Scalable primordial germ cell-like-cell platform for functional genomics identifies epigenetic fertility modifiers

Liangdao Li[1#], Jingyi Gao[1], Dain Yi[1], Alex P. Sheft[1], John C. Schimenti[1,2*] and Xinbao Ding[1#]

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

2 Cornell University, College of Agriculture and Life Sciences, Department of Molecular Biology and Genetics, Ithaca, NY, 14853, USA

* 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 large-scale genetic screens has been a limitation. Here, we demonstrated that simultaneous overexpression of 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. *Nanog* overexpression enhanced the PGC regulatory network, suppressed differentiation of somatic lineages, and maintained PGC fate. PGCLCs generated in this manner have transcriptomes similar to *in vivo* PGCs and are more advanced than cytokine-induced PGCLCs. These differentiated PGCLCs could be sustained over prolonged periods of culture and could differentiate into spermatogonia-like cells *in vitro*. We exploited this platform to conduct a bulk epigenome-wide CRISPRi screen to identify genes that decreased the efficiency of PGCLC formation when downregulated. One of the identified candidates was *Ncor2*, a transcriptional repressor that acts via recruitment of Class I and Class IIa histone deacetylases (HDACs). Consistently with this finding, the HDAC inhibitors valproic acid (VPA; an anti-convulsant) and sodium butyrate (SB; a widely-used dietary supplement) suppressed PGCLC differentiation. Furthermore, exposure of developing mouse embryos to SB or VPA caused hypospermatogenesis. This work demonstrates the feasibility of this platform to efficiently conduct large-scale functional screens of genes, chemicals, or other agents that may impact germline development.

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 (also known as 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; “2i+LIF” medium). EpiSCs self-renew when exposed to Activin A and 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). In this 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 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 (8,11). The overexpression of *Nanog* can promote efficient PGCLC differentiation by binding endogenous enhancers of *Prdm1* and *Prdm14*, even without BMP4 and associated cytokines (12). Furthermore, Nanog overexpression induces cytokine-free PGCLC specification by repressing *Otx2*, encoding a TF essential for anterior neuroectoderm specification (13,14).

Promoting cell proliferation is another strategy for producing a large number of PGCs/ PGCLCs. However, extended culture of mouse PGCs promotes dedifferentiation into pluripotent embryonic germ cells (EGCs) (15,16). 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 (17–19). 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).

While several genetic causes of infertility have been identified (20,21), pinpointing specific epigenetic factors that perturb gene expression is more challenging. Sperm counts have been reported to have declined worldwide since the mid 1900s - possibly due to environmental factors introduced after the industrial revolution - but there are major concerns regarding data analysis and interpretations underlying these claims (22). Exposure to certain chemicals during fetal development, such as endocrine disruptors, can cause transgenerational alterations in epigenetic landscapes and hinder reproductive function and fertility in both males and females (23–26). During PGC development, there is a global loss of repressive histone marks such as H3K9me2 and an increase of active histone marks such as H3K4me3 and H3K27ac (27,28). These dynamic epigenetic landscapes are crucial, and exposure to specific environmental toxins or chemicals that interfere with the epigenetic profile may impact PGC development (29,30). Histone acetylation and deacetylation regulate gene expression and repression, respectively. Nuclear receptor co-repressor 2 (NCOR2) is a well-studied nuclear receptor co-repressor that facilitates histone deacetylation (31,32). There are 18 histone deacetylases (HDACs) that are evolutionally conserved across all eukaryotes and can be divided into four classes (33). Both Class I (HDAC1, 2, and 3) and Class IIa (HDAC4, 5, 7, and 9) are known to interact with NCOR2 complexes to help down-regulate gene expression (31,32). Trichostatin A, an inhibitor of Class I and II HDACs (HDACi), promotes PGC dedifferentiation into EGCs, suggesting that histone deacetylation is important for maintaining the PGC state (34).

Here, we report a simple, scalable and economical approach to induce the differentiation of PGCLCs from formative ESCs and EpiLCs at scale by overexpressing four TFs (*Prdm1*, *Prdm14*, *Tfap2c* and *Nanog*). The resulting PGCLCs exhibit transcriptional and epigenetic profiles resembling those of late (E11.5) PGCs, and could be further differentiated into spermatogonia-like cells. Using this platform, we performed

143 a CRISPR inhibition (CRISPRi) screen and identified several epigenetic factors that
144 impacted the efficiency of PGCLC specification and proliferation, including *Ncor2*.
145 Furthermore, we found that the exposure to Class I and Class IIa HDAC inhibitors
146 attenuated PGCLC formation *in vitro* and germ cell formation in mice exposed during
147 development. Our *in vitro* differentiation system holds promise for facilitating high-
148 throughput functional genomic studies of germ cell development, characterization of
149 potential infertility-causing genetic variants, and improving IVG.
150

151 ## Results

152

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

155 To establish an efficient, cytokine-free system for generating PGCLCs, we began
156 with a mouse ESC line containing enhanced GFP (eGFP) under the regulatory control
157 of *Stella* (also called *Dppa3/Pgc7*), a PGC marker (35), and validated the ability of this
158 line to form germline-transmitting chimeric mice after blastocyst microinjection (Fig.
159 S1A-D), demonstrating its pluripotency. The Stella-eGFP reporter enables real-time
160 visualization of PGCLC development from ESCs *in vitro*. Next, based on the finding that
161 *Nanog* promotes transition of ESCs to PGCLCs by binding the enhancer regions of
162 *Prdm1* and *Prdm14* and inducing their expression (12), we introduced a stably-
163 integrated doxycycline (Dox)-inducible *Nanog*. To test the effectiveness of this
164 transgene, we utilized a high efficiency enhancer activity detection system (36) to derive
165 ESC lines containing targeted integrations of a lacZ reporter under the control of *Prdm1*
166 or *Prdm14* enhancers into the safe harbor H11 locus. Dox treatment of embryoid bodies
167 (EBs; ESCs>EpiLCs>EBs) derived from these lines triggered β-galactosidase
168 production, indicating the effectiveness of NANOG-stimulated transcription (Fig. S1E-H).
169 Besides promoting expression of *Prdm1* and *Prdm14*, NANOG represses *Otx2,* a TF
170 that promotes differentiation of pluripotent cells towards a neuroectodermal fate (12,14).
171 Mutant cells lacking OTX2 enter the germline from an EpiLC state with high efficiency,
172 even in the absence of the essential PGCLC-promoting cytokine signals (37). In

173 summary, NANOG plays a crucial role in PGCLC differentiation by promoting the
174 expression of PGC core regulatory TF-encoding genes *Prdm1* and *Prdm14*, while
175 simultaneously inhibiting the somatic fate inducer *Otx2* (Fig. S1I).

176 Overexpression of *Prdm1*, *Prdm14* and *Tfap2c* (hereafter called "3TF") can induce
177 PGCLCs without cytokines, however, the efficiency of this differentiation system is
178 relatively low with no increase in the fraction of cells becoming PGCLCs after day 4
179 post-induction (9,38). Because *Nanog* can induce PGCLCs independently of BMP
180 signaling and performs better when combined with cytokines (12), we hypothesized that
181 simultaneous overexpression of NANOG and downstream targets (i.e. *Prdm1*, *Prdm14*,
182 *Tfap2c* and *Nanog*, hereafter called "4TF") in EpiLCs might render PGCLCs robustly.
183 We isolated several Dox-inducible clonal cell lines harboring combinations of these TFs
184 in Stella-eGFP ESCs that also contained rtTAM2 (reverse tetracycline-controlled
185 transactivator mutant; Fig. 1A and B). The resulting ESCs, cultured in 2i+LIF medium,
186 were differentiated to EpiLCs by 2 days of Activin A and bFGF exposure. To induce
187 PGCLC differentiation, the EpiLCs aggregates were treated with Dox instead of
188 cytokines. We first tested whether ectopic expression of one or more TFs drove EpiLC-
189 to-PGCLC differentiation in U-bottom 96-well plates. Overexpression of *Nanog* alone led
190 to ~35% *Stella-eGFP*$^+$ cells in day 6 EBs. This percentage increased when *Nanog* was
191 combined with one of the core TF transgenes (i.e., *Nanog*+*Prdm1*, *Nanog*+*Prdm14* and
192 *Nanog*+*Tfap2c*), even though these genes are upregulated by NANOG during PGCLC
193 differentiation (12). Remarkably, simultaneous overexpression of all 4TFs caused ~80%
194 of cells to become eGFP$^+$ (Figs. 1A-E).

195 *In vivo*, PGCs undergo substantial epigenetic reprogramming, characterized by the
196 loss of H3K9me2 and the acquisition of H3K27me3, as well as global erasure of DNA
197 methylation (2). To test the fidelity of epigenetic changes in PGCLCs produced in our
198 4TF system, we examined histone modifications, CpG methylation status of imprinted
199 genes, and capacity for dedifferentiation into pluripotent EGCs. Differentiated cells with
200 higher eGFP fluorescence displayed lower H3K9me2 and higher H3K27me3 labeling
201 (Fig. 1F). The mRNA levels of epigenetic modifiers *Dnmt3a* and *Dnmt3b* were increased
202 in EpiLCs and subsequent downregulated in EBs (Fig. 1G), which aligns with previous
203 findings (8). Bisulfite sequencing of PGCLCs (eGFP$^+$ cells) revealed reduced

204 methylation in differentially methylated regions for both paternally (*H19*) and maternally
205 (*Snrpn*) imprinted loci, while eGFP$^-$ cells (presumably somatic-like) exhibited a more
206 methylated state (Fig. 1H). In addition, PGCLCs could de-differentiated to EGCs *in vitro*
207 when transferred to 2i+LIF medium (Fig. S2). Collectively, our data indicate that
208 simultaneously overexpression of 4TFs effectively induces the differentiation of EpiLCs
209 into PGCLCs.

