

# Single Cell Sequencing Reveals Early PGC-like Intermediates During Mouse Primed to Naïve Transition

Jing Liu (✉ [liu\\_jing@gibh.ac.cn](mailto:liu_jing@gibh.ac.cn))

Guangzhou Institutes of Biomedicine and Science, Chinese Academic and Science

<https://orcid.org/0000-0003-1600-7744>

**Shengyong Yu**

Guangzhou Institutes of Biomedicine and Health

**Chunhua Zhou**

Guangzhou Institutes of Biomedicine and Health

**Jiangping He**

Guangzhou Regenerative Medicine and Health-Guangdong Laboratory (GRMH-GDL)

<https://orcid.org/0000-0002-1425-3530>

**Xingnan Huang**

Guangzhou Institutes of Biomedicine and Health

**Zhaokai Yao**

School of Basic Medical Sciences, Southern Medical University

**Hong Zhu**

Guangzhou Institutes of Biomedicine and Science, Chinese Academic and Sciences

**Baomei Cai**

Center for Cell Lineage and Atlas, Bioland Laboratory

**Lizhan Xiao**

Guangzhou Institutes of Biomedicine and Health

**He Liu**

Center for Cell Lineage and Atlas, Bioland Laboratory

**Yue Qin**

Guangzhou Institutes of Biomedicine and Science, Chinese Academic and Sciences

**Jing Guo**

Guangzhou Institutes of Biomedicine and Health

**Yuhan Chen**

School of Basic Medical Sciences, Southern Medical University

**Yixuan Wang**

University of Science and Technology of China

**Shuyan Chen**

Guangzhou Institutes of Biomedicine and Science, Chinese Academic and Sciences

**Man Zhang**

Center for Cell Lineage and Atlas, Bioland Laboratory <https://orcid.org/0000-0002-4704-4356>

**Xiaoyang Zhao**

Southern Medical University <https://orcid.org/0000-0003-2544-8293>

**Jiekai Chen**

Guangzhou Institutes of Biomedicine and Health <https://orcid.org/0000-0001-5168-7074>

**Shangtao Cao**

Center for Cell Lineage and Atlas, Bioland Laboratory

**Duanqing Pei**

Chinese Academy of Sciences <https://orcid.org/0000-0002-5222-014X>

---

**Article**

**Keywords:** Single cell analysis, spermatogenesis

**Posted Date:** February 23rd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-151340/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Single Cell Sequencing Reveals Early PGC-like Intermediates During Mouse Primed to Naïve Transition

Shengyong Yu<sup>1,3,6</sup>, Chunhua Zhou<sup>1,6</sup>, Jiangping He<sup>2,6</sup>, Xingnan Huang<sup>1</sup>, Zhaokai Yao<sup>5</sup>, Hong Zhu<sup>1,3</sup>, Baomei Cai<sup>2</sup>, Lizhan Xiao<sup>1,3</sup>, He Liu<sup>2</sup>, Yue Qin<sup>1,3</sup>, Jing Guo<sup>1</sup>, Yuhan Chen<sup>5</sup>, Yixuan Wang<sup>1,4</sup>, Shuyan Chen<sup>1,3</sup>, Man Zhang<sup>2</sup>, Xiaoyang Zhao<sup>5</sup>, Jiekai Chen<sup>1,2</sup>, Shangtao Cao<sup>2,\*</sup>, Duanqing Pei<sup>1,2,\*</sup> and Jing Liu<sup>1,2,7,\*</sup>

## Affiliations:

1. CAS Key Laboratory of Regenerative Biology, South China Institutes for Stem Cell Biology and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Science, Chinese Academic and Sciences, Guangzhou 510530, China
2. Center for Cell Lineage and Atlas, Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory), 510005 Guangzhou, China
3. University of Chinese Academy of Science, Beijing, 100049, China
4. School of Life Sciences, University of Science and Technology of China, Hefei 230026, China
5. Department of Developmental Biology, School of Basic Medical Sciences, Southern Medical University, Guangzhou, 510515, China
6. These authors contributed equally to this work
7. Lead Contact

\*Correspondence to:

[cao\\_shangtao@grmh-gdl.cn](mailto:cao_shangtao@grmh-gdl.cn); [pei\\_duanqing@gibh.ac.cn](mailto:pei_duanqing@gibh.ac.cn); [liu\\_jing@gibh.ac.cn](mailto:liu_jing@gibh.ac.cn)

## Highlights:

- Single cell atlas for BMP4 driven Primed-Naïve transition.
- *c-Kit* marks the early naïve branch.
- A PGC-like program is essential for naïve cell fate.
- Gata2 knockout hampers placenta-like fate while facilitates PGC-like fate.

## SUMMARY

Single cell analysis provides clarity unattainable with bulk approaches. Here we apply single cell RNA-seq to a newly established BMP4 induced mouse primed to naive transition (Bi-PNT) system and show that the reset is not a direct reversal of cell fate but through developmental intermediates. We first show that mEpiSCs bifurcate into c-Kit<sup>+</sup> naïve and c-Kit<sup>-</sup> placenta-like cells, among which, the naïve branch undergoes further transition through a primordial germ cell-like cells (PGCLCs) intermediate capable of spermatogenesis in vivo. Indeed, deficiency of Prdm1/Blimp1, the key regulator for PGC specification, blocks the induction of PGCLCs and naïve cells. Instead, Gata2 knockout arrests placenta-like fate, but facilitates the generation of PGCLCs. Our results not only reveal a newly cell fate dynamics between primed and naïve states at single-cell resolution, but also provide a model system to explore mechanisms involved in regaining germline competence from primed pluripotency.

## INTRODUCTION

Pluripotency refers to the ability to generate all cell types in an individual. In mice, it has been well established that two distinct, yet interchangeable, states exist, namely naïve and primed pluripotency represented by ESCs (Embryonic Stem Cells) and EpiSCs (Epiblast Stem Cells)<sup>1-3</sup>. Despite the similarities between ESCs and EpiSCs in gene expression profile, in vitro differentiation potentials, they differ markedly in their morphology, ability to colonize blastocysts and conditions for maintenance in vitro<sup>1</sup>. From an embryological perspective, naïve and primed pluripotency appear to represent the developmental potentials of pre- and post- implantation embryos respectively<sup>1</sup>. Therefore, a detailed understanding of both states may inform our general understanding of early embryo development.

The interconvertibility between naïve and primed states has been well documented<sup>4-8</sup>. For example, naïve ESCs can be differentiated into primed state closely resembling EpiSCs<sup>9, 10</sup>. Likewise, EpiSCs have been converted to the naïve state similar to mESCs<sup>4, 5</sup>. Unlike the differentiation of ESCs to the primed state, the reprogramming

of EpiSCs towards naive state has been largely accomplished with the help of transcription factors or TFs<sup>5, 11-16</sup>, similar to the iPSC(Induced pluripotent stem cells) strategy pioneered by Yamanaka and colleagues<sup>17</sup>. TFs are powerful cell fate regulators that can overcome barriers established during development through multiple processes involving chromatin remodeling, metabolic switches, cell morphology changes. Recently, we have achieved a BMP4-dependent primed to naive transition (PNT) in a similar way as the reprogramming and show that a 3-day exposure of EpiSCs in BMP4, Dot1L inhibitor and EZH2 inhibitor followed by a 5-day culture in 2iL can convert the majority (~80%) of the primed cells into chimera-competent naive colonies<sup>18</sup>. Through RNA-seq and ATAC-seq bulk analyses, we demonstrated that many downstream targets of BMP4, including the previously unrecognized Zbtb7 family members, are activated through chromatin remodeling events for the resetting of primed to naïve states<sup>18</sup>. Even though, it is still unclear if there are intermediates governing the resetting of primed fate into a naive one by BMP4 or the precise cell dynamics during this resetting process, as distinct molecular routes have been shown to mediate the acquisition of naïve pluripotency with different TFs and signals<sup>19</sup>. In this report, by taking advantage of single cell RNA-sequencing (scRNA-seq), we report the cell fate continuum between primed and naive states with intermediates previously masked in bulk analysis. Specifically, our results reveal a Prdm1-dependent and germline competent PGC-like cell state mediate the transition of PNT, which provide new insight into the acquisition of naïve pluripotency and germline from primed pluripotency.

## **RESULTS**

### **Single cell atlas of BMP4 induced primed to naive transition (BiPNT)**

We have established a robust BMP4-dependent primed to naive transition (PNT) system, which consisting of two stages, a 3-day incubation with BMP4, EPZ5676, EPZ6438 and 5-day culturing with 2iL<sup>18</sup>, that can convert ~80% EpiSCs to chimera competent naive colonies. To map the cell fate continuum, we performed scRNA-seq from cells that are at D0, 1, 2, 3, 5 and 8 (Fig. 1a). Analysis with t-SNE methods yielded well segregated clusters for cells for each day (Fig. 1b), indicating distinct stages of

cell fate during PNT. Indeed, distinct gene expression profiles can be observed from D0 to D8 using selective genes from primed to naive states (Fig. 1c).

To further identify the cell population at two stages, we firstly analyzed single cells from D0-3 at stage 1 and showed that many cells emerged at D3 have naive pluripotency program genes (Fig. 1d). Among them, *Klf2*, *Dppa5a*, and *Dppa3* are almost exclusively expressed in D3 cluster (Fig. 1d, lower panels). Interestingly, we also found that some cells in D3 also have a placenta-like program, expressing genes like *Plac1*, *Gata2*, *Krt8* and *Peg10* (Fig. 1d, middle panels). These results suggest that mEpiSCs undergo fate transitions towards placenta-like and naïve-like pluripotency cells at the stage1.

2iL is a well-known condition for maintaining naive pluripotent state<sup>20</sup>. When the BMP4-treated cells were switched to 2iL condition, it is expected that they can be further matured into naive pluripotency. Indeed, scRNA-seq reveals a wide spread presence of naive pluripotency program among D5 and D8 cells (Fig. 1e), suggesting that 2iL has further reprogrammed the cells in D3 into naive state. We can also observe minor populations in both D5 and D8 cultures that are placenta-like (Fig. 1e).

We then plotted the cells with naive- and placenta-like programs to show that BMP4 triggers almost equal numbers of cells for each fate up to D2 when the naive fate continues to increase while the placental fate levels off (Fig. 1f). It is apparent that 2iL continues to support the increase of the naive fate as expected but begins to diminish the placental ones at D5 and D8 (Fig. 1f). Therefore, we can conclude that two primary fates emerge in BMP4-treated mEpiSCs and 2iL favors the naive fate while partially suppresses the placenta-like one (Fig. 1f). *Gata2*, which is highly expressed in placental cells but not in naïve pluripotent cells, plays an important role for the development of trophoblast<sup>21, 22</sup>. In order to trace both naïve- and placenta-like cells, we constructed an EpiSC cell line carrying the *Oct4*-GFP/*Gata2*-Tdtomato double transgene reporters (Extended Data Fig. 1a). We found that the naive pluripotent reporter *Oct4*-GFP is incompatible with placental signaling *Gata2*-Tdtomato at the late stage of PNT (Fig.1g and Extended Data Fig. 1b ). These results suggest that BMP4 triggers primed EpiSCs to bifurcate into two distinct fates towards naïve-like or placenta-like cells.

### **c-Kit marks the naive branch**

To further resolve the apparent naive versus placental fate choices in D2 and D3, we re-analyzed the scRNA-seq data with Harmony<sup>23</sup>, a program designed to resolve branches or trajectories along cell fate continuum. Harmony resolves the cells into two main branches, the placental branch (PB) or naïve branch (NB) as expected (Fig. 2a). While the PB is enriched with imprinting genes such as *Igf2*, *Peg10*, *H19* and placental genes *Gata2*, *Tead3*, the NB expresses *Dppa5a*, *Klf2* and *Nanog*, mostly naive-related genes (Fig. 2b). Consistent with the naive- and placenta-dichotomy, the PB is enriched with genes from placenta-like fate while the NB with those from the naive-like fate (Fig. 2c).

