Scalable and Efficient Generation of Mouse Primordial Germ Cell-like Cells

Xinbao Ding[1#], Liangdao Li[1#], Jingyi Gao[1], Dain Yi[1] and John C Schimenti[1*]

1 Cornell University, College of Veterinary Medicine, Department of Biomedical Sciences, Ithaca, NY 14853.

* Corresponding author: jcs92@cornell.edu

[#] Contributed equally

Abstract

Primordial germ cells (PGCs) are the founder cells of the germline. The ability to generate PGC-like cells (PGCLCs) from pluripotent stem cells has advanced our knowledge of gametogenesis and holds promise for developing infertility treatments. However, generating an ample supply of PGCLCs for demanding applications such as high-throughput genetic screens has been a limitation. Here, we demonstrated that simultaneous overexpressing 4 transcriptional factors - *Nanog* and three PGC master regulators *Prdm1*, *Prdm14* and *Tfap2c* - in suspended mouse epiblast like cells (EpiLCs) and formative embryonic stem cells (ESCs) results in efficient and cost-effective production of PGCLCs. The overexpression of *Nanog* enhances the PGC regulatory network and suppresses differentiation of somatic lineages, enabling a significant improvement in the efficiency of PGCLC production. Transcriptomic analysis reveals that differentiated PGCLCs exhibit similarities to *in vivo* PGCs and are more advanced compared to cytokine-induced PGCLCs. These differentiated PGCLCs could be sustained over prolonged periods of culture and could differentiate into spermatogonia-like cells *in vitro*. Importantly, the ability to produce PGCLCs at scale, without using costly cytokines, enables biochemical and functional genomic screens to dissect mechanisms of germ cell development and infertility.

Introduction

In mammals, gametogenesis is a complex and long process that is initiated by the specification of primordial germ cells (PGCs), the precursors of oocytes and sperm. Unraveling the mechanisms of PGC development is crucial for understanding key processes such as sexually dimorphic epigenetic inheritance and the bases of infertility, a relatively common health condition impacting ~15% of all couples [1]. In mammals, PGCs specified in the embryonic epiblast respond to extrinsic and intrinsic signaling pathways including bone morphogenetic protein (BMP) and WNT signaling. In mice, BMP4 activates WNT3, and subsequently a master mesodermal transcription factor (TF), brachyury (T). T activates critical early germline markers *Prdm1* (also known as *Blimp1*), *Prdm14*, and *Tfap2c* (also known as *AP2γ*) [2]. *Prdm1* is expressed in the precursors of PGCs, which then stimulates *Prdm14* and *Tfap2c* transcription to activate the germline pathway. The core TF network containing *Prdm1*, *Prdm14*, and *Tfap2c* is critical for mouse PGC specification; it represses somatic differentiation while driving reacquisition of a pluripotency network and epigenetic reprogramming [2]. After migration to and colonization of the genital ridges, male germ cells differentiate into gonocytes. Thereafter, they become spermatogonial stem cells that subsequently seed recurrent rounds of mitotic expansion yielding spermatogonia that enter meiosis, and eventually, haploid spermatozoa.

Characterization of these gene regulatory pathways in PGCs formed the basis for creating mouse primordial germ cell-like cells (PGCLCs) from pluripotent stem cells *in vitro* [2]. Pluripotency has at least three distinct states that mirror early embryonic development: naïve or ground state ESCs resembling the preimplantation epiblast; formative ESCs, resembling early post-implantation gastrulating epiblast; and primed epiblast stem cells (EpiSCs), resembling late post-implantation gastrulating epiblast [3–5]. Naïve ESCs typically are cultured with inhibitors of fibroblast growth factor and mitogen-activated protein kinase (FGF/MEK) as well as glycogen synthase kinase 3β (GSK3β) in the presence of leukemia inhibitory factor (LIF) (common known as 2i+LIF medium). EpiSCs self-renew when exposed to Activin A and fibroblast growth factor 2 (FGF2). Naïve ESCs do not readily respond to germ cell inductive stimuli, while primed EpiSCs

display limited germline competence. By modulating activin and WNT signaling, a formative state has been stably captured *in vitro* [4–7]. As an intermediate pluripotency state, formative ESCs possess dual competency for mouse chimera formation (when introduced into host blastocysts) and direct responsiveness to PGC specification.

The ability to model germ cell development in culture holds great potential for applications such as dissecting the genomic regulation of gametogenesis, and identifying genetic causes of infertility that would be important for assisted reproductive technologies (ART) involving gene corrections and optimization of *in vitro* gametogenesis (IVG). However, these applications remain challenging due to limited scalability of PGCLC generation. Conventional methods involve a transient conversion from naïve ESCs to germline-competent epiblast-like cells (EpiLCs), which presumably represent a formative state. Aggregated EpiLC clusters are then exposed to high concentrations of BMPs and supporting cytokines, leading to their differentiation into PGCLCs [8]. Alternatively, PGCLC specification can be induced by overexpressing three TFs (*Prdm1*, *Prdm14*, and *Tfac2c*), and to a lesser extent, two TFs (*Prdm1* + *Tfap2c* or *Prdm14* + *Tfap2c*) or even a single TF (*Prdm14*), in EpiLC aggregates [9]. *Nanog*, a core pluripotency factor expressed in PGCs, is not essential for emergence of PGCs or germline transmission, but its deletion results in a significant reduction in PGC numbers both *in vitro* and *in vivo* [10]. *Nanog* is also expressed in naïve ESCs, and its expression undergoes a transient downregulation as cells transition into the EpiLC state. The overexpression of *Nanog* can promote efficient PGCLC differentiation by binding endogenous enhancers of *Prdm1* and *Prdm14*, even without BMP4 and associated cytokines [11]. Furthermore, Nanog overexpression induces cytokine-free PGCLC specification by repressing *Otx2*, encoding a TF essential for anterior neuroectoderm specification [12,13].

Promoting cell proliferation is another strategy for producing a large number of PGCs/ PGCLCs. However, long-term culture of mouse PGCs *in vitro* is prone to dedifferentiation to pluripotent embryonic germ cells (EGCs)[14,15]. Saitou and colleagues found that Rolipram and Forskolin, which stimulate cAMP signaling, enables expansion of PGCLCs up to ~50-fold in the presence of SCF (Stem cell factor) on m220 feeders

(stromal cells expressing a membrane-bound form of human SCF), and the expanded PGCLCs can contribute to spermatogenesis after transplantation to neonatal testes [16–18]. Nevertheless, PGCLCs cultured under these conditions cannot be expanded for more than 7 days. Although these methods have been instrumental in understanding and improving PGCLC formation, they are associated with either low efficiency or high costs (for reagents such as cytokines).

Here, we report a simple and economical approach to induce the differentiation of PGCLCs from both formative ESCs and EpiLCs at scale by overexpressing four TFs (*Prdm1*, *Prdm14*, *Tfap2c* and *Nanog*). The resulting PGCLCs exhibit transcriptional and epigenetic profiles similar to those of late (E11.5) PGCs, and these could be further differentiated into spermatogonia-like cells. This system holds promise for facilitating high-throughput functional genomic studies of germ cell development, characterization of potential infertility-causing genetic variants, and improving IVG.

Results

Establishing a cytokine-free platform for efficient production of PGCLCs from EpiLCs.

To establish an efficient, cytokine-free system for generating PGCLCs, we began with a mouse ESC line containing enhanced GFP (eGFP) under the regulatory control of *Stella* (also called *Dppa3/Pgc7*), a PGC marker [19], and validated the ability of this line to form germline-transmitting chimeric mice after blastocyst microinjection (Fig. S1). The Stella-eGFP reporter enables real-time visualization of PGCLC development from ESCs *in vitro*. Next, based on the finding that *Nanog* promotes transition of ESCs to PGCLCs by binding the enhancer regions of *Prdm1* and *Prdm14* and inducing their expression [11], we introduced a stably-integrated doxycycline (Dox)-inducible *Nanog*. To test the effectiveness of this transgene, we utilized a high efficiency enhancer activity detection system [20] to derive ESC lines containing a targeted integrations of lacZ reporter under the control of *Prdm1* or *Prdm14* enhancers into the safe harbor H11 locus. Dox treatment of embryoid bodies (EBs) derived from these lines triggered β-galactosidase production, indicating the effectiveness of NANOG-stimulated transcription (Fig. S2A-D).

Besides promoting expression of *Prdm1* and *Prdm14*, NANOG represses *Otx2,* a TF that promotes differentiation of pluripotent cells towards a neuroectodermal fate [11,13]. Mutant cells lacking OTX2 enter the germline from an EpiLC state with high efficiency, even in the absence of the essential PGCLC-promoting cytokine signals [21]. In summary, Nanog plays a crucial role in PGCLC differentiation by promoting the expression of PGC core regulatory TF-encoding genes *Prdm1* and *Prdm14*, while simultaneously inhibiting the somatic fate inducer *Otx2* (Fig. S2E).

Based on observations that overexpression of *Prdm1*, *Prdm14* and *Tfap2c* can induce PGCLCs without cytokines [9], we hypothesized that simultaneous overexpression of four TFs (i.e. *Prdm1*, *Prdm14*, *Tfap2c* and *Nanog*, hereafter called "4TF") in EpiLCs might robustly render PGCLCs *in vitro*. We isolated several Dox-inducible clonal cell lines harboring combinations of these TFs in Stella-eGFP ESCs that also contained rtTAM2 (reverse tetracycline-controlled transactivator mutant; Fig. 1A and B). The resulting ESCs, cultured in 2i+LIF medium, were differentiated to EpiLCs by two-day culture in the presence of Activin A and bFGF. To induce PGCLC differentiation, the EpiLCs aggregates were treated with Dox instead of cytokines. We first tested whether ectopic expression of one or more TFs drove EpiLC-to-PGCLC differentiation in U-bottom 96-well plates. Consistent with previous observations [11], induction of *Nanog* alone led to ~1/3 of cells becoming eGFP$^+$ (Fig. 1A). The efficiency rose to ~80% by simultaneous overexpression of all 4TFs. Cell lines harboring *Nanog* plus other TF combinations also showed substantial increases in the fraction of eGFP$^+$ cells compared to Nanog alone (Figs. 1A, D and E).