210 **Transcriptomic properties of 4TF-induced PGCLCs**

211 To examine the fidelity of our 4TF-induced PGCLCs, i.e., Stella-eGFP$^+$ cells (Fig.
212 S3), RNA sequencing (RNA-seq) was performed and the transcriptomes were
213 compared to those of *in vivo* germ cells (encompassing PGCs at E9.5, E10.5, and
214 E11.5 and male/female germ cells at E12.5 and E13.5) and cytokine (Ck)-induced
215 PGCLCs (those at days 2, 4 and 6 after induction, referred to as Ck_D2/4/6PGCLC)
216 (17,39,40). Principal component analysis (PCA) showed that 4TF PGCLCs at days 2, 4
217 and 6 (4TF_D2/4/6PGCLC) clustered tightly as a group separate from naïve ESCs,
218 EpiLCs and Ck_D2PGCLCs, representing a transient state towards the acquisition of a
219 PGC-like state from EpiLC/ epiblast states (9) (Fig. 2A). Notably, 4TF-induced PGCLCs
220 grouped closely with PGCs at E10.5 and E11.5, while Ck_D4/6PGCLCs grouped
221 closely with E9.5 and E10.5 PGCs (Fig. 2A). These data suggest that 4TF-induced
222 PGCLCs are more akin to late PGCs than Ck_D2/4/6PGCLCs. We next applied two-
223 sided rank–rank hypergeometric overlap (RRHO) analysis to compare the gene
224 expression signatures between samples in a threshold-free manner (41). We observed
225 robust overlap in both up- and down-regulated transcripts amongst differentiating cells
226 in both Ck- and 4TF-induced groups relative to either ESCs or EpiLCs (Fig. 2B). In the
227 Ck-induced group, ~60% of genes exhibited concordant differential expression (53%
228 when compared to ESCs and 67% when compared to EpiLCs) between D2 and D6
229 PGCLCs. The concordance was higher for the D2 and D6 4TF-induced group (74% and
230 78%, respectively). In both induction systems, D4 and D6 PGCLC shared around 84%
231 concordantly regulated genes (Fig. 2C). These data demonstrated that 4TF-induced
232 PGCLCs at early stages exhibited greater consistency in their transcriptomic changes
233 compared those in Ck-induced PGCLCs. Given that only microarray (not RNA-seq)
234 datasets were available for 3TF- and Nanog-induced PGCLCs, we performed a

235 Spearman correlation coefficient analysis to compare their similarities with Ck-induced
236 PGCLCs. Microarray datasets of 3TF- and Nanog-induced PGCLCs (referred to as
237 3TF_D2/4PGCLC and Nanog_D4PGCLC; datasets for 3TF_D6PGCLC and
238 Nanog_D2/6PGCLC are not available) were compared to the corresponding microarray
239 datasets of Ck-PGCLCs generated using the same platforms. The 3TF_D2/4PGCLCs
240 showed high correlation with Ck_D4/6PGCLCs but not Ck_D2PGCLCs;
241 Nanog_D4PGCLC showed a similar correlation to Ck_D4PGCLCs (Fig. 2D). Consistent
242 with the PCA, 4TF_D2/4/6PGCLCs are highly correlated with Ck_D4/6PGCLCs but not
243 Ck_D2PGCLCs (Fig. 2D).

244 *In vivo*, PGC specification follows WNT-dependent induction, which activates the
245 master mesodermal TF, T. This triggers the core TF network for PGCs (i.e., *Prdm1*,
246 *Prdm14*, and *Tfap2c*). Whereas the Ck induction system relies on WNT signaling, the
247 3TF and Nanog induction systems do not. To investigate WNT signaling activity during
248 4TF-induced PGCLC differentiation, we compared the transcriptomes of Ck- and TF-
249 induced PGCLCs. Gene set variation analysis (GSVA) revealed strong enrichment of
250 Wnt signaling pathways in Ck_D2PGCLCs (including downstream target genes *Wnt3*,
251 *Wnt5a*, *Wnt5b*, *Axin2*, *Mixl1*, *Cdx2*, and *T*), as previously reported (9), but not in
252 3TF_D2PGCLCs or in our 4TF_D2PGCLCs (Fig. 2E and F). Overall, unlike Ck
253 induction, the 4TF, 3TF and Nanog induce PGCLCs from EpiLCs independently of the
254 BMP-WNT-T pathway.

255 Unlike our 4TF system, the 3TF system exhibits low differentiation efficiency that did
256 not increase after day 4 after induction (9,38). To explore the impact of *Nanog*
257 overexpression in driving efficient differentiation into PGCLCs, we performed Gene set
258 enrichment analysis (GSEA). This revealed that pathways predominantly related to
259 neurodevelopment were suppressed in D2PGCLCs compared to EpiLCs (Figs. S4A and
260 B), and *Otx2* was involved in the majority of these pathways (Fig. S4C). *Otx2* plays a
261 crucial role in brain development and exhibits reciprocal antagonism with *Nanog*.
262 Induction of transgenic *Nanog* in EpiLCs shifts the balance between these mutual
263 antagonists, favoring *Nanog* (42). When *Otx2* levels decrease sufficiently, this enables
264 NANOG to influence the regulatory elements controlling genes associated with PGC-
265 specific TFs (42). GSVA revealed that neurodevelopment-related pathways and genes

266 were highly repressed in 4TF_D2PGCLCs, while 3TF_D2PGCLCs showed a modest
267 level of suppression. Notably, these pathways and genes were enriched in
268 Ck_D2PGCLCs (Fig. 2G and H, and S4D).

269 To better characterize the transcriptome landscapes during PGCLC differentiation,
270 we applied k-Means analysis of top 500 differentially expressed genes (DEGs) and
271 defined one cluster (C4) representing genes expressed in 4TF_D2/4/6PGCLCs (Fig.
272 3A). Gene Ontology (GO) analysis revealed that 11 of the top 20 pathways in this
273 cluster were associated with reproductive processes (Fig. 3B). GSVA confirmed
274 enrichment in germ cell development terms such as gamete generation, oogenesis,
275 piRNA metabolic process, meiosis and reproduction were enriched in 4TF-induced but
276 not Ck- and 3TF- induced PGCLCs compared to ESCs and EpiLCs (Fig. 3C and S5).
277 GSEA revealed upregulation of genes related to gamete generation, meiotic cell cycle
278 and reproduction in 4TF-induced PGCLCs. Processes related to piRNA processing and
279 meiotic cell cycle also enriched in 3TF-induced D4PGCLCs. However, upregulation of
280 these gene sets were barely observed in Ck- and Nanog-induced PGCLCs (Fig. 3D and
281 Fig. S6). We next compared the DEGs among Ck-, 3TF- and 4TF-induced PGCLCs.
282 Across all three groups, the naïve pluripotency gene *Klf4* was highly expressed in
283 ESCs, while genes related to the formative state (*Otx2*, *Pou3f1*, *Fgf5*, *Dnmt3a* and
284 *Dnmt3b*) were highly expressed in EpiLCs (Fig. 3E). Early PGC-related genes (e.g.,
285 *Prdm1*, *Prdm14*, *Nanog*, *Dnd1*, *Tfap2c*, *Dppa3* and *Itgb3*) were notably expressed in
286 Ck_D4PGCLC, 3TF_D2PGCLC and 4TF_D2PGCLC, indicating that both 3TF and 4TF
287 induction systems cause molecular differentiation more rapidly than cytokine induction.
288 Some PGC genes, including *Rhox9, Rhox6, Rhox5 and Gm9*, were highly expressed in
289 Ck_D6PGCLC, 3TF_D4PGCLC and 4TF_D4PGCLC. Interestingly, late PGC, meiosis
290 and piRNA-related genes were consistently expressed in 4TF-induced PGCLCs but not
291 in Ck-induced PGCLCs (Fig. 3E and S7). Network analysis identified a regulatory axis
292 facilitates the transition from early to late PGCLCs (Fig. S8; see Discussion).

293 **Further differentiation of 4TF-induced PGCLCs *in vitro***

294 We next examined the potential of PGCLCs to differentiate into advanced stages *in*
295 *vitro*. D6PGCLCs (*Stella-eGFP*$^{+}$ cells) were aggregated with dissociated neonatal

296 testicular cells from germ cell deficient $Kit^{W/Wv}$ mice in U bottom 96-well plates. Then, the
297 cell aggregates were transferred to a trans-well plate and cultured in medium containing
298 GNDF and bFGF that are essential for promoting spermatogonial growth (43) (Fig. 4A).
299 After a two-week induction period, we conducted immunofluorescence (IF) staining to
300 assess whether the aggregates express late PGC or spermatogonia proteins. We
301 detected cell clusters that expressed NANOS3, DAZL and DDX4 within the aggregates
302 (Figs. 4B). Over 70% of DDX4-positive cells also expressed GCNA (Germ cell nuclear
303 antigen) (Fig. 4C), a robust marker of germ cells from E11 onwards (44,45).
304 Interestingly, certain DDX4-positive cells also contained the spermatogonial marker
305 PLZF (formally ZBTB16; Zinc finger and BTB domain-containing protein 16) (46) but not
306 the Sertoli cell marker SOX9 (Fig. 4C). To test whether the germ cells can developed
307 into advanced stages, the aggregates were cultured using a gas-liquid interphase
308 method for an additional week in the medium containing KSR (knockout serum
309 replacement) at 34°C (Fig. 4A) (47). Interestingly, cells expressing both GCNA and
310 STRA8 were identified in the aggregates (Fig. 4D). These data suggest that the 4TF-
311 induced PGCLCs have the capacity of further development *in vitro*.

312 **Scale-up differentiation and long-term culture of PGCLCs**

313 To validate the Stella-eGFP transgene as an accurate marker of PGCLCs, we
314 analyzed differentiating cultures with two additional PGC surface markers, ITGB3 (also
315 known as CD61) and SSEA1 (8). Doubly positive cells were shown to differentiate into
316 functional gametes (48). As expected, day 6 EBs were significantly enriched for
317 $ITGB3^+/SSEA1^+$ and $ITGB3^+$/Stella-$eGFP^+$ cells compared to ESCs and EpiLCs (Fig.
318 5A). Next, we sought to determine the scalability of 4TF-induction system. We seeded
319 2×10^3 cells/ well to U bottom 96-well plates, 8×10^4 cells/well to untreated 12-well plates
320 (suitable for suspension culture) and 1×10^6 cells to a single 100mm bacteriological Petri
321 dish, then assessed differentiation in day 6 EBs. The percentages of both
322 $ITGB3^+/SSEA1^+$ and $ITGB3^+$/Stella-$eGFP^+$ populations were highest in 12-well plate
323 and lowest in 96-well plates (Fig. 5A and B). Suspecting that EB size influences
324 differentiation efficiency, we measured the diameters of day 6 EBs. In the U bottom 96-
325 well plates, EB sizes were homogeneous and by far the largest (~2.5× on average)
326 amongst the culture formats (Fig. 5C). The results indicate an inverse relationship

327 between efficiency of PGCLC differentiation and EB size, which is consistent with
328 observation in the human PGCLC differentiation system (49). The scalability of the
329 Nanog-only induction system was also observed by seeding varying numbers of EpiLCs
330 into different plates/ dish. However, the percentages of both ITGB3$^+$/SSEA1$^+$ and
331 ITGB3$^+$/Stella-eGFP$^+$ populations were 30% and 60% lower, respectively, than those in
332 the 4TF induction system (Fig. S9A; 12 well plate cultures). EpiLCs represent a
333 transient, germline-competent formative state, with day 2 EpiLCs demonstrating the
334 highest efficiency in response to cytokine induction compared to day 1 or 3 EpiLCs (8).
335 We found that 4TFs can also induce PGC fate in day 1 EpiLCs, albeit with lower
336 efficiency than in day 2 EpiLCs (Fig. 5D and E). *Nanog* overexpression alone was
337 insufficient to differentiate day 1 EpiLCs into PGCLCs (Fig. S9B and C).