We then searched for cell surface markers that may distinguish the NB from the PB. Among a group of cell surface marks (Extended Data Fig. 2a), we found that *c-Kit* is a good candidate (Fig. 2d). *c-Kit* begins to be expressed at D1 and persists in the NB significantly (Fig. 2d). When plotted along the pseudo time, it becomes very clear that *c-Kit* begins to diverge at early phase of PNT that the NB maintains its expression while the PB diminishes (Fig. 2e), suggesting that *c-Kit* may help mark the NB. To test this idea, we sorted D3 cells with *c-Kit*-APC and show that ~ 50% of D3 cells are *c-Kit* positive (*c-Kit*<sup>+</sup>) (Fig. 2f). We then replated the *c-Kit*<sup>+</sup> and *c-Kit*<sup>-</sup> cells separately and continued their culturing under 2iL to show that almost all (94.2%) of *c-Kit*<sup>+</sup> cells become GFP positive compared to 0.7% for *c-Kit*<sup>-</sup> cells (Fig. 2g). We further performed RNA-seq to probe the difference between *c-Kit*<sup>+</sup> and *c-Kit*<sup>-</sup> population (Fig. 2h), and showed that 1102 genes are highly expressed in *c-Kit*<sup>-</sup> cells, including the placenta markers *Plac1*, *Gata2* and *Krt8*, while 1065 genes in *c-Kit*<sup>+</sup> cells including naïve markers *Pou5f1*, *Sox2*, *Nanog*, *Esrrb*, *Klf2*, mesoderm marker *T* and primordial germ cells (PGCs) marker *Prdm1* (Fig. 2h). We then performed qRT-PCR analysis to further confirm the expression of these naïve pluripotent markers in *c-Kit*<sup>+</sup> cells and placenta markers in *c-Kit*<sup>-</sup> cells (Extended Data Fig. 2b). Consistently, GO (gene ontology) analysis reveals that programs such as embryonic placenta development, endothelial cell migration and placenta development are highly enriched in *c-Kit*<sup>-</sup> cells, while those involved in gastrulation, BMP signaling pathway, non-canonical Wnt signaling pathway are highly enriched in *c-Kit*<sup>+</sup> cells (Fig. 2i). Intriguingly, since the PGC markers such as

*T* and *Prdm1* are specifically enriched in *c-Kit*<sup>+</sup> (Fig. 2h), we further examined the expression of other PGC related markers by qRT-PCR and showed that *Dppa3*, *Nanos3*, *Ifitm1*, *Ifitm3* are also highly expressed in the *c-Kit*<sup>+</sup> cells (Fig. 2j). We further performed ATAC-seq to investigate the chromatin open or close state between the *c-Kit*<sup>+</sup> or *c-Kit* population as we described before<sup>24, 25</sup>. We identified 16806 peaks specifically opened in *c-Kit* cells, among which are gene loci such as *Gata2*, *Plac1*, *Elf5*, *Krt18* and *Igf2* (Placenta markers), (Extended Data Fig. 2c,d), and 13048 peaks specifically opened in *c-Kit*<sup>+</sup> cells, including gene loci such as *Prdm1*, *Prdm14*, *Nanos3* (PGC markers) and *Klf2*, *Nanog*, *Nr0b1* and *Nr5a2* (Naïve markers) ( and Extended Data Fig. 2c). Motif discovery from those open peaks indicates that, motifs from AP-1 family members, such as JunB, Fra1, TEAD family members such as TEAD4, TEAD2, are specifically enriched in *c-Kit* cells, while motifs for OCT6/4/2, SOX9/3/2/17, ZIC3 in *c-Kit*<sup>+</sup> cells (Extended Data Fig. 2d, e). Together, these results suggest that *c-Kit* marks the naive branch quite early among BMP4-treated cells.

### **Activation of a PGC-like program during Bi-PNT**

Since the *c-Kit*<sup>+</sup> cells are also enriching for PGC markers (Fig. 2j), we further examined the detailed cell fate continuum along the naive branch. We generated a pseudo time plot of NB during PNT and found it can be divided into 4 distinct phases (I-IV) (Fig. 3a). The first phase (I) is mainly cells expressing genes of the primed fate, such as *Otx2*, *Dnmt3b* and *Fgf5*. The second phase (II) includes cells that appear to express genes enriched for mesoderm like, PGC precursor cells (Aramaki, 2013), such as *Msx1/2*, *Cdx2*, *T* and *Prdm1*. The third phase (III) are cells expressing genes enriched in an early PGC-like fate, such as *Prdm14*, *Tfap2c*, *Dppa3*, *Nanos3* and *Dnd1*. And the last phase (IV) contains cells with naive features, i.e., with *Nanog*, *Esrrb*, *Sox2*, *Klf2* and *Zfp42* (Fig. 3a). Consistently, heat map generated with Harmony also indicates a continuous cell fate transition from primed pluripotency, PGC precursor, early PGC, then end with naïve pluripotency along the NB (Fig. 3b and Extended Data Fig. 3a). To distinguish the putative PGC-like cells (PGCLCs) from naïve pluripotent cells, we detected the expression of KLF4 in *Oct4*-GFP positive cells by immunofluorescence during BiPNT at Day6 and Day8<sup>26</sup> (Fig. 3c). Since PGC and

naïve pluripotent cells are both positive for *Oct4*-GFP, the emergence of GFP<sup>+</sup>/KLF4<sup>+</sup> cells suggest the existence of a PGCLC fate before reaching to naïve pluripotency during BiPNT (Fig. 3c). We further used the PGC surface markers SSEA1 and CD61 to probe the PGCLC with fluorescence activated cell sorting (FACS) and showed that 16.1% cells are SSEA1<sup>+</sup>/CD61<sup>+</sup> at biPNT day6 (Fig. 3d). The generation of PGCLC at the second stage of PNT is surprisingly as 2iL medium is well-known to mediate the induction and maintenance of naïve pluripotency. We further investigated the effects of the individual component in 2iL to the expression of PGC and naïve pluripotency markers and showed by qRT-PCR that both CHIR and PD are required for the activation of PGC genes such as *Prdm14*, *Dppa3* and *Dnd1*, and naïve pluripotent genes *Klf4* and *Tbx3*, though CHIR is dispensable for the expression of *Prdm1*, while LIF is mainly responsible for the expression of naïve pluripotent genes such as *Klf4* and *Tbx3* (Extended Data Fig. 3b). These results are consistent with previous study that ERK inhibition can induced an incomplete PGCLC cell fate and LIF/Stat3 is required for naïve pluripotency<sup>27, 28</sup>. To further investigate the PGCLC emerged in BiPNT, we generated EpiSC cell line from a ESC bearing *Blimp1*-mVenus(BV) and *Stella*-ECFP(SC) transgenes to mimic the PGCLC fate<sup>29</sup>. FACS analysis and fluorescence microscope showed that, the expression of *Prdm1/Blimp1* was significantly activated when stimulated with BMP4, and reached to a peak at 89% at Day1, while the expression of *Stella* was specifically upregulated when changed to 2iL medium (Fig. 3e). Furthermore, the percentage of double positive for BVSC (BV<sup>+</sup>SC<sup>+</sup>) cells reached to a peak at 13% at Day6 (Fig. 3e and Extended Data Fig. 3c), indicating an efficient generation of PGCLC during BiPNT. In addition, inhibition of JAK/STAT3 signaling in 2i medium increases the BV<sup>+</sup>SC<sup>+</sup> cells significantly (Fig. 3f), suggesting that inhibition of LIF/STAT3 retains the PGCLCs while restricts the derivation of naïve pluripotency. We further tested the GK15 medium which used for PGC specification previously<sup>30, 31</sup> for its ability in PGCLC induction during BiPNT, and showed that when switching to GK15 medium with or without cytokines at Day3, BV<sup>+</sup>SC<sup>+</sup> cells emergent in both condition, albeit at lower efficiency (GK15~4.32%, GK15+cytokines~9.52%) (Extended Data Fig. 3d), indicating that the Day3 cells have acquired the potential for germline specification. We then collected the BV<sup>+</sup>SC<sup>+</sup> cells at Day6 by FACS and

compared their expression of PGC or pluripotency markers with EpiSCs and ESCs. Comparing to EpiSCs and ESCs, the BV<sup>+</sup>SV<sup>+</sup> cells express higher levels of PGC markers such as *Prdm1*, *Prdm14*, *Tfap2c*, *Dppa3*, but lower levels of naïve pluripotency markers such as *Klf4* and *Tbx3*, while both of the BV<sup>+</sup>SC<sup>+</sup> cells and ESCs express the pluripotency genes *Oct4*, *Nanog* and *Dppa5a* (Fig. 3g and Extended Data Fig. 3e). Immunofluorescence analysis further confirmed the absent expression of KLF4 in BV<sup>+</sup> cells (Extended Data Fig. 3f). These data indicated a PGCLC state may exist during BiPNT.

### Characteristics of BV<sup>+</sup>SC<sup>+</sup> PGCLCs

To further characterize the Day6 BV<sup>+</sup>SC<sup>+</sup> PGCLC cells, we performed RNA-seq experiments and compared our RNA-seq data to other published datasets containing Day4/Day6 in vivo PGCLC or primary PGC (E9.5,11.5,13.5)<sup>32, 33</sup>. PCA analysis revealed that the transcriptomes of the BV<sup>+</sup>SC<sup>+</sup> cells are similar to D4 and D6 PGCLCs, and to a lesser extent of E9.5 PGCs (Fig. 4a). We then evaluated the epigenetic profile of Day6 BV<sup>+</sup>SC<sup>+</sup> cells. IF analysis revealed a reduced H3K9me2 and increased H3K27me3 (Fig. 4b), which was further confirmed by western blot (Fig. 4c), and consist with previous report<sup>31</sup>. We further determined the imprinting states of maternally (*Snrpn*) and paternally (*H19*) imprinted genes in Day6 BV<sup>+</sup>SC<sup>+</sup> cells and showed that whereas the DNA methylation in *Snrpn* loci was retained, the methylation of *H19* loci reduced significantly, suggesting that Day6 BV<sup>+</sup>SC<sup>+</sup> cells may undergo the process of imprint erasure (Fig. 4d). Consistently, qRT-PCR analysis also showed the downregulation of de novo DNA methyltransferases *Dnmt3a/b* as well as *Uhrf1*, a critical regulator which recruits *Dnmt1* to hemi-methylated DNA (Extended Data Fig. 4a). We further performed blastocysts microinjection experiments for the sorted BV<sup>+</sup>SC<sup>+</sup> cells to determine the chimeras formation ability of those cells (Extended Data Fig. 4c). Notably, BV<sup>+</sup>SC<sup>+</sup> cells cultured in 2iL medium for one passage showed a quick loss of *Blimp1* expression and can further form naïve-like colonies (Extended Data Fig. 4b). In contrast, Day6 sorted BV<sup>+</sup>SC<sup>+</sup> cells without passage showed a poor chimeras formation ability (12.5% versus 66.7%) (Extended Data Fig. 4c), consistent with previous observations that PGC cannot contribute to chimeras, despite

expressing naïve pluripotency genes<sup>26, 34, 35</sup>. To further determine the capacity of the Day6 sorted BV<sup>+</sup>SC<sup>+</sup> cells in spermatogenesis, we transplanted them into the seminiferous tubules of neonatal *W/W<sup>v</sup>* mice lacking endogenous germ cells<sup>31, 36, 37</sup>. Notably, three months after transplantation, tubules with normal spermatogenesis, which express the spermatogonia marker (DDX4 and PLZF), spermatocytes markers (SYCP3 and γH2AX) and spermatids markers (DDX4 and PNA) can be detected in two out of six testes sections transplanted with BV<sup>+</sup>SC<sup>+</sup> cells (Fig. 4e, f and Extended Data Fig. 4d), while only Sertoli cells in the tubules of control group. Together, these data indicated that the Day6 BV<sup>+</sup>SC<sup>+</sup> cells are germline competent PGC-like cells.

