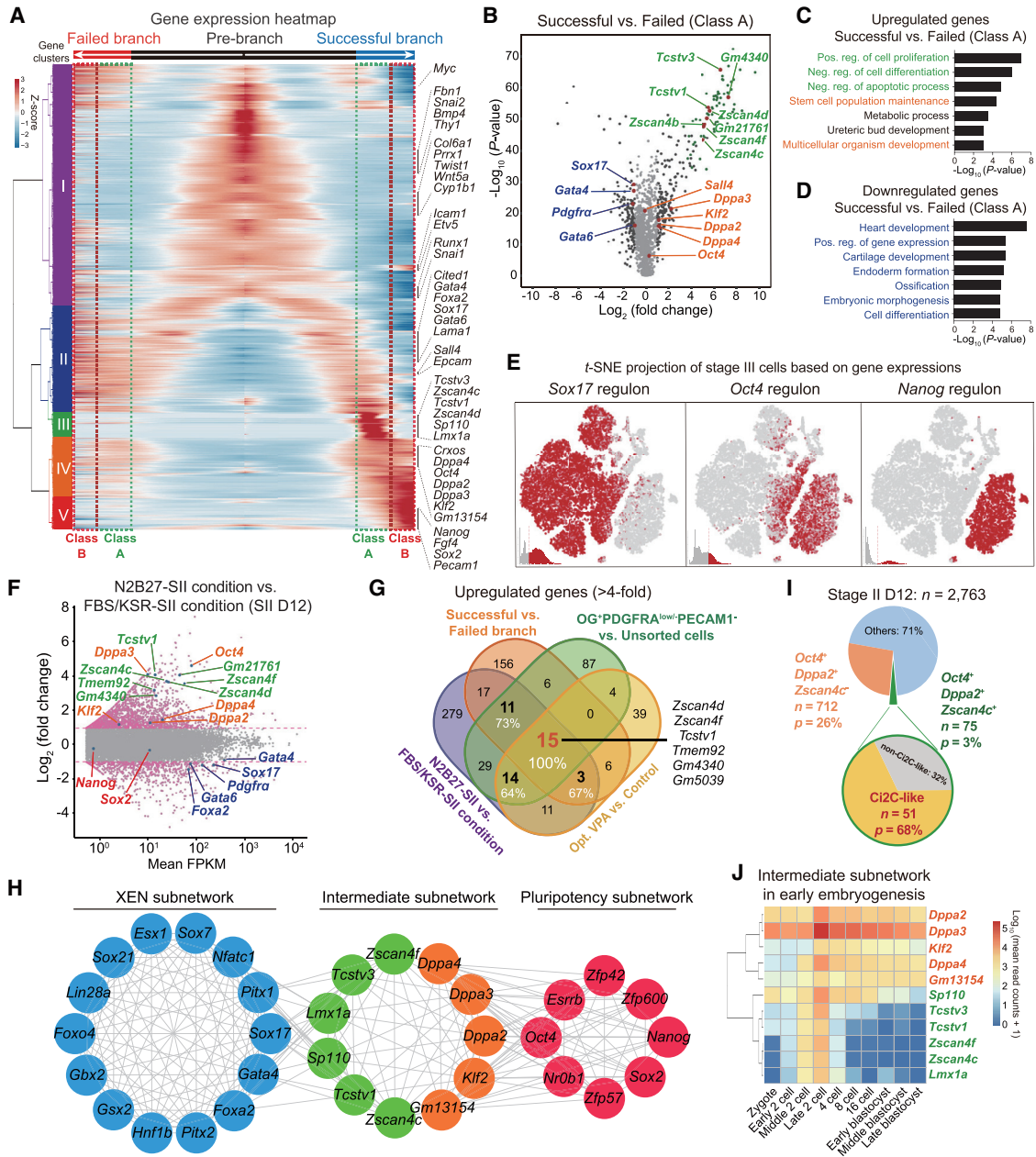See also Figure S3.

Figure 4. 2C-like Program Activation, Pluripotency Program Upregulation, and XEN-like Program Downregulation Represent Late Molecular Series in Successful Reprogramming

(A) Gene expression heatmap of 1,518 top DEGs (cataloged in five clusters) in a pseudo-temporal order. Successful and failed reprogramming trajectories (including pre-branch) are shown on the right and left, respectively. Class A and B indicate cell populations at earlier stage and later stage after bifurcation, respectively. Representative genes are shown on the right.

(B) Volcano plot displaying the DEGs in class A cells between the successful and failed branch. Representative 2C (green), early pluripotency (orange), and XEN (blue) genes are indicated. 2C-like program genes are highlighted as green dots. Gray dots represent non-DEGs (<2-fold change).

(C) GO analysis of upregulated genes in class A cells comparing the successful with failed branches.

(D) GO analysis of downregulated genes in class A cells comparing successful with failed branches. 2C, early pluripotency and XEN genes are included in green, orange, and blue GO terms, respectively.

(E) t-SNE plot displaying Sox17, Oct4, and Nanog regulons (red dots, active; gray dots, inactive) as well as their intensity distributions in stage III.

(F) MA plot displaying DEGs comparing the N2B27-SII with FBS/KSR-SII conditions at the end of stage II. Pink dots represent genes that are differentially expressed by >2-fold.

(G) Venn diagram showing the overlap of upregulated (>4-fold) gene numbers and corresponding percentages of gene number in 2C-like program (in at least 3 overlapping groups) in 4 groups with differential reprogramming potential. All (15 genes) of the commonly upregulated genes in the 4 groups are included in the 2C-like program, some of which are highlighted.

pluripotency gene in *Zscan4⁻* cells (Figures 2E, 4I, S2H, and S2I). To further examine whether 2C genes were capable of regulating early pluripotency genes, we performed genetic overexpression using a doxycycline (dox)-inducible system in a fast but inefficient reprogramming system (Figure S5A). Transient ectopic expression of *Zscan4c* or *Zscan4d* was sufficient to enhance the expression of *Oct4*, *Dppa* family and consequently boosted reprogramming efficiency (Figures 5C and 5D). Similarly, *DPPA2*, *DPPA4*, or *OCT4* overexpression stimulated the expressions of both 2C genes and early pluripotency genes and enhanced iPS generation (Figures 5C and S5C). Collectively, these data verified the mutual regulatory relationships of *Zscan4*, *Dppa2*, *Dppa4*, and *Oct4* and suggested that they formed a critical intermediate gene network bridging XEN-like cells to CiPSCs.

Next, we investigated the regulatory relationships controlling cell fate transitions between major subnetworks. If the defined intermediate subnetwork served as the next transcriptomic event after XEN-like program, then *Zscan4* (part of the network) expression should be diminished by interfering with the XEN-like state. Indeed, knockdown of XEN genes dramatically blocked ZSCAN4⁺ cells and subsequent CiPSC formation (Figure S5D). As per the requirements of XEN-like program silencing and pluripotency program establishment at the end of successful reprogramming (Figure 3B), *SOX2* or *NANOG* overexpression in stage III significantly enhanced CiPSC formation, in contrast to the blockade of reprogramming by *GATA4* or *GATA6* overexpression (Figure S5E), further confirming our bioinformatics analysis. Taken together, all the results of genetic loss and gain-of-function studies revealed the chronological occurrence of major transcriptional events from XEN-like state to pluripotency (Figure 4H).

Due to the use of demethylation modulating compounds in stage II, we asked whether the DNA methylation was affected, especially in Ci2C-like cells. We therefore established a XEN-like cell line with a p*Zscan4c*-Emerald (Em) reporter, which generated ∼4% positive (p*Zscan4c*-Em⁺) cells at the end of stage II (Figures 5E and S5F), in accordance with the scRNA-seq result (Figure 2F). As expected, p*Zscan4c*-Em⁺ cells expressed higher levels of 2C genes as well as some early pluripotency genes (Figures 5F and S5G). Moreover, p*Zscan4c*-Em⁺ cells were able to generate CiPSCs (Figure 5G). Next, we analyzed p*Zscan4c*-Em⁺ and p*Zscan4c*-Em⁻ cells via whole-genome bisulfite sequencing (WGBS) at a single-base resolution. Interestingly, we found that both p*Zscan4c*-Em⁺ and p*Zscan4c*-Em⁻ cells exhibited a widespread loss of DNA methylation (3.6% in p*Zscan4c*-Em⁺ cells and 4.7% in p*Zscan4c*-Em⁻ cells) compared with the >50% methylation level in XEN-like cells and CiPSCs (Figure 5H). The genome-wide DNA demethylation occurred in all of the genome regions, including promoters, CpG islands, and specific genes (Figures 5I and 5J). These results were further confirmed by an ELISA detecting the 5mC levels in the unsorted cell population (Figure S5H). Together, these findings suggested that chemical in-duction triggered notable global DNA hypomethylation of cells in intermediate-late transitions, conferring a possible open chromatin status and unique epigenetic signature resembling that of early embryogenesis and ES-2C-like cells (Baker and Pera, 2018; Eckersley-Maslin et al., 2016; Smith et al., 2012).

Finally, we tried to enrich cells with the intermediate subnetwork by cell sorting and evaluate their reprogramming potential. We sorted OG⁺PDGFRA^low/⁻PECAM1⁻ cells and found that they expressed higher levels (2.5- to 10-fold higher) of intermediate subnetwork genes than unsorted cell population, suggesting the substantial enrichment of the gene network (Figures 4G and S5I; Table S5). Importantly, single cell or bulk reprogramming showed that the OG⁺PDGFRA^low/⁻PECAM1⁻ population exhibited greater CiPSC generation potential compared with the unsorted cell population (Figures S5J and S5K). These results indicated that enrichment of the intermediate subnetwork could improve CiPSC generation potential.