338 Next, we investigated the potential for maintaining PGCLCs over long periods. Day 6
339 EBs from both 4TF- and Nanog-induction systems were dispersed, and 8×10^4 cells/ well
340 were seeded into a 12-well plate. The cells were passaged every 4 days in suspension
341 (Fig. 5F). We successfully maintained the EBs for 10 continuous passages with Dox-
342 induced transgene expression. However, cell numbers declined without Dox in both
343 systems, indicating that TFs, particularly Nanog, are essential for sustaining EB cell
344 proliferation (Fig. 5G). While the doubling times between 4TF- and Nanog-induced cells
345 were comparable (Fig. 5H), only the 4TF system consistently maintained the PGC
346 identity (>70% of cells), whereas nearly all cells in the Nanog-only system lost PGC
347 marker expression (Fig. 5I; S9D and E). Lastly, we investigated the impact of culture
348 plate type on PGCLC differentiation by seeding 8×10^4 cells/well onto treated 12-well
349 plates designed for adherent culture. After differentiation, we observed the presence of
350 attached cells, along with both floating and attached EBs (Fig. S10A). While PGCLCs
351 were robustly induced, the differentiation efficiency was lower compared to those
352 cultured in untreated plates (Fig. S10B and C), indicating that suspension culture is
353 more effective in promoting PGCLC differentiation compared to adherent culture.

354 **High throughput CRISPR perturbation screen identifies epigenetic regulators**
355 **impacting PGC development.**

356 A major motivation for developing an inexpensive, simple and scalable germ cell

357 culture system was to conduct genome-scale functional genomic screens for genes
358 involved in various aspects of germ cell biology and development. Epigenetic
359 reprogramming is an especially important aspect of normal PGC development. To
360 identify epigenetic regulators impacting PGC development, we designed a CRISPRi
361 screening strategy to target all the genes with known or putative functions in epigenetic
362 regulation (Fig. 6A). We built a custom lentiviral library encoding 7,360 sgRNAs
363 targeting 701 "epigenetic" genes, plus 350 non-targeting sgRNAs as controls (Tables S1
364 and S2), and introduced these into a 4TF-inducible ESC line containing an integrated
365 dCas9-KRAB (Krüppel-associated box) transgene (Fig. S11A-D). The library-infected
366 ESCs contained 89% of all the sgRNAs in the library, with at least 5 sgRNAs present for
367 694 genes (Fig. S11E and Table S3). PGCLC differentiation was initiated in 100mm
368 bacteriological Petri dishes (Fig. 6A), and both *Stella-eGFP*$^{+}$ and *Stella-eGFP*$^{-}$ cells
369 were isolated from dispersed day 6 EBs (Fig. S11F). In each resulting population and
370 ESCs, sgRNAs were PCR amplified and then sequenced. Quality control measures
371 indicated a high degree of uniformity and coverage of the sgRNA library in each cell
372 population (Fig. S12). The MAGeCKFlute algorithm was utilized to calculate a beta
373 score for every potential gene by comparing each population to the ESCs (50). Genes
374 with a negative beta score suggests negative selection of the epigenetic factor within
375 the cell population, whereas a positive beta score signifies positive selection. Using one
376 standard deviation from the mean as a threshold, we identified 53 and 61 under- and
377 over-represented genes, respectively, from the 701 queried genes (Fig. S13A and B;
378 Table S4).

379 Transduction of sgRNAs targeting *Ncor2*, which encodes a nuclear receptor co-
380 repressor that facilitates histone deacetylation (31,32), had the most negative impact on
381 4TF-induced PGCLC in our screen (Fig. S13 and Table S4). Treatment with *Ncor2*-
382 targeting siRNA caused significant reductions in *ITGB3*$^{+}$*/Stella-eGFP*$^{+}$ cells (Fig. S14A-
383 D), consistent with a previous observation from a Ck-induction system (51). NCOR2
384 contains three repression domains (RD1, RD2, and RD3) capable of interacting with
385 various classes of HDACs (52) (Fig. S14E). Because one motivation of this project was
386 to identify environmental agents (including drugs and dietary supplements) that might
387 cause deleterious epigenetic germline perturbations, we tested the effects of two HDACi

388 on PGCLC development: sodium butyrate (SB) and valproic acid (VPA). Both are short-
389 chain fatty acids that inhibit HDACs by binding to the zinc-containing domains of these
390 enzymes (Fig. S14E) (53). Treatment with VPA or SB resulted in a reduction in PGCLC
391 and increased H3K27ac levels (Fig. S14F-I). To investigate why HDACi treatment
392 dramatically decreased the efficiency of PGCLC differentiation, we conducted RNA-seq
393 to analyze the transcriptomes of *Stella-GFP*$^{+}$ and *Stella-GFP*$^{-}$ cells on day 6 of
394 differentiation from control and HDACi-treated groups. Our findings revealed that HDACi
395 treatment result in a loss of germline identity because the genes associated with
396 somatic lineages were upregulated (Fig. S15 and S16; Text S1).

397 To assess the impact of histone deacetylation on *in vivo* PGC development, HDACi
398 was administered to pregnant mice and their offspring were examined for potential
399 reproductive defects. From E5.5 to E11.5, pregnant females received daily injections of
400 either 600mg/kg or 800mg/kg of SB per body weight or 300mg/kg of VPA per body
401 weight (Fig. 6B). All treated dams were viable and produced a comparable number of
402 pups to the water-injected control group. Quantification of follicle numbers in 1-month-
403 old ovaries revealed no significant differences compared to the controls (Fig. S17). In 2-
404 month-old males, testis weights were unaffected by HDACi treatment during pregnancy
405 (Fig. 6C). However, these animals exhibited a 30% reduction in sperm numbers in the
406 SB groups and an 18% decrease in the VPA group (Fig. 6D). Histological analysis
407 identified atrophic seminiferous tubules (1-3 per section of a whole testis), despite
408 normal morphologies in the cauda epididymides (Fig. 6E). TUNEL (terminal
409 deoxynucleotidyl transferase dUTP nick end labeling) staining revealed no apoptotic
410 cells within atrophic tubules (Fig. S18). These findings suggest that fetal exposure to SB
411 and VPA leads to PGC loss, or possibly spermatogonial stem cell loss, contributing to
412 hypospermatogenesis in adulthood.

413 **Direct differentiation of PGCLCs from formative 4TF-inducible ESCs.**

414 To determine if 4TFs can efficiently induce PGCLCs from formative ESCs, a
415 pluripotent stem cell that directly responds to PGC specification, the 4TF-inducible
416 ESCs were transformed to a formative state through exposure to bFGF, Activin A and
417 GSK3β inhibitor (CHIR99021) (Fig. 7A) as described previously (5). Naïve ESCs

418 typically grow as tightly packed colonies with a dome shape. Upon EpiLC induction, the
419 cells quickly underwent a morphological conversion that includes flattening, diminished
420 cell-cell interactions, and the formation of cellular protrusions. However, formative ESCs
421 displayed distinctive morphologies (Fig. 7B). Immunolabeling of ZO1 (zonula occludens-
422 1, also known as tight junction protein 1) revealed that formative ESCs and EpiLCs
423 were more similar to one another than naïve ESCs (Fig. 7B). The regulation of *Oct4*
424 expression in naïve and primed pluripotent cells is differentially controlled by distal (DE)
425 and proximal enhancers, respectively (54). The DE also exhibits activity in the formative
426 state, albeit at a lower level compared to that in the naïve state (5). To monitor the exit
427 from the naïve state during naïve-to-formative conversion, we compared the GFP signal
428 intensity of Oct4-DE-GFP ("OG2") iPSCs during this transition. Consistent with previous
429 observations (5), the majority of formative iPSCs and EpiLCs were GFP^+ but the
430 average signal intensity was lower than that in naïve iPSCs (Fig. 7B). Collectively, these
431 results confirm the establishment of formative ESCs/iPSCs from naïve cells.

432 We next tested if formative 4TF-inducible ESCs respond directly to PGC
433 specification. After removing the feeder layer, PGCLC differentiation was conducted by
434 seeding 8×10^4 ESCs/well to untreated 12-well plates (Fig. 7A). FACS analysis of day 6
435 EBs revealed distinct $ITGB3^+/SSEA1^+$ and $ITGB3^+/Stella\text{-}eGFP^+$ populations with
436 induction by Dox; such cells were not observed in undifferentiated formative ESCs and
437 without induction by Dox (Fig. 7C and D). However, the efficiency was lower compared
438 to that observed when EpiLC were induced to PGCLCs (Fig. 5A and B). Treatment with
439 an FGF receptor (FGFR) inhibitor PD173074 was reported to significantly increase Ck-
440 induced PGCLC formation from formative ESC via unknown mechanisms (5). However,
441 this enhancement was not observed in 4TF-induced PGCLCs (Fig. S19). Interestingly,
442 overexpression of Nanog alone is insufficient to direct formative ESC toward a PGC fate
443 (Fig. S20). We next aggregated the D6PGCLC (*Stella-eGFP*$^+$ cells) with somatic cells
444 from $Kit^{W/Wv}$ testes to assess further differentiation potential. Similar to EpiLC derived
445 PGCLC (Fig. 4A), the formative ESC-derived PGCLCs demonstrated the capacity for
446 further development (Fig. 7E). Collectively, these findings demonstrated that
447 overexpressing 4TFs can directly and effectively induce differentiation of formative
448 ESCs into PGCLCs.