In vivo, PGCs undergo substantial epigenetic reprogramming, characterized by the loss of H3K9me2 and the acquisition of H3K27me3, as well as global erasure of DNA methylation [2]. To test the fidelity of epigenetic changes in PGCLCs produced in our 4TF system, we examined histone modifications, CpG methylation status of imprinted genes, and capacity for dedifferentiation into pluripotent EGCs. Differentiated cells with higher eGFP fluorescence displayed lower H3K9me2 and higher H3K27me3 labeling (Fig. 1F). The mRNA levels of epigenetic modifiers *Dnmt3a* and *Dnmt3b* were increased in EpiLCs and subsequent downregulated in EBs (Fig. 1G), which aligns with previous findings [8]. Bisulfite sequencing of PGCLCs (eGFP$^+$ cells) revealed reduced methylation

in differentially methylated regions for both paternally (*H19*) and maternally (*Snrpn*) imprinted loci, while eGFP$^-$ cells (presumably somatic-like) exhibited a more methylated state (Fig. 1H). In addition, PGCLCs could de-differentiated to EGCs *in vitro* when transferred to 2i+LIF medium (Fig. S3). Collectively, our data indicate that simultaneously overexpression of 4TFs effectively induces the differentiation of EpiLCs into PGCLCs.

Transcriptomic properties of 4TF-induced PGCLCs

To examine the fidelity of our 4TF-induced PGCLCs, RNA sequencing (RNA-seq) was performed (Fig. S4) and the transcriptomes were compared to those of *in vivo* germ cells (encompassing PGCs at E9.5, E10.5, and E11.5 and male/female germ cells at E12.5 and E13.5) and cytokine (Ck)-induced PGCLCs (those at day 2,4 and 6 after induction, referred to as Ck_D2/4/6PGCLC) [16,22,23]. Principal component analysis (PCA) showed that 4TF PGCLCs at days 2, 4 and 6 (4TF_D2/4/6PGCLC) clustered tightly as a group separate from naïve ESCs, EpiLCs and Ck_D2PGCLCs (Fig. 2A). Notably, 4TF-induced PGCLCs grouped closely with PGCs at E10.5 and E11.5, while Ck_D4/6PGCLCs grouped closely with E9.5 and E10.5 PGCs (Fig. 2A). These data suggest that 4TF-induced PGCLCs are more akin to late PGCs than Ck_D2/4/6PGCLCs. Pearson correlation coefficient analysis revealed that 4TF_D2/4/6PGCLCs are highly correlated with Ck_D4/6PGCLCs but not Ck_D2PGCLCs (Fig. 2B), consistent with the PCA. We next applied two-sided rank–rank hypergeometric overlap (RRHO) analysis to compare the gene expression signatures between samples in a threshold-free manner [24]. We observed robust overlap in both up- and down-regulated transcripts amongst differentiating cells in both Ck- and 4TF-induced groups relative to either ESCs or EpiLCs (Fig. 2C). In the Ck-induced group, ~60% of genes exhibited concordant differential expression (53% when compared to ESCs and 67% when compared to EpiLCs) between D2 and D6 PGCLCs. Interestingly, the concordance was higher for the D2 and D6 4TF-induced group (74% and 78%, respectively). In both induction systems, D4 and D6 PGCLC shared around 84% concordantly regulated genes (Fig. 2D and Fig. S5). These data demonstrated that 4TF-induced PGCLCs at early stages exhibited greater consistency in their transcriptomic changes compared those in Ck-induced PGCLCs.

In vivo, PGC specification follows WNT-dependent induction, which activates the master mesodermal TF, T. This triggers the core TF network for PGCs (i.e., *Prdm1*, *Prdm14*, and *Tfap2c*). To investigate the activity of WNT signaling in various PGCLC differentiation conditions, gene expression datasets from 3TF-induced PGCLCs at days 2 and 4 (referred to as 3TF_D2/4PGCLC) were retrieved [9] and compared to Ck- and 4TF-induced PGCLCs. Gene set variation analysis (GSVA) revealed strong enrichment of Wnt signaling pathways in Ck_D2PGCLCs (including downstream target genes *Wnt3*, *Wnt5a*, *Wnt5b*, *Axin2*, *Mixl1*, *Cdx2*, and *T*) as previously reported [9], but not in 3TF_D2PGCLCs or in our 4TF_D2PGCLCs (Fig. 2E and F).

To better characterize the transcriptome landscapes during PGCLC differentiation, we applied k-Means analysis of top 500 differentially-expressed genes and defined one cluster (C4) representing genes expressed in 4TF_D2/4/6PGCLCs (Fig. 3A). Gene Ontology (GO) analysis revealed that 11 of the top 20 pathways in this cluster were associated with reproductive processes (Fig. 3B). GSVA confirmed enrichment in germ cell development terms such as gamete generation, oogenesis, piRNA metabolic process, meiosis and reproduction, were enriched in PGCLCs compared to ESCs and EpiLCs (Fig. 3C). D2PGCLCs were predominantly enriched for Biological Processes including “S-adenosylmethionine metabolic process”, “leukocyte migration” and “inflammatory response” (Fig. S6). In contrast, the upregulated genes at days 4 and 6 were primarily enriched in processes related to gamete generation, meiotic cell cycle and reproduction (Fig. 3D and S7). We next compared the differentially expressed genes among Ck-, 3TF- and 4TF-induced PGCLCs. Across all three groups, the naïve pluripotency gene *Klf4* was highly expressed in ESCs, while genes related to the formative state (*Otx2*, *Pou3f1*, *Fgf5*, *Dnmt3a* and *Dnmt3b*) were highly expressed in EpiLCs (Fig. 3E). Early PGC-related genes (e.g., *Prdm1*, *Prdm14*, *Nanog*, *Dnd1*, *Tfap2c*, *Dppa3* and *Itgb3*) were notably expressed in Ck_D4PGCLC, 3TF_D2PGCLC and 4TF_D2PGCLC, indicating that both 3TF and 4TF induction systems cause molecular differentiation more rapidly than cytokine induction. Some PGC genes, including *Rhox9, Rhox6, Rhox5 and Gm9*, were highly expressed in Ck_D6PGCLC, 3TF_D4PGCLC and 4TF_D4PGCLC. Interestingly, late PGC, meiosis and piRNA-related genes were consistently expressed in 4TF-induced PGCLCs but not in Ck-

induced PGCLCs (Fig. 3E and S7), explaining the enrichment of meiotic and piRNA metabolic pathways in 4TF-induced PGCLCs. Since there is no transcriptome data for 3TF_D6PGCLC, it uncertain whether they exhibit a similar phenomenon as 4TF_D6PGCLC (Fig. 3E). To comprehend the expression of meiotic and piRNA genes in 4TF-induced PGCLCs, we conducted a network analysis on the differentially upregulated genes and identified a sub-network constructed by germ cell specific genes (Fig. S8). During germ cell development, *Tfap2c* was found to be activated by *Prdm1* and *Prdm14* [2], direct targeting *Dmrt1* in PGCLCs [25,26]. Human DMRT1 upregulates late PGC markers (such as DAZL and PIWIL2), while DAZL enhances DMRT1's impact by repressing pluripotency and promoting other late PGC genes (such as DDX4 and SYCP3) [27]. The regulatory axis involving *Prdm1*, *Prdm14*, *Tfap2c*, *Dmrt1*, *Dazl*, *Ddx4* and *Sycp3* genes facilitated the transition from early to late PGCLCs (Fig. S8).

In the 3TF induction system, ~30% of PGCLCs (Blimp1-mVenus^{+}/Stella-eCFP^{+} cells) were differentiated, and this efficiency declined after day 4 [9]. Conversely, in our 4TF induction system, we consistently observed a high level of differentiation efficiency beyond day 4. This variation prompted us to investigate the distinctive role of *Nanog*. GSEA revealed that pathways predominantly related to neurodevelopment were suppressed in D2PGCLCs compared to EpiLCs (Figs. S9A and B), and *Otx2* was involved in the majority of these pathways (Fig. S9C). *Otx2* plays a crucial role in brain development and exhibits reciprocal antagonism with *Nanog*. Induction of transgenic *Nanog* in EpiLCs shifts the balance between these mutual antagonists, favoring *Nanog* [28]. When *Otx2* levels decrease sufficiently, this enables NANOG to influence the regulatory elements controlling genes associated with PGC-specific TFs [28]. These observations, together with the evidence that *Nanog* enhances the expression of *Prdm1* and *Prdm14*, illustrates its dual role in both strengthening the PGC program and preventing the development of non-germline lineages.

We next examined the potential of PGCLCs to differentiate into spermatogonia-like cells *in vitro*. D6PGCLCs were aggregated with dissociated neonatal testicular cells from germ cell deficient Kit$^{W/Wv}$ mice and co-cultured for 2 days in U bottom 96-well plates. Then, the cell aggregates were transferred to a trans-well plate and cultured in medium containing GNDF and bFGF that are essential for promoting spermatogonial

growth [29]. After a two-week induction period, we conducted IF staining to assess whether the aggregates express late PGC or spermatogonia proteins. We detected cell clusters that expressed NANOS3 and DDX4 within the aggregates (Figs. 3F and S10). Over 70% of DDX4-positive cells also express GCNA (Germ cell nuclear antigen). Interestingly, certain DDX4-positive cells also contained PLZF (formally ZBTB16; Zinc finger and BTB domain-containing protein 16), a marker indicative of spermatogonia (Fig. 3F)[30]. DDX4-positive cells did not express the Sertoli cell marker, SOX9 (Fig. 3F). These data suggest that the 4TF-induced PGCLCs have the capacity of further development *in vitro*.

Scale-up differentiation and long-term culture of PGCLCs

To validate the Stella-eGFP transgene as an accurate marker of PGCLCs, we analyzed differentiating cultures with two additional PGC surface markers, ITGB3 (also known as CD61) and SSEA1 [8]. Doubly positive cells were shown to differentiate into functional gametes [31]. As expected, the cell populations characterized by ITGB3$^+$/SSEA1$^+$ and ITGB3$^+$/Stella-eGFP$^+$ were significantly enriched in the day 6 EBs compared to ESCs and EpiLCs (Fig. 4A). Next, we sought to determine the scalability this 4TF-induction system. We seeded 2×10^3 cells/ well to U bottom 96-well plates, 8×10^4 cells/well to untreated 12-well plates and 1×10^6 cells to a single 100mm bacteriological Petri dish, then assessed differentiation in day 6 EBs. The percentages of both ITGB3$^+$/SSEA1$^+$ and ITGB3$^+$/Stella-eGFP$^+$ populations were highest in 12-well plate and lowest in 96-well plates (Fig. 4A and B). Suspecting that EB size influences differentiation efficiency, we measured the diameters of day 6 EBs. In the U bottom 96-well plates, EB sizes were homogeneous and by far the largest (~2.5× on average) amongst the culture formats (Fig. 4C). The results indicate an inverse relationship between efficiency of PGCLC differentiation and EB size.