### ***Prdm1*-KO blocks the generation of PGCLC and naïve pluripotency fate**

To test whether the PGCLCs are intermediates for the acquisition of naïve pluripotency, we generated OG2-EpiSC deficient in *Prdm1* (*Prdm1*-KO or *Prdm1*<sup>-/-</sup> mEpiSC) (Extended Data Fig. 5a,b), a factor which is obligatory for PGC specification, but not for the derivation of pluripotency<sup>38, 39</sup>. *Prdm1*<sup>-/-</sup> EpiSC exhibited indistinguishable morphology from wildtype (WT) EpiSC and expressed comparable levels of primed pluripotency markers such as *Oct4*, *Sox2* and *Otx2* (Fig. 5a and Extended Data Fig. 5c). However, when undergoing BiPNT, no GFP<sup>+</sup> cells can be generated from *Prdm1*<sup>-/-</sup> cells, while expression of exogenous *Prdm1* can rescue this defect (Fig. 5b and Extended Data Fig. 5d). Furthermore, the *Prdm1*<sup>-/-</sup> cells showed upregulation of somatic marker *Hoxa1* and *Hoxb1* (Extended Data Fig. 5e). In addition, they failed to activate PGC specific gene *Nanos3* and naïve pluripotent markers (*Klf2*, *Nr5a2*), while increasing expression of placenta related markers (*Plac1*, *Peg10*) (Extended Data Fig. 5f). These data indicated that *Prdm1* is required for generating both PGCLC and naïve pluripotency during BiPNT.

To further investigate the role of *Prdm1* in the cell trajectory during BiPNT, we performed single cell analysis. To this end, we collected scRNA-seq data for *Prdm1*<sup>-/-</sup> EpiSCs undergoing BiPNT at corresponding time point as wildtype (WT) EpiSC (Fig. 5c). By comparing WT and *Prdm1*<sup>-/-</sup> samples using UMAP, we show that the separation between WT and *Prdm1*<sup>-/-</sup> populations mainly occurs at Day2-3 (Fig. 5d). We then analyzed the expression of cell fate markers for PGC, placenta, or naïve

pluripotency by UMAP in these populations (Fig. 5e), and show that: 1) PGC markers are mainly present in subpopulations of WT samples, especially *Nanos3* and *Dppa3* (Fig. 5e, the first row); 2) the naïve markers are exclusively expressed in the subgroup of WT subpopulation but not in *Prdm1*<sup>-/-</sup> (Fig. 5e, the second row); 3) the placenta markers, such as *Plac1*, *Igf2*, *Ahnak* and *Peg10*, are in fact more enriched in the subpopulation of *Prdm1*<sup>-/-</sup> samples compared to the WT (Fig. 5e, the bottom row). These data demonstrated at single cell resolution that when loss of *Prdm1*, EpiSC failed to enter a PGCLC fate and naïve pluripotent state but retained a placenta-like fate. In addition, we examined the DNA methylation state of the naïve-like cells (Two naïve colonies: Naïve-1# and 2#) and showed that they had loss the DNA methylation in the imprinted loci *Peg1*, *Peg3β* and *Snrpn* (Extended Data Fig. 5g). Therefore, these data suggest that *Prdm1*-dependent PGCLCs are an intermediate for successful BiPNT.

### **GATA2 gates placenta-like program whereas hampers PGCLC program**

To determine the key event for early cell fate choice during BiPNT, we reanalyzed the scRNAseq data at D0, D1 and D2 by UMAP (). We found a significant cell diversity at D2 (Fig. 6a). Heatmap for representative genes show an activation of placenta signaling (*Gata2*, *Plac1*) as early as Day1 (Fig. 6b), and a diversity of pluripotent signaling (*Dppa3*, *Nanog*) at Day2 (Fig. 6b), suggesting an early cell fate choice. To further investigate the key factor(s) regulate this cell fate choice, we use pySCENIC<sup>40</sup> to predict the Regulon of each cell subgroup (Extended Data Fig. 6a). Heatmap showed a widely activation of early BMP4 response regulons, such as *Id1*, *Cdx2*, *Tfap2c*, and a separate pattern for regulons *Prdm1* (PGC) and *Gata2* (Placenta) at D2 populations (Fig. 6c), corresponding to the final placenta-like trajectory and naïve pluripotent trajectory. Consistently, the expression of *Gata2* and trophoblast marker *Elf5* was restricted to the placenta branch (Extended Data Fig. 6b). The *Prdm1*-KO almost totally abolish PNT, indicating *Prdm1* as a key Regulon for PGC-Naïve trajectory. To test whether *Gata2* is a regulon specific for placenta trajectory, we generated two independent *Gata2*-KO EpiSC cell lines in BVSC EpiSC with Crispr/cas9 (Extended Data Fig. 6c, d). These two cell lines express comparable levels

of primed pluripotency markers ([Extended Data Fig. 6e](#)). When performing BiPNT, FACS analysis showed a significant increasing of BV<sup>+</sup> cells at Day3 and BV<sup>+</sup>SC<sup>+</sup> cells at Day6 in *Gata2*-KO groups than in WT during BiPNT ([Fig. 6d](#)). Consistently, qRT-PCR analysis further shows a failed activation of placenta signaling such as *Plac1*, *Peg10* and *Phlda2*, but a significant increasing in PGC-like signaling such as *Prdm1*, *Prdm14*, *Dppa3* and *Nanos3* in *Gata2*-KO cells comparing to WT cells ([Fig. 6e](#)). These data suggested that *Gata2*-KO turns off the placenta-like program while facilitates the PGC-like program.

## DISCUSSION

In this report, by performing scRNA-seq to a previously well-established PNT process driven by BMP4 (BiPNT)<sup>18</sup>, we generated a cell fate continuum between primed and naïve pluripotency states. Surprisingly, we found that BiPNT, in single cell level, goes through cell fate stages previously unexpected, i.e., a *Prdm1*<sup>+</sup> PGC-like stage, before the establishment of naïve pluripotent cell fate. We also identified several key regulators, *Prdm1/Blimp1*, a primordial germ cell (PGC) determinant, whose depletion totally blocks the derivation of naïve pluripotency; *Gata2*, a transcription factor necessary for regulation of trophoblast program, whose depletion turns off the placenta-like program while facilitates the PGC-like program ([Fig. 7](#)). Interestingly, through re-analyzing our published data, we found that *Zbtb7a/b*, the novel targets of BMP4<sup>18</sup>, also regulated the expression of some PGC genes during PNT, such as *c-kit*, *Prdm1* and *Nanos3*, in RNA and chromatin level ([Extended Data Fig. 7a, b](#)). Our data showed a unique path for resetting the stable mouse primed pluripotent state into a naïve pluripotent one by linking several quite important developmental intermediate stages therefore sheds light on the mechanism of cell reprogramming and differentiation.

*Prdm1* has been demonstrated to be dispensable for the derivation and maintenance of ESC as well as the reprogramming of EpiSC into rESC<sup>39, 41</sup>. The obligatory role of *Prdm1* in our study may due to the differences of induction systems and may further underline the PGCLC intermediates for recapturing naïve pluripotency in BiPNT. Actually, the BV<sup>+</sup>SC<sup>+</sup> cells induced during BiPNT exhibited PGC features, including

upregulation of early PGC markers, increase of H3K27me3 and reduction of H3K9me2, erasure of imprinting, poor contribution to blastocyte chimeras and contribution to spermatogenesis, indicating their PGC-like cells fate.

In mice, permissiveness for PGC specification is unique to formative epiblast cells (i.e., E5.0-E6.0 epiblast, EpiLCs) or PSCs with formative pluripotency<sup>31, 42-44</sup>. The primed EpiSCs, which corresponding to E6.5 epiblast, have lost the competence of PGC responsiveness (Ref). Nevertheless, in the BMP4 driven PNT system, a PGC-like program was robustly activated from the primed EpiSCs. In addition, this process seems to recapitulate the PGC specification pathway, which firstly enters a T and Prdm1 double positive PGC precursor and then activates the PGC-like program (Fig. 3a and Fig. 7). One explanation for this paradox may be the use of epigenetic inhibitors including EPZ5676, a DOT1L inhibitor, and EPZ6438, an EZH2 inhibitor in the BiPNT system<sup>18</sup>, which may reduce the epigenetic barrier(s) for cell fate transition. Consistent with this hypothesis, Hans R Schöler and colleagues previously showed that reducing DOT1L-mediated H3K79/K27 methylation endows reprogramming competence of almost all OCT factors during human somatic cell reprogramming<sup>45</sup>. Therefore, the further studies involved in EpiSCs to PGC-like cells induction may improve our understanding of the epigenetic mechanism governing PGC competence.

In this study, we also identified another cell bifurcation which expresses the trophoblast associated genes such as *Gata2*, *Plac1*, *Peg10* during BiPNT, and we termed it placenta-like cells. Indeed, BMP4 has been shown to induce the expression of genes associated with trophoblast in EpiSC and primed human ESC<sup>2, 3, 46, 47</sup>. However, studies also proposed that the BMP4-induced cells in EpiSC and hESC may actually correspond to a mesoderm identity<sup>48</sup>. Intriguingly, the placenta-like cells induced during BiPNT express *Elf5*, a key transcription factor in trophoblast lineage determination (Extended Data Fig. 6b), suggesting that these cells may be trophoblast like cells. It would be quite interesting to investigate whether the placental-like cells can function beyond what we attempted here.

## ACKNOWLEDGEMENTS

We thank M. Saitou for BV<sup>+</sup>SC<sup>+</sup> ESC cell lines. The work was supported by grants from National Natural Science Foundation of China (31421004, 31401264, 31530038, 31461143011, 91419310, 21907095), National Basic Research Program of China (2014CB965200), The national Key Research and Development Program of China, Stem cell and Translational Research (2016YFA0101800), Science and Technology Planning Project of Guangdong Province (2015B020228003, 2014B020225002, 2014B030301058, 2014B05052012, 2014B050504008, 2019A1515011024). Bureau of Science and Technology of Guangzhou Municipality (Grants 2014J2200077).

## AUTHOR CONTRIBUTIONS

J.L., S.Y., designed the project, S.Y., C.Z., S.C. performed the experiments, J.H., X.H. analyzed the data. S.Y. and C.Z. isolated the primary EpiSC cell lines. S.C. and J.G. performed the blastocyst injection, H.L. and Q.Y. performed the scRNA-seq experiments, H.Z. and B.C. performed the gene editing and qRT-PCR experiments. Z.Y., and Y.C., performed the PGCLC transplantation experiments. M.Z., and X.Z., supervised the germ cell experiments. J.C. supervised the data analysis. J.L. and D.P. supervised the whole study, wrote the manuscript, and approved the final version.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

## REFERENCES

1. Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4**, 487–492 (2009).
2. Tesar, P.J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196–199 (2007).
3. Brons, I.G. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191–195 (2007).
4. Bao, S. *et al.* Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. *Nature* **461**, 1292–1295 (2009).
5. Guo, G. *et al.* Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* **136**, 1063–1069 (2009).
6. Guo, G. & Smith, A. A genome-wide screen in EpiSCs identifies Nr5a nuclear receptors as potent inducers of ground state pluripotency. *Development* **137**, 3185–3192 (2010).