To elucidate the differences in efficiency in generating PGCLCs from formative 4TF ESCs vs. day 2 EpiLCs, and to explore why Nanog alone is insufficient to drive differentiation of formative ESCs to PGCLCs, we compared the transcriptomes of our derived EpiLCs to those in a published dataset of formative ESCs cultured and differentiated under identical conditions (55). RRHO analysis indicated substantial similarities in gene expression changes during differentiation and a significant correlation between 48h EpiLC and naïve ESC transcriptomes, indicating a shared cell state (Fig. 7F). Interestingly, while formative ESCs overlapped with EpiLCs at various stages in co-upregulated genes, they did not show overlap in co-downregulated genes (Fig. 7F). This suggests that formative ESCs represent distinct sub-stages, leading to differences in PGCLC induction 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 (9,12,37,56–60), current methodologies have shortcomings in terms of efficiency, scalability and long-term culture. Our 4TF-induction system addresses these limitations, having several advantages compared to Ck-, 3TF-, and Nanog-only inducible systems (Fig. S21). We demonstrated that simultaneous overexpression of *Nanog*, *Prdm1*, *Prdm14* and *Tfap2c* in mouse EpiLCs or formative ESCs can render PGCLCs in a highly efficient and cost-effectively manner (Fig. S21). The induced PGCLCs exhibit late PGC hallmarks, and they can develop into spermatogonium-like cells *in vitro*. Overexpression of *Nanog* had the important effect of enhancing the PGC regulatory network, suppressing the formation of non-germline lineages and maintaining PGC fate, significantly improving PGCLC production.

Specification of mouse PGCs requires coordinated actions of *Prdm1*, *Prdm14*, and *Tfap2c*. Here, we found that forced overexpression of *Nanog* combined with one of these core TFs can drive PGCLC specification more efficiently that *Nanog* alone, and 4TFs yielded the highest efficiency, indicating additive or synergistic effects. Given that

479 NANOG can bind and activate enhancers of *Prdm1* and *Prdm14* in EpiLCs *in vitro* (12)
480 as we confirmed here, we conclude that *Nanog* overexpression strengthens the PGC
481 core regulatory network in the 4TF system. NANOG also plays a role in suppressing
482 somatic lineage induction development. For example, OTX2, a crucial TF for anterior
483 neuroectoderm specification, shows reciprocal antagonism with NANOG and thus
484 restricts mouse germline entry (37). 4TF-induced PGCLCs exhibited *Otx2*
485 downregulation, *Nanog* upregulation, and significant downregulation of
486 neurodevelopment pathways in D2PGCLCs *vs* EpiLCs. However, these pathways were
487 activated in Ck-induced PGCLCs or not fully repressed in 3TF-induced PGCLCs. Using
488 *Nanog-eGFP* transgene iPSCs, Hayashi et al. demonstrated that *Nanog* was
489 upregulated as early as day 2 in Ck-induced PGCLCs (8). The distinct kinetics of
490 NANOG expression in the 4TF and Ck induction systems correlated with the repression
491 and activation of neurodevelopment pathways, respectively. Lastly, NANOG is essential
492 for maintaining PGC fate. Depletion of *Nanog* in PGCLCs showed decreased
493 proliferation and increased apoptosis (10). Those findings indicate multifunctional roles
494 of NANOG in driving differentiation into PGCLCs.

495 The transcriptional profiles of Ck- and 4TF-induced PGCLCs resembled PGCs at
496 different stages. This disparity partially arises from the essential role of WNT signaling
497 and the mesodermal program in the Ck-induction system, whereas the 4TF-induction
498 system circumvents these requirements. Combining insights from previous studies
499 (9,12), all PGCLC systems using TF overexpression (Nanog, 3TF and 4TF) drive
500 differentiation independently of the BMP4-WNT-T pathway. The gene regulatory axis of
501 early-to-late PGC development established in 4TF-induced PGCLCs propels their
502 differentiation beyond what Ck achieved within the same timeframe. Consequently, in
503 the 4TF but not Ck-induced PGCLCs, the late PGC regulatory network was triggered as
504 indicated by expression of genes involved in piRNA processes (such as *Piwil2/Mili*,
505 *Tdrd5*, *Mov10l1*, *Tdrd7* and *Tex19.2*).

506 It was surprising that several meiosis genes were expressed in the 4TF-induced
507 PGCLCs, including *Stra8*, *Sycp1/2/3*, *Meiob* and *Smc1b*. One possible explanation is
508 that although the engineered ESCs were male, some fraction of cells may have initiated
509 a meiotic transcription program as would normally occur in females absent sex

510 determination signals that emanate from somatic cells *in vivo*. We also observed
511 expression of *Mael*, involved in pachytene piRNA processes (namely LINE-element
512 suppression) a gene expressed strongly in mouse spermatocytes and which is essential
513 for normal spermatogenesis in mice (61). However, MAEL protein is present at
514 substantial levels in fetal oocytes and primordial follicles in both humans and mice
515 (62,63) where it is also involved in suppressing LINE-1 activity that contributes to fetal
516 oocyte attrition. Regarding the mechanisms by which these meiosis genes are
517 transcribed before clear evidence of meiocytes in the D6PGCLC cultures, network
518 analysis gives a clue. During germ cell development, *Tfap2c* is activated by *Prdm1* and
519 *Prdm14* (2), directly targeting *Dmrt1* in PGCLCs (38,64). Human DMRT1 upregulates
520 late PGC markers (such as DAZL and PIWIL2), while DAZL enhances DMRT1's impact
521 by repressing pluripotency and promoting expression of late PGC (65) and meiosis
522 genes. This beginning of the meiosis transcriptional program in these cultures may
523 precede developmental or cellular hallmarks of meiosis.

524 At present, gonadal and testicular somatic cells are needed for further differentiation
525 of male PGCLCs (18,66,67). Seminiferous tubule-like structures can be constructed by
526 aggregating Ck-induced PGCLCs with gonadal somatic cells under a gas-liquid
527 interphase culture condition (18,67). Given that 4TF-induced PGCLCs exhibit more
528 advanced development compared to Ck-induced PGCLCs, we opted for the readily
529 accessible neonatal testicular somatic cells to facilitate PGCLC differentiation, along
530 with key cytokines (i.e., GDNF and bFGF) crucial for spermatogonial proliferation. While
531 no seminiferous tubule-like structures were observed, clusters of germ cells were
532 formed and surrounded by somatic cells. Similar to previous observations (18,67), a
533 subset of germ cells underwent further differentiation into cells with spermatogonium-
534 like characteristics. Some germ cells displayed the early meiosis marker STRA8 after
535 culture at 34°C, but further work is necessary to explore whether these are truly meiotic
536 cells that can complete meiosis and progress to later stages of spermatogenesis *in vivo*
537 or *in vitro*.

538 Unraveling the genetic and epigenetic factors governing gametogenesis is crucial for
539 understanding causes of developmental or reproductive abnormalities. Though genomic
540 analyses have characterized gene expression patterns and epigenetic modifications

541 during germ cell development (68–70), the field has been hampered by the lack of high-
542 throughput platforms for genomewide functional studies. Massively parallel CRISPR-
543 based screens require large number of cells to accommodate libraries containing tens
544 of thousands of sgRNAs. Previously, a screen of 199 epigenetic modifiers was
545 performed using individual RNAi knockdowns under standard conditions, i.e., exposure
546 of individual EBs to cytokines in 96-well plates (51). By contrast, the 4TF-inducible
547 system enabled *en masse* gene knockdowns using Petri dishes and sequencing-based
548 screening of 701 "epigenetic" genes for key roles in PGC development. Our study not
549 only identified epigenetic events influencing the generation and development of PGCs,
550 but also to draw links between epigenetic modifications and environmental factors (such
551 as medications and dietary supplements) that might influence gametogenesis and
552 possibly fertility.

553 The 4TF-induction system bypasses the need for BMP4 and WNT3, enabling direct
554 programming of EpiLCs into PGCLCs. Consequently, our CRISPRi screen likely did not
555 query epigenetic regulators involved during the period of BMP signaling and the
556 activation of germline-specific TFs. As mentioned earlier, an RNAi-based epigenetic
557 screen was described that utilized a Blimp1 reporter for PGCLC identification. *Blimp1*
558 expression is equivalent to ~E6.25, and they likely preferentially identified epigenetic
559 genes that regulate the transition of epiblast cells to the PGC fate, as well as the
560 repression of somatic networks. In contrast, our platform preferentially queried
561 epigenetic factors that regulate PGC differentiation downstream of germline TF
562 expression. Disruption of these factors holds the potential to impede or override the
563 germline transcription network.

564 The dietary supplement SB is beneficial for intestinal homeostasis and energy
565 metabolism by enhancing intestinal barrier function and mucosal immunity with its anti-
566 inflammatory properties (71). However, the potential effects of maternal SB
567 supplementation upon fetal development are poorly explored. VPA is a widely
568 prescribed drug used to treat epilepsy, bipolar disorder, and migraines. There is an
569 elevated risk of birth defects (such as cleft lip, heart and neural tube defects) in children
570 born to women taking VPA during pregnancy (72), but possible impacts on germ cell
571 development remain unknown. Our *in vitro* and *in vivo* results raise the possibility that

572 fetal exposure to HDAC inhibitors can deplete PGCs, manifesting as atrophic
573 seminiferous tubules and hypospermatogenesis in adulthood. However, the ovarian
574 reserve was unaffected even though treatment occurred before sex determination. It is
575 possible that compromised cells were preferentially lost in the process of fetal oocyte
576 attrition, which normally eliminates up to 80% of oocytes (73,74). Alternatively, the
577 molecular impacts of treatments may only manifest in prospermatogonia or
578 spermatogonial stem cells. Although the recommended clinical doses of SB and VPA
579 are lower than those utilized in our study, we cannot say if human fetuses are more or
580 less sensitive to these drugs than mouse embryos with respect to germline
581 development. Our findings contribute to an expanding body of evidence indicating that
582 epigenetic factors and dietary choices can impact male fertility (75,76).

583 In mice, competence for germline specification distinguishes the formative state of
584 pluripotency from the naïve and primed states (77). Studies using Ck- ,3TF- and Nanog-
585 induction systems have demonstrated that a formative state is essential for robust
586 induction of a PGC-like state (8,9,12). Induction of 3TF in naïve ESC resulted in a
587 unique phenotype characterized by intense activation of the Stella reporter without
588 forming a PGC-like state. The transcriptomes of these Stella^{+} cells were close to those
589 of ESCs, but not PGCLCs/PGCs (9). Since *Nanog* abundantly expressed in naïve ESC,
590 the above result also indicates that induction of 4TFs in naïve ESCs fails to form
591 PGCLCs. Given that both formative ESCs and EpiLCs respond directly to PGC
592 specification, it was unclear whether exposure of formative ESCs to 4TF overexpression
593 was compatible with efficient differentiation into PGCLCs. Recent advancements have
594 identified four distinct sub-stages of formative pluripotent stem cells, exhibiting
595 functional characteristics and WNT/ β-catenin signaling modulation (4–7). In our study,
596 we cultivated formative ESCs using one of these protocols (5), and revealed that 4TFs
597 can drive both formative ESCs and EpiLCs into PGCLCs. PGCLCs (Stella-eGFP^{+} cells)
598 differentiated from formative ESCs exhibit similar further developmental potential with
599 those from EpiLCs. Importantly, enables simplification of the PGCLC differentiation
600 process, as formative ESCs can be stably maintained and would not require preliminary
601 induction into the transient EpiLC-like state from naïve ESCs.