Next, we investigated the potential for maintaining PGCLCs over long periods. Day 6 EBs from a 12-well plate were dispersed and seeded into one well of an untreated 12-well plate in the presence of Dox. Cells were passaged every 4th day, and after culturing EBs for 20 and 43 days, >70% of cells were ITGB3$^+$/SSEA1$^+$ and ITGB3$^+$/Stella-eGFP$^+$ (Fig. 4D).

Direct differentiation of PGCLCs from formative ESCs

We next asked whether 4TFs can efficiently induce PGCLCs from formative ESCs, a pluripotent stem cell that directly responds to PGC specification. The 4TF Stella-eGFP ESCs were transformed to a formative state through exposure to bFGF, Activin A and GSK3β inhibitor (CHIR99021), as described previously [5]. Naïve ESCs typically grow as tightly packed colonies with a dome shape. Upon EpiLC induction, the cells quickly underwent a morphological conversion that includes flattening, diminished cell-cell interactions, and the formation of cellular protrusions. However, formative ESCs displayed distinctive morphologies (Fig. 5A). Immunolabeling of ZO1 (zonula occludens-1, also known as tight junction protein 1) revealed that formative ESCs and EpiLCs were more similar to one another than naïve ESCs (Fig. 5B). The regulation of *Oct4* expression in naïve and primed pluripotent cells is differentially controlled by distal (DE) and proximal (PE) enhancers, respectively [32]. The DE enhancer also exhibits activity in the formative state, albeit at a lower level compared to that in the naïve state [5]. To monitor the exit from the naïve state during naïve-to-formative conversion, we compared the GFP signal intensity of Oct4-DE-GFP ("OG2") iPSCs during this transition. Consistent with previous observations [5], the majority of formative iPSCs and EpiLCs were GFP^+ but the average signal intensity was lower than that in naïve iPSCs (Fig. 5C). Collectively, these results confirm the establishment of formative ESCs/iPSCs from naïve cells.

We next tested if formative 4TF ESCs respond directly to PGC specification. Following the removal of the feeder layer, Dox was added to promote PGCLC differentiation. FACS analysis of EB cells revealed distinct $ITGB3^+/SSEA1^+$ and $ITGB3^+/Stella\text{-}eGFP^+$ populations; such cells were not observed in undifferentiated formative ESCs (Fig. 5D). This differentiation system proved to be scalable (Fig. 5D), although the efficiencies were lower compared to those observed in EpiLC-to-PGCLC differentiation system described earlier (Fig. 4A and B). Collectively, these findings demonstrated that overexpressing 4TFs can directly and effectively induce differentiation of formative ESCs into PGCLCs.

To elucidate the differences in differentiation efficiencies between formative ESC and

EpiLC, we compared the transcriptomes of our EpiLCs to a published formative ESC dataset [5]. The RRHO analysis revealed significant overlap in co-regulated genes between formative ESC and EpiLC when compared to naïve ESC, indicating a shared cell state (Fig. S11A and B). Interestingly, genes associated with the formative state (*Otx2*, *Pou3f1*, *Fgf5*, *Dnmt3a* and *Dnmt3b*) exhibited higher expression in EpiLCs than those in formative ESCs (Fig. S11C). Thus, formative ESCs and EpiLCs represent distinct sub-stages, leading to variations in differentiation efficiencies.

Discussion

PGC development entails important regulatory and epigenetic events that lay the groundwork for subsequent sex-specific gametogenesis programs. Although *in vitro* PGCLC induction systems have facilitated molecular studies of the specification and developmental events of the germ cell lineage [9,11,21,33–37], current techniques have shortcomings in terms of efficiency, scalability and long-term culture. Our 4TF-induction system addresses these limitations. We demonstrated that simultaneous overexpression of *Nanog*, *Prdm1*, *Prdm14* and *Tfap2c* in suspended mouse EpiLCs and formative ESCs can render PGCLCs in an efficiently and cost-effectively manner. The induced PGCLCs exhibit late PGC hallmarks, and they can be further developed into spermatogonium-like cells *in vitro*. Overexpression of *Nanog* had the important effect of enhancing the PGC regulatory network while suppressing the formation of non-germline lineages, significantly improving PGCLC production.

Sequential TFs govern PGC differentiation, and specification of mouse PGCs requires coordinated actions of *Prdm1*, *Prdm14*, and *Tfap2c*. Here, we found that forced expression of *Nanog* alone, or in addition to combinations of these core TFs can drive PGCLC specification. The 4TF-induced PGCLC exhibits similar epigenetic reprogramming with those of Ck- and 3TF-induced PGCLCs. Interestingly, the 4TF-induction system exhibits superior performance compared to the 3TF (i.e., *Prdm1*, *Prdm14* and *Tfap2c*) induction system, underscoring the pivotal role of *Nanog*. Given that NANOG can bind and activate enhancers of *Prdm1* and *Prdm14* in EpiLCs *in vitro*

[11], which we confirmed here, we conclude that *Nanog* overexpression strengthens the PGC core regulatory network in the 4TF system. NANOG also plays a role in suppressing somatic lineage induction development. OTX2, for example, shows reciprocal antagonism with NANOG, restricting mouse germline entry [21]. In 4TF-induced PGCLCs, we observed downregulation of *Otx2* and simultaneous upregulation of *Nanog*, mirroring the pattern seen in Ck- and 3TF-induced PGCLCs. OTX2 functions as a crucial TF for anterior neuroectoderm specification. Comparing 4TF_D2PGCLCs to EpiLCs, we observed significant downregulation of neurodevelopment pathways associated with *Otx2*. Therefore, the dual role of NANOG contributes to high efficiency PGCLC generation.

The transcriptional profiles of Ck- and 4TF-induced PGCLCs proved to be different and resembled PGCs at distinct stages. This disparity partially arises from the essential role of WNT signaling and mesodermal program in the Ck-induction system, whereas the 4TF-induction system circumvents these requirements. Additionally, the gene regulatory axis of early-to-late PGC development established in 4TF-induced PGCLCs propels their differentiation beyond what Ck achieved within the same timeframe. Consequently, in the 4TF but not Ck-induced PGCLCs, the late PGC regulatory network was triggered as indicated by expression of genes involved in piRNA processes (such as *Piwil2*, *Mael*, *Tdrd5*, *Mov10l1*, *Tdrd7* and *Tex19.2*) and meiosis (such as *Stra8*, *Sycp1/2/3*, *Meiob* and *Smc1b*).

Gonadal and testicular somatic cells are essential to support the further differentiation of male PGCLCs *in vitro* [17,38,39]. Seminiferous tubule-like structures can be constructed by aggregating Ck-induced PGCLCs with gonadal somatic cells under a gas-liquid interphase culture condition [17,39]. Given that 4TF-induced PGCLCs exhibit more advanced development compared to Ck-induced PGCLCs, we opted for the readily accessible neonatal testicular somatic cells to facilitate PGCLCs differentiation, along with key cytokines (i.e., GDNF and bFGF) crucial for spermatogonial proliferation. While no seminiferous tubule-like structure was observed, clusters of DDX4$^+$ cells were formed and surrounded by somatic cells. Similar to previous observations [17,39], a subset of DDX4$^+$ cells underwent further differentiation into cells with spermatogonium-like characteristics. However, further work is necessary to explore whether these cells can

progress to later stages of spermatogenesis.

Efficient, scalable and simple PGCLC differentiation systems, such as that we report here, will be useful for optimizing IVG and facilitating high-throughput screening platforms. One interesting parameter we observed was a correlation between the EB size and differentiation efficiency. EB size likely impacts differentiation via factors such as the diffusion of soluble molecules and the degree of interactions in both extracellular matrix-cell and cell-cell adhesive contexts [40]. In addition to scalability and efficiency, a useful aspect of the 4TF system I that the derived PGCLCs persisted for prolonged periods in culture, in contrast to a 3TF-induction system where efficiency declined after day 4 [9]. However, further work is necessary to characterize the properties of these long-term cultured or differentiated cells.

In mice, competence for germline specification distinguishes the formative state of pluripotency from the naïve and primed states [41]. Given that both formative ESCs and EpiLCs respond directly to PGC specification, it was unclear whether exposure of formative ESCs to 4TF overexpression was compatible with efficient differentiation into PGCLCs. Recent advancements have identified four distinct sub-stages of formative pluripotent stem cells, exhibiting functional characteristics and WNT/ β-catenin signaling modulation [4–7]. In our study, we cultivated formative ESCs using one of these protocols [7], and revealed that 4TFs can drive both formative ESCs and EpiLCs into PGCLCs. Importantly, enables simplification of the PGCLC differentiation process, as formative ESCs can be stably maintained and would not require preliminary induction into the transient EpiLC-like state from naïve ESCs.

We envision that the high efficiency 4TF system we described here will be useful for several applications. One such application is for conducting high-throughput screening of genetic factors that influence gametogenesis. For example, massively parallel CRISPR-based whole genome screens require large number of cells to accommodate libraries containing tens of thousands of guide RNAs. Scalability is also important for screening small molecule libraries for chemicals that may further improve differentiation, or libraries of drugs or compounds in the environment that may impair or negatively impact germ cell development. Considering that genetic causes underlie ~50% of

human infertility [1], parallelized screening of genetic variants, particularly rare or minor alleles that are classified as variants of uncertain significance (VUS), for impacts on IVG. Functionally interpreting the VUS within genes that essential for gametogenesis is important for genetic counseling, clinical applications, and potentially genome editing to enable people with identified genetic causes of infertility to produce viable gametes *in vitro* or *in vivo* [1,42,43].

Materials and Methods

Mice and cell lines

All animal usage was approved by Cornell University's Institutional Animal Care and Use Committee, under protocol 2004-0038 to J.C.S. Four- to eight-day-old $Kit^{W/Wv}$ male mice generated by crossing C57BL/6J-Kit^{Wv}/J and WB/ReJ Kit^{W}/J were used in this study. C57BL/6J-Kit^{Wv}/J (#000049), WB/ReJ Kit^{W}/J (#000692) and B6(Cg)-Tyr^{c-2J}/J (#000058) mice were purchased from The Jackson Laboratory. The Stella-eGFP transgenic ESC line (Stella-eGFP ESCs) with Stella flanking sequences coupled to eGFP as the reporter was kindly provided by Prof. M. Azim Surani [19].