7. Li, W. & Ding, S. Converting mouse epiblast stem cells into mouse embryonic stem cells by using small molecules. *Methods Mol Biol* **1074**, 31–37 (2013).
8. Weinberger, L., Ayyash, M., Novershtern, N. & Hanna, J.H. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat Rev Mol Cell Biol* **17**, 155–169 (2016).
9. Hanna, J. *et al.* Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell Stem Cell* **4**, 513–524 (2009).
10. Najm, F.J. *et al.* Isolation of epiblast stem cells from preimplantation mouse embryos. *Cell Stem Cell* **8**, 318–325 (2011).
11. Festuccia, N. *et al.* Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. *Cell Stem Cell* **11**, 477–490 (2012).
12. Okashita, N. *et al.* PRDM14 Drives OCT3/4 Recruitment via Active Demethylation in the Transition from Primed to Naive Pluripotency. *Stem Cell Reports* **7**, 1072–1086 (2016).
13. Pastor, W.A. *et al.* TFAP2C regulates transcription in human naive pluripotency by opening enhancers. *Nat Cell Biol* **20**, 553–564 (2018).
14. Stuart, H.T. *et al.* NANOG amplifies STAT3 activation and they synergistically induce the naive pluripotent program. *Curr Biol* **24**, 340–346 (2014).
15. Tai, C.I. & Ying, Q.L. Gbx2, a LIF/Stat3 target, promotes reprogramming to and retention of the pluripotent ground state. *J Cell Sci* **126**, 1093–1098 (2013).
16. Ye, S., Li, P., Tong, C. & Ying, Q.L. Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1. *EMBO J* **32**, 2548–2560 (2013).
17. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
18. Yu, S. *et al.* BMP4 resets mouse epiblast stem cells to naive pluripotency through ZBTB7A/B-mediated chromatin remodelling. *Nat Cell Biol* (2020).
19. Stuart, H.T. *et al.* Distinct Molecular Trajectories Converge to Induce Naive Pluripotency. *Cell Stem Cell* **25**, 388–406 e388 (2019).
20. Ying, Q.L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).
21. Krendl, C. *et al.* GATA2/3–TFAP2A/C transcription factor network couples human pluripotent stem cell differentiation to trophectoderm with repression of pluripotency. *Proc Natl Acad Sci U S A* **114**, E9579–E9588 (2017).
22. Ng, Y.K., George, K.M., Engel, J.D. & Linzer, D.I. GATA factor activity is required for the trophoblast-specific transcriptional regulation of the mouse placental lactogen I gene. *Development* **120**, 3257–3266 (1994).

23. Nowotschin, S. *et al.* The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* **569**, 361–367 (2019).
24. Cao, S. *et al.* Chromatin Accessibility Dynamics during Chemical Induction of Pluripotency. *Cell Stem Cell* **22**, 529–542 e525 (2018).
25. Li, D. *et al.* Chromatin Accessibility Dynamics during iPSC Reprogramming. *Cell Stem Cell* **21**, 819–833 e816 (2017).
26. Murakami, K. *et al.* NANOG alone induces germ cells in primed epiblast in vitro by activation of enhancers. *Nature* **529**, 403–407 (2016).
27. Kimura, T. *et al.* Induction of primordial germ cell-like cells from mouse embryonic stem cells by ERK signal inhibition. *Stem Cells* **32**, 2668–2678 (2014).
28. Leitch, H.G. *et al.* Rebuilding pluripotency from primordial germ cells. *Stem Cell Reports* **1**, 66–78 (2013).
29. Ohinata, Y., Sano, M., Shigeta, M., Yamanaka, K. & Saitou, M. A comprehensive, non-invasive visualization of primordial germ cell development in mice by the Prdm1-mVenus and Dppa3-ECFP double transgenic reporter. *Reproduction* **136**, 503–514 (2008).
30. Sasaki, K. *et al.* Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells. *Cell Stem Cell* **17**, 178–194 (2015).
31. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532 (2011).
32. Yamaguchi, S. *et al.* Tet1 controls meiosis by regulating meiotic gene expression. *Nature* **492**, 443–447 (2012).
33. von Meyenn, F. *et al.* Comparative Principles of DNA Methylation Reprogramming during Human and Mouse In Vitro Primordial Germ Cell Specification. *Dev Cell* **39**, 104–115 (2016).
34. Hackett, J.A., Kobayashi, T., Dietmann, S. & Surani, M.A. Activation of Lineage Regulators and Transposable Elements across a Pluripotent Spectrum. *Stem Cell Reports* **8**, 1645–1658 (2017).
35. Leitch, H.G. *et al.* On the fate of primordial germ cells injected into early mouse embryos. *Dev Biol* **385**, 155–159 (2014).
36. Chuma, S. *et al.* Spermatogenesis from epiblast and primordial germ cells following transplantation into postnatal mouse testis. *Development* **132**, 117–122 (2005).
37. Nakaki, F. *et al.* Induction of mouse germ-cell fate by transcription factors in vitro. *Nature* **501**, 222–226 (2013).
38. Ohinata, Y. *et al.* Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* **436**, 207–213 (2005).
39. Bao, S. *et al.* The Germ Cell Determinant Blimp1 Is Not Required for Derivation of Pluripotent Stem Cells. *Cell Stem Cell* **11**, 110–117 (2012).
40. Van de Sande, B. *et al.* A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc* **15**, 2247–2276 (2020).

41. Gillich, A. *et al.* Epiblast stem cell-based system reveals reprogramming synergy of germline factors. *Cell Stem Cell* **10**, 425–439 (2012).
42. Yu, L. *et al.* Derivation of Intermediate Pluripotent Stem Cells Amenable to Primordial Germ Cell Specification. *Cell Stem Cell* (2020).
43. Kinoshita, M. *et al.* Capture of Mouse and Human Stem Cells with Features of Formative Pluripotency. *Cell Stem Cell* (2020).
44. Ohinata, Y. *et al.* A signaling principle for the specification of the germ cell lineage in mice. *Cell* **137**, 571–584 (2009).
45. Kim, K.P. *et al.* Permissive epigenomes endow reprogramming competence to transcriptional regulators. *Nat Chem Biol* (2020).
46. Li, Y. & Parast, M.M. BMP4 regulation of human trophoblast development. *Int J Dev Biol* **58**, 239–246 (2014).
47. Xu, R.H. *et al.* BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nat Biotechnol* **20**, 1261–1264 (2002).
48. Bernardo, Andreia S. *et al.* BRACHYURY and CDX2 Mediate BMP-Induced Differentiation of Human and Mouse Pluripotent Stem Cells into Embryonic and Extraembryonic Lineages. *Cell Stem Cell* **9**, 144–155 (2011).
49. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
50. Lun, A.T., McCarthy, D.J. & Marioni, J.C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* **5**, 2122 (2016).
51. Wolf, F.A., Angerer, P. & Theis, F.J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15 (2018).
52. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902 e1821 (2019).

## Figure legends

### Figure 1. Single cell analysis for BiPNT

- (a) Flow chart for the analysis of BMP induced Primed-Naïve Transition (BiPNT). scRNA-seq analysis was adopted from 10x Genomics.
- (b) t-SNE projection of all 50406 individual cells during the whole PNT process.
- (c) Expression of genes in 2 different categories at the indicated time points.
- (d) t-SNE projection of cells and typical gene expressions in stage I.
- (e) t-SNE projection of cells and typical gene expressions in stage II.
- (f) Percentages of naive cells and placenta-like cells at different time points.
- (g) Representative images of *Oct4*-GFP/*Gata2*-Tdtomato EpiSC induced PNT at

day5 and day8. Scale bars, 100  $\mu$ m.

### **Figure 2. c-Kit is a Key marker for successful BiPNT**

- (a) Trajectory reconstruction of all single cells throughout PNT, colored by indicated time points (left). The bifurcating branches defined as NB (naïve branch) and PB (placenta branch) separately. Circles indicate cells at the terminus of the NB and PB, respectively (right).
- (b) Identification of the differentially expressed genes between NB and PB branches.
- (c) Violin plot displaying the expression of representative naive pluripotency genes and placental genes in indicated cell types.
- (d) Plots showing the expression patterns of cell surface marker *c-kit* in NR and RP branch.
- (e) Expression trends of *c-kit* along pseudo-time during NB and PB branch specification.
- (f) c-Kit positive and negative cells were separated at day3 of PNT.
- (g) Images of *Oct4*-GFP colonies (upper panel) and FCAS result (lower panel) for c-Kit positive and negative cells sorted in (f) and re-plated in stage2 medium(2iL) for another 5 days.
- (h) Heat maps showing gene expression differences between c-Kit positive and negative cell samples by RNA-seq. Representative genes are shown on the right.
- (i) Go analysis for differentiated expression genes at (h).
- (j) Expression of primordial germ cell (PGC) markers measured by Q-PCR in day3 cells of PNT (unsorting), c-Kit positive and negative cells after sorting. Data are mean  $\pm$  s.d., n=2 independent experiments.

### **Figure 3. Activation of PGC-like program during successful BiPNT**

- (a) Gene expression heatmap of 3219 dynamics expressed genes along NB branch in a pseudo-temporal order. Representative genes are shown on the right.
- (b) Plots showing the expression patterns of the indicated genes during BiPNT.
- (c) Immunofluorescence analysis of KLF4 in OCT4-GFP positive cells at day 6 and 8

of PNT. Scale bars, 100  $\mu$ m.

- (d) FACS analysis for surface expression of SSEA1 and CD61 at day 6 of PNT.
- (e) FACS analysis of BVSC induction at distinct time points during PNT.
- (f) FACS analysis of BVSC induction at day6 induced in 2i medium with 5 $\mu$ M Jak inhibitor I (2i+Jaki).
- (g) Q-PCR analysis for expression of indicated genes in EpiSC, BV+SC+ cells from day 6 of PNT and ESC. Data are mean  $\pm$  s.d., n=3 independent experiments.

#### **Figure 4. Characterization of PGC-like cells.**

- (a) PCA analysis of BV<sup>+</sup>SC<sup>+</sup> Cells, EpiSC, ESC, PGCLC-D4/6, E9.5 PGC, E11.5 PGC, E13.5 PGC and EpiLC.
- (b) Immunofluorescence analyses of H3K9me2 and H3K27me3 in BV+ cells at day6 of PNT.
- (c) Western blot analyses (left) of H3K9me2 and H3K27me3 in ESC, EpiSC and BV+ cells at day6 of PNT. Quantification of H3K9me2 and H3K27me3 normalized to H3 levels (right). Data are mean  $\pm$  s.d..
- (d) Bisulfite sequence analysis of 5mC of differentially methylated regions (DMRs) of the imprinted genes *Snrpn* and *H19* in EpiSC and BV+SC+ cells. White and black circles indicate unmethylated and methylated CpGs, respectively.
- (e) Testis sections from *W/W<sup>o</sup>* mice and another one that was transplanted with Day6 BV<sup>+</sup>SC<sup>+</sup> cells stained by hematoxylin and eosin. Scale bars, 100  $\mu$ M.
- (f) Immunofluorescence analysis of DDX4 and SOX9 expression in testis sections from *W/W<sup>o</sup>* mice and another one that was transplanted with BVSC BV<sup>+</sup>SC<sup>+</sup>. Scale Bars, 100  $\mu$ m.

#### **Figure 5. PGCLC mediate the successful BiPNT.**

- (a) Representative images of wild-type (WT) and Prdm1 knockout (Prdm1-KO) EpiSCs. Scale bars, 250  $\mu$ m.
- (b) FACS analysis of Oct4-GFP<sup>+</sup> cells induced with WT, Prdm1-KO and rescue (overexpression of Prdm1) EpiSCs at day8.