Ci2C-like Cells Shared Similar, but Not Identical, Gene Expression Profiles to Those of the 2C Embryos and 2C-like Cells in ESC Culture

To better understand the Ci2C-like cells, we compared their gene expression profile with those of ES-2C-like cells and 2C embryos. First, we sequenced the transcriptomes of 7,748 ESCs at the single-cell level and identified 123 ES-2C-like cells that highly expressed 2C genes (∼1.6%) (Figure S6A), consistent with the reported frequency of Zscan4c::eGFP⁺ cells in ESCs (Eckersley-Maslin et al., 2016). Through unsupervised hierarchical clustering analysis, Ci2C-like cells were found to exhibit gene expression profiles that were remarkably more similar to ES-2C-like cells than to major ESCs, CiPSCs, MEFs, or XEN-like cells, in contrast with the non-Ci2C-like cells that clustered together with XEN-like cells (Figure 6A). Then, we assessed the transcriptional similarity of the Ci2C-like cells using reported datasets for ES-2C-like cells (Eckersley-Maslin et al., 2016) and found that ∼52% of ES-2C-like cell-specific genes were upregulated (>2-fold) in the Ci2C-like cells compared with the non-Ci2C-like cells (Figure S6B). Moreover, the expression pattern of Ci2C-like cells also exhibited a higher similarity with the reported ES-2C-like cells (Eckersley-Maslin et al., 2016; Macfarlan et al., 2012) than with other cell types (Figures 6A and S6C).

Next, we benchmarked the expression of 2C-like program against the single-cell RNA sequencing (RNA-seq) data of early embryo at several preimplantation stages (Deng et al., 2014). Surprisingly, most of the genes in chemically induced 2C-like program (>95%) culminated their expression levels at the 2C stage, suggesting striking enrichment of the 2C-like program in 2C embryos (Figure 6B). In addition, ∼33% of the 110 genes that were specifically expressed in 2C embryos (Deng et al., 2014) were expressed higher (>2-fold) in Ci2C-like cells compared with non-Ci2C-like cells, and most of the remaining genes were not differentially expressed (Figure 6C). Moreover, Ci2C-like cells clustered together with 2C embryos (Deng

---

(H) Gene correlation network during reprogramming progression from XEN-like state to pluripotency.

(I) Pie chart displaying the number (*n*) and percentage (*p*) of cells co-expressing indicated panel of genes at the end of stage II.

(J) Expression of intermediate subnetwork genes in early embryo development (Deng et al., 2014).
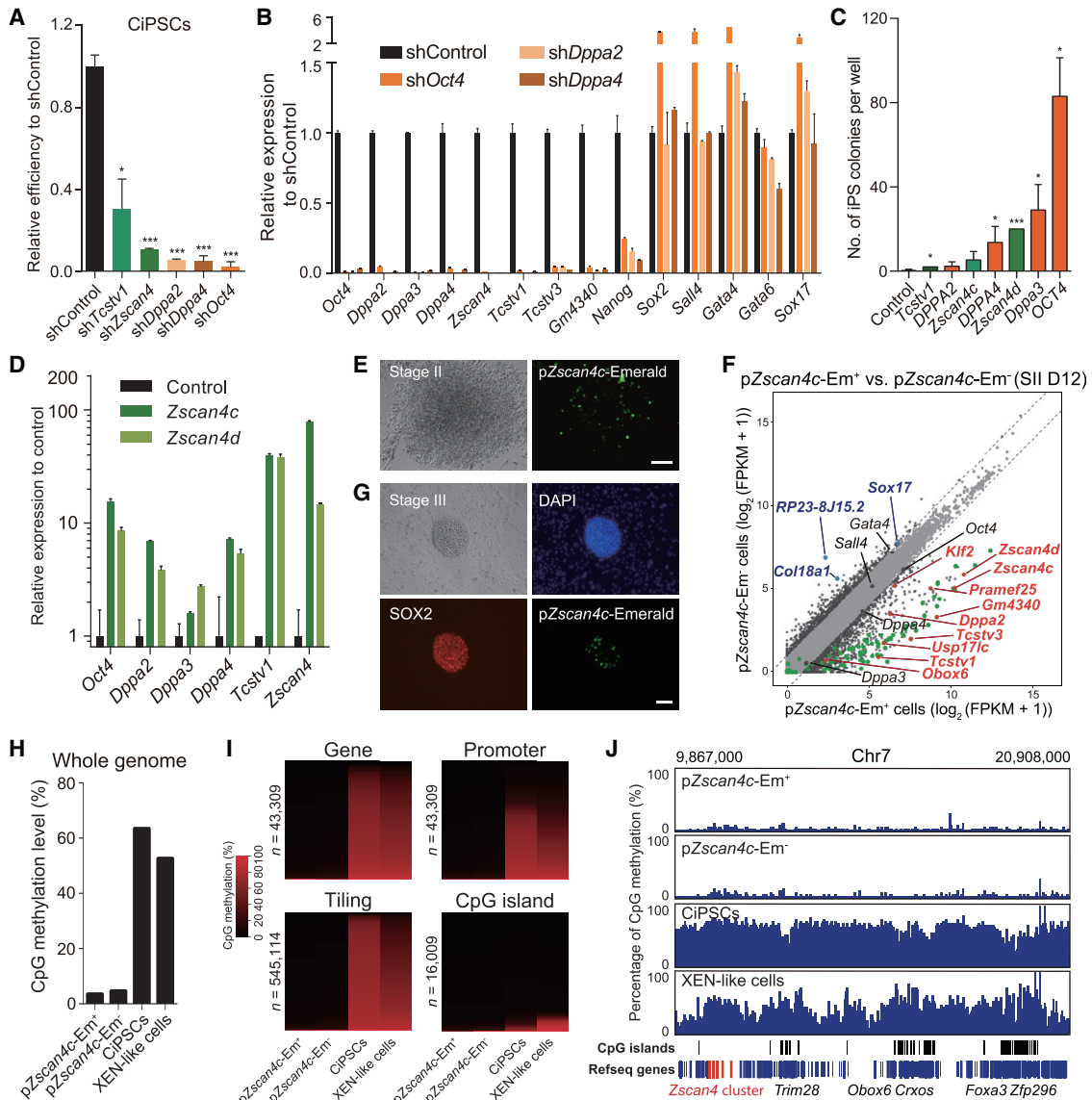
See also Figure S4.

Figure 5. Transcriptomic and Epigenetic Signatures of Early Embryos Mediate Late Reprogramming Process

(A) Relative CiPSC generation efficiency with the knockdowns of intermediate subnetwork genes (n ≥ 2). The reprogramming efficiency of shControl (non-targeting vector) is 20%–50%.

(B) The relative expression of representative marker genes detected at stage III D5 after the knockdowns with the indicated shRNA (n = 2).

(C) Number of iPS colonies generated with transient overexpression of intermediate genes (stage II plus the early 4 days of stage III) during a fast reprogramming system from XEN-like cells to CiPSCs (n ≥ 2).

(D) The relative expression of intermediate subnetwork genes after overexpression of *Zscan4c* and *Zscan4d* (n = 2).

(E) A typical p*Zscan4c*-Em$^+$ colony at the end of stage II. Scale bar, 100 μm.

(F) Scatterplots comparing the global gene expression profile of p*Zscan4c*-Em$^+$ and p*Zscan4c*-Em$^−$ cells analyzed by RNA-seq. 2C-like program is highlighted as green dots and some red dots.

(G) A typical CiPS colony (positive for SOX2 protein, with scattered p*Zscan4c*-Em$^+$ cells) generated from p*Zscan4c*-Em$^+$ cells sorted at the end of stage II. Scale bar, 100 μm.

(H) Global CpG methylation level measured by WGBS in four indicated cell types.

(I) Heatmap for CpG methylation at genomic elements in four indicated cell types. DNA methylation percentages of individual CpG sites are shown. "*n*" represent the numbers of the region sets.

(J) CpG methylation across part of chromosome 7 (~11 Mb) in four indicated cell types analyzed by WGBS. Each bar indicates a CG site, and the height of each bar represents the methylation percentage (0%–100%). Annotations below show the location of CpG islands and representative genes (*Zscan4a/b/c/d/f* are highlighted).

Experiments were performed on purified XEN-like cells unless otherwise indicated. The data are presented as the mean ± SD. Significance was assessed compared with the controls using a one-tailed Student's t test. ***p < 0.001; *p < 0.05. See also Figure S5.
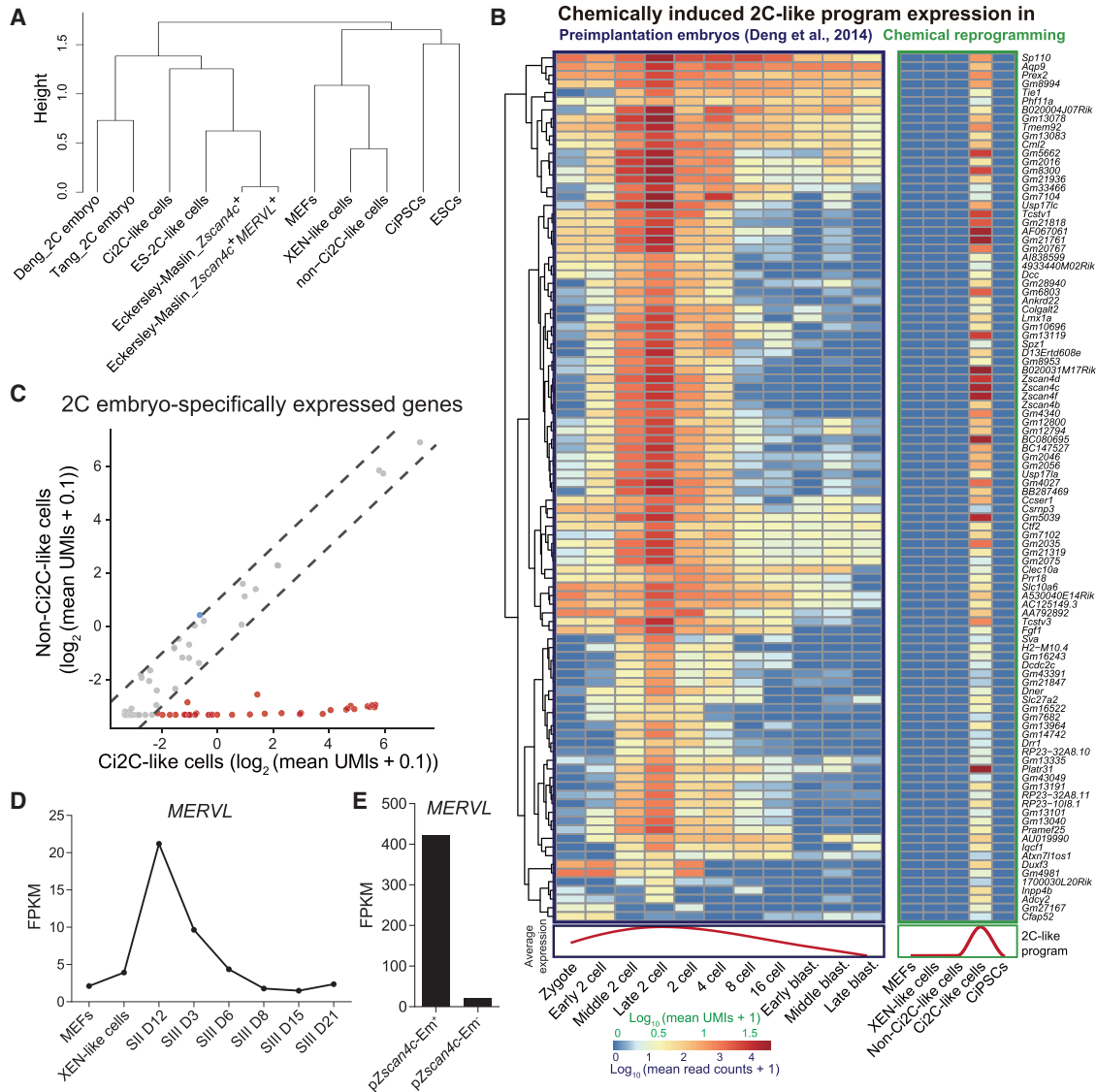
Figure 6. Ci2C-like Cells Shared Similar, but Not Identical, Gene Expression Profiles to Those of the 2C Embryos and 2C-like Cells in ESC Culture