Vectors

To construct pPBhCMV1-Prdm1-pA and pPBhCMV1-Prdm14-pA, mouse *Prdm1* and *Prdm14* coding sequences were cloned into pPBhCMV1-Nanog-pA (a gift from M. Azim Surani) by PCR flanked with *Xho*I and *Not*I restriction sites. To construct targeting vectors containing enhancer-reporter transgenes, PCR-amplified sequences of mouse *Prdm1* and *Prdm14* enhancers [11] bearing terminal *Not*I restriction sites were cloned into PCR4-*Shh::lacZ*-H11 (Addgene, # 139098). Plasmid clones were validated by Sanger sequencing. pPyCAG-Pbase and pPB-CAG-rtTA-IN (Addgene, #60612) were kindly provided by M. Azim Surani. pSpCas9(BB)-2A-PuroR (px459) V2.0 (Addgene, #62988), psPAX2 (Addgene, #12260), pMD2.G (Addgene, #12259) and pLV-tetO-Tfap2c (Addgene, #70269) were obtained from Addgene. Primer sequences are listed in Supplementary Table 1.

gRNA cloning

For CRISPR-assisted insertion of PCR4-*Shh::lacZ*-H11 vector into the H11 locus, we used a gRNA that was previously designed for this purpose [20]. Synthesized oligonucleotides were annealed and cloned into px459 vector. The sequences of all gRNAs are listed in Supplementary Table 1.

Lentiviral constructs

To isolate pLV-tetO-Tfap2c lentivirus, HEK293T cells (cultured in DMEM with 10%

FBS, 1mM NEAA and penicillin–streptomycin) were transfected with 500μl Opti-MEM containing 30 μl TransIT®-LT1 Transfection Reagent (Mirus, MIR2305), 10μg pLV-tetO-Tfap2c, 7.5μg psPAX2 and 2.5μg pMD2.G. Viral supernatants were collected at 48 and 72 hours, concentrated using Amicon Ultra-15 centrifugal filter units with an Ultracel-30 membrane (Millipore, Billerica, MA, USA), filtered through a 0.45 μM filter, and stored at -80°C until used.

ESC and iPSC culture

Mouse ESCs and iPSCs maintained on γ-irradiated feeder MEFs (mouse embryonic fibroblast cells) in DMEM containing 15% FBS, penicillin-streptomycin, NEAA, β-mercaptoethanol, 1,000U/ml LIF (Peprotech, 300-05). Naïve ESCs were cultured in 2i+ LIF medium containing 1×N2B27, 1 μM PD0325901 (Reprocell, 04-0006-02), 3 μM CHIR99021 (Reprocell, 04-0004-02) and 1,000 U/ml LIF on wells coated with 0.01% poly-L-ornithine (Sigma-Aldrich, P3655) and 10 ng/ml laminin (Corning, 354232), as described previously [31].

For the formative state conversion, naïve ESCs and iPSCs were dissociated into single cells and plated on MEFs plate in 2i +LIF medium. Follow the protocol describe previously [5], the next day, medium was changed to formative medium (N2B27 medium containing 10 ng/ml bFGF, 10 ng/ml Activin A and 3 μM CHIR99021) for further culture. It usually takes about 2-4 passages for fully conversion. Formative ESCs and iPSCs were maintained in formative medium.

Derivation of genetically modified ESC clones

A *piggyBac* transposon expression system was used to stably integrate single copies of various vectors that were doxycycline (Tet) inducible. To establish the Tet-on expression system, pPBhCMV1-Prdm1-pA, pPBhCMV1-Prdm14-pA and pPBhCMV1-Nanog-pA vectors were transfected using TransIT®-LT1 Transfection Reagent (Mirus Bio™, MIR2304) into ESCs together with pPBCAG-rtTA and pPyCAG-PBase vectors. After 1 week of Geneticin™ (80 μg/ml; Life Technologies, 10131035) selection (all vectors except pPyCAG-PBase contain neomycin resistance genes), clones containing desired inserts were identified. Cells infected with pLV-tetO-Tfap2c lentivirus were selected by Zeocin® (InvivoGen, ant-zn-05) to identify transformants.

To construct ESC lines containing enhancer-reporters, the targeting vector PCR4-*Shh::lacZ*-H11 and H11 gRNA-expression plasmid px459 were transfected into cells bearing Tet-inducible *Nanog* along. After puromycin selection for 48 hours, single colonies were expanded and correctly targeted clones were identified using primers listed in Supplementary Table 1.

Microinjection for chimera production

The ESCs were microinjected to B6(Cg)-Tyr^{c-2J}/J blastocysts by standard methods in Cornell's Transgenic Core Facility [44].

PGCLC differentiation

For PGCLC differentiation, EpiLCs were induced as described previously [31]. Briefly, 1×10^5 ESCs were placed per well of a 12-well plate coated with 16.7 μg/mL bovine plasma fibronectin (Sigma-Aldrich, F1141) in N2B27 medium containing 20 ng/mL Activin A (Peprotech, 120-14E), 12 ng/mL bFGF (Gibco, 13256-029), and 1% knockout serum replacement (KSR, ThermoFisher, 10828010). The medium was changed the following day. A total of 48 hours after plating, 2×10^3 EpiLCs per well were seeded in a U-bottom 96-well plate (ThermoFisher Scientific, 268200) in GK15 medium (GMEM with 15% KSR, 0.1mM NEAA, 1mM sodium pyruvate, 0.1mM β-mercaptoethanol, penicillin-streptomycin, and 2mM L-glutamine) with Dox (Sigma, D9891). To induce PGCLCs from formative ESCs, after removing the feeder cells, 2×10^3 cells per well were seeded in the U-bottom 96-well plate. Alternatively, 8×10^4 cells per well were seeded in an untreated 12-well plate, or 1×10^6 cells were seeded into a 100mm Petri dish in GK15 medium. Transgenes were induced by addition of 2 μg/ml Dox at day 0 of PGCLC induction. Cells were cultured at 37 °C in a 5% CO_2-95% air atmosphere.

Flow cytometry

To flow sort PGCLCs, single-cell suspensions were stained with 1 μg/ml propidium iodide (PI, Invitrogen; P3566) and filtered through a 70 μm cell strainer. eGFP$^+$ cells were sorted using a BD FACSMelody 4-Way Sorter. Suspended cells were washed with PBS and fixed in 70% ethanol. For staining, cells were washed in MACS buffer (1×PBS, 0.5%BSA and 2mM EDTA) and blocked with 5% goat serum (Sigma; NS02L) in MACS

buffer at 4°C for 1 hour. Cells were then stained with PE (Phycoerythrin) mouse/rat anti-CD61 (1:500, Invitrogen; 12-0611-82) and Alexa Fluor® 647 mouse/human anti-SSEA1 antibodies (1:200, Biolegend; 125608) at 4°C for 1 hour. Cells were then washed in MACS buffer twice and filtered through a 35 μm cell strainer (Falcon; 352235). The stained cells were analyzed using a BD FACSymphony A3 cytometer. FACS data were analyzed using FlowJo™ v10.4 or flowCore package in R.

Cell aggregate cultures for differentiation into spermatogonia

Sorted PGCLCs were mixed with dissociated neonatal testicular cells from $Kit^{W/Wv}$ mice with 10:1 ratio and co-cultured for 2 days in a U bottom 96-well using GK15 medium. The aggregates were subsequently transferred to a trans-well plate in N2B27 medium containing 20 ng/ml bFGF, 40 ng/ml GDNF (PeproTech, 450-44), 1,000 U/ml LIF, 1×Insulin-Transferrin-Selenium (ThermoFisher Scientific, 41400045) and 15% KSR for additional 2 weeks. The aggregates were cultured at 37 °C in a 5% CO_2-95% air atmosphere.

Cell staining

ALP staining was performed using the Vector Red Alkaline Phosphatase Substrate Kit (Vector Laboratories, Burlingame, CA, USA) according to the manufacturer's directions. *LacZ* staining was performed the Senescence β-Galactosidase Staining Kit (Cell Signaling Technology, #9860) according to the manufacturer's directions.

For cell IF staining, the cells growing on cover slides were washed with PBS, fixed in 4% paraformaldehyde (PFA) for 10 min at RT, washed twice with PBS. For section IF staining, cell aggregates were fixed in 4% PFA and embedded in paraffin. The 6 μm sections were deparaffinized and performed antigen retrieval using sodium citrate buffer. The slides were incubated for 10 min at 37 °C in blocking buffer (PBS containing 10% normal goat serum). Then, they were incubated overnight in a humidified chamber at 4 °C with the following antibodies: Ddx4 (Abcam, ab13840, 1:500), Sox9 (Abcam, 1:100), Plzf (Thermo Fisher Scientific, 39987, 1:100), Nanos3 (Abcam, ab70001, 1:500), Zo1 (Thermo Fisher Scientific, 40-2200, 1:100), E-cadherin (1:100), Ssea1 (Millipore, MAB4301, 1:100) and Sox2 (Thermo Fisher Scientific, MA1-014, 1:100). After washing twice with PBS, the slides were incubated at 37 °C for 30 min with a 1:500

dilution of secondary antibody goat anti-rabbit IgG 594 (Molecular Probes, A11012) and/ or goat anti-mouse IgG 488 (Molecular Probes, A11001), then incubated at 37 °C for 5 min with 500 ng/mL of DAPI and mounted using Vectashield antifade mounting medium (Vector, H-1000).

Bisulfite sequencing

Methylation patterns in the differentially methylated domains (DMDs) of the paternally imprinted gene *H19* and maternally imprinted gene *Snrpn* were determined using bisulfite genomic sequencing as previously described [45]. Genomic DNA was isolated and bisulfite treatment was conducted with an EZ DNA methylation-lightning kits (ZYMO Research, D5030) according to the manufacturer's protocol. Bisulfate-converted DNA was subjected to PCR amplification of the *H19* and *Snrpn* DMDs. The primer sequences and PCR conditions for amplification are previously described [45]. PCR products were sub-cloned into the pGEM-T Easy vector (Promega, A1360) and sequenced. Sequences were determined and analyzed using BiQ Analyzer software [46].

Reverse transcription and qRT-PCR

Total RNA extraction from cells was performed by E.Z.N.A.® Total RNA Kit I (Omega BIO-TEK) according to the manufacturer's instructions or extracted with TRIzol® reagent (Thermo Fisher Scientific, 15596018). Reverse transcription was performed using a qScript™ cDNA SuperMix kit (Quantabio, 95048). qRT-PCR was performed with a C1000 Touch™ Thermal Cycler (Bio-Rad) amplification system using RT² SYBR® Green qPCR Mastermixes (Qiagen) and primer sets specific for each gene (Table S1). The relative levels of transcripts were calculated using the $2^{-\Delta\Delta C_T}$ method and gene expression levels were normalized to *Gapdh*.