- (c) Schematic diagram of scRNA-seq for Prdm1-KO EpiSC.
- (d) UMAP plots of WT and Prdm1-KO cells. WT cells are colored by the day (d) during PNT while Prdm1-KO cells are marked with gray (left) or vice versa (right). The black arrow indicates the initial depletion of D2-3 cells in Prdm1-KO EpiSC induced PNT.
- (e) UMAP as (d), but cells are colored by the expression of indicated genes for PGC, naïve pluripotency and placenta.

### **Figure 6. Deficiency of Gata2 promotes PGCLC.**

- (a) t-SNE representation of 24233 cells during the first three days of whole process, colored by time point of samples.
- (b) Cells colored by expression pattern of Gata2, Plac1, Dppa3 and Nanog in t-SNE representation.
- (c) t-SNE plot displaying Id1, Cdx2, Tfap2c, Prdm1 and Gata2 regulons (red dots, active; black dots, inactive) in the first three days, respectively.
- (d) FACS analysis of BVSC induction at day3 and day6 during PNT in WT or two independent Gata2-KO EpiSCs.
- (e) Q-PCR analysis for expression of placenta and PGC associated genes in WT and two Gata2-KO cells from day 6 of PNT. Data are mean  $\pm$  s.d., n=3 independent experiments.

### **Figure 7. A model for the cell fate trajectories of BiPNT**

#### **Methods**

##### **Mice**

129Sv/Jae and ICR mice were purchased from Beijing Vital River Laboratory and Oct4-GFP (OG2) transgenic allele-carrying mice (CBA/Caj X C57BL/6J) were purchased from The Jackson Laboratory. All animal experiments were operated under the Animal Protection Guidelines of Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Science, Guangzhou, China.

## **EpiSCs derivation and cell culture**

Mouse EpiSCs were derived from E5.5 mouse embryos by crossing male homozygous *Oct4*-GFP transgenic allele-carrying (CBA/CaJ X C57BL/6J) with 129/Sv female mice, following the protocol described in (Brons et al., 2007) and in (Greber et al., 2010). After culturing on feeder for the first 4-8 passages, stable EpiSCs were maintained feeder-free on fetal bovine serum (FBS)-coated dishes in N2B27 + bFGF (15 ng/ml) + ActivinA (20 ng/ml) + XAV939 (1 $\mu$ M) (FAX) medium. N2B27 medium comprised equivalent DMEM/F12 (GIBCO) and Neurobasal (GIBCO), 0.5% N2 (GIBCO), 1% B27 (GIBCO), 1% GlutaMAX (GIBCO), 1% non-essential amino acids (NEAA) (GIBCO) and 0.1mM  $\beta$ -mercaptoethanol (GIBCO). EpiSCs were dissociated with Accutase (Sigma) and passaged as singular cell of approximately 40,000 cells in a well of 6-well plate every 3 days. The medium was refreshed daily.

Mouse ES cells were maintained feeder-free on 0.1% gelatin in N2B27 medium supplemented with 2i/LIF medium as described (Yu et al., 2020b).

## **Derivation of EpiSCs from mouse BVSC-ESCs**

Mouse BVSC-ESCs were dissociated into single cells using 0.05% Trypsin-EDTA and were plated at a density of  $2-5 \times 10^4$  in a well of 12-well plate coated with FBS in N2B27-2iL medium. The next day, medium was changed to FAX medium for 3 days and then re-plated at a ratio of 1:10. Cells were passaged with Accutase every 3 days. Experiments were performed between P8 and P20.

## **BMP4 induction of EpiSCs into naïve state**

Reprogramming of EpiSCs into naïve state was performed with our protocol published previously with minor modification (Yu et al., 2020b)

### 1. Reagents setup:

Stage 1 medium: iCD1 medium (Chen et al., 2011a) comprised Vitamin C (50  $\mu$ g/ml), bFGF (10 ng/ml), LiCl (5 mM), LIF (1000 U/ml) and CHIR99021 (1  $\mu$ M), with the addition of BMP4 (5-10 ng/ml), EPZ5676 (2.5  $\mu$ M) and EPZ6438 (1  $\mu$ M). In some experiments, we also added 0.05 $\mu$ M GSK-Lsd1 or 2 $\mu$ M Parnate.

Stage 2 medium: N2B27-2iL medium comprised N2B27 medium with the supplement of 50  $\mu$ g/ml Vitamin C, 1  $\mu$ M PD0325901, 3  $\mu$ M CHIR99021 and 1000 U/ml LIF.

2. EpiSCs were passaged with Accutase into singular cell and seeded at a density of 5,000 cells per well of 24-well plate coated with FBS in N2B27+FAX medium with the addition of 5  $\mu$ M Y27632. Next day, the medium was replaced with Stage 1 medium and changed daily for the first 3 days. The Stage 2 medium was refreshed daily for the last 5 days.

### **Flow cytometry**

Cells were dissociated with 0.05% trypsin-EDTA and neutralized with DMEM containing 10%FBS. After the dissociation, cells were collected by centrifugation and resuspended with flow cytometry buffer (DPBS with 0.1% BSA and 1mM EDTA) at a density of  $1 \times 10^6$  cells per 100ul buffer. Cells were then incubated with fluorophore-conjugated antibodies for 30 min at 4°C. After incubation, cells were washed with DPBS for three times, resuspended with flow cytometry buffer and filtered with 40um cell strainer to remove large clumps of cells. Cells were then analyzed with Fortessa (BD Biosciences) or sorted with Arial (BD Biosciences). The FACS data was analyzed with FlowJo software.

The following antibodies were used: APC anti-c-Kit (eBioscience, 17-1171-83,1:200), Alexa Fluor 647 anti-mouse/human CD15 (SSEA-1) (Biolegend, 125607, 1:200) and PE anti-mouse/rat CD61 (Biolegend, 104307, 1:500).

### **Immunofluorescence**

The cells were seeded on coverslips and the fixed with 4% paraformaldehyde for 30 min at room temperature. The cells were washed with PBS once and subsequently soaked in the mixing buffer of equal volume of 0.1% Triton X-100 and 3% BSA for 1 hour at room temperature. The cells were washed with PBS once, then incubated with primary antibody at 4°C for overnight. The cells were washed with PBS five times next day, then incubated with appropriate secondary antibody for 1 hour at room temperature. The cells were washed with PBS five times to remove the secondary antibody and counterstained with DAPI for 3 min and washed with PBS once. In the end, the coverslips were mounted on the slide for observation under the confocal microscope (Zeiss 710 NLO). Primary antibodies and secondary antibodies were diluted in 3% BSA. Primary antibodies used in this study were anti-KLF4 (R&D system, AF3156, 1:100), anti-H3K9me2 (Abcam, ab1220, 1:100), anti-H3K27me3 (Merck, 17-622) and anti-PRDM1 (Invitrogen, 14-5963-82, 1:50).

### Western blot

1x10<sup>6</sup> cells were harvested and lysed with 100uL lysis buffer which containing 62.5 mM Tris-HCl (pH 6.8), 2% SDS, 10% glycerol, 0.025% bromophenol blue and 50mM DTT with protease inhibitor cocktail (Roche) on ice for 5 minutes and boiled for 10 minutes at 100°C. The samples were separated by 10-12.5% SDS-PAGE and transferred to a PVDF membrane (Millipore) using a wet transfer system, and then incubated with the primary antibodies and secondary antibodies. The following antibodies were used: anti-H3K9me2 (Abcam, ab1220,1:500), anti-H3K27me3 (Merck,17-622,1:2000), anti-H3 (Abcam, ab1791, 1:2000).

### Bisulfite Sequencing

The genomic purification kit (Promega) was used to isolate genomic DNAs, and an EpiTect Bisulfite Kit (QIAGEN) was used to perform the bisulfite reactions. LA Taq (TaKaRa) was used to amplify differentially methylated regions of *Snrpn*, *H19*, *Peg1*, and *Peg3*. The PCR products were subcloned into the pMD-18T vector (TaKaRa), and then sequenced. The results were analyzed in [quma.cdb.riken.jp](http://quma.cdb.riken.jp).

### Quantitative RT-PCR

Total RNA was extracted using the TRIzol. cDNA was reversed transcribed by HiScript II Reverse Transcriptase (Vazyme biotech R222-01) from 2ug of total RNA. Realtime quantitative PCR was performed with ChamQ™ SYBR® qPCR Master Mix (Vazyme biotech Q311-02). Relative expression values were normalized to *Gapdh*. Each experiment was operated in technical triplicate.

### Gene editing in EpiSCs

Knockout EpiSCs lines were obtained by the CRISPR/Cas9 system (Cong et al., 2013; Mali et al., 2013; Slaymaker et al., 2016). Guide RNAs were designed in website: [crispr.mit.edu](http://crispr.mit.edu), and cloned downstream of human U6 promoter of pX330 and pXP (modified from pX330, with the resistance of puromycin). For *Prdm1*, *Gata2* knockout, two pairs of gRNAs were designed to delete the critical exons. For establishing *Gata2*-tdTomato reporter, two pairs gRNAs were used to make cleavages and a tdTomato cassette was targeted into exon1 of *Gata2* by homologous recombination. The gRNA

and linearized donor plasmids were transfected with Lipofectamine 3000 (Thermo Fisher Scientific) in EpiSCs followed the guide of manufacturer.  $1 \times 10^5$  cells were resuspended in FAX medium with Y27632 in one plate of the 12-well plate. After 8-10 hours, the medium was changed with fresh FAX medium. 48 hours post transfection, the cells were selected with puromycin for 3 days and 1000 cells were seeded on one well of the 6-well plate for colony formation. Singular colony were picked up and evaluated by genotyping.

### **RNA-seq and data analysis**

Total RNA was isolated. Sequencing libraries were prepared according to the VAHTS™ mRNA-seq V3 Library Prep Kit for Illumina® (NR611-01/02) and sequenced using a NextSeq 500 High Output Kit V2 (75 cycles) (FC-404-1005, Illumina).

RNS-seq was proceeded as previously described (Zhou et al., 2017). Briefly, sequenced reads were aligned to the GENCODE annotations transcriptome (M13) using Bowtie2 (version 2.2.5) (Langmead et al., 2012) and RSEM (version 2.4.1) (Li et al., 2011), and normalized using EDASeq (version 2.4.1) (Risso et al., 2011). Differentially expressed genes were obtained using DESeq2 (version 1.10.1) (Love et al., 2014) with  $q$ -value $<0.05$  and fold change  $>2$  as a threshold. Gene ontology analysis was performed using goseq (version 1.22.0) (Young et al., 2010). PCA was performed using function prcomp.

### **ATAC-seq**

ATAC-seq was executed as previously described (Buenrostro et al., 2013; Buenrostro et al., 2015a). Briefly, a total of 50,000 cells were washed with 50  $\mu$ L cold PBS once, then centrifuged for 5 min at 500g at 4°C, then resuspended in 50  $\mu$ L lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.2% (v/v) IGEPAL CA-630) and incubated at 4°C for 10 min. The suspension of nuclei was centrifuged for 5 min at 500 g at 4°C, followed by the addition of 50  $\mu$ L transcription reaction mix (10  $\mu$ L 5X TTBL, 5  $\mu$ L TTE Mix V50 and 35  $\mu$ L nuclease-free H<sub>2</sub>O) of TruePrep™ DNA Library Prep Kit V2 for Illumina® (TD501-503, Vazyme biotech) and incubated at 37°C for 30 min. DNA fragments were purified by MinElute Kit (QIAGEN). ATAC-seq libraries were amplified by 13 cycles of amplification according to the TruePrep™ DNA Library Prep Kit V2 for Illumina® (TD501-503, Vazyme biotech), then the libraries were purified by a Qiaquick

PCR (QIAGEN) column. Library concentration was detected by a VAHTS™ Human Genomic DNA Quantification and QC Kit (NQ201, Vazyme biotech) according to the manufacturer's instructions. The concentration of ATAC library peaks was measured with Real-Time Quantitative PCR by detecting the differential expression of housekeeping gene (GAPDH) and negative genes. In the end, the ATAC library was sequenced on a NextSeq 500 using a NextSeq 500 High Output Kit V2 (150 cycles) (FC-404-2002, Illumina) according to the manufacturer's instructions.