602 This scalable and efficient 4TF-induction system offers a valuable tool to conduct

603 unbiased genetic screens impacting PGCLC formation. Besides CRISPRi as described
604 here, this platform enables other CRISPR-based perturbation screens including
605 knockouts, upregulation via CRISPR activation (CRISPRa), combinatorial disruptions,
606 and single-cell transcriptomics as a phenotypic readout (78,79). Furthermore, a useful
607 aspect of the 4TF system is that the derived PGCLCs persisted for prolonged periods in
608 culture, enabling various studies involving, for example, large amounts of material for
609 biochemical analyses, genetic manipulations initiated after cells reach this advanced
610 stage, responses to drug or biologic exposures, or as a starting point for exploring
611 further differentiation. Finally, considering that genetic causes underlie ~50% of human
612 infertility (1), parallelized screening of genetic variants, particularly rare or minor alleles
613 that are classified as variants of uncertain significance (VUS), is another potential
614 application. Functionally interpreting the VUS within genes that essential for
615 gametogenesis, along with achieving a comprehensive understanding of epigenetic
616 networks, are important for genetic counseling, clinical applications, and potentially
617 gene correction therapies for individuals to produce viable gametes *in vitro or in vivo*
618 (1,80,81).

619

620 ## Materials and Methods

621

622 **Mice and cell lines**

623 All animal usage was approved by Cornell University's Institutional Animal Care and
624 Use Committee, under protocol 2004-0038 to J.C.S. Four- to eight-day-old $Kit^{W/Wv}$ male
625 mice generated by crossing C57BL/6J-Kit^{Wv}/J and WB/ReJ Kit^{W}/J were used in this
626 study. C57BL/6J-Kit^{Wv}/J (#000049), WB/ReJ Kit^{W}/J (#000692), B6(Cg)-Tyr^{c-2J}/J
627 (#000058) and C57BL/6J (# 000664) mice were purchased from The Jackson
628 Laboratory. The Stella-eGFP transgenic ESC line (Stella-eGFP ESCs) with Stella
629 flanking sequences coupled to eGFP as the reporter was kindly provided by Prof. M.
630 Azim Surani (35).

631

632 **Vectors**

633 To construct pPBhCMV1-Prdm1-pA and pPBhCMV1-Prdm14-pA, mouse *Prdm1* and
634 *Prdm14* coding sequences were cloned into pPBhCMV1-Nanog-pA (a gift from M. Azim
635 Surani) by PCR flanked with *Xho*I and *Not*I restriction sites. To construct targeting
636 vectors containing enhancer-reporter transgenes, PCR-amplified sequences of mouse
637 *Prdm1* and *Prdm14* enhancers (12) bearing terminal *Not*I restriction sites were cloned
638 into PCR4-*Shh::lacZ*-H11 (Addgene, #139098). Plasmid clones were validated by
639 Sanger sequencing. pPyCAG-Pbase and pPB-CAG-rtTA-IN (Addgene, #60612) were
640 kindly provided by M. Azim Surani. pSpCas9(BB)-2A-PuroR (px459) V2.0 (Addgene,
641 #62988), psPAX2 (Addgene, #12260), pMD2.G (Addgene, #12259), pLV-tetO-Tfap2c
642 (Addgene, #70269), pLX_311-KRAB-dCas9 (Addgene, #96918) and pXPR_050
643 (Addgene, # 96925) were obtained from Addgene. Primer sequences are listed in Table
644 S5.

645

646 **Microinjection for chimera production**

647 The ESCs were microinjected to B6(Cg)-Tyr^{c-2J}/J blastocysts by standard methods in
648 Cornell's Transgenic Core Facility (82).

649

650 **Derivation of genetically modified ESC clones**

651 A *piggyBac* transposon expression system was used to stably integrate single
652 copies of various vectors that were doxycycline (Tet) inducible. To establish the Tet-on
653 expression system, pPBhCMV1-Prdm1-pA, pPBhCMV1-Prdm14-pA and pPBhCMV1-
654 Nanog-pA vectors were transfected using TransIT®-LT1 Transfection Reagent (Mirus
655 Bio™, MIR2304) into ESCs together with pPBCAG-rtTA and pPyCAG-PBase vectors.
656 After 1 week of Geneticin™ (80 μg/ml; Life Technologies, 10131035) selection (all
657 vectors except pPyCAG-PBase contain neomycin resistance genes), clones containing
658 desired inserts were identified. Cells infected with pLV-tetO-Tfap2c lentivirus (described
659 below) were selected by Zeocin® (InvivoGen, ant-zn-05) to identify transformants.
660 To construct ESC lines containing enhancer-reporters, the targeting vector PCR4-
661 *Shh::lacZ*-H11 and H11 gRNA-expression plasmid px459 were transfected into cells
662 bearing Tet-inducible *Nanog* along. After puromycin selection for 48 hours, single
663 colonies were expanded and correctly targeted clones were identified using primers
664 listed in Table S5.
665
666 **ESC, iPSC culture and PGCLC differentiation**
667 Mouse ESCs maintained on γ-irradiated feeder MEFs (mouse embryonic fibroblast
668 cells) in DMEM containing 15% FBS, penicillin-streptomycin, NEAA, β-mercaptoethanol,
669 1,000U/ml LIF (Peprotech, 300-05). Naïve ESCs were cultured in 2i+ LIF medium
670 containing 1×N2B27, 1 μM PD0325901 (Reprocell, 04-0006-02), 3 μM CHIR99021
671 (Reprocell, 04-0004-02) and 1,000 U/ml LIF on wells coated with 0.01% poly-L-ornithine
672 (Sigma-Aldrich, P3655) and 10 ng/ml laminin (Corning, 354232).
673 For the formative state conversion, naïve ESCs and iPSCs were dissociated into
674 single cells and plated on MEFs plate in 2i +LIF medium. Follow the protocol describe
675 previously (5), the next day, medium was changed to formative medium (N2B27
676 medium containing 10 ng/ml bFGF, 10 ng/ml Activin A and 3 μM CHIR99021) for further
677 culture. It usually takes about 2-4 passages for fully conversion. Formative ESCs and
678 iPSCs were maintained in formative medium.
679 For PGCLC differentiation, EpiLCs were induced by naïve ESCs as described
680 previously (48). Briefly, 1×10^5 ESCs were placed per well of a 12-well plate coated with
681 16.7 μg/mL bovine plasma fibronectin (Sigma-Aldrich, F1141) in N2B27 medium

682 containing 20 ng/mL Activin A (Peprotech, 120-14E), 12 ng/mL bFGF (Gibco, 13256-
683 029), and 1% knockout serum replacement (KSR, ThermoFisher, 10828010). The
684 medium was changed the following day. A total of 48 hours after plating, 2×10^3 EpiLCs
685 per well were seeded in a U-bottom 96-well plate (ThermoFisher Scientific, 268200) in
686 GK15 medium (GMEM with 15% KSR, 0.1mM NEAA, 1mM sodium pyruvate, 0.1mM β-
687 mercaptoethanol, penicillin-streptomycin, and 2mM L-glutamine) with Dox (Sigma,
688 D9891). To induce PGCLCs from formative ESCs, after removing the feeder cells,
689 2×10^3 cells per well were seeded in the U-bottom 96-well plate. Alternatively, 8×10^4 cells
690 per well were seeded in an untreated 12-well plate (ThermoFisher Scientific, 150200),
691 or 1×10^6 cells were seeded into a 100mm Petri dish in GK15 medium. To sustain the
692 PGCLCs, day 6 EBs in a 12-well plate were dissociated into single cells, and 8×10^4
693 cells per well were seeded in an untreated 12-well plate. The cells were passaged every
694 4 days. Transgenes were induced by addition of 2 μg/ml Dox at day 0 of PGCLC
695 induction. Cells were cultured at 37 °C in a 5% CO_2-95% air atmosphere.
696

697 **Flow cytometry**

698 To flow sort PGCLCs, single-cell suspensions were stained with 1 μg/ml propidium
699 iodide (PI, Invitrogen, P3566) and filtered through a 70 μm cell strainer. $eGFP^+$ cells
700 were sorted using a BD FACSMelody 4-Way Sorter. To analyze the PGCLCs,
701 suspended cells were washed with PBS and fixed in 70% ethanol. For staining, cells
702 were washed in MACS buffer (1×PBS, 0.5%BSA and 2mM EDTA) and blocked with 5%
703 goat serum (Sigma, NS02L) in MACS buffer at 4°C for 1 hour. Cells were then stained
704 with PE (Phycoerythrin) mouse/rat anti-CD61 and Alexa Fluor® 647 mouse/human anti-
705 SSEA1 antibodies (Table S6) at 4°C for 1 hour. Cells were then washed in MACS buffer
706 twice and filtered through a 35 μm cell strainer (Falcon, 352235). The stained cells were
707 analyzed using a BD FACSymphony A3 cytometer. FACS data were analyzed using
708 FlowJo™ v10.4 or flowCore package in R.
709

710 **Cell aggregate cultures for further differentiation**

711 Sorted D6PGCLCs were mixed with dissociated neonatal testicular cells from $Kit^{W/Wv}$
712 mice with 10:1 ratio and co-cultured for 2 days in a U bottom 96-well using GK15

713 medium. The aggregates were subsequently transferred to a trans-well plate in N2B27
714 medium containing 20 ng/ml bFGF, 40 ng/ml GDNF (PeproTech, 450-44), 1,000 U/ml
715 LIF, 1×Insulin-Transferrin-Selenium (ThermoFisher Scientific, 41400045) and 15% KSR
716 for additional 2 weeks. The aggregates were cultured at 37 °C in a 5% CO_2-95% air
717 atmosphere. Two weeks later, the aggregates were cultured on agarose gel stand in
718 MEMα (Minimum essential medium α; ThermoFisher Scientific, 12571063) containing
719 15% KSR for an additional week, as described previously (47). The aggregates in this
720 stage were cultured at 34 °C in a 5% CO_2-95% air atmosphere.
721

722 **Cell staining**

723 ALP staining was performed using the Vector Red Alkaline Phosphatase Substrate
724 Kit (Vector Laboratories, Burlingame, CA, USA) according to the manufacturer's
725 directions. *LacZ* staining was performed the Senescence β-Galactosidase Staining Kit
726 (Cell Signaling Technology, #9860) according to the manufacturer's directions. For
727 TUNEL staining, sections were deparaffinized and performed TUNEL staining using the
728 DeadEnd™ Fluorometric TUNEL System (Promega; G3250) following manufacturer's
729 instructions.
730 For cell IF staining, the cells growing on cover slides were washed with PBS, fixed in
731 4% paraformaldehyde (PFA) for 10 min at RT, washed twice with PBS. For section IF
732 staining, cell aggregates were fixed in 4% PFA and embedded in paraffin. The 6 μm
733 sections were deparaffinized and performed antigen retrieval using sodium citrate
734 buffer. The slides were incubated for 10 min at 37 °C in blocking buffer (PBS containing
735 10% normal goat serum). Then, they were incubated overnight in a humidified chamber
736 at 4 °C with the antibodies listed in Table S6. After washing twice with PBS, the slides
737 were incubated at 37 °C for 30 min with a 1:500 dilution of secondary antibodies (Table
738 S6), then incubated at 37 °C for 5 min with 500 ng/mL of DAPI and mounted using
739 Vectashield antifade mounting medium (Vector, H-1000).
740

741 **Bisulfite sequencing**

742 Methylation patterns in the differentially methylated domains (DMDs) of the
743 paternally imprinted gene *H19* and maternally imprinted gene *Snrpn* were determined

using bisulfite genomic sequencing as previously described (83). 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 and listed in Table S5 (83). PCR products were sub-cloned into the pGEM-T Easy vector (Promega, A1360) and sequenced. Sequences were determined and analyzed using BiQ Analyzer software (84).