RNA-seq and data analysis

Total RNAs were purified from naïve ESCs, EpiLCs, and PGCLCs (Stella-eGFP$^+$) at days 2, 4 and 6 (three biological replicates for each) using TRIzol® reagent. Purified RNAs were used for Illumina RNA-seq library preparation with NEBNext Ultra II Directional RNA Library Prep Kit (NEB, E7765), and a minimum of 20 million raw reads were obtained. The raw counts were then aligned to mouse mm10 genome with

ENSEMBL gene annotations. Differential analysis of RNA-seq data was done using the R package DESeq2. RRHO2 analysis was performed also using an R package [24]. k-Means clustering analysis was performed using iDEP 1.12 [47]. GSEA was performed to explore enriched pathways and interpret RNA-seq data using predefined gene sets from the Molecular Signatures Database (v.7.1) in R packages [48]. GSVA was carried out using the GSVA package in R [49]. Gene sets enriched in the GSEA were organized into a graphical network produced using EnrichmentMap [50] plugins in Cytoscape (v.3.10.1) [51], as described previously [52]. The online STRING database (version 12.0) was used for generation of protein interaction [53]. The upregulated differentially expressed genes (log2FC>2, $p<0.05$ and count>250) from PGCLCs on days 2, 4 and 6 compared to ESC were integrated for the network analysis.

Accession numbers

The RNA-seq data generated in this study have been deposited at GEO (Gene Expression Omnibus) and are publicly available as of the date of publication. The RNA-seq data of ESCs/EpiLCs/Ck_D2/4/6PGCLCs (GSE67259) and i*n vivo* germ cells (GSE74094 and GSE87644) were downloaded from the GEO database. The ESCs/EpiLCs/3TF_D2/4PGCLCs microarray dataset is accession # GSE43775. The GEO number for RNA-seq data of formative ESCs is GSE135989.

Software and websites

DESeq2: https://bioconductor.org/packages/release/bioc/html/DESeq2.html

SARTools: https://github.com/PF2-pasteur-fr/SARTools

RRHO2: https://github.com/RRHO2/RRHO2

GSVA: https://alexslemonade.github.io/refinebio-examples/

GSEA: https://www.gsea-msigdb.org/gsea/index.jsp

iDEP 1.1: http://149.165.154.220/idep11/

g:Profiler: https://biit.cs.ut.ee/gprofiler/gost

ShinyGo 0.80: http://bioinformatics.sdstate.edu/go80/

ggplot2: https://ggplot2.tidyverse.org

STRING 12.0: https://string-db.org

FlowJo: https://www.flowjo.com

BioRender: www.biorender.com

Cytoscape: https://cytoscape.org

EnrichmentMap: https://apps.cytoscape.org/apps/enrichmentmap

Image process and data statistics

Fluorescent images were captured by an Olympus XM10 camera. Bright-field images were captured by the Olympus SC30 camera. Cropping, color, and contrast adjustments were made with Adobe Photoshop CC 2024, using identical background adjustments for all images. All data were expressed as mean ± SEM. Statistical tests were carried out using one-way analysis of variance (ANOVA) followed by the Tukey's *post hoc* test with R. Graph generation was performed using R. The schematic diagrams were draw by BioRender.

Acknowledgement

The authors would like to thank R. Munroe and C. Abratte of Cornell's transgenic facility for generating the mice, A. Surani for Stella-eGFP ESCs and several plasmids, and J.C. Bloom for OG2 iPSCs. This work is supported by a grant from the National Institutes of Health (R01 HD082568 to JCS) and contract CO29155 from the NY State Stem Cell Program (NYSTEM). X.D. was supported by a postdoctoral fellowship from the Empire State Stem Cell Fund through New York State Department of Health contract no. C30293GG.

Competing interests

The authors declare no competing or financial interests.

References

1. Ding, X. & Schimenti, J. C. Strategies to identify genetic variants causing infertility. *Trends Mol. Med.* **27**, 792–806 (2021).

2. Saitou, M. & Hayashi, K. Mammalian in vitro gametogenesis. *Science* **374**, eaaz6830 (2021).

3. Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4**, 487–492 (2009).

4. Kinoshita, M. *et al.* Capture of Mouse and Human Stem Cells with Features of Formative Pluripotency. *Cell Stem Cell* **28**, 453-471.e8 (2021).

5. Yu, L. *et al.* Derivation of intermediate pluripotent stem cells amenable to primordial germ cell specification. *Cell Stem Cell* **28**, 550-567.e12 (2021).

6. Luo, Q. *et al.* Epiblast-like stem cells established by Wnt/β-catenin signaling manifest distinct features of formative pluripotency and germline competence. *Cell Rep.* **42**, 112021 (2023).

7. Wang, X. *et al.* Formative pluripotent stem cells show features of epiblast cells poised for gastrulation. *Cell Res.* **31**, 526–541 (2021).

8. 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).

9. Nakaki, F. *et al.* Induction of mouse germ-cell fate by transcription factors in vitro. *Nature* **501**, 222–226 (2013).

10. Zhang, M. *et al.* Esrrb Complementation Rescues Development of Nanog-Null Germ Cells. *Cell Rep.* **22**, 332–339 (2018).

11. Murakami, K. *et al.* NANOG alone induces germ cells in primed epiblast in vitro by activation of enhancers. *Nature* **529**, 403–407 (2016).

12. Vojtek, M., Zhang, J., Sun, J., Zhang, M. & Chambers, I. Differential repression of *Otx2* underlies the capacity of NANOG and ESRRB to induce germline entry. *BioRxiv* (2021) doi:10.1101/2021.06.14.448276.

13. Acampora, D. *et al.* Forebrain and midbrain regions are deleted in Otx2-/- mutants due to a defective anterior neuroectoderm specification during gastrulation. *Development* **121**, 3279–3290 (1995).

14. Shamblott, M. J. *et al.* Derivation of pluripotent stem cells from cultured human primordial germ cells. *Proc Natl Acad Sci USA* **95**, 13726–13731 (1998).

15. Matsui, Y., Zsebo, K. & Hogan, B. L. Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* **70**, 841–847 (1992).

16. Ohta, H. *et al.* In vitro expansion of mouse primordial germ cell-like cells recapitulates an epigenetic blank slate. *EMBO J.* **36**, 1888–1907 (2017).

17. Ishikura, Y. *et al.* In vitro reconstitution of the whole male germ-cell development from mouse pluripotent stem cells. *Cell Stem Cell* **28**, 2167-2179.e9 (2021).

18. Murase, Y. *et al.* Long-term expansion with germline potential of human primordial germ cell-like cells in vitro. *EMBO J.* **39**, e104929 (2020).

19. Payer, B. *et al.* Generation of stella-GFP transgenic mice: a novel tool to study germ cell development. *Genesis* **44**, 75–83 (2006).

20. Kvon, E. Z. *et al.* Comprehensive in vivo interrogation reveals phenotypic impact of human enhancer variants. *Cell* **180**, 1262-1271.e15 (2020).

21. Zhang, J. *et al.* OTX2 restricts entry to the mouse germline. *Nature* **562**, 595–599 (2018).

22. Yamashiro, C. *et al.* Persistent requirement and alteration of the key targets of PRDM1 during primordial germ cell development in mice. *Biol. Reprod.* **94**, 7 (2016).

23. Sasaki, K. *et al.* Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells. *Cell Stem Cell* **17**, 178–194 (2015).

24. Cahill, K. M., Huo, Z., Tseng, G. C., Logan, R. W. & Seney, M. L. Improved identification of concordant and discordant gene expression signatures using an updated rank-rank hypergeometric overlap approach. *Sci. Rep.* **8**, 9588 (2018).

25. Schemmer, J. *et al.* Transcription factor TFAP2C regulates major programs required for murine fetal germ cell maintenance and haploinsufficiency predisposes to teratomas in male mice. *PLoS ONE* **8**, e71113 (2013).

26. Magnúsdóttir, E. *et al.* A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nat. Cell Biol.* **15**, 905–915 (2013).

27. Irie, N. *et al.* DMRT1 regulates human germline commitment. *Nat. Cell Biol.* **25**, 1439–1452 (2023).

28. Vojtek, M., Zhang, J., Sun, J., Zhang, M. & Chambers, I. Differential repression of Otx2 underlies the capacity of NANOG and ESRRB to induce germline entry. *Stem Cell Reports* **17**, 35–42 (2022).

29. Kubota, H., Avarbock, M. R. & Brinster, R. L. Growth factors essential for self-renewal and expansion of mouse spermatogonial stem cells. *Proc Natl Acad Sci USA* **101**, 16489–16494 (2004).

30. Buaas, F. W. *et al.* Plzf is required in adult male germ cells for stem cell self-renewal. *Nat. Genet.* **36**, 647–652 (2004).

31. Hayashi, K. & Saitou, M. Generation of eggs from mouse embryonic stem cells and induced pluripotent stem cells. *Nat. Protoc.* **8**, 1513–1524 (2013).

32. Choi, H. W. *et al.* Distinct enhancer activity of oct4 in naive and primed mouse pluripotency. *Stem Cell Reports* **7**, 911–926 (2016).

33. Aramaki, S. *et al.* A mesodermal factor, T, specifies mouse germ cell fate by directly activating germline determinants. *Dev. Cell* **27**, 516–529 (2013).

34. Tischler, J. *et al.* Metabolic regulation of pluripotency and germ cell fate through α-ketoglutarate. *EMBO J.* **38**, (2019).

35. Hackett, J. A. *et al.* Tracing the transitions from pluripotency to germ cell fate with CRISPR screening. *Nat. Commun.* **9**, 4292 (2018).

36. Naitou, Y. *et al.* Dual role of Ovol2 on the germ cell lineage segregation during gastrulation in mouse embryogenesis. *Development* **149**, (2022).

37. Yin, Y. *et al.* A noncanonical role of NOD-like receptor NLRP14 in PGCLC differentiation and spermatogenesis. *Proc Natl Acad Sci USA* **117**, 22237–22248 (2020).

38. Eguizabal, C. *et al.* Complete meiosis from human induced pluripotent stem cells. *Stem Cells* **29**, 1186–1195 (2011).

39. Ishikura, Y. *et al.* In Vitro Derivation and Propagation of Spermatogonial Stem Cell Activity from Mouse Pluripotent Stem Cells. *Cell Rep.* **17**, 2789–2804 (2016).