(A) Unsupervised hierarchical clustering of the indicated cell types: cells in this study (Ci2C-like cells, non-Ci2C-like cells, CiPSCs, XEN-like cells in chemical reprogramming, ESCs, ES-2C-like cells, MEFs) and in reported datasets (2C embryo [Deng et al., 2014; Tang et al., 2011] and ES-2C-like cells [*Zscan4c*+ and *Zscan4c*+*MERVL*+ cells] [Eckersley-Maslin et al., 2016]).

(B) Heatmap displaying the expression of 2C-like program in mouse preimplantation embryos across different stages (Deng et al., 2014) and five cell types in chemical reprogramming.

(C) Scatterplot showing the expression of 110 2C embryo-specific genes (>8-fold higher expression level at the 2C stage than other developmental stages, analyzed from the dataset of (Deng et al., 2014)) in Ci2C-like cells and non-Ci2C-like cells.

(D) The dynamic expression of *MERVL* at multiple time points during chemical reprogramming analyzed by bulk RNA-seq.

(E) The expression of *MERVL* in p*Zscan4c*-Em+ and p*Zscan4c*-Em− cells at the end of stage II.

See also Figure S6.

et al., 2014; Tang et al., 2011) in an unsupervised hierarchical clustering analysis, suggesting similarity of their global expression profile (Figure 6A).

In addition, we examined another feature of 2C embryos, the expression of transposable element mouse endogenous retroviral element (*MERVL*) (Macfarlan et al., 2012) via bulk RNA-seq. *MERVL* exhibited a expression pattern similar to that of the 2C-like program, with its expression level peaking at the

end of stage II (Figures 6D and S3E). Moreover, *MERVL* expression was significantly enriched in p*Zscan4c*-Em+ cells (Figure 6E), suggesting the co-expression of *MERVL* with 2C-like program. Another unique phenomenon observed in ES-2C-like cells and 2C embryos is the presence of transcript for *Oct4* but absence for the OCT4 protein (Eckersley-Maslin et al., 2016; Macfarlan et al., 2012). Similarly, we found that ZSCAN4 was co-stained with p*Oct4*-GFP at the end of stage II in chemical
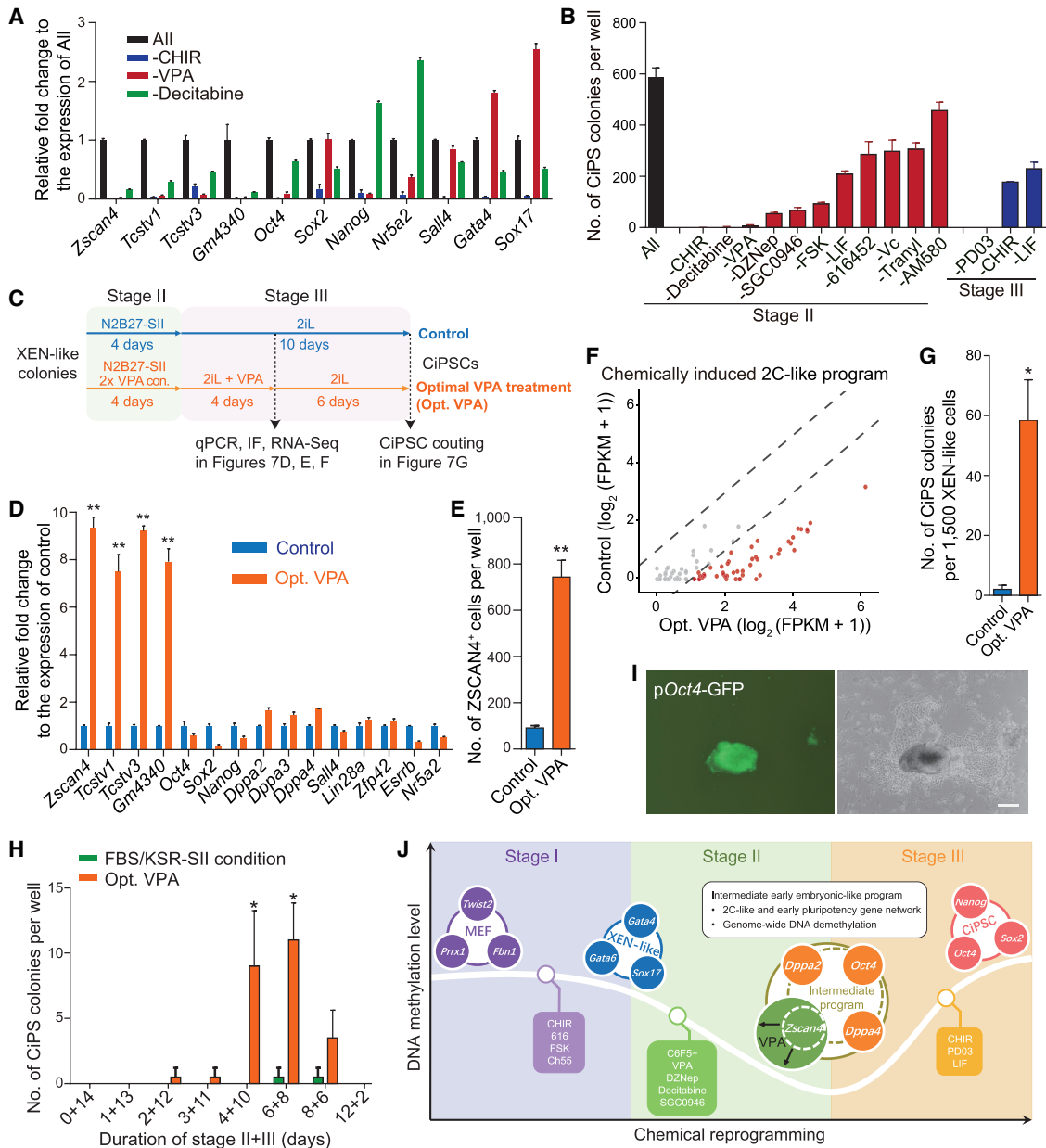
Figure 7. Substantial Acceleration of Chemical Reprogramming by Enhancing the 2C-like Program

(A) qPCR analyses of gene expression at stage III D4 induced by removing essential small molecules in N2B27-SII condition (n = 3).

(B) Number of CiPS colonies under the indicated conditions at stage III D12 (n = 3).

(C) Schematic diagram of the 14-day reprogramming system with N2B27-SII condition (control), or with doubled VPA concentration (con.) in stage II and additional VPA treatment in the early 4 days of stage III (optimal VPA treatment, opt. VPA for short).

(D) qPCR analyses of 2C and pluripotency gene expression at stage III D4 induced with control or opt. VPA (n = 3).

(E) Number of ZSCAN4+ cells at stage III D4 induced with control or opt. VPA analyzed by immunostaining (n = 2).

(F) Scatterplot displaying the expression of the 2C-like program genes in cells at stage III D4 with control or opt. VPA.

(G) Number of CiPS colonies induced with control or opt. VPA in 14-day reprogramming system from XEN-like colonies (n = 3).

(H) Number of CiPS colonies induced by opt. VPA or FBS/KSR-SII condition under a different combination of stage II and III durations (n = 2).

(I) Typical primary CiPS colonies generated with the optimized 16-day reprogramming protocol beginning from MEFs. Scale bar, 200 μm.

(J) Model for chemical reprogramming. Four major programs (MEF, XEN-like, early embryonic-like, CiPSC) are successively induced by small molecules treatment at three stages. White line indicates the DNA methylation level throughout the process. We highlight the intermediate early embryonic-like program as the bridge linking XEN-like state to pluripotency. Optimal VPA treatment augments 2C-like program (e.g., *Zscan4* expression) and consequently enhances CiPSC generation.

For (D)–(G), reprogramming system was shown in (C). The data are presented as the mean ± SD. Significance was assessed compared with the controls using a one-tailed Student's t test. ***p < 0.001; **p < 0.01; *p < 0.05. See also Figure S7.

reprogramming (Figure S6D), confirming the co-expression of *Oct4* with *Zscan4c* at the mRNA level (Figures 4I, S4G, and S4H). Moreover, ZSCAN4⁺ cells were dispersedly present in ~90% of colonies (Figure S6E), which agreed well with the high percentage of colonies that expressed OG (Figure S6F). The OG⁺ZSCAN4⁺ cells lacked OCT4 protein (Figure S6D), further supporting the similarity of Ci2C-like cells with ES-2C-like cells and 2C embryos.

Taken together, our comprehensive comparison revealed that Ci2C-like cells exhibit molecular features that are highly similar to those of 2C embryo and ES-2C-like cells, including transcriptomic similarity, the expression of endogenous retrotransposons, and widespread DNA hypomethylation. It was worthwhile to mention that, compared with the ES-2C-like cells, Ci2C-like cells expressed lower levels of late pluripotency genes and higher levels of XEN genes as well as early pluripotency gene *Dppa3*, suggesting a partially reprogrammed state with the residual XEN memory and incomplete establishment of the pluripotency network (Figure S6G).