### **Single cell sequencing**

Cells at different induction time points of PNT were collected and resuspended in DPBS with 0.04% BSA. Then, cells suspensions (500-1000 cells per microliter) were loaded on a Chromium Single Cell Controller (10x Genomics) to obtain single-cell gel beads in emulsion (GEMs) by using Single Cell 3' Library and Gel Bead Kit V2 (10x genomics, 120237). Captured cDNAs were lysed and the released RNA were barcoded through reverse transcription in singular GEMs (Zheng et al., 2017). Barcoded cDNAs were pooled and cleaned by DynaBeads® MyOne Silane Beads (Invitrogen, 37002D). Single-cell RNA-seq libraries were prepared by Single Cell 3' Library Gel Bead Kit V2 (10x Genomics, 120237) following the manufacturer's instruction. Sequencing was operated on an Illumina HiSeq X Ten with pair end 150bp (PE150).

### **Single-Cell RNA-seq Data Analysis**

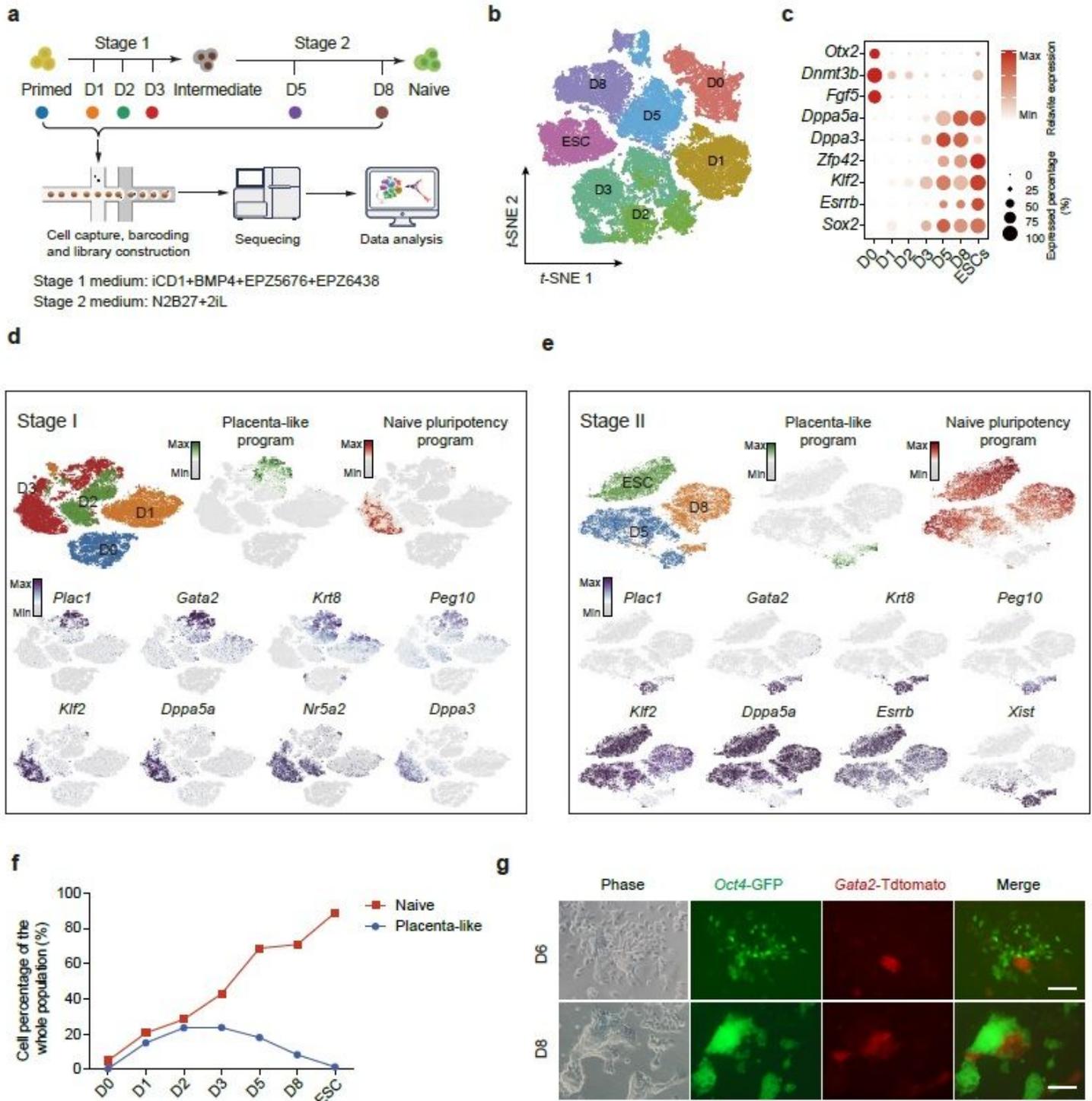
Fastq reads were aligned to the genome using STARsolo<sup>49</sup> with the setting '--outSAMattributes NH HI AS nM CR CY UR UY --readFilesCommand zcat --outFilterMultimapNmax 100 --winAnchorMultimapNmax 100 --outMultimapperOrder Random --runRNGseed 777 --outSAMmultNmax 1'. The count matrix was lightly filtered to exclude cell barcodes with low numbers of counts: Cells with less than 1000 UMIs, less than 500 genes or more than 20% fraction of mitochondrial counts were removed. The filtered matrix was normalized using scran<sup>50</sup>. The top 4000 most highly variable genes were used for PCA, and the first 50 PCs (principle components) were used for downstream analysis with SCANPY<sup>51</sup>. Single cell trajectory was analyzed by Harmony<sup>23</sup> and the top 1000 highly variable genes were used for PCA, and the force directed layout was computed using first 150 PCs. The genes expression trajectories

on pseudotemporal orderings of cells (Figure 3A) were analyzed by LineagePulse (<https://github.com/YosefLab/LineagePulse>) according to the pseudotime generated from Harmony. The Prmd1 KO and WT scRNA-seq data batch effect was corrected by Seurat<sup>52</sup>.

### **Data availability**

The sequencing data supporting the conclusions of this article are available at GEO under accession code: GSE147088  
- with reviewer token; ejczmiuujlerfyj

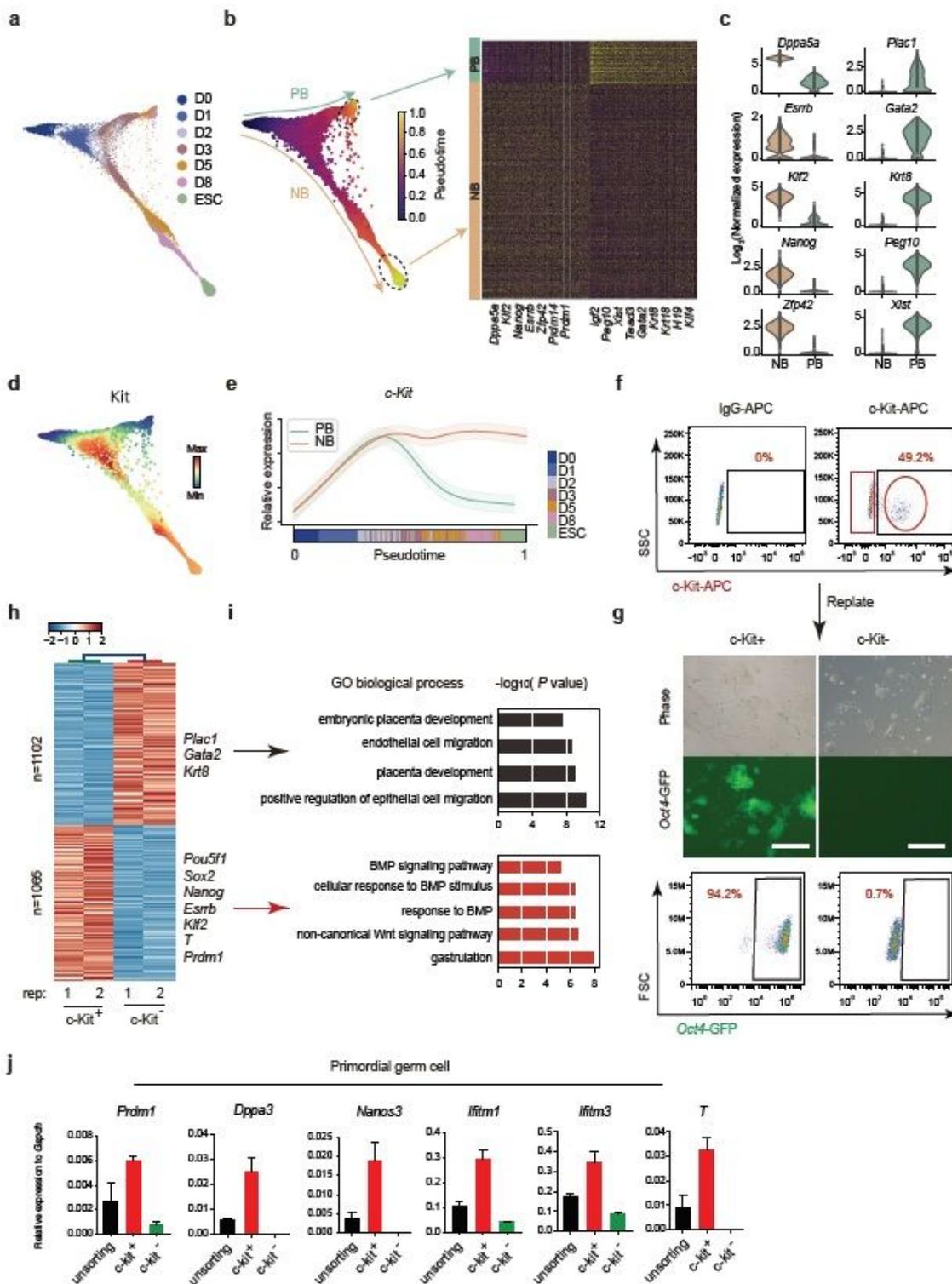
# Figures



**Figure 1**

Single cell analysis for BiPNT (a) Flow chart for the analysis of BMP induced Primed-Naïve Transition (BiPNT). scRNA-seq analysis was adopted from 10x Genomics. (b) t-SNE projection of all 50406 individual cells during the whole PNT process. (c) Expression of genes in 2 different categories at the indicated time points. (d) t-SNE projection of cells and typical gene expressions in stage I. (e) t-SNE