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 S5). The relative levels of transcripts were calculated using the $2^{-\Delta\Delta C_T}$ method and gene expression levels were normalized to *Gapdh*.

CRISPRi library design and gRNA cloning

Seven hundred and one epigenetic factors were selected from Epifactors, and 10 unique sgRNAs were designed for each gene using the GPP sgRNA Design tool. The CRISPRi library sgRNA oligos with *Bsm*b1 overhangs were synthesized by GenScript (Piscataway, NJ) and cloned into the lentiviral vector *pXPR_050* by Golden Gate cloning following published protocols (85). Briefly, oligo pools were PCR amplified using primers listed in Table S5. *pXPR_050* plasmids were digested using the Bsmb1 restriction enzyme (NEB, R0580). The synthesized double-strand oligo pools and digested *pXPR_050* plasmids were ligated using T7 ligase (NEB, M0318S). The *pXPR_050* plasmids were then transformed into Stbl4 competent cells (Invitrogen, 11635018). The transformed bacteria were then plated on LB plates with ampicillin resistance and grown for 24 hours at 30°C. To ensure high coverage of the CRISPRi library, bacterial colonies

775 over 100× of the CRISPRi oligo library (~800,000) were collected and grown in SOC
776 medium (Sigma, S1797). Plasmid DNA was isolated using a midi-prep kit (Invitrogen,
777 K210005). The integrity of the CRISPRi library was validated by Illumina Miseq through
778 Cornell Genomics Facility (detailed library preparation steps are presented below in
779 "Library validation and CRISPRi dropout screening").
780 For CRISPR-assisted insertion of PCR4-*Shh::lacZ*-H11 vector into the H11 locus,
781 we used a gRNA that was previously designed for this purpose (36). Synthesized
782 oligonucleotides were annealed and cloned into px459 vector. The sequences of all
783 gRNAs are listed in Table S5.
784
785 **Lentiviral packaging and transduction**
786 Lentiviral particles were produced by transfecting HEK293T cells with the following
787 plasmids and amounts: 2.5μg pMD2.G, 7.5μg psPAX2, and 10μg vectors (from
788 *pLX_311* lentiviral expression of *dCas9-KRAB*, the entire constructed *pXPR_050*
789 CRISPRi sgRNA expression library and pLV-tetO-Tfap2c) using TransIT-LT1 (Mirus,
790 MIR2305). HEK293T culture media was collected at both 48 and 72 hours post
791 transfection and concentrated using Amicon Ultra-15 columns (Millipore, UFC903024).
792 ESCs were first transduced with lentiviral packaged *pLX_311* and selected with
793 Blasticidin (Gibco, A1113903) to generate stable *dCas9-KRAB* expressing cell lines,
794 which were validated by PCR (Table S5) and Western blotting as described below.
795 Approximately 1.6×10^7 ESCs with stable *dCas9-KRAB* expression were then
796 transduced with the *pXPR_050* CRISPRi sgRNA expression library at a multiplicity of
797 infection of 0.3 to ensure at least 500× coverage of the library. The cells were selected
798 in puromycin (Gibco, A1113803) (1.2mg/ml) for 2 days. All cells were infected with
799 lentivirus for 48 hours.
800
801 **Library validation and CRISPRi dropout screening.**
802 Genomic DNAs from cells carrying CRISPRi sgRNA library were extracted using a
803 Genomic DNA Purification Kit (Thermo Scientific, K0722). To recover sgRNA
804 sequences, both CRISPRi plasmid pools and genomic DNAs were PCR amplified with
805 Miseq_Adaptor primers (Table S5) using Q5 High-Fidelity DNA Polymerase (NEB,

806 M0491L). PCR samples were purified from 2% agarose gels using a Gel Extraction Kit
807 (Qiagen, 28704), then sequenced on an Illumina Miseq (Cornell Genomics Facility)
808 using the seqRead_Miseq primer (Table S5). To validate the integrity of the CRISPRi
809 library, the raw counts were mapped to the original CRISPRi library using
810 ScreenProcessing Tools (86). To identify CRISPRi dropout candidates, sgRNA
811 representations from ESCs, *Stella-eGFP*$^{+}$, and *Stella-eGFP*$^{-}$ cells were analyzed using
812 MAGeCK, MAGeCK-VISPR, and MAGeCKFlute Tools (50).
813
814 **Knockdown and drug treatment during PGCLC differentiation.**
815 Twenty-four hours before PGCLC differentiation, EpiLCs were transfected with
816 control siNT (ON-TARGETplus Non-targeting Control siRNA, Dharmacon, D-001810-10-
817 05) and siNcor2 (ON-TARGETplus siRNA SMARTPool, Dharmacon, L-045364-00-0005)
818 short dsRNA molecules using jetPRIME transfection reagent (Polyplus, 101000046) at a
819 final concentration of 50nM. EpiLCs were then differentiated into PGCLCs over 6 days.
820 Knockdown efficiencies were measured at day 2. Sodium butyrate (Sigma; B5887) and
821 valproic acid (Sigma, P4543) were dissolved in water at stocks concentrations of
822 100mg/ml and 1M, respectively. To treat cells, dilutions were added to the PGCLC
823 culture media at the start of PGCLC differentiation. Cells were collected 2, 4, or 6 days
824 later. To prepare the cells for RNA-seq, the EBs were treated with 0.05mg/ml SB or
825 0.5mM VPA.
826
827 **Western blotting**
828 Cells were lysed and homogenized using RIPA buffer (Thermo Fisher Scientific,
829 89900) with protease inhibitors (Sigma, 04693159001) and phosphatase inhibitors
830 (Sigma, 4906845001). The samples were centrifuged at 13,000rpm (~15,000G), and
831 supernatants were used for western blot analysis. The protein concentrations were
832 measured using a BCA Protein Assay Kit (Pierce, 23250). Samples were run through a
833 gradient polyacrylamide gel (Bio-Rad, 4561083EDU) and transferred to PVDF
834 membranes (Millipore, IPVH00010). Membranes were blocked by 5% nonfat milk for 1
835 hour. The antibodies used for western blotting were listed in Table S6. The HRR
836 substrate (Millipore, WBLUC0500) and ChemiDoc Imaging System (Bio-Rad) were used

837 for protein visualization.

838

839 **Mouse treatment and reproductive phenotyping**

840 For administration of drugs during pregnancy, E0.5 is the morning in which a
841 vaginal plug was detected. Pregnant females received daily injections of water, SB
842 (600mg/kg and 800mg/kg), or VPA (300mg/kg) from E5.5 to E11.5. Histology analysis,
843 sperm and follicle counts were performed as described previously (80,87). To prepare
844 the paraffin blocks, tissues were fixed overnight in Bouin's solution or 4% PFA at room
845 temperature and were embedded by Histology Lab at Cornell University. Paraffin
846 sections at 6μm thick were deparaffinized and stained with hematoxylin and eosin. One
847 cauda epididymis per male was used for sperm counting. To quantified follicle number,
848 the ovary block was serially sectioned and stained by H&E. Every fifth section was
849 scored for follicles at various stages.

850

851 **RNA-seq and data analysis**

852 Total RNAs were purified from cells using TRIzol® reagent (Invitrogen, 15596018).
853 Purified RNAs were used for Illumina RNA-seq library preparation with NEBNext Ultra II
854 Directional RNA Library Prep Kit (NEB, E7765), and a minimum of 20 million raw reads
855 were obtained. The raw counts were then aligned to mouse mm10 genome with
856 ENSEMBL gene annotations. Differential analysis of RNA-seq data was done using the
857 R package DESeq2. RRHO2 analysis was performed also using an R package (41). k-
858 Means clustering analysis was performed using iDEP 1.12 (88). GSEA was performed
859 to explore enriched pathways and interpret RNA-seq data using predefined gene sets
860 from the Molecular Signatures Database (v.7.1) in R packages (89). GSVA was carried
861 out using the GSVA package in R (90). Gene sets enriched in the GSEA were organized
862 into a graphical network produced using EnrichmentMap (91) plugins in Cytoscape
863 (v.3.10.1) (92), as described previously (93). The online STRING database (version
864 12.0) was used for generation of protein interaction (94). The upregulated DEGs
865 (log2FC>2, p<0.05 and count>250) from PGCLCs on days 2, 4 and 6 compared to ESC
866 were integrated for the network analysis.

867

868 **Accession numbers**

869 The RNA-seq data generated in this study have been deposited at GEO (Gene
870 Expression Omnibus) (GSE256104, reviewer token: mpgbukuqdnmrxgv; GSE259290,
871 reviewer token: ivwzgkugrbijlyz). The RNA-seq data of
872 ESCs/EpiLCs/Ck_D2/4/6PGCLCs (GSE67259) and i*n vivo* germ cells (GSE74094 and
873 GSE87644) were downloaded from the GEO database. The
874 ESCs/EpiLCs/3TF_D2/4PGCLCs microarray dataset is accession # GSE43775.
875 CRISPRi (GSE259289; reviewer token: gdadeiwgvzafxmb). The GEO number for RNA-
876 seq data of formative ESCs is GSE135989. The GEO number for microarray dataset of
877 ESCs/Nanog_D4PGCLC/Ck_D4PGCLC is GSE71933. The GEO number for RNA-seq
878 datasets of 6h, 12h, 24h, and 48h EpiLCs is GSE256381.