Substantial Acceleration of Chemical Reprogramming by Enhancing the 2C-like Program

Based on the intermediate gene network identified above, we asked whether the 2C-like program could be further activated, consequently enhancing the reprogramming process. First, to identify the key factors responsible for 2C gene activation, we removed small molecules individually from the N2B27-SII condition. We found that the removal of CHIR99021 (GSK3β inhibitor), decitabine (DNA methyltransferase inhibitor), or VPA (histone deacetylase inhibitor, HDACi) exclusively reduced the 2C gene expressions at the end of stage II (Figure 7A) and subsequently significantly impaired the generation of CiPSCs (Figure 7B). In stage III, removal of PD0325901 (MEK inhibitor) resulted in maintained expressions of *Zscan4*, little activation of *Nanog* (Figure S7A), and failed generation of CiPSCs (Figure 7B), suggesting requirement of the transcriptional switch from the 2C-like program to the late pluripotency program in final reprogramming stage.

Although reprogramming was highly efficient under the N2B27-SII condition, the entire process consumed ~40 days, of which more than 24 days involved the induction from XEN-like cells to CiPSCs. Thus, we tried to accelerate the reprogramming process by stimulating the expression of 2C-like program. VPA was found to be indispensable for upregulating 2C genes during chemical reprogramming (Figure 7A). Moreover, HDACi has been reported to activate the 2C::tdTomato reporter (Macfarlan et al., 2012) and *Zscan4* expression in ESCs (Ishiuchi and Torres-Padilla, 2013). We therefore focused on VPA and tested different concentrations and durations of VPA treatment in a 14-day reprogramming system beginning with XEN-like colonies. Surprisingly, the optimal modulation of VPA treatment involved in concentration increase in stage II followed by a transient treatment in early 4 days of stage III (coincident with the peak of 2C-like program transcriptional wave, Figures 2C and S3E) (see Method Details; Figure 7C), which resulted in selective upregulation of 2C genes (by ~10-fold) rather than pluripotency genes (Figure 7D), and downregulation of XEN genes (Figure S7B). The upregulation of *Zscan4* expression was confirmed at the protein level by immunofluorescence (Figure 7E). More-

over, RNA-seq revealed that optimal VPA treatment stimulated the expressions of almost all genes in 2C-like program, with 57% of genes upregulated >2-fold (Figures 4G and 7F). Consequently, the CiPSC efficiency was remarkably improved by up to 30-fold (Figure 7G). Strikingly, only 2 days of optimal VPA treatment before switching to stage III medium was sufficient to generate CiPSCs from XEN-like cells, unlike the FBS/KSR-SII condition, which required a minimum of 6 days for induction (Figure 7H).

Finally, we applied the N2B27-SII condition with optimal VPA treatment to reprogram XEN-like cells directly induced from a modified stage I medium (see Method Details) without re-plating and feeder cells (Figures S7C and S7D). We robustly obtained CiPSCs in ~20 days, and the minimum requirement for yielding chimera-competent CiPSCs was 16 days (Figures 7I and S7E–S7H). Collectively, these data not only corroborated the role of the intermediate gene network during the transition from XEN-like state to pluripotency but also allowed us to discover an accelerated chemical reprogramming system by fine-tuning the chemical treatment.

DISCUSSION

In chemical reprogramming, mature cells are induced to undergo remarkable transcriptomic and epigenetic changes prior to achieving pluripotency. A key question in reprogramming is what molecular dynamics may lead to the establishment of full pluripotency network. To answer this question, we applied high-throughput single-cell RNA-seq to dissect the cell heterogeneity, identified the cell progression trajectory, and uncovered the ordered reactivation of molecular events throughout chemical reprogramming. Interestingly, we identified an intermediate gene network with a specific DNA methylation status resembling early embryonic signatures before full reprogramming. More importantly, based on these findings, we remarkably accelerated the chemical reprogramming kinetics by boosting the 2C-like program with HDACi.

Remarkably, our study discovered that early embryonic-like programs mediated the cell fate transitions from XEN-like state to pluripotency (Figure 7J). First, single-cell transcriptomic analysis revealed that the concomitant 2C-like program and early pluripotency program were the earliest events differentially expressed in two trajectory branches shortly after XEN-like state, forming an intermediate subnetwork that determined successful reprogramming (Figures 4A–4C and 4H). The expression level of the 2C-like program was well correlated with the reprogramming potential across various comparisons (Figures 1B, 4B, 4F, 4G, 7D–7H, and S5I–S5K). Second, we experimentally verified the important role of the gene network, which included *Zscan4*, *Oct4*, *Dppa2*, and *Dppa4*, in the late transition from XEN-like state to pluripotency (Figures 5A–5D and S5A–S5C). Importantly, all of these genes in the intermediate subnetwork exhibited expression patterns and regulatory relationships in late reprogramming stage that were similar to those in early embryogenesis from the 2C stage to blastocyst (Figures 4H and 4J) (Deng et al., 2014; Huang et al., 2017). Moreover, we were able to greatly boost CiPSC generation by enhancing the 2C-like program through precisely controlling chemical treatment (Figures 7C–7I). Third, a widespread loss of DNA methylation was uncovered

in the intermediate cells before complete reprogramming (Figures 5H–5J and S5H), similar to the hallmark of global epigenetic reprogramming observed in early embryogenesis (Baker and Pera, 2018; Smith et al., 2012). This unique DNA methylation pattern may indicate an open chromatin state with global erasure of epigenetic memory and lay a critical epigenetic foundation for achieving naive pluripotency, which exhibits global DNA hypomethylation (Leitch et al., 2013). Interestingly, several previous studies have also reported upregulation of preimplantation transcripts during OSKM-induced reprogramming; however, it remains to be further investigated at the single-cell level whether these programs were concurrent in a defined cell population, how similar these programs might be to that of early embryos, as well as their contributions to the establishment of pluripotency and the epigenetic status (Cacchiarelli et al., 2015; Eckersley-Maslin et al., 2016; Friedli et al., 2014). Our discoveries clearly revealed that the nature of chemical reprogramming was a transcriptomic and epigenetic signature reminiscent of early embryogenesis and highlighted "dynamic early embryonic-like programs" as key molecular events occurring at the late chemical reprogramming stage.

2C embryos are totipotent at the early developmental stage following fertilization when the landmark zygotic genome activation (ZGA) event occurs, which results from extensive epigenome reprogramming and represents the switch from a maternal to a zygotic transcriptome (Liu et al., 2016; Minami et al., 2007; Zhang et al., 2016a). Interestingly, a rare subpopulation known as 2C-like cells was identified in ESC culture that expressed 2C genes and was distinguished from the major ESC population by a special chromatin decompaction state (Akiyama et al., 2015; Dan et al., 2017; Eckersley-Maslin et al., 2016; Macfarlan et al., 2012). In this study, we unexpectedly discovered that somatic cells can also be chemically induced into 2C-like cells that exhibit key molecular features of 2C stage: a similar global expression profile (including a 2C-like program with over 95% of the genes specifically expressed in 2C embryos) (Figures 6A–6C and S6A–S6C), the expression of the 2C embryo-specific retrotransposable element *MERVL* (Figures 6D, 6E, S3E, and S5I), and global DNA hypomethylation (Figures 5H–5J). 2C genes *Zscan4* and *Tcstv1*, whose knockdown impaired CiPSC formation, worked in concert as a network with the core reprogramming factor *Oct4* and kinetic-booster *Dppa* family (Xu et al., 2015b) to mediate the pluripotency transition. In addition, 2C genes may perform multiple critical functions that aid the late reprogramming, because they are significantly enriched for the GO terms "negative regulation of cell differentiation," "positive regulation of cell proliferation," and "negative regulation of apoptotic process" (Figures 3E, 4C, and S5G), and are reported to facilitate telomere elongation (Akiyama et al., 2015; Dan et al., 2017; Eckersley-Maslin et al., 2016), maintain genomic stability in ESCs (Zalzman et al., 2010; Zhang et al., 2016b), and improve the OSKM-iPSC quality (Jiang et al., 2013). Collectively, the similar principle of cell fate determination in various systems including two-cell embryo, ESCs, and somatic cell reprogramming suggests a general role of the 2C-like program in mediating cell fate determination involving the early developmental transcriptome.

More importantly, using the 2C-like program as an indicator, we were able to remarkably accelerate the chemical reprogramming process, reducing the time required from 40 days to a minimum of 16 days, by enhancing the HDACi treatment (Figures 7C–7I and S7C–S7H). Coincident with our findings, HDACi has also been reported to open chromatin and activate 2C-like transcriptome in ESCs (Ishiuchi and Torres-Padilla, 2013; Macfarlan et al., 2012) and in zygote-2C transition during the SCNT process (Liu et al., 2018), suggesting the same principles of regulating 2C program across various systems. Accordingly, it is worthwhile to attempt further improvement of the reprogramming process by manipulating other 2C-like state boosters such as CAF-1 (Ishiuchi et al., 2015), which also functioned as a booster in OSKM-reprogramming system (Cheloufi et al., 2015). Finally, the discovery of Ci2C-like cells and their reprogramming conditions also suggested the possibility of, and laid the preliminary foundation for, obtaining cells *in vitro* that may represent an earlier developmental stage than that of ESCs.

In summary, through dissection of cell fate dynamics during chemical reprogramming using scRNA-seq, we deciphered the concomitant early pluripotency program and 2C-like program as the intermediate molecular event bridging the XEN-like state toward pluripotency, which consequently allows the development of a much faster chemical reprogramming system. With further optimization of chemical reprogramming, we expect it to be a unique and promising strategy for manipulating cell fate conversion and interpreting pluripotency and development in the future.

STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mice
  - Cell Culture
- METHOD DETAILS
  - MEF isolation
  - Generation of CiPSCs in high efficiency
  - Generation of CiPSCs from XEN-like colonies in about 2 weeks
  - Generation of CiPSCs from MEFs in about 20 days
  - Calculation of cell number, efficiency and percentage
  - Establishment of XEN-like cell lines
  - Single cell reprogramming
  - Flow cytometry and cell sorting
  - Reverse transcription (RT)-quantitative PCR (qPCR)
  - RNA sequencing
  - Whole-genome bisulfite sequencing (WGBS)
  - Enzyme-linked immunosorbent assay (ELISA)
  - Immunofluorescence
  - Plasmid construction and lentivirus production
  - Chimera construction
  - Single cell cDNA library preparation and sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Single cell data preprocess
  - Single cell trajectory analysis
  - PCA and *t*-SNE analysis

- ○ Gene regulatory network analysis
- ○ Transcriptome analysis of Ci2C-like cells with reported datasets
- ○ WGBS analysis
- ○ Ontology Annotation
- ● DATA AND SOFTWARE AVAILABILITY

AUTHOR CONTRIBUTIONS

T.Z. and H.D. conceived this project and wrote the paper. H.D. supervised this project. T.Z., Y.F., J.Z., Z.Y., S.C., Z.J., and S.D. performed the experiments. Y.L., Q.Z., and C.L. conducted the bioinformatics analyses. J.X. performed scRNA-seq experiments. X.X., C.T., and Y.B. participated in the design of the experiments.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Shu et al. (2013).

REFERENCES

Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086.

Akiyama, T., Xin, L., Oda, M., Sharov, A.A., Amano, M., Piao, Y., Cadet, J.S., Dudekula, D.B., Qian, Y., Wang, W., et al. (2015). Transient bursts of Zscan4 expression are accompanied by the rapid derepression of heterochromatin in mouse embryonic stem cells. DNA Res. *22*, 307–318.

Assenov, Y., Müller, F., Lutsik, P., Walter, J., Lengauer, T., and Bock, C. (2014). Comprehensive analysis of DNA methylation data with RnBeads. Nat. Methods *11*, 1138–1140.

Baker, C.L., and Pera, M.F. (2018). Capturing totipotent stem cells. Cell Stem Cell *22*, 25–34.

Cacchiarelli, D., Trapnell, C., Ziller, M.J., Soumillon, M., Cesana, M., Karnik, R., Donaghey, J., Smith, Z.D., Ratanasirintrawoot, S., Zhang, X., et al. (2015). Integrative analyses of human reprogramming reveal dynamic nature of induced pluripotency. Cell *162*, 412–424.

Cao, S., Yu, S., Li, D., Ye, J., Yang, X., Li, C., Wang, X., Mai, Y., Qin, Y., Wu, J., et al. (2018). Chromatin accessibility dynamics during chemical induction of pluripotency. Cell Stem Cell *22*, 529–542.

Cheloufi, S., Elling, U., Hopfgartner, B., Jung, Y.L., Murn, J., Ninova, M., Hubmann, M., Badeaux, A.I., Euong Ang, C., Tenen, D., et al. (2015). The histone chaperone CAF-1 safeguards somatic cell identity. Nature *528*, 218–224.

Dan, J., Rousseau, P., Hardikar, S., Veland, N., Wong, J., Autexier, C., and Chen, T. (2017). Zscan4 inhibits maintenance DNA methylation to facilitate telomere elongation in mouse embryonic stem cells. Cell Rep. *20*, 1936–1949.

Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science *343*, 193–196.

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.

Eckersley-Maslin, M.A., Svensson, V., Krueger, C., Stubbs, T.M., Giehr, P., Krueger, F., Miragaia, R.J., Kyriakopoulos, C., Berrens, R.V., Milagre, I., et al. (2016). MERVL/Zscan4 network activation results in transient genome-wide DNA demethylation of mESCs. Cell Rep. *17*, 179–192.

Falco, G., Lee, S.L., Stanghellini, I., Bassey, U.C., Hamatani, T., and Ko, M.S. (2007). Zscan4: a novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. Dev. Biol. *307*, 539–550.

Friedli, M., Turelli, P., Kapopoulou, A., Rauwel, B., Castro-Díaz, N., Rowe, H.M., Ecco, G., Unzu, C., Planet, E., Lombardo, A., et al. (2014). Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. Genome Res. *24*, 1251–1259.

Gurdon, J.B. (1962). Adult frogs derived from the nuclei of single somatic cells. Dev. Biol. *4*, 256–273.

Hou, P., Li, Y., Zhang, X., Liu, C., Guan, J., Li, H., Zhao, T., Ye, J., Yang, W., Liu, K., et al. (2013). Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. Science *341*, 651–654.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Huang, Y., Kim, J.K., Do, D.V., Lee, C., Penfold, C.A., Zylicz, J.J., Marioni, J.C., Hackett, J.A., and Surani, M.A. (2017). Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition. eLife *6*, e22345.

Ishiuchi, T., and Torres-Padilla, M.E. (2013). Towards an understanding of the regulatory mechanisms of totipotency. Curr. Opin. Genet. Dev. *23*, 512–518.

Ishiuchi, T., Enriquez-Gasca, R., Mizutani, E., Bošković, A., Ziegler-Birling, C., Rodriguez-Terrones, D., Wakayama, T., Vaquerizas, J.M., and Torres-Padilla, M.E. (2015). Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. Nat. Struct. Mol. Biol. *22*, 662–671.

Jiang, J., Lv, W., Ye, X., Wang, L., Zhang, M., Yang, H., Okuka, M., Zhou, C., Zhang, X., Liu, L., and Li, J. (2013). Zscan4 promotes genomic stability during reprogramming and dramatically improves the quality of iPS cells as demonstrated by tetraploid complementation. Cell Res. *23*, 92–106.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559.

Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. Nat. Struct. Mol. Biol. 20, 311–316.

Li, Y., Zhang, Q., Yin, X., Yang, W., Du, Y., Hou, P., Ge, J., Liu, C., Zhang, W., Zhang, X., et al. (2011). Generation of iPSCs from mouse fibroblasts with a single gene, Oct4, and small molecules. Cell Res. 21, 196–204.

Li, X., Liu, D., Ma, Y., Du, X., Jing, J., Wang, L., Xie, B., Sun, D., Sun, S., Jin, X., et al. (2017). Direct Reprogramming of fibroblasts via a chemically induced XEN-like state. Cell Stem Cell 21, 264–273.

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.

Liu, X., Wang, C., Liu, W., Li, J., Li, C., Kou, X., Chen, J., Zhao, Y., Gao, H., Wang, H., et al. (2016). Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. Nature 537, 558–562.

Liu, Z., Cai, Y., Wang, Y., Nie, Y., Zhang, C., Xu, Y., Zhang, X., Lu, Y., Wang, Z., Poo, M., and Sun, Q. (2018). Cloning of Macaque monkeys by somatic cell nuclear transfer. Cell 172, 881–887.

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.

Minami, N., Suzuki, T., and Tsukamoto, S. (2007). Zygotic gene activation and maternal factors in mammals. J. Reprod. Dev. 53, 707–715.

Ping, W., Hu, J., Hu, G., Song, Y., Xia, Q., Yao, M., Gong, S., Jiang, C., and Yao, H. (2018). Genome-wide DNA methylation analysis reveals that mouse chemical iPSCs have closer epigenetic features to mESCs than OSKM-integrated iPSCs. Cell Death Dis. 9, 187.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods 14, 979–982.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. 29, 24–26.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504.

Shu, J., Wu, C., Wu, Y., Li, Z., Shao, S., Zhao, W., Tang, X., Yang, H., Shen, L., Zuo, X., et al. (2013). Induction of pluripotency in mouse somatic cells with lineage specifiers. Cell 153, 963–975.

Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature 484, 339–344.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676.

Takahashi, K., and Yamanaka, S. (2016). A decade of transcription factor-mediated reprogramming to pluripotency. Nat. Rev. Mol. Cell Biol. 17, 183–193.

Tang, F., Barbacioru, C., Nordman, E., Bao, S., Lee, C., Wang, X., Tuch, B.B., Heard, E., Lao, K., and Surani, M.A. (2011). Deterministic and stochastic allele specific gene expression in single mouse blastomeres. PLoS ONE 6, e21208.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. 31, 46–53.

Xu, J., Du, Y., and Deng, H. (2015a). Direct lineage reprogramming: strategies, mechanisms, and applications. Cell Stem Cell 16, 119–134.

Xu, X., Smorag, L., Nakamura, T., Kimura, T., Dressel, R., Fitzner, A., Tan, X., Linke, M., Zechner, U., Engel, W., and Pantakani, D.V. (2015b). Dppa3 expression is critical for generation of fully reprogrammed iPS cells and maintenance of Dlk1-Dio3 imprinting. Nat. Commun. 6, 6008.

Ye, J., Ge, J., Zhang, X., Cheng, L., Zhang, Z., He, S., Wang, Y., Lin, H., Yang, W., Liu, J., et al. (2016). Pluripotent stem cells induced from mouse neural stem cells and small intestinal epithelial cells by small molecule compounds. Cell Res. 26, 34–45.

Zalzman, M., Falco, G., Sharova, L.V., Nishiyama, A., Thomas, M., Lee, S.L., Stagg, C.A., Hoang, H.G., Yang, H.T., Indig, F.E., et al. (2010). Zscan4 regulates telomere elongation and genomic stability in ES cells. Nature 464, 858–863.

Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., Ming, J., Wu, X., Zhang, Y., Xu, Q., et al. (2016a). Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature 537, 553–557.

Zhang, Q., Dan, J., Wang, H., Guo, R., Mao, J., Fu, H., Wei, X., and Liu, L. (2016b). Tcstv1 and Tcstv3 elongate telomeres of mouse ES cells. Sci. Rep. 6, 19852.

Zhao, Y., Yin, X., Qin, H., Zhu, F., Liu, H., Yang, W., Zhang, Q., Xiang, C., Hou, P., Song, Z., et al. (2008). Two supporting factors greatly improve the efficiency of human iPSC generation. Cell Stem Cell 3, 475–479.

Zhao, Y., Zhao, T., Guan, J., Zhang, X., Fu, Y., Ye, J., Zhu, J., Meng, G., Ge, J., Yang, S., et al. (2015). A XEN-like state bridges somatic cells to pluripotency during chemical reprogramming. Cell 163, 1678–1691.

Zhao, T., Li, Y., and Deng, H. (2016). Cell fate conversion-from the viewpoint of small molecules and lineage specifiers. Diabetes Obes. Metab. 18 (Suppl 1), 3–9.

Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 8, 14049.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rat monoclonal anti-EpCAM PE-Cy7	eBioscience	Cat#25-5791-80 (G8.8); RRID: AB_1724047
Rat anti-PECAM-1 PE	BioLegend	Cat#102407 (390); RRID: AB_312902
Rat anti-PDGFRA APC	BioLegend	Cat#135908 (APA5); RRID: AB_2043970
Rabbit polyclonal anti-Zscan4	Millipore	Cat#AB4340; LOT: 2864843
Rabbit polyclonal anti-Nanog	Millipore	Cat#AB5731; RRID: AB_2267042
Rabbit polyclonal anti-Sox2	Stemgent	Cat#09-0024; RRID: AB_2195775
Mouse monoclonal anti-Oct4	BD Biosciences	Cat#611203; RRID: AB_398737
Cy3-AffiniPure Donkey Anti-Rabbit IgG (H+L)	Jackson ImmunoResearch	Cat#711-165-152; RRID: AB_2307443
Chemicals, Peptides, and Recombinant Proteins		
VPA (V)	Sigma-Aldrich	Cat#P4543
CHIR99021 (CHIR, C)	WUXI APPTEC	N/A
616452 (616, 6)	WUXI APPTEC	N/A
Tranylcypromine (Tranyl, T)	Enzo Life Sciences	Cat#BML-EI217
Forskolin (FSK, F)	Enzo Life Sciences	Cat#BML-CN100
AM580 (A)	Tocris	Cat#0760
EPZ004777 (EPZ, E)	Selleckchem	Cat#S7353
Ch55 (5)	Tocris	Cat#2020
3-Deazaneplanocin A (DZNep, Z)	WUXI APPTEC	N/A
SGC0946 (SGC, S)	Selleckchem	Cat#S7079
Decitabine (D)	Enzo Life Sciences	Cat#ALX-480-096-M005
PD0325901 (PD03)	WUXI APPTEC	N/A
L-Ascorbic acid 2-phosphate (vitamin C, Vc)	Sigma-Aldrich	Cat#A8960
Recombinant Human FGF-basic (bFGF, b)	Origene	Cat#TP750002
Mouse LIF (LIF, L)	Miltenyi Biotec	Cat#130-099-895
AlbuMAX-II	Thermo Fisher Scientific	Cat#11021-037
Critical Commercial Assays		
Direct-zol RNA MiniPrep Plus	Zymo Research	Cat#R2072
EasyScript One-Step gDNA Removal and cDNA Synthesis SuperMix	Transgen BIOTECH	Cat#AT311-03
KAPA SYBR FAST qPCR Master	Kapa Biosystems	Cat#KM4101
Single Cell 3′ Library and Gel Bead Kit V2	10x Genomics	Cat#120237
Chromium single cell chip kit V2	10x Genomics	Cat#120236
EasyPure Genomic DNA Kit	TransGen BIOTECH	Cat#EE101-11
EZ DNA Methylation-Gold Kit	Zymo Research	Cat#D5005
Accel-NGS Methyl-Seq DNA Library Kit	Swift	Cat#30024
NEBNext Ultra RNA Library Prep Kit for Illumina	NEB England BioLabs	Cat#E7530L
5-mC DNA ELISA Kit	Zymo Research	Cat#D5325
Deposited Data		
RNA-seq data	This paper	BIG: CRA000932; GEO: GSE114952
Experimental Models: Cell Lines		
ESC	(Zhao et al., 2015)	N/A
XEN-like cell line (carrying pOct4-GFP reporter)	This paper	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
XEN-like cell line (carrying p*Oct4*-GFP and p*Fsp1*-Tdtomato reporter)	This paper	N/A
p*Zscan4*-Emerald ESC	(Zalzman et al., 2010)	N/A
Experimental Models: Organisms/Strains		
Mouse: C57BL/6J− Tg(GOFGFP)11Imeg/Rbrc	RIKEN BioResource Center	Mouse strain datasheet: RBRC00771
Mouse: CD-1(ICR)	Beijing Vital River Laboratory Animal Technology Co., Ltd	Mouse strain datasheet: 201
Oligonucleotides		
Zscan4 shRNA	Sigma-Aldrich	SHPM01
Tcstv1 shRNA	Sigma-Aldrich	SHPM01
Gata4 shRNA	Sigma-Aldrich	SHPM01
Gata6 shRNA	Sigma-Aldrich	SHPM01
Sox17 shRNA	Sigma-Aldrich	SHPM01
Oct4 shRNA	Sigma-Aldrich	SHPM01
Dppa2 shRNA	Sigma-Aldrich	SHPM01
Dppa4 shRNA	Sigma-Aldrich	SHPM01
Non-targeting shRNA	Sigma-Aldrich	SHPM01
Recombinant DNA		
FUW-tetO-*OCT4*	This paper	N/A
FUW-tetO-*SOX2*	This paper	N/A
FUW-tetO-*NANOG*	This paper	N/A
FUW-tetO-*GATA4*	This paper	N/A
FUW-tetO-*GATA6*	This paper	N/A
FUW-tetO-*DPPA2*	This paper	N/A
FUW-tetO-*DPPA4*	This paper	N/A
FUW-tetO-*Tcstv1*	This paper	N/A
FUW-tetO-*Zscan4c*	This paper	N/A
FUW-tetO- *Zscan4d*	This paper	N/A
FUW-tetO-*Dppa3*	This paper	N/A
Software and Algorithms		
Cell Ranger 2.0.1	10x Genomics	https://support.10xgenomics.com/
R 3.4.1	R Project	https://www.r-project.org
Monocle 2.4.0	Qiu et al., 2017	https://github.com/cole-trapnell-lab/monocle-release
TopHat 2.1.1	Kim et al., 2013	https://ccb.jhu.edu/software/tophat
Cufflinks 2.2.1	Trapnell et al., 2013	https://github.com/cole-trapnell-lab/cufflinks
STAR 2.5.1b	Dobin et al., 2013	https://github.com/alexdobin/STAR
featureCounts 1.5.3	Liao et al., 2014	http://bioinf.wehi.edu.au/featureCounts
SCENIC 0.1.6	Aibar et al., 2017	https://github.com/aertslab/SCENIC
WGCNA 1.62	Langfelder and Horvath, 2008	https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/
Cytoscape 3.6.0	Shannon et al., 2003	http://www.cytoscape.org
Bismark 0.19.0	Krueger and Andrews, 2011	https://github.com/FelixKrueger/Bismark
RnBeads 1.10.7	Assenov et al., 2014	https://rnbeads.org
IGV 2.4.8	Robinson et al., 2011	http://software.broadinstitute.org/software/igv/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Hongkui Deng (hongkui_deng@pku.edu.cn).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

All animal experiments were performed according to the Animal Protection Guidelines of Peking University, China. The C57BL/6J-Tg(GOFGFP)11Imeg/Rbrc (p*Oct4*-GFP, OG) mice were mated with ICR mice to generate offspring carrying *Oct4* promoter-driven GFP and provides primary mouse embryonic fibroblasts (MEFs).

Cell Culture

MEFs were cultured in high glucose DMEM (GIBCO) supplemented with 15% fetal bovine serum (FBS, Pan-Biotech), 1% GlutaMAX-I (GIBCO), 1% nonessential amino acids (NEAA, GIBCO), 0.1 mM β-mercaptoethanol (Sigma-Aldrich) and 1% penicillin-streptomycin (GIBCO).

Mouse ESCs and CiPSCs were maintained on feeder layers (the mitomycin C-treated MEFs) in ESC culture medium containing KnockOut DMEM (GIBCO), 10% knockout serum replacement (KSR, GIBCO), 10% FBS, 1% GlutaMAX-I, 1% NEAA, 0.1 mM β-mercaptoethanol, 1% penicillin-streptomycin supplemented with 2iL (3 μM CHIR99021, 1 μM PD0325901 and 10 ng/ml LIF) or with additional 50 μg/ml vitamin C (for scRNA-seq in Figure S6A). MEFs, ESCs and CiPSCs were incubated at 37°C with 5% CO$_2$.

METHOD DETAILS

MEF isolation

MEFs were isolated from E13.5 embryos. After the removal of head, limbs, visceral tissues and gonads, remaining embryos were washed twice with PBS (containing 2% penicillin-streptomycin), minced by scissors and dissociated with 0.25% trypsin-EDTA (GIBCO) (1-2 ml per embryo) at 37°C for 5-10 min. After trypsinization, MEF medium was added and cells were pipetted up and down several times for dissociation. Centrifuge cells at 400 g for 5 min and resuspend cells in MEF medium. Generally, 1×10^7 cells can be obtained from a single embryo and are plated in a 100-mm dish (P0). The next day, change fresh MEF medium to remove non-adherent cells. Primary MEFs usually become confluent in 2 days and were ready to passage for reprogramming. To remove cell clusters and obtain single cell suspension, cells were passed through a 40 μm cell strainer before seeding for chemical induction (P1).

Generation of CiPSCs in high efficiency

1. Reagents setup:

   Throughout the chemical reprogramming process, cells were maintained at 37°C with 5% CO$_2$.

   Small molecules: VPA, CHIR99021, 616452, Tranylcypromine, Forskolin, AM580, EPZ004777, DZNep, Decitabine, SGC0946 and PD0325901 were described in key resources table. The schematic diagram of the protocol is shown in Figure S1A.

   Stage I medium: For day 0-12, KnockOut DMEM supplemented with 10% KSR, 10% FBS, 1% GlutaMAX-I, 1% NEAA, 0.1 mM 2-mercaptoethanol, 1% penicillin-streptomycin, 100 ng/ml bFGF and the small-molecule cocktail VC6TFAE (500 μM VPA, 20 μM CHIR99021, 10 μM 616452, 5 μM Tranylcypromine, 10 μM Forskolin, 0.05 μM AM580 and 5 μM EPZ004777). For day 12-16, reduce the concentrations of bFGF and CHIR99021 to 25 ng/ml and 10 μM, respectively.

   Stage II medium (FBS/KSR-SII condition): FBS/KSR-based medium (KnockOut DMEM with 10% KSR, 10% FBS, 1% GlutaMAX-I, 1% NEAA, 0.1 mM 2-mercaptoethanol, 1% penicillin-streptomycin) supplemented with 25 ng/ml bFGF and the small-molecule cocktail VC6TFAZDS (500 μM VPA, 10 μM CHIR99021, 10 μM 616452, 5 μM Tranylcypromine, 10 μM Forskolin, 0.05 μM AM580, 0.05 μM DZNep, 0.5 μM Decitabine and 5 μM SGC0946).