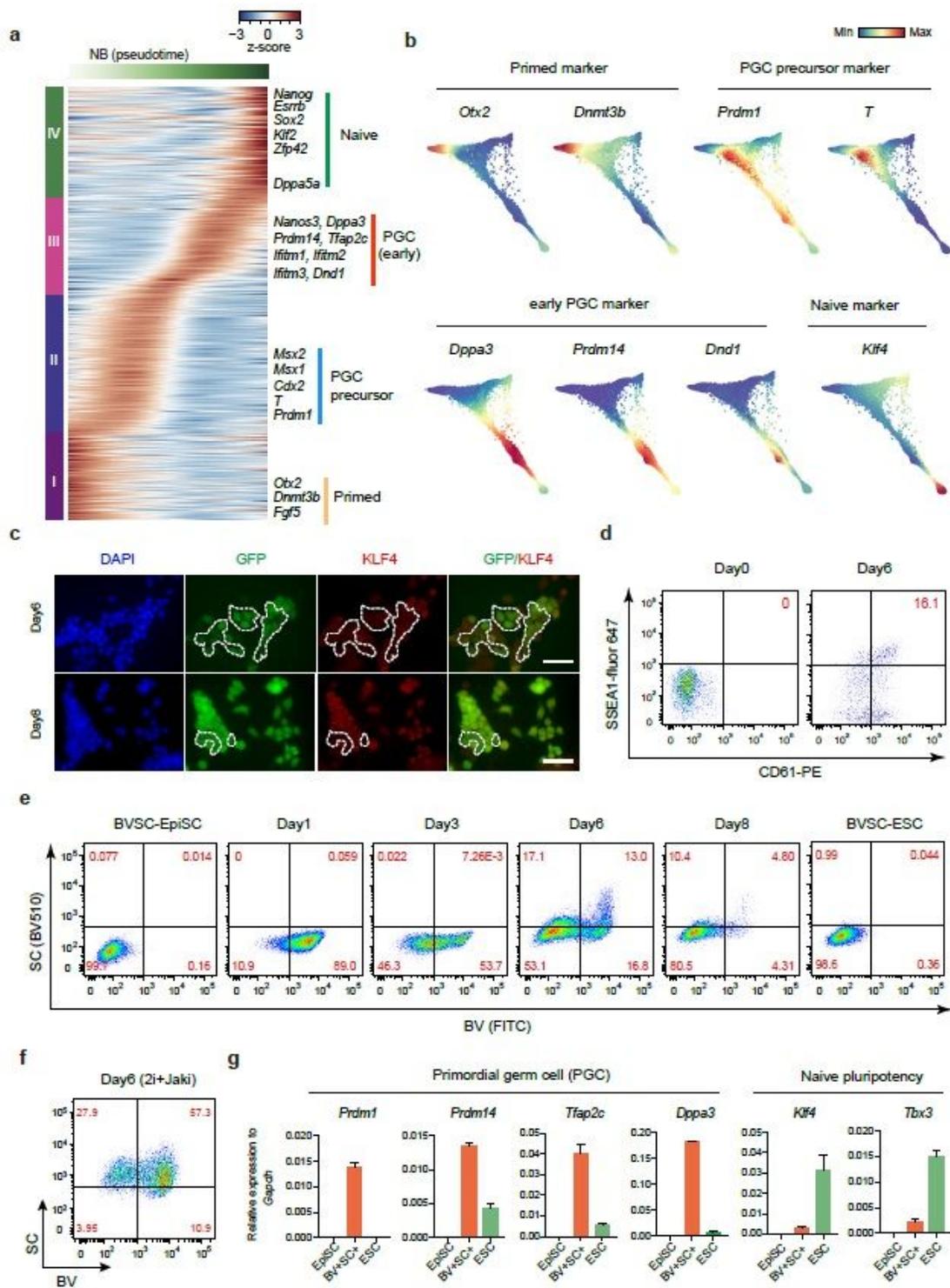
projection of cells and typical gene expressions in stage II. (f) Percentages of naïve cells and placenta-like cells at different time points. (g) Representative images of Oct4-GFP/Gata2-Tdtomato EpiSC induced PNT at day5 and day8. Scale bars, 100  $\mu$ m.



**Figure 2**

c-Kit is a Key marker for successful BiPNT (a) Trajectory reconstruction of all single cells throughout PNT, colored by indicated time points (left). The bifurcating branches defined as NB (naïve branch) and PB

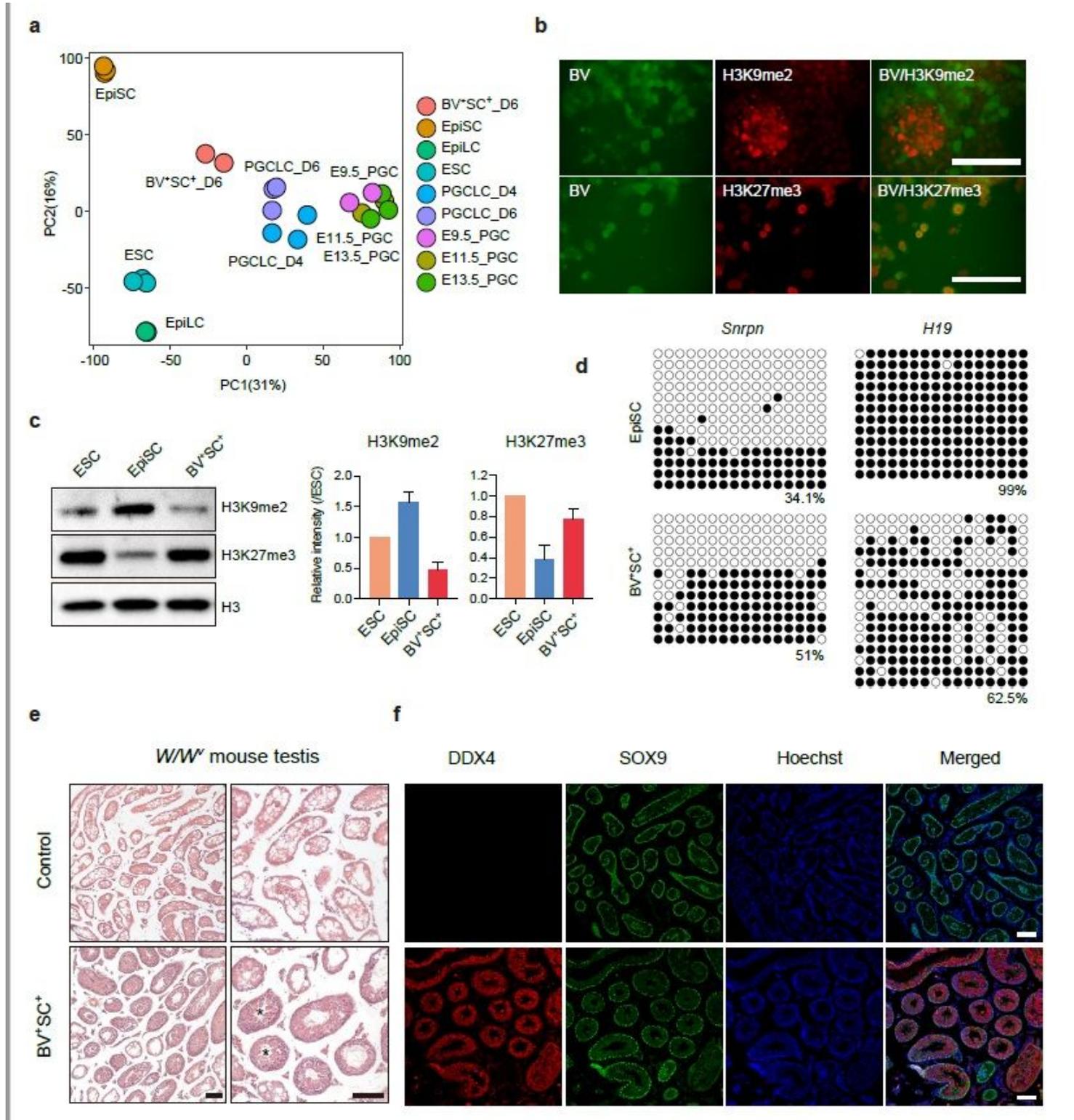
(placenta branch) separately. Circles indicate cells at the terminus of the NB and PB, respectively (right). (b) Identification of the differentially expressed genes between NB and PB branches. (c) Violin plot displaying the expression of representative naive pluripotency genes and placental genes in indicated cell types. (d) Plots showing the expression patterns of cell surface marker c-kit in NR and RP branch. (e) Expression trends of c-kit along pseudo-time during NB and PB branch specification. (f) c-Kit positive and negative cells were separated at day3 of PNT. (g) Images of Oct4-GFP colonies (upper panel) and FCAS result (lower panel) for c- Kit positive and negative cells sorted in (f) and re-plated in stage2 medium(2iL) for another 5 days. (h) Heat maps showing gene expression differences between c-Kit positive and negative cell samples by RNA-seq. Representative genes are shown on the right. (i) Go analysis for differentiated expression genes at (h). (j) Expression of primordial germ cell (PGC) markers measured by Q-PCR in day3 cells of PNT (unsorting), c-Kit positive and negative cells after sorting. Data are mean  $\pm$  s.d., n=2 independent experiments.



**Figure 3**

Activation of PGC-like program during successful BiPNT (a) Gene expression heatmap of 3219 dynamics expressed genes along NB branch in a pseudo-temporal order. Representative genes are shown on the right. (b) Plots showing the expression patterns of the indicated genes during BiPNT. (c) Immunofluorescence analysis of KLF4 in OCT4-GFP positive cells at day 6 and 8 of PNT. Scale bars, 100  $\mu$ m. (d) FACS analysis for surface expression of SSEA1 and CD61 at day 6 of PNT. (e) FACS analysis of

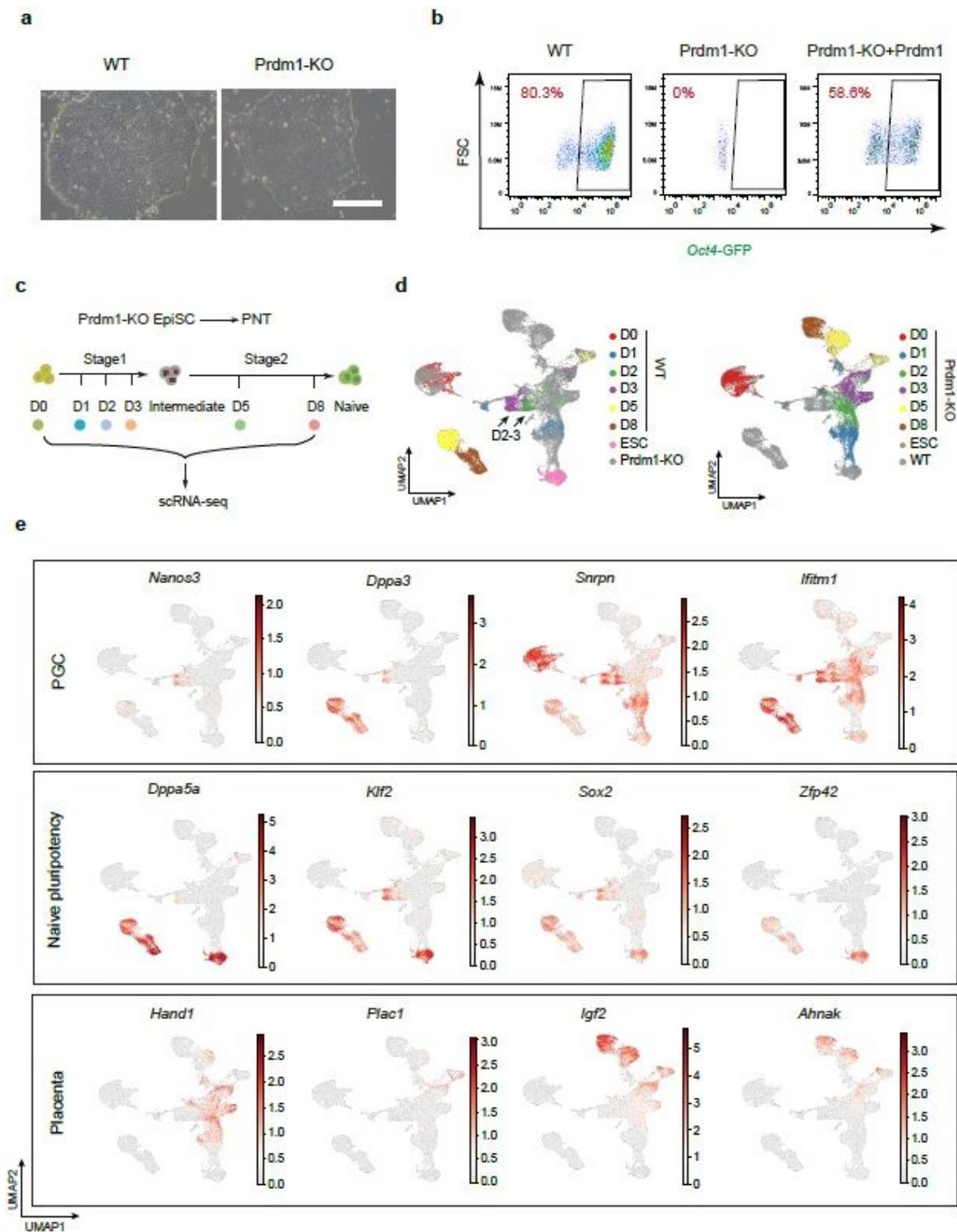
BVSC induction at distinct time points during PNT. (f) FACS analysis of BVSC induction at day6 induced in 2i medium with 5uM Jak inhibitor I (2i+Jaki). (g) Q-PCR analysis for expression of indicated genes in EpiSC, BV+SC+ cells from day 6 of PNT and ESC. Data are mean  $\pm$  s.d., n=3 independent experiments.