879

880 **Image process and data statistics**

881 Fluorescent images were captured by an Olympus XM10 camera. Bright-field
882 images were captured by the Olympus SC30 camera. Cropping, color, and contrast
883 adjustments were made with Adobe Photoshop CC 2024, using identical background
884 adjustments for all images. Fluorescence intensity was measured by ImageJ. All data
885 were expressed as mean ± SEM. Statistical tests were carried out using a paired
886 student's *t-test* or one-way analysis of variance (ANOVA) followed by the Tukey's *post*
887 *hoc* test with R. Graph generation was performed using R. The schematic diagrams
888 were created using BioRender.

889

890 **Software and websites**

891 Epifactors: https://epifactors.autosome.org/

892 GPP sgRNA Design tool: https://portals.broadinstitute.org/gpp/public/analysis-
893 tools/sgrna-design

894 ScreenProcessing Tools: https://github.com/mhorlbeck/ScreenProcessing

895 MAGeCK, MAGeCK-VISPR, and MAGeCKFlute Tools:
896 https://github.com/WubingZhang/MAGeCKFlute

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

898 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

913 **Acknowledgements**

914 The authors would like to thank R. Munroe and C. Abratte of Cornell's transgenic
915 facility for generating the mice, A. Surani for Stella-eGFP ESCs and several plasmids,
916 and J.C. Bloom for OG2 iPSCs. This work is supported by a grant from the National
917 Institutes of Health (R01 HD082568 to JCS), a Center grant (1P50HD096723) from the
918 National Institute of Child Health and Human Development (K. Orwig: PI; JCS: Lead
919 Project 2), and contract CO29155 from the NY State Stem Cell Program (NYSTEM).
920 X.D. was supported by a postdoctoral fellowship from the Empire State Stem Cell Fund
921 through New York State Department of Health contract no. C30293GG.

922

923 **Competing interests**

924 The authors declare no competing or financial interests.

925

References

1. Ding X, Schimenti JC. Strategies to identify genetic variants causing infertility. Trends Mol Med. 2021 Aug;27(8):792–806.

2. Saitou M, Hayashi K. Mammalian in vitro gametogenesis. Science. 2021 Oct 1;374(6563):eaaz6830.

3. Nichols J, Smith A. Naive and primed pluripotent states. Cell Stem Cell. 2009 Jun 5;4(6):487–92.

4. Kinoshita M, Barber M, Mansfield W, Cui Y, Spindlow D, Stirparo GG, et al. Capture of Mouse and Human Stem Cells with Features of Formative Pluripotency. Cell Stem Cell. 2021 Mar 4;28(3):453-471.e8.

5. Yu L, Wei Y, Sun H-X, Mahdi AK, Pinzon Arteaga CA, Sakurai M, et al. Derivation of intermediate pluripotent stem cells amenable to primordial germ cell specification. Cell Stem Cell. 2021 Mar 4;28(3):550-567.e12.

6. Luo Q, Pui H-P, Chen J, Yu L, Jannig PR, Pei Y, et al. Epiblast-like stem cells established by Wnt/β-catenin signaling manifest distinct features of formative pluripotency and germline competence. Cell Rep. 2023 Jan 31;42(1):112021.

7. Wang X, Xiang Y, Yu Y, Wang R, Zhang Y, Xu Q, et al. Formative pluripotent stem cells show features of epiblast cells poised for gastrulation. Cell Res. 2021 May;31(5):526–41.

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. 2011 Aug 19;146(4):519–32.

9. Nakaki F, Hayashi K, Ohta H, Kurimoto K, Yabuta Y, Saitou M. Induction of mouse germ-cell fate by transcription factors in vitro. Nature. 2013 Sep 12;501(7466):222–6.

10. Zhang M, Leitch HG, Tang WWC, Festuccia N, Hall-Ponsele E, Nichols J, et al. Esrrb Complementation Rescues Development of Nanog-Null Germ Cells. Cell Rep. 2018 Jan 9;22(2):332–9.

11. Acampora D, Di Giovannantonio LG, Garofalo A, Nigro V, Omodei D, Lombardi A, et al. Functional Antagonism between OTX2 and NANOG Specifies a Spectrum of Heterogeneous Identities in Embryonic Stem Cells. Stem Cell Reports. 2017 Nov 14;9(5):1642–59.

958 12. Murakami K, Günesdogan U, Zylicz JJ, Tang WWC, Sengupta R, Kobayashi T,
959 et al. NANOG alone induces germ cells in primed epiblast in vitro by activation of
960 enhancers. Nature. 2016 Jan 21;529(7586):403–7.

961 13. Vojtek M, Zhang J, Sun J, Zhang M, Chambers I. Differential repression of *Otx2*
962 underlies the capacity of NANOG and ESRRB to induce germline entry. BioRxiv.
963 2021 Jun 14;

964 14. Acampora D, Mazan S, Lallemand Y, Avantaggiato V, Maury M, Simeone A, et
965 al. Forebrain and midbrain regions are deleted in Otx2-/- mutants due to a
966 defective anterior neuroectoderm specification during gastrulation. Development.
967 1995 Oct;121(10):3279–90.

968 15. Shamblott MJ, Axelman J, Wang S, Bugg EM, Littlefield JW, Donovan PJ, et al.
969 Derivation of pluripotent stem cells from cultured human primordial germ cells.
970 Proc Natl Acad Sci USA. 1998 Nov 10;95(23):13726–31.

971 16. Matsui Y, Zsebo K, Hogan BL. Derivation of pluripotential embryonic stem cells
972 from murine primordial germ cells in culture. Cell. 1992 Sep 4;70(5):841–7.

973 17. Ohta H, Kurimoto K, Okamoto I, Nakamura T, Yabuta Y, Miyauchi H, et al. In
974 vitro expansion of mouse primordial germ cell-like cells recapitulates an
975 epigenetic blank slate. EMBO J. 2017 Jul 3;36(13):1888–907.

976 18. Ishikura Y, Ohta H, Sato T, Murase Y, Yabuta Y, Kojima Y, et al. In vitro
977 reconstitution of the whole male germ-cell development from mouse pluripotent
978 stem cells. Cell Stem Cell. 2021 Dec 2;28(12):2167-2179.e9.

979 19. Murase Y, Yabuta Y, Ohta H, Yamashiro C, Nakamura T, Yamamoto T, et al.
980 Long-term expansion with germline potential of human primordial germ cell-like
981 cells in vitro. EMBO J. 2020 Nov 2;39(21):e104929.

982 20. Gajbhiye R, Fung JN, Montgomery GW. Complex genetics of female fertility.
983 NPJ Genom Med. 2018 Oct 12;3:29.

984 21. Krausz C, Riera-Escamilla A. Genetics of male infertility. Nat Rev Urol. 2018
985 Jun;15(6):369–84.

986 22. Jørgensen N, Lamb DJ, Levine H, Pastuszak AW, Sigalos JT, Swan SH, et al.
987 Are worldwide sperm counts declining? Fertil Steril. 2021 Dec;116(6):1457–63.

988 23. Brieño-Enríquez MA, Larriba E, Del Mazo J. Endocrine disrupters, microRNAs,
989 and primordial germ cells: a dangerous cocktail. Fertil Steril. 2016 Sep
990 15;106(4):871–9.

991 24. Rattan S, Zhou C, Chiang C, Mahalingam S, Brehm E, Flaws JA. Exposure to
992 endocrine disruptors during adulthood: consequences for female fertility. J
993 Endocrinol. 2017 Jun;233(3):R109–29.

994 25. Pivonello C, Muscogiuri G, Nardone A, Garifalos F, Provvisiero DP, Verde N, et
995 al. Bisphenol A: an emerging threat to female fertility. Reprod Biol Endocrinol.
996 2020 Mar 14;18(1):22.

997 26. Skinner MK. Endocrine disruptors and epigenetic transgenerational disease
998 etiology. Pediatr Res. 2007 May;61(5 Pt 2):48R-50R.

999 27. Kurimoto K, Yabuta Y, Hayashi K, Ohta H, Kiyonari H, Mitani T, et al.
1000 Quantitative Dynamics of Chromatin Remodeling during Germ Cell Specification
1001 from Mouse Embryonic Stem Cells. Cell Stem Cell. 2015 May 7;16(5):517–32.

1002 28. Tang WWC, Dietmann S, Irie N, Leitch HG, Floros VI, Bradshaw CR, et al. A
1003 unique gene regulatory network resets the human germline epigenome for
1004 development. Cell. 2015 Jun 4;161(6):1453–67.

1005 29. Lombó M, Getino-Álvarez L, Depincé A, Labbé C, Herráez MP. Embryonic
1006 Exposure to Bisphenol A Impairs Primordial Germ Cell Migration without
1007 Jeopardizing Male Breeding Capacity. Biomolecules. 2019 Jul 25;9(8).

1008 30. Li Z, Fang F, Zhao Q, Li H, Xiong C. Supplementation of vitamin C promotes
1009 early germ cell specification from human embryonic stem cells. Stem Cell Res
1010 Ther. 2019 Nov 15;10(1):324.

1011 31. Nagy L, Kao HY, Chakravarti D, Lin RJ, Hassig CA, Ayer DE, et al. Nuclear
1012 receptor repression mediated by a complex containing SMRT, mSin3A, and
1013 histone deacetylase. Cell. 1997 May 2;89(3):373–80.

1014 32. Huang EY, Zhang J, Miska EA, Guenther MG, Kouzarides T, Lazar MA. Nuclear
1015 receptor corepressors partner with class II histone deacetylases in a Sin3-
1016 independent repression pathway. Genes Dev. 2000 Jan 1;14(1):45–54.

1017 33. Park S-Y, Kim J-S. A short guide to histone deacetylases including recent
1018 progress on class II enzymes. Exp Mol Med. 2020 Feb 19;52(2):204–12.

1019 34. Durcova-Hills G, Tang F, Doody G, Tooze R, Surani MA. Reprogramming
1020 primordial germ cells into pluripotent stem cells. PLoS ONE. 2008 Oct
1021 27;3(10):e3531.

1022 35. Payer B, Chuva de Sousa Lopes SM, Barton SC, Lee C, Saitou M, Surani MA.
1023 Generation of stella-GFP transgenic mice: a novel tool to study germ cell
1024 development. Genesis. 2006 Feb;44(2):75–83.