   Stage II medium (N2B27-SII condition): N2B27-based medium (1:1 mixture of DMEM/F12 (GIBCO) and Neurobasal (GIBCO), with 1% N2 supplement (GIBCO), 2% B27 supplement (GIBCO), 1% GlutaMAX-I, 1% NEAA, 0.1 mM 2-mercaptoethanol, 1% penicillin and streptomycin) supplemented with 10 ng/ml LIF, 50 μg/ml vitamin C, 2 mg/ml Albumax-II (dispensable), 25 ng/ml bFGF (dispensable) and the small-molecule cocktail VC6TFAZDS (500 μM VPA, 10 μM CHIR99021, 10 μM 616452, 5 μM Tranylcypromine, 10 μM Forskolin, 0.05 μM AM580, 0.05 μM DZNep, 0.5 μM Decitabine and 5 μM SGC0946).

   Stage III medium: N2B27-based medium supplemented with 3 μM CHIR99021, 1 μM PD0325901 and 10 ng/ml LIF.

2. MEFs were seeded at a density of 50,000 cells per well of a 6-well plate or 300,000 cells per 100 mm dish with MEF culture medium.
3. The next day (day 0), change the medium into stage I medium. Change the medium every 4 days.
4. On day 12, for the induction with FBS/KSR-SII condition, re-plate the cells at a density of 50,000-200,000 cells per well of a 6-well plate with stage I medium (reduced concentrations of bFGF and CHIR99021). For the induction with N2B27-SII condition, feeder cells (MEFs treated with mitomycin C and seeded at 150,000 cells per well of 6-well plate) should be prepared one

day before re-plating. Re-plate the cells on feeder cells at a density of 2,000-10,000 cells per well of 6-well plate with stage I medium (reduced concentrations of bFGF and CHIR99021).

5. On day 16, XEN-like epithelial colonies are formed and the culture is changed into stage II medium. Change the medium every 4 days.
6. On day 28, change the culture into stage III medium. Change the medium every 4 days.
7. 2i-competent, ESC-like and p*Oct4*-GFP positive CiPS colonies should become visible approximately 3 days (for N2B27-SII condition) or 6 days (for FBS/KSR-SII condition) in stage III medium.

Generation of CiPSCs from XEN-like colonies in about 2 weeks

Optimized N2B27-SII condition: N2B27-based medium supplemented with 10 ng/ml LIF, 50 μg/ml vitamin C, 2 mg/ml Albumax-II (dispensable), 25 ng/ml bFGF (dispensable) and the small-molecule cocktail VC6TFAZDS (VPA at 1 mM).

1. On day −4, re-plate 2,000-10,000 XEN-like cells or induced cells at stage I D12 per well of 6-well plate on feeder cells with stage I medium (reduced concentrations of bFGF and CHIR99021).
2. On day 0, switch the medium into optimized N2B27-SII condition.
3. From day 4-8, incubate cells in N2B27 medium containing 2iL, 50 μg/ml vitamin C and 500 μM VPA.
4. On day 8, VPA was removed and cells were cultured in N2B27 medium containing 2iL and 50 μg/ml vitamin C.
5. After another 4–8 days, CiPS colonies emerged.

Generation of CiPSCs from MEFs in about 20 days

1. Reagents setup:

Small molecules: VPA, CHIR99021, 616452, Tranylcypromine, Forskolin, Ch55, EPZ004777, DZNep, Decitabine, SGC0946 and PD0325901 were described in key resources table.

Optimized stage I medium: FBS/KSR-based medium supplemented with the small-molecule cocktail VC6TF5E (100 μM VPA, 40 μM CHIR99021, 10 μM 616452, 5 μM Tranylcypromine, 10 μM Forskolin, 1 μM Ch55 and 5 μM EPZ004777).

Optimized N2B27-SII medium: N2B27-based medium supplemented with 10 ng/ml LIF, 50 μg/ml vitamin C, 25 ng/ml bFGF, 2 mg/ml Albumax-II (recommended for enhancing efficiency) and the small-molecule cocktail VC6TF5ZDS (VPA at 1 mM).

Stage III medium: N2B27-based medium with 3 μM CHIR99021, 1 μM PD0325901, 10 ng/ml LIF and 50 μg/ml vitamin C.

2. On day −1, MEFs were seeded at 50,000 cells per well of 6-well plate with MEF culture medium.
3. The next day (day 0), change the medium into optimized stage I medium and small XEN-like colonies emerged at day 4-6.
4. From day 4-8 or 6-12, cells were cultured in optimized N2B27-SII medium
5. During day 8-12 or 12-16, cells were cultured in stage III medium with 500 μM VPA.
6. On day 12 or 16, VPA was removed and cells were cultured in stage III medium.
7. After another 4–10 days, CiPS colonies emerged.

Notes: In order to obtain robust chemical reprogramming in about 20 days, it is strongly recommended to use freshly isolated MEFs at passage 1 (P1) (described above). The minimal requirement for obtaining CiPSCs using this protocol is 16-day (referred to as 16-day reprogramming protocol in Figure 7). The schematic diagram of the protocol is shown in Figure S7C.

Calculation of cell number, efficiency and percentage

In general, 1,000 XEN-like cells or cells from stage I were re-plated into each well of 12-well plate. 4 days after re-plating, approximate 300 XEN-like colonies were formed. XEN-like colonies rapidly proliferated with N2B27-SII induction and generated a total of 500,000 – 1,000,000 cells with about 30,000 Ci2C-like cells (~4% of the whole cell) at the end of stage II (D12). Finally, approximate 500 CiPS colonies were generated at the end of stage III (50% colony efficiency relative to 1,000 re-plated cells). The CiPS colonies were typically present as large colonies with thousands of cells per colony, accounting to a large percent of cells at the end of stage III (~77% of the whole cell population).

Establishment of XEN-like cell lines

The establishment of purified XEN-like cell lines capable of long-term expansion was described before (Li et al., 2017; Zhao et al., 2015). Briefly, single XEN-like colony induced from MEFs with stage I induction was picked, trypsinized to single cells and re-plated on feeder cells (only used for the first passage) for expansion. Routinely, the purified XEN-like cell lines were long-term cultured in stage I medium (with reduced concentrations of bFGF, CHIR99021) (AM580 and EPZ004777 were not essential) on gelatin-coated (dispensable) plate, refreshed with medium every day and passaged at 1:20 to 1:30 every 3 days. p*Zscan4c*-Emerald XEN-like cell line was established through differentiating from p*Zscan4c*-Emerald ESCs (Zalzman et al., 2010). In brief, p*Zscan4c*-Emerald ESCs were passaged into stage I medium (with reduced concentrations of bFGF and CHIR99021) on gelatin-coated plate and passaged when cells were confluent. We differentiated the cells for at least 9 passages and confirmed the XEN-like peoporty by detecting XEN

gene mRNA and protein expression before using for chemical reprogramming. p*Zscan4c*-Emerald XEN-like cells were typically cultured in stage I medium (with reduced concentrations of bFGF and CHIR99021) and passaged at 1:20 every 3 days.

Single cell reprogramming
Feeder cells (treated with mitomycin C, 10,000 cells per well of 96-well plate) were prepared before sorting. The next day, individual XEN-like cells (cell lines induced from MEFs isolated from p*Fsp1*-Cre: Rosa26R^tdTomato: p*Oct4*-GFP progeny mice as described before [Ye et al., 2016]) (constitutively expressed tdTomato was used for quantifying re-plating efficiency and tracking individual cells) or individual cells on stage III after antibody staining were sorted into each well of 96-well plate and reprogrammed as described above.

Flow cytometry and cell sorting
Single-cell suspensions of chemically induced cells or CiPSCs were dissociated by accutase at 37°C for 10 min (for antibody staining) or by 0.25% trypsin-EDTA at 37°C for 3-5 min (for analysis without antibody staining) followed by filtration through 40 μm cell strainers.

(1) For analysis of OG percentage, individual XEN-like cell or p*Zscan4c*-Emerald cells sorting, flow cytometry was performed using Aria III (BD) or FACSVerse (BD) or MoFlo XDP (BeckMan Coulter).

(2) For OG⁺PDGFRA^low/-PECAM1⁻ staining, cell pellets were washed with PBS and incubated with PDGFRA, PECAM1 and EPCAM (EPCAM was used for excluding feeder cells) antibodies in PBS containing 2% BSA for 30 min at 4°C in the dark. Then cells were washed with PBS and re-suspended in PBS containing 2% BSA for fluorescence-activated cell sorting analysis. Unsorted cell population presented the control group which were stained with antibody but without sorting.

Reverse transcription (RT)-quantitative PCR (qPCR)
Total RNA was isolated using Direct-zol RNA MiniPrep Kit (Zymo Research, R2072). Extracted RNA was treated with DNase and cDNA was synthesized from 0.5-1 μg of total RNA using TransScript First-Strand cDNA Synthesis SuperMix (TransGen Biotech, AT311-03). qPCR was performed by using KAPA SYBR® FAST qPCR Kit Master Mix (KAPA Biosystems, KM4101) on a CFX Connect™ Real-Time System (Bio-Rad). The data were analyzed using the delta-delta Ct method. *Gapdh* was used as a control to normalize the expression of target genes. Primer sequences for qPCR in this study are listed in Table S7. *Zscan4* was detected by a common primer for all *Zscan4* paralogs.

RNA sequencing
Total RNA was isolated using Direct-zol RNA MiniPrep Kit (Zymo Research). RNA sequencing libraries were constructed using the NEBNext® Ultra RNA Library Prep Kit for Illumina® (NEB England BioLabs). Fragmented and randomly primed 2 × 150 bp paired-end libraries were sequenced using Illumina HiSeq X Ten.

Whole-genome bisulfite sequencing (WGBS)
Genomic DNA of sorted p*Zscan4c*-Em⁺, p*Zscan4c*-Em⁻ cells at stage II D12, XEN-like cell with p*Zscan4c*-Emerald reporter and CiPSCs were extracted using EasyPure Genomic DNA Kit (TransGen, EE101-11). DNA were denatured and bisulfite converted using EZ DNA Methylation-Gold™ Kit (Zymo Research, D5005). In brief, 20 μl DNA (300 ng) was mixed with 130 μl CT Conversion Reagent and incubated at 98°C for 10 minutes, 64°C for 2.5 hours. Samples and 600 μl M-Binding Buffer were added into a Zymo-Spin IC Column, followed by inverting the column several times and centrifuging at full speed (> 10,000 g) for 30 s. 100 μl M-Wash Buffer was added to the column and centrifuged at full speed for 30 s. Add 200 μl M-Desulphonation Buffer to the column, incubate at room temperature (20-30°C) for 15-20 minutes and then centrifuging at full speed for 30 s. Wash the column with 200 μl M-Wash Buffer twice and elute the DNA with 10 μl M-Elution Buffer. Recovered bisulfite-converted DNAs were constructed into sequencing libraries using Accel-NGS Methyl-Seq DNA Library Kit (Swift, 30024) following the manufacturer's instructions and each libraries were sequenced 90G raw data by using Illumina HiSeq X Ten.