40. Bratt-Leal, A. M., Carpenedo, R. L. & McDevitt, T. C. Engineering the embryoid body microenvironment to direct embryonic stem cell differentiation. *Biotechnol. Prog.* **25**, 43–51 (2009).

41. Smith, A. Formative pluripotency: the executive phase in a developmental continuum. *Development* **144**, 365–373 (2017).

42. Ding, X. *et al.* In vivo versus in silico assessment of potentially pathogenic missense variants in human reproductive genes. *Proc Natl Acad Sci USA* **120**, e2219925120 (2023).

43. Ding, X. & Schimenti, J. C. Female infertility from oocyte maturation arrest: assembling the genetic puzzle. *EMBO Mol. Med.* **15**, e17729 (2023).

44. Luo, Y. *et al.* Hypersensitivity of primordial germ cells to compromised replication-associated DNA repair involves ATM-p53-p21 signaling. *PLoS Genet.* **10**, e1004471 (2014).

45. Zhou, Q. *et al.* Complete Meiosis from Embryonic Stem Cell-Derived Germ Cells In Vitro. *Cell Stem Cell* **18**, 330–340 (2016).

46. Bock, C. *et al.* BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics* **21**, 4067–4068 (2005).

47. Ge, S. X., Son, E. W. & Yao, R. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics* **19**, 534 (2018).

48. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**, 15545–15550 (2005).

49. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).

50. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS ONE* **5**, e13984 (2010).

51. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

52. Reimand, J. *et al.* Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.* **14**, 482–517 (2019).

53. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).

Figure legends

Figure 1. Construction of 4TF-induced PGCLC differentiation platform

A. Percentages of day 6 PGCLCs (Stella-eGFP$^+$ cells) differentiated from Dox-inducible ESCs harboring distinct combinations of TFs. Each individual clonal line is represented by a dot. B. The process for constructing the 4TF-inducible ESCs. The bolded fonts highlight the 4TFs and rtTA. C. Schema illustrating the 4TF-induced PGCLC differentiation from ESCs harboring the Stella-eGFP reporter. D. Induction of PGCLCs (Stella-eGFP$^+$ cells) in the presence or absence of Dox in EBs during the 6-day period. E. FACS patterns of eGFP$^+$ cells (indicated by dashed lines) during differentiation. F. Relative fluorescence density of H3K9me2 and K3K27me3 in GFP$^-$, GFP dim and GFP bright cells from Dox-induced day 6 EBs. G. Relative mRNA expression of *Dnmt3a* and *Dnmt3b* in ESCs, EpiLCs and day 6 EBs. H. Bisulphite sequence analysis of methylated cytosine in the DMRs of the imprinted genes in GFP$^+$ and GFP$^-$ cells in day 6 EBs. White and black circles represent unmethylated and methylated cytosines, respectively. Data in G are represented as the mean ± SEM. Data in F and G were analyzed using one-way ANOVA with *Tukey's post hoc* test. Scale bar in D represents 100 μm.

Figure 2. 4TF-induced PGCLCs share similar transcriptomes with cytokine- and 3TF-induced PGCLCs

A. PCA of indicated transcriptomes from 4TF- and cytokine (Ck)-induced PGCLCs, *in vivo* germ cells at indicated times of gestation, EpiLCs and ESCs. "D" = days of induction. PGCLCs were isolated by flow sorted for diagnostic markers. B. Heatmap of the Pearson correlation coefficients between PGCLCs. C. Comparison of differential expression profiles between D6PGCLC (x-axis) and D2/4PGCLC (y-axis) with RRHO2 maps. The extent of overlap transcriptome was represented by each heatmap colored based on -log10(p-value) from a hypergeometric test. Each map shows the extent of overlapped upregulated genes, versus ESCs or EpiLCs, in the bottom left corner,

whereas shared downregulated genes are displayed in the top right corners. D. Percentage of concordant genes encompassing both up and down regulated genes compared with D6PGCLCs. E. GSVA score of Wnt signaling pathways in D2PGCLC between Ck-, 3TF- and 4TF-induced conditions (Ck_D2, 3TF_D2 and 4TF_D2). ESC, EpiLC and PGCLC in each group were used to calculate GSVA score and only the scores of D2PGCLC were displayed. F. Heatmap of selected genes associated with WNT signaling and targets.

Figure 3. Gene expression profiling of 4TF-induced PGCLCs

A. Heatmap of k-means clustering of variably expressed genes in ESCs, EpiLCs, and D2-D6 4TF-induced PGCLCs (n = 500, k = 4). Genes were grouped into four clusters ("C1-4") on the basis of expression similarity. B. Biological process GO term analysis of Cluster 4 (C4). Terms of germ cell development (highlighted by red) was highly enriched for terms in C4. C. Heatmaps showing the average GSVA enrichment score of selected germ cell development pathways. D. Representative GSEA enrichment plots depicting significant enrichment of germ cell signatures in D6PGCLCs compared with ESCs. E. Heatmap of representative gene expression of pluripotency and germ cells among ESCs, EpiLCs and PGCLCs. F. Further development of 4TF_D6PGCLC by aggregated with testicular somatic cells isolated from $Kit^{w/wv}$ mice. The cultured aggregates were subjected to IF and quantification. Data in lower panel of F are represented as the mean ± SEM, and the single dot represents the percentage of double positive cells in a cluster of $DDX4^{+}$ cells. Scale bar in F represents 50 μm.

Figure 4. Scale-up and long-term differentiation of PGCLCs

A. Representative FACS pattern of $Itgb3^{+}/Ssea1^{+}$ and $Itgb3^{+}$/Stella-$eGFP^{+}$ cells in ESCs, EpiLCs and day 6 EBs from U bottom 96-well plates, 12-well non-treated plates and 100mm petri dishes. B. Quantification of PGCLC populations in day 6 EBs at different plates/dish. C. Boxplots showing the distribution of the diameter of day 6 EBs in different plates/dish. D. FACS pattern of $Itgb3^{+}/Ssea1^{+}$ and $Itgb3^{+}$/Stella-$eGFP^{+}$ cells

in day 20 and 43 EBs cultured in 12-well non-treated plates. Data in B and D are represented as the mean ± SEM. Data in B and C were analyzed using one-way ANOVA with *Tukey's post hoc* test. Scale bar in C represents 200 μm.

Figure 5. Direct differentiation of PGCLCs from formative ESCs

A. Representative morphologies of naïve ESCs, formative ESCs and EpiLCs. B. Immunofluorescence of tight junction protein ZO1 on naïve ESCs, formative ESCs and EpiLCs. C. GFP expression in naïve and formative OG2 iPSCs, and EpiLC differentiated from naïve OG2 iPSCs. D. Representative FACS pattern of Itgb3^{+}/Ssea1^{+} and Itgb3^{+}/Stella-eGFP^{+} cells in formative ESCs, EBs with Dox treatment. Scale bars in A and C represent 100 μm and in B represents 50 μm.