**Figure 4**

Characterization of PGC-like cells. (a) PCA analysis of BV+SC+ Cells, EpiSC, ESC, PGCLC-D4/6, E9.5 PGC, E11.5 PGC, E13.5 PGC and EpiLC. (b) Immunofluorescence analyses of H3K9me2 and H3K27me3 in BV+

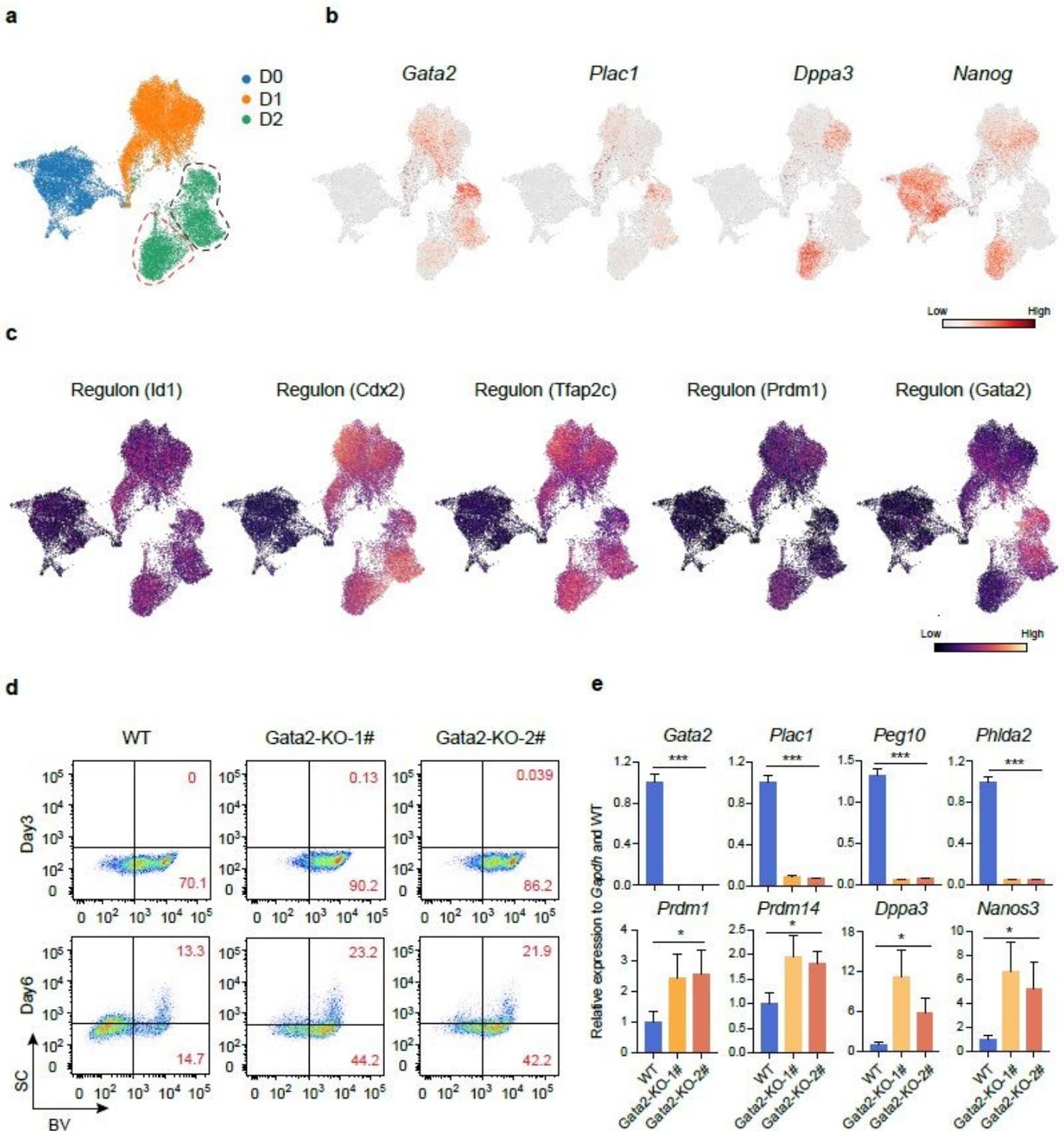
cells at day6 of PNT. (c) Western blot analyses (left) of H3K9me2 and H3K27me3 in ESC, EpiSC and BV+ cells at day6 of PNT. Quantification of H3K9me2 and H3K27me3 normalized to H3 levels (right). Data are mean  $\pm$  s.d.. (d) Bisulfite sequence analysis of 5mC of differentially methylated regions (DMRs) of the imprinted genes Snrpn and H19 in EpiSC and BV+SC+ cells. White and black circles indicate unmethylated and methylated CpGs, respectively. (e) Testis sections from W/Wv mice and another one that was transplanted with Day6 BV+SC+ cells stained by hematoxylin and eosin. Scale bars, 100  $\mu$ M. (f) Immunofluorescence analysis of DDX4 and SOX9 expression in testis sections from W/Wv mice and another one that was transplanted with BVSC BV+SC+. Scale Bars, 100  $\mu$ m.



**Figure 5**

PGCLC mediate the successful BiPNT. (a) Representative images of wild-type (WT) and Prdm1 knockout (Prdm1-KO) EpiSCs. Scale bars, 250  $\mu$ m. (b) FACS analysis of Oct4-GFP<sup>+</sup> cells induced with WT, Prdm1-KO and rescue (overexpression of Prdm1) EpiSCs at day 8. (c) Schematic diagram of scRNA-seq for Prdm1-KO EpiSC. (d) UMAP plots of WT and Prdm1-KO cells. WT cells are colored by the day (d) during PNT while Prdm1-KO cells are marked with gray (left) or vice versa (right). The black arrow indicates the

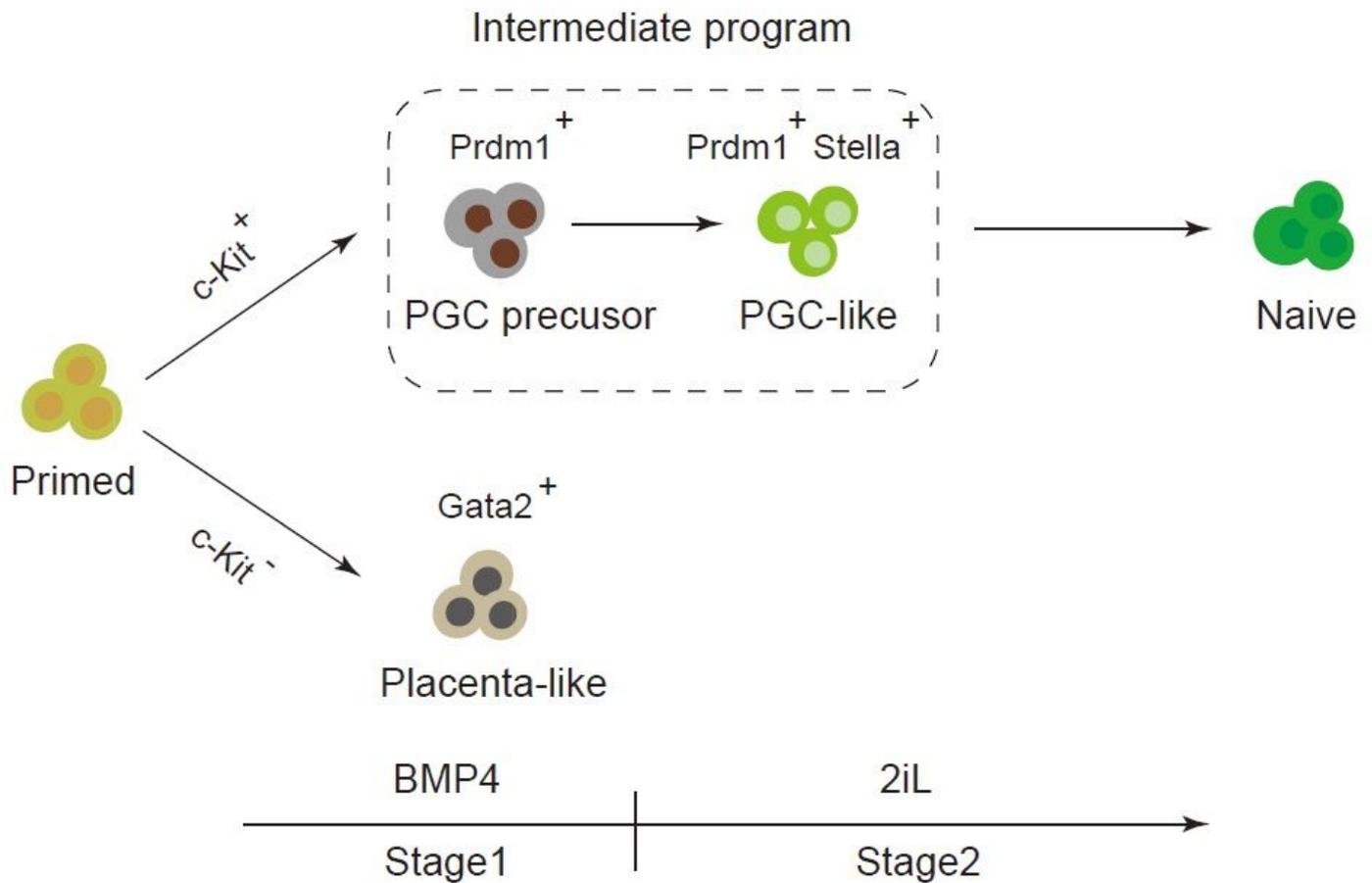
initial depletion of D2-3 cells in Prdm1-KO EpiSC induced PNT. (e) UMAP as (d), but cells are colored by the expression of indicated genes for PGC, naïve pluripotency and placenta.



**Figure 6**

Deficiency of *Gata2* promotes PGCLC. (a) t-SNE representation of 24233 cells during the first three days of whole process, colored by time point of samples. (b) Cells colored by expression pattern of *Gata2*, *Plac1*, *Dppa3* and *Nanog* in t-SNE representation. (c) t-SNE plot displaying *Id1*, *Cdx2*, *Tfap2c*, *Prdm1* and

Gata2 regulons (red dots, active; black dots, inactive) in the first three days, respectively. (d) FACS analysis of BVSC induction at day3 and day6 during PNT in WT or two independent Gata2-KO EpiSCs. (e) Q-PCR analysis for expression of placenta and PGC associated genes in WT and two Gata2-KO cells from day 6 of PNT. Data are mean  $\pm$  s.d., n=3 independent experiments.



**Figure 7**

A model for the cell fate trajectories of BiPNT

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [NCBscBiPNTSup.pdf](#)