Enzyme-linked immunosorbent assay (ELISA)
MEFs, XEN-like cells, reprogrammed cells at stage II D12 and stage III D4 were collected and the genomic DNAs were extracted using EasyPure Genomic DNA Kit (TransGen, EE101-11). The 5-methylcytosine (5-mC) of genomic DNAs were detected and quantitated using 5-mC DNA ELISA Kit (Zymo Research, D5325). In brief, 100 ng DNA of each samples and controls were added to PCR tubes in a final volume to 100 μl with 5-mC Coating Buffer and incubated at 98°C for 5 minutes. Denatured DNAs were added to the wells of ELISA plate and incubated for at 37°C for 1 hour. After wash, 200μl of 5-mC ELISA buffer were added to each well, and incubate at 37°C for 30 minutes. Next, Anti-5-Methylcytosine (1:2,000) and Secondary Antibody (1:1,000) were mixed with 5-mC ELISA Buffer and added to each wells, and then incubated at 37°C for 1 hour. After wash, added 100 μl of HRP Developer per well and incubated at room temperature for 35 minutes, and the absorbances were measured at 405 nm. Two technical replicates were used for each condition.

Immunofluorescence

Immunofluorescence was performed as previously described (Hou et al., 2013). After fixation with 4% paraformaldehyde (DingGuo, AR-0211) at room temperature for 30 min, cells were permeabilized with PBS-0.1% Triton X-100 (Sigma-Aldrich, T8787) and blocked with PBS containing 2.5% donkey serum (Jackson Immuno Research, 017-000-121) at 37°C for 1 hour. Primary antibodies incubation with appropriate dilutions (anti-Oct4, 1:200, BD Biosciences, 611203; anti-Sox2, 1:500, Stemgent, 09-0024; anti-Nanog, 1:500, Millipore, AB5731; anti-Zscan4, 1:1000, Millipore, AB4340) were performed at 4°C overnight in blocking buffer. On the next day, cells were washed with PBS for three times and probed with secondary antibodies at 37°C for 1 hour in blocking buffer (Jackson ImmunoResearch). Cells were then washed with PBS for three times and DNA was stained with DAPI solution (Roche Life Science, 10236276001). Antibodies are listed in the key resources table.

Plasmid construction and lentivirus production

Genetic knockdown was performed using shRNAs (SIGMA Mission® shRNA, SHPM01) according to the manufacturer's protocol. Genetic overexpression and lentivirus production, collection and infection were the same as described (Hou et al., 2013; Li et al., 2011; Zhao et al., 2008).

Chimera construction

Chimeric mice were obtained by the injection of CiPSCs (ICR X C57) into blastocysts or eight-cell embryos. For blastocyst injection, 10-15 CiPS cells were injected into the E3.5 blastocysts and transferred to the uterus of 2.5 day post-coitum pseudopregnant females (ICR). For eight-cell injection, 7-10 CiPS cells were injected into each E2.5 embryos and transferred into 0.5 d post-coitum pseudopregnant females (ICR). Approximately 15 injected embryos were transferred into each uterine horn of pseudopregnant mice.

Single cell cDNA library preparation and sequencing

Cells at different time points throughout chemical reprogramming process were harvested and resuspended at 1×10^6 cells per milliliter in $1 \times$ PBS with 0.04% BSA. Then, cell suspensions (300-1000 living cells per microliter determined by trypan blue staining) were loaded on a Chromium Single Cell Controller (10x Genomics) to generate single-cell gel beads in emulsion (GEMs) by using Single Cell 3′ Library and Gel Bead Kit V2 (10x Genomics, 120237). Captured cells were lysed and the released RNA were barcoded through reverse transcription in individual GEMs (Zheng et al., 2017). Barcoded cDNAs were pooled and cleanup by using DynaBeads® MyOne Silane Beads (Invitrogen, 37002D). Single-cell RNA-seq libraries were prepared using Single Cell 3′ Library Gel Bead Kit V2 (10x Genomics, 120237) following the manufacture's introduction. Sequencing was performed on an Illumina HiSeq X Ten with pair end 150bp (PE150).

QUANTIFICATION AND STATISTICAL ANALYSIS

Single cell data preprocess

We checked the quality of raw sequencing data by FastQC software. The fastq sequence was trimmed by FASTX-Toolkit. We used cellranger count pipeline 2.0.1 to analyze the sequencing data and generated the single cell information, once for each individual sample. All the cells in different batches were merged together by cellranger aggr pipeline and normalized by equalizing the read depth among libraries. The final result was the matrix of all cells and their global gene expressions.

Single cell trajectory analysis

Single cell trajectory was analyzed using matrix of cells and gene expressions by Monocle 2 for Figures 3A and S3A. Before monocle analysis, we removed the cells beyond UMI counts ± 2-fold SD of the average total sample counts ($\log_{10}$), which were regarded as dead cells or cell doublets in one gel bead. 1,518 differentially expressed genes were identified for trajectory study by comparing the transcriptomes among 4 groups of cells within the solid line in Figures S2E and S2F. Then, we used monocle to construct lineage trajectory and branch points by analyzing 1,518 DE genes. The state representing fibroblast was chosen as the root. Monocle constructed the final tree structure by using max_components parameter equal 5 for Figure 3A. As for Figure S3A, procedures are same as mentioned above but with top 2,000 DE genes from Stage I cell clusters and max_components equal 2. Gene expression heatmap in Figure 4A is based on lineage trajectory in Figure 3A and performed by "plot_genes_branched_heatmap" function of monocle 2. The core idea of this function is to merge all cells into total 200 bins in the order of pseudotime. Data of volcano plots in Figures 4B and S4B is extracted from heatmap matrix in Figure 4A.

PCA and *t*-SNE analysis

Principal component analysis (PCA) and *t*-SNE (*t*-distributed Stochastic Neighbor Embedding) for Figures 2B, 2D, 2E, S2E–S2I, S4G, and S6A were performed using the prcomp and Rtsne package of the R software (Version 3.4.1). For Figures 2B and S2E, we ran PCA using the expression matrix of the top 1,000 most variable genes selected from previous 1,518 DE genes. The number of principal components was determined from a scree plot. *t*-SNE was then performed on the first 10 principal components with default parameters to visualize cells in a two-dimensional space. All other figures follow the same procedure but with different parameters. And their parameters varies with the data.

Gene regulatory network analysis

The analysis of regulon activity for Figures 4E, S3C, and S4E were performed by following standard SCENIC pipeline. The input to SCENIC is an expression matrix, in which rows correspond to genes and columns correspond to cells. The cells in stage III (for Figures 4E and S4E) or terminal cells in two branches (for Figure S3C) were selected as the input cells. The genes with at least 186 UMIs across all cells and detected in at least 62 cells were selected as the input genes. The expression matrix was loaded into GENIE3 and the co-expressed genes to each TF was constructed. This step is time consuming and we ran it on high performance computing (HPC). The TF co-expression modules were then analyzed by RcisTarget. The Normalized Enrichment Score (NES) of the transcription factor binding motifs (TFBS) was calculated and NES > 3.0 were considered as significantly enriched. The filtered potential targets by RcisTarget mouse mm9 database from the co-expression module were used to build the regulons. The regulons activity (Area Under the Curve) was analyzed by AUCell and the active regulons are determined by AUCell default threshold. The active regulons were then mapped to SIII t-SNE for Figure 4E. Regulon-based t-SNE clustering analysis was performed for Figure S4E. Binary regulon activity (with any other regulon correlated with abs (correlation) > 0.30 and active in at least 30% of cells) was calculated by using terminal cells in two branches for Figure S3C.

The gene correlation network for Figure 4H was constructed by WGCNA analysis of the gene expression correlation. Genes were selected as the overlapping of the TFs in top 1,518 differentially expressed genes with intermediate cell programs (XEN-like, 2C-like and pluripotency programs) plus some known master regulators for intermediate cell types. The gene correlation network was visualized by Cytoscape.

Transcriptome analysis of Ci2C-like cells with reported datasets

The expression profile of mouse 2C::tdTomato cells (Macfarlan et al., 2012) for Figure S6C was obtained as raw sequencing reads from EBI database (GSM838738, GSM838739, GSM905088) and mapped to mouse mm10 using TopHat alignment software tools. The expression values of each gene were normalized to RPKM using Cufflinks. The expression profile of preimplantation embryo cells (Deng et al., 2014) for Figures 6A–6C was obtained as raw sequencing reads from EBI database (GSE45719) and mapped against mouse mm10 using STAR. The read counts were calculated using featureCounts. To compare gene expression profile among different datasets and platforms, the expression values from the published data and our own data were normalized by scaled to the expression of ESCs in each study for Figures 6A and S6C.

WGBS analysis

WGBS analysis was performed using Bismark/Bowtie2 algorithm. The clean fastq format reads were obtained from the sequencer and aligned to mouse mm10 reference with Bismark in directional mode using the default parameters. The deduplication tool from Bismark was used to remove duplicate reads. The methylation calls were extracted from the unique read alignment using bismark_methylation_extractor and bismark2bedGraph command. The methylation calls from both Watson and Crick strand were combined for the following analysis. The whole genome CpG methylation level for Figure 5H was calculated by bismark2summary command. The CpG methlation level for each region set for Figure 5I was performed by RnBeads R package. The visualization of each track was using IGV for Figure 5J.

Ontology Annotation

Gene ontology (GO) term enrichment analyses for Figures 3E, 4C, 4D, S4C, S4D, S4F, and S5G were performed using DAVID 6.8 functional annotation tool (Huang et al., 2009). Terms that had a P-value < 0.05 was defined as significantly enriched.

DATA AND SOFTWARE AVAILABILITY

The accession number for the RNA-seq data reported in this paper is NCBI GEO: GSE114952 or Beijing institute of Genomics (BIG) Data Center: CRA000932.