Figure 1

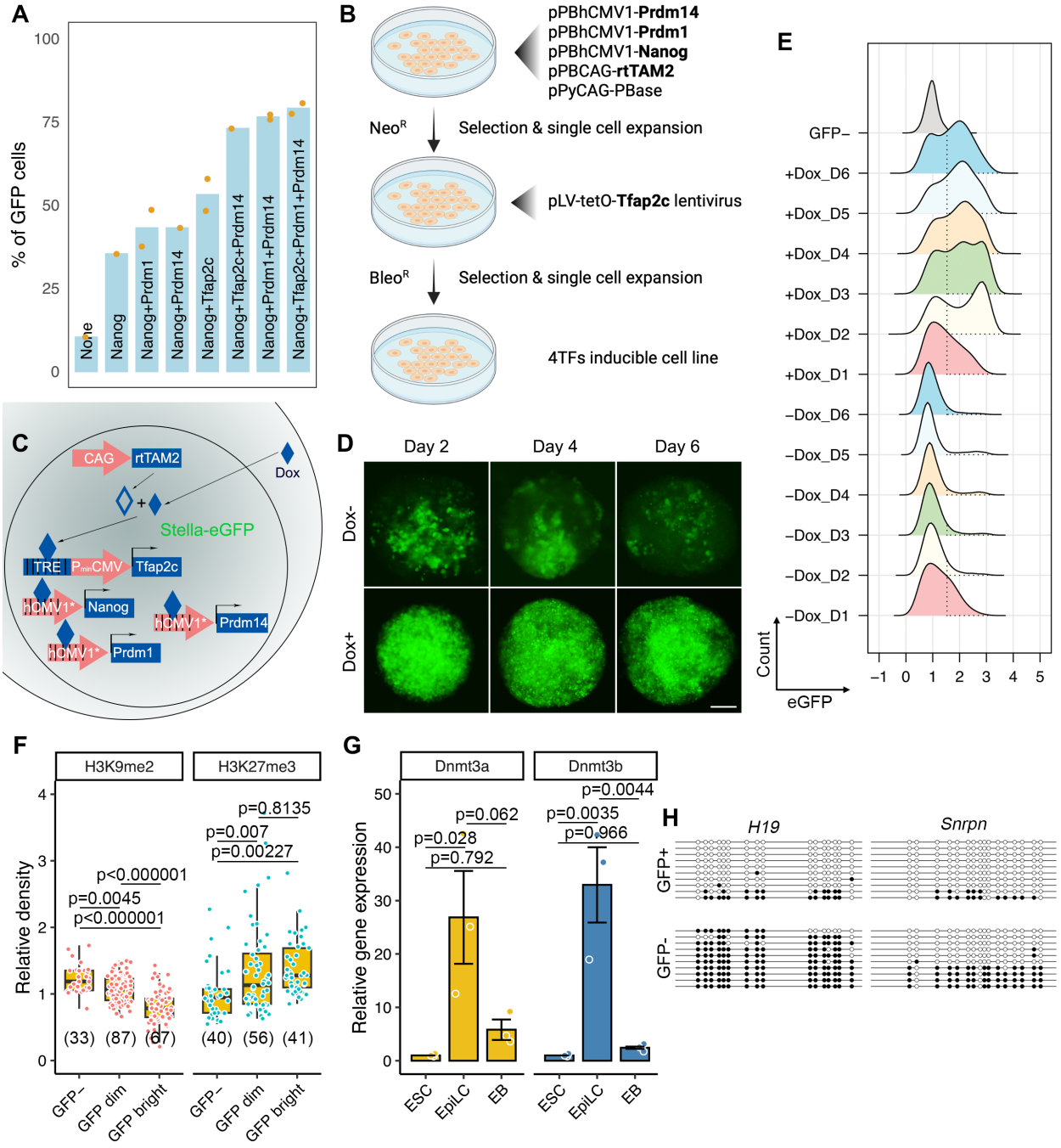

A
% of GFP cells
100
75
50
25
0
None
Nanog
Nanog+Prdm1
Nanog+Prdm14
Nanog+Tfap2c
Nanog+Tfap2c+Prdm14
Nanog+Prdm1+Prdm14
Nanog+Tfap2c+Prdm1+Prdm14
B
pPBhCMV1-Prdm14
pPBhCMV1-Prdm1
pPBhCMV1-Nanog
pPBCAG-rtTAM2
pPyCAG-PBase
NeoR
Selection & single cell expansion
pLV-tetO-Tfap2c lentivirus
BleoR
Selection & single cell expansion
4TFs inducible cell line
C
CAG
rtTAM2
Dox
Stella-eGFP
TRE
PminCMV
Tfap2c
hCMV1*
Nanog
Prdm14
Prdm1
D
Day 2
Day 4
Day 6
Dox-
Dox+
E
GFP–
+Dox_D6
+Dox_D5
+Dox_D4
+Dox_D3
+Dox_D2
+Dox_D1
–Dox_D6
–Dox_D5
–Dox_D4
–Dox_D3
–Dox_D2
–Dox_D1
Count
eGFP
–1 0 1 2 3 4 5
F
H3K9me2
H3K27me3
Relative density
p<0.000001
p=0.0045
p<0.000001
p=0.8135
p=0.007
p=0.00227
(33) (87) (67)
(40) (56) (41)
GFP– GFP dim GFP bright
GFP– GFP dim GFP bright
G
Dnmt3a
Dnmt3b
Relative gene expression
p=0.062
p=0.028
p=0.792
p=0.0044
p=0.0035
p=0.966
ESC EpiLC EB
ESC EpiLC EB
H
H19
Snrpn
GFP+
GFP-

Figure 2

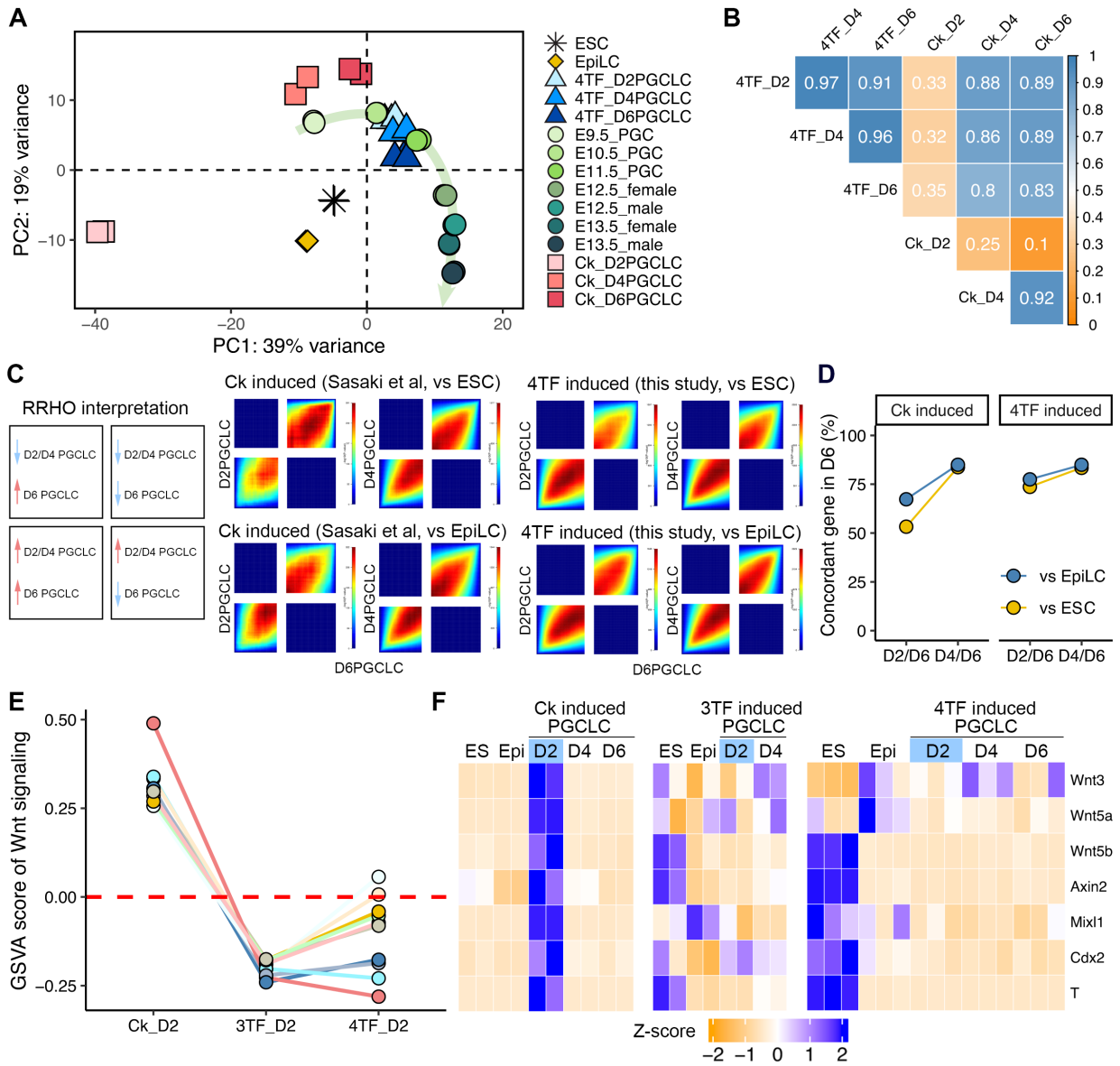

A
PC2: 19% variance
PC1: 39% variance
ESC
EpiLC
4TF_D2PGCLC
4TF_D4PGCLC
4TF_D6PGCLC
E9.5_PGC
E10.5_PGC
E11.5_PGC
E12.5_female
E12.5_male
E13.5_female
E13.5_male
Ck_D2PGCLC
Ck_D4PGCLC
Ck_D6PGCLC
B
4TF_D4
4TF_D6
Ck_D2
Ck_D4
Ck_D6
4TF_D2 0.97 0.91 0.33 0.88 0.89
4TF_D4 0.96 0.32 0.86 0.89
4TF_D6 0.35 0.8 0.83
Ck_D2 0.25 0.1
Ck_D4 0.92
C
RRHO interpretation
D2/D4 PGCLC
D6 PGCLC
Ck induced (Sasaki et al, vs ESC)
4TF induced (this study, vs ESC)
Ck induced (Sasaki et al, vs EpiLC)
4TF induced (this study, vs EpiLC)
D2PGCLC
D4PGCLC
D6PGCLC
D
Ck induced
4TF induced
Concordant gene in D6 (%)
vs EpiLC
vs ESC
D2/D6 D4/D6
E
GSVA score of Wnt signaling
Ck_D2
3TF_D2
4TF_D2
F
Ck induced PGCLC
3TF induced PGCLC
4TF induced PGCLC
ES Epi D2 D4 D6
Wnt3
Wnt5a
Wnt5b
Axin2
Mixl1
Cdx2
T
Z-score

Figure 3

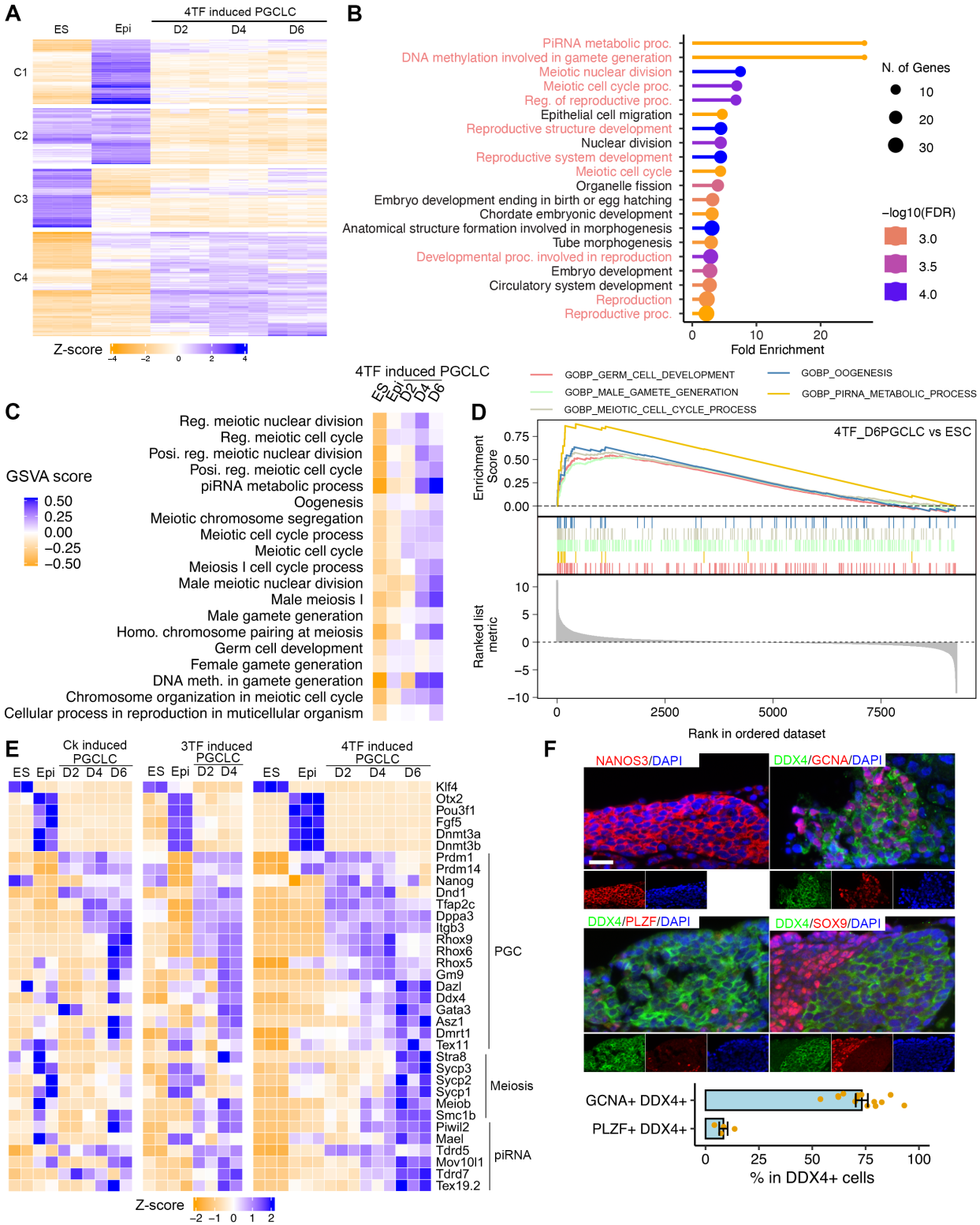

A
ES
Epi
4TF induced PGCLC
D2
D4
D6
C1
C2
C3
C4
Z-score
−4
−2
0
2
4
B
PiRNA metabolic proc.
DNA methylation involved in gamete generation
Meiotic nuclear division
Meiotic cell cycle proc.
Reg. of reproductive proc.
Epithelial cell migration
Reproductive structure development
Nuclear division
Reproductive system development
Meiotic cell cycle
Organelle fission
Embryo development ending in birth or egg hatching
Chordate embryonic development
Anatomical structure formation involved in morphogenesis
Tube morphogenesis
Developmental proc. involved in reproduction
Embryo development
Circulatory system development
Reproduction
Reproductive proc.
0
10
20
Fold Enrichment
N. of Genes
10
20
30
−log10(FDR)
3.0
3.5
4.0
C
4TF induced PGCLC
ES
Epi
D2
D4
D6
GSVA score
0.50
0.25
0.00
−0.25
−0.50
Reg. meiotic nuclear division
Reg. meiotic cell cycle
Posi. reg. meiotic nuclear division
Posi. reg. meiotic cell cycle
piRNA metabolic process
Oogenesis
Meiotic chromosome segregation
Meiotic cell cycle process
Meiotic cell cycle
Meiosis I cell cycle process
Male meiotic nuclear division
Male meiosis I
Male gamete generation
Homo. chromosome pairing at meiosis
Germ cell development
Female gamete generation
DNA meth. in gamete generation
Chromosome organization in meiotic cell cycle
Cellular process in reproduction in muticellular organism
D
GOBP_GERM_CELL_DEVELOPMENT
GOBP_MALE_GAMETE_GENERATION
GOBP_MEIOTIC_CELL_CYCLE_PROCESS
GOBP_OOGENESIS
GOBP_PIRNA_METABOLIC_PROCESS
4TF_D6PGCLC vs ESC
Enrichment Score
0.75
0.50
0.25
0.00
Ranked list metric
10
5
0
−5
−10
0
2500
5000
7500
Rank in ordered dataset
E
Ck induced PGCLC
3TF induced PGCLC
4TF induced PGCLC
ES
Epi
D2
D4
D6
Klf4
Otx2
Pou3f1
Fgf5
Dnmt3a
Dnmt3b
Prdm1
Prdm14
Nanog
Dnd1
Tfap2c
Dppa3
Itgb3
Rhox9
Rhox6
Rhox5
Gm9
Dazl
Ddx4
Gata3
Asz1
Dmrt1
Tex11
Stra8
Sycp3
Sycp2
Sycp1
Meiob
Smc1b
Piwil2
Mael
Tdrd5
Mov10l1
Tdrd7
Tex19.2
PGC
Meiosis
piRNA
Z-score
−2
−1
0
1
2
F
NANOS3/DAPI
DDX4/GCNA/DAPI
DDX4/PLZF/DAPI
DDX4/SOX9/DAPI
GCNA+ DDX4+
PLZF+ DDX4+
0
25
50
75
100
% in DDX4+ cells

Figure 4

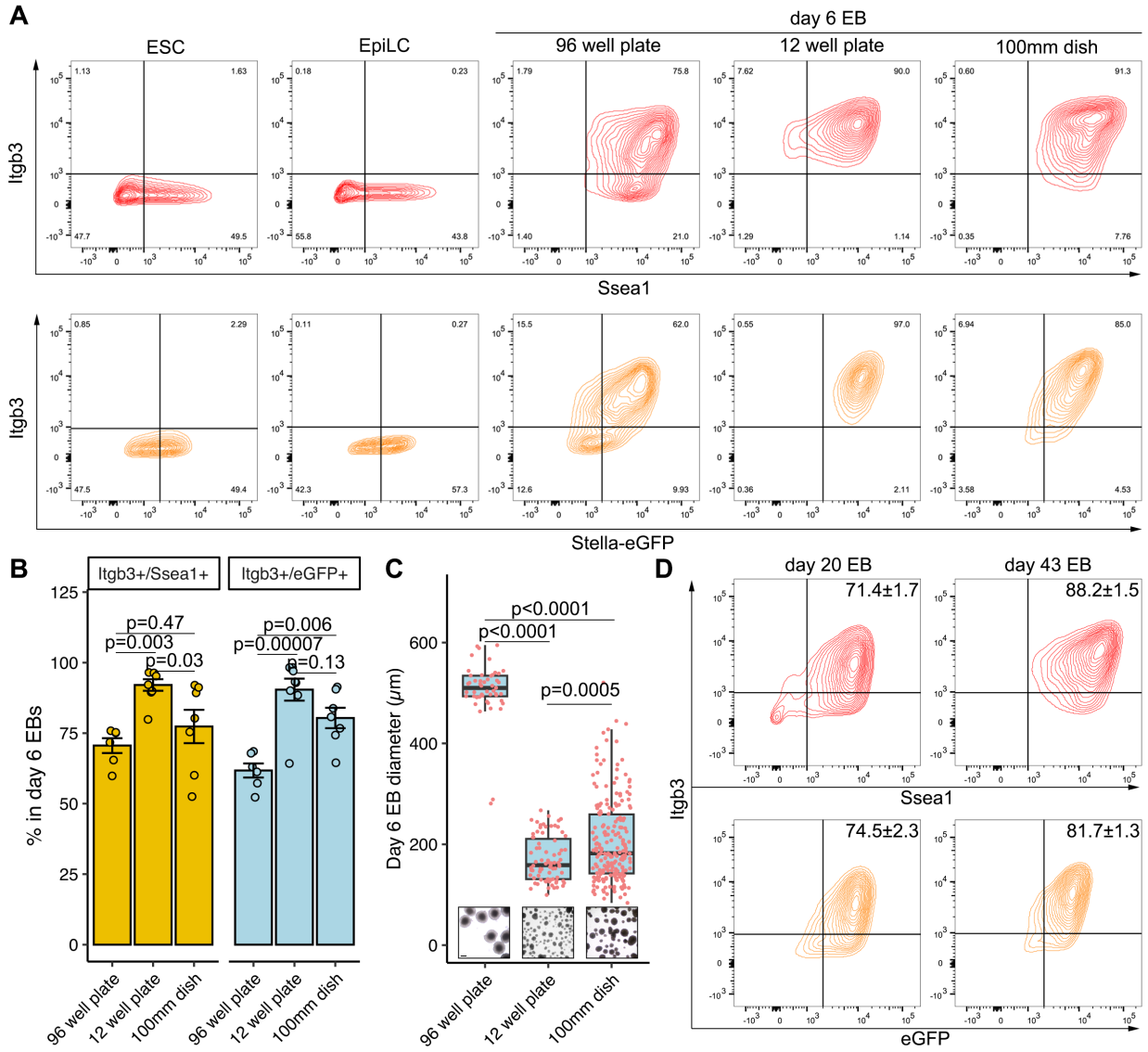
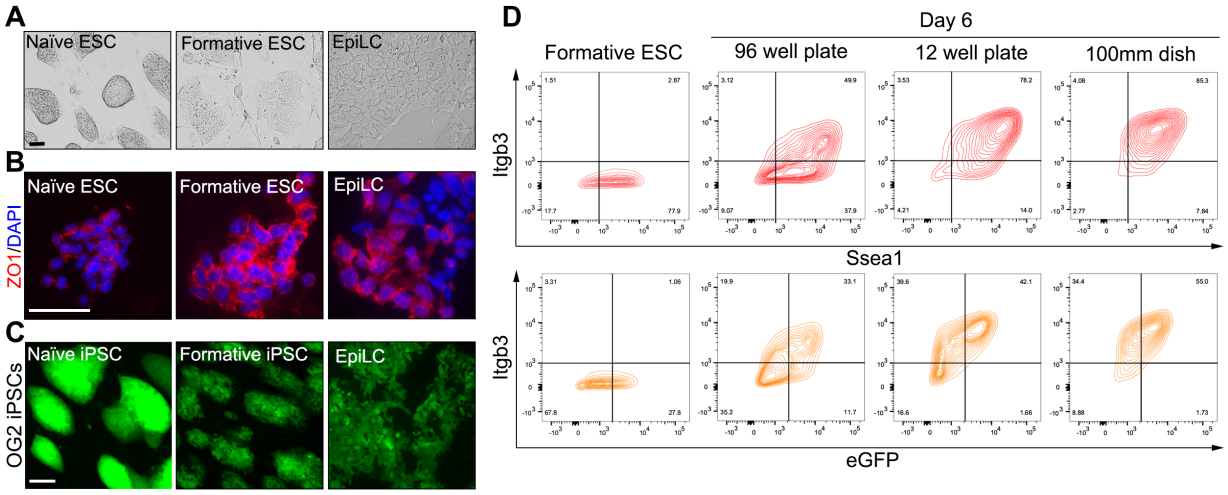

A
day 6 EB
ESC
EpiLC
96 well plate
12 well plate
100mm dish
Itgb3
Ssea1
Stella-eGFP
B
Itgb3+/Ssea1+
Itgb3+/eGFP+
% in day 6 EBs
p=0.47
p=0.003
p=0.03
p=0.006
p=0.00007
p=0.13
96 well plate
12 well plate
100mm dish
C
Day 6 EB diameter (μm)
p<0.0001
p<0.0001
p=0.0005
D
day 20 EB
day 43 EB
71.4±1.7
88.2±1.5
74.5±2.3
81.7±1.3
eGFP

Figure 5


A
Naïve ESC
Formative ESC
EpiLC
B
ZO1/DAPI
Naïve ESC
Formative ESC
EpiLC
C
OG2 iPSCs
Naïve iPSC
Formative iPSC
EpiLC
D
Day 6
Formative ESC
96 well plate
12 well plate
100mm dish
Itgb3
Ssea1
Itgb3
eGFP