1025 36. Kvon EZ, Zhu Y, Kelman G, Novak CS, Plajzer-Frick I, Kato M, et al.
1026 Comprehensive in vivo interrogation reveals phenotypic impact of human
1027 enhancer variants. Cell. 2020 Mar 19;180(6):1262-1271.e15.

1028 37. Zhang J, Zhang M, Acampora D, Vojtek M, Yuan D, Simeone A, et al. OTX2
1029 restricts entry to the mouse germline. Nature. 2018 Oct 3;562(7728):595–9.

1030 38. Magnúsdóttir E, Dietmann S, Murakami K, Günesdogan U, Tang F, Bao S, et al.
1031 A tripartite transcription factor network regulates primordial germ cell specification
1032 in mice. Nat Cell Biol. 2013 Aug;15(8):905–15.

1033 39. Yamashiro C, Hirota T, Kurimoto K, Nakamura T, Yabuta Y, Nagaoka SI, et al.
1034 Persistent requirement and alteration of the key targets of PRDM1 during
1035 primordial germ cell development in mice. Biol Reprod. 2016 Jan;94(1):7.

1036 40. Sasaki K, Yokobayashi S, Nakamura T, Okamoto I, Yabuta Y, Kurimoto K, et al.
1037 Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells.
1038 Cell Stem Cell. 2015 Aug 6;17(2):178–94.

1039 41. Cahill KM, Huo Z, Tseng GC, Logan RW, Seney ML. Improved identification of
1040 concordant and discordant gene expression signatures using an updated rank-
1041 rank hypergeometric overlap approach. Sci Rep. 2018 Jun 25;8(1):9588.

1042 42. Vojtek M, Zhang J, Sun J, Zhang M, Chambers I. Differential repression of Otx2
1043 underlies the capacity of NANOG and ESRRB to induce germline entry. Stem
1044 Cell Reports. 2022 Jan 11;17(1):35–42.

1045 43. Kubota H, Avarbock MR, Brinster RL. Growth factors essential for self-renewal
1046 and expansion of mouse spermatogonial stem cells. Proc Natl Acad Sci USA.
1047 2004 Nov 23;101(47):16489–94.

1048 44. Enders GC, May JJ. Developmentally regulated expression of a mouse germ cell
1049 nuclear antigen examined from embryonic day 11 to adult in male and female
1050 mice. Dev Biol. 1994 Jun;163(2):331–40.

1051 45. Carmell MA, Dokshin GA, Skaletsky H, Hu Y-C, van Wolfswinkel JC, Igarashi
1052 KJ, et al. A widely employed germ cell marker is an ancient disordered protein
1053 with reproductive functions in diverse eukaryotes. eLife. 2016 Oct 8;5.

1054 46. Buaas FW, Kirsh AL, Sharma M, McLean DJ, Morris JL, Griswold MD, et al. Plzf
1055 is required in adult male germ cells for stem cell self-renewal. Nat Genet. 2004
1056 Jun;36(6):647–52.

1057 47. Sato T, Katagiri K, Gohbara A, Inoue K, Ogonuki N, Ogura A, et al. In vitro
1058 production of functional sperm in cultured neonatal mouse testes. Nature. 2011
1059 Mar 24;471(7339):504–7.

1060 48. Hayashi K, Saitou M. Generation of eggs from mouse embryonic stem cells and
1061 induced pluripotent stem cells. Nat Protoc. 2013 Aug;8(8):1513–24.

1062 49. Jo K, Teague S, Chen B, Khan HA, Freeburne E, Li H, et al. Efficient
1063 differentiation of human primordial germ cells through geometric control reveals a
1064 key role for Nodal signaling. eLife. 2022 Apr 8;11.

1065 50. Wang B, Wang M, Zhang W, Xiao T, Chen C-H, Wu A, et al. Integrative analysis
1066 of pooled CRISPR genetic screens using MAGeCKFlute. Nat Protoc. 2019
1067 Mar;14(3):756–80.

1068 51. Mochizuki K, Hayashi Y, Sekinaka T, Otsuka K, Ito-Matsuoka Y, Kobayashi H, et
1069 al. Repression of somatic genes by selective recruitment of HDAC3 by BLIMP1 is
1070 essential for mouse primordial germ cell fate determination. Cell Rep. 2018 Sep
1071 4;24(10):2682-2693.e6.

1072 52. Guenther MG, Lane WS, Fischle W, Verdin E, Lazar MA, Shiekhattar R. A core
1073 SMRT corepressor complex containing HDAC3 and TBL1, a WD40-repeat
1074 protein linked to deafness. Genes Dev. 2000 May 1;14(9):1048–57.

1075 53. Kim H-J, Bae S-C. Histone deacetylase inhibitors: molecular mechanisms of
1076 action and clinical trials as anti-cancer drugs. Am J Transl Res. 2011
1077 Feb;3(2):166–79.

1078 54. Choi HW, Joo JY, Hong YJ, Kim JS, Song H, Lee JW, et al. Distinct enhancer
1079 activity of oct4 in naive and primed mouse pluripotency. Stem Cell Reports. 2016
1080 Nov 8;7(5):911–26.

1081 55. Lyu X, Cui Y, Kong Y, Yang M, Shen H, Liao S, et al. A transient transcriptional
1082 activation governs unpolarized-to-polarized morphogenesis during embryo
1083 implantation. Mol Cell. 2024 Jul 25;84(14):2665-2681.e13.

1084 56. Aramaki S, Hayashi K, Kurimoto K, Ohta H, Yabuta Y, Iwanari H, et al. A
1085 mesodermal factor, T, specifies mouse germ cell fate by directly activating
1086 germline determinants. Dev Cell. 2013 Dec 9;27(5):516–29.

1087 57. Tischler J, Gruhn WH, Reid J, Allgeyer E, Buettner F, Marr C, et al. Metabolic
1088 regulation of pluripotency and germ cell fate through α-ketoglutarate. EMBO J.
1089 2019 Jan 3;38(1).

1090 58. Hackett JA, Huang Y, Günesdogan U, Gretarsson KA, Kobayashi T, Surani MA.
1091 Tracing the transitions from pluripotency to germ cell fate with CRISPR
1092 screening. Nat Commun. 2018 Oct 16;9(1):4292.

1093 59. Naitou Y, Nagamatsu G, Hamazaki N, Shirane K, Hayashi M, Hayashi M, et al.
1094 Dual role of Ovol2 on the germ cell lineage segregation during gastrulation in
1095 mouse embryogenesis. Development. 2022 Feb 15;149(4).

1096 60. Yin Y, Cao S, Fu H, Fan X, Xiong J, Huang Q, et al. A noncanonical role of
1097 NOD-like receptor NLRP14 in PGCLC differentiation and spermatogenesis. Proc
1098 Natl Acad Sci USA. 2020 Sep 8;117(36):22237–48.

1099 61. Soper SFC, van der Heijden GW, Hardiman TC, Goodheart M, Martin SL, de
1100 Boer P, et al. Mouse maelstrom, a component of nuage, is essential for
1101 spermatogenesis and transposon repression in meiosis. Dev Cell. 2008
1102 Aug;15(2):285–97.

1103 62. Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, et al. A
1104 draft map of the human proteome. Nature. 2014 May 29;509(7502):575–81.

1105 63. Malki S, van der Heijden GW, O'Donnell KA, Martin SL, Bortvin A. A role for
1106 retrotransposon LINE-1 in fetal oocyte attrition in mice. Dev Cell. 2014 Jun
1107 9;29(5):521–33.

1108 64. Schemmer J, Araúzo-Bravo MJ, Haas N, Schäfer S, Weber SN, Becker A, et al.
1109 Transcription factor TFAP2C regulates major programs required for murine fetal
1110 germ cell maintenance and haploinsufficiency predisposes to teratomas in male
1111 mice. PLoS ONE. 2013 Aug 13;8(8):e71113.

1112 65. Irie N, Lee S-M, Lorenzi V, Xu H, Chen J, Inoue M, et al. DMRT1 regulates
1113 human germline commitment. Nat Cell Biol. 2023 Oct;25(10):1439–52.

1114 66. Eguizabal C, Montserrat N, Vassena R, Barragan M, Garreta E, Garcia-Quevedo
1115 L, et al. Complete meiosis from human induced pluripotent stem cells. Stem
1116 Cells. 2011 Aug;29(8):1186–95.

1117 67. Ishikura Y, Yabuta Y, Ohta H, Hayashi K, Nakamura T, Okamoto I, et al. In Vitro
1118 Derivation and Propagation of Spermatogonial Stem Cell Activity from Mouse
1119 Pluripotent Stem Cells. Cell Rep. 2016 Dec 6;17(10):2789–804.

1120 68. Bleckwehl T, Crispatzu G, Schaaf K, Respuela P, Bartusel M, Benson L, et al.
1121 Enhancer-associated H3K4 methylation safeguards in vitro germline
1122 competence. Nat Commun. 2021 Oct 1;12(1):5771.

1123 69. Li L, Li L, Li Q, Liu X, Ma X, Yong J, et al. Dissecting the epigenomic dynamics
1124 of human fetal germ cell development at single-cell resolution. Cell Res. 2021
1125 Apr;31(4):463–77.

1126 70. Alexander AK, Rice EJ, Lujic J, Simon LE, Tanis S, Barshad G, et al. A-MYB
1127 and BRDT-dependent RNA Polymerase II pause release orchestrates

1128 transcriptional regulation in mammalian meiosis. Nat Commun. 2023 Mar
1129 29;14(1):1753.

1130 71. Liu H, Wang J, He T, Becker S, Zhang G, Li D, et al. Butyrate: A Double-Edged
1131 Sword for Health? Adv Nutr. 2018 Jan 1;9(1):21–9.

1132 72. Jentink J, Loane MA, Dolk H, Barisic I, Garne E, Morris JK, et al. Valproic acid
1133 monotherapy in pregnancy and major congenital malformations. N Engl J Med.
1134 2010 Jun 10;362(23):2185–93.

1135 73. Kurilo LF. Oogenesis in antenatal development in man. Hum Genet.
1136 1981;57(1):86–92.

1137 74. Baker TG. A quantitative and cytological study of germ cells in human ovaries.
1138 Proc R Soc Lond, B, Biol Sci. 1963 Oct 22;158:417–33.

1139 75. Babakhanzadeh E, Nazari M, Ghasemifar S, Khodadadian A. Some of the
1140 factors involved in male infertility: A prospective review. Int J Gen Med. 2020 Feb
1141 5;13:29–41.

1142 76. Mann U, Shiff B, Patel P. Reasons for worldwide decline in male fertility. Curr
1143 Opin Urol. 2020 May;30(3):296–301.

1144 77. Smith A. Formative pluripotency: the executive phase in a developmental
1145 continuum. Development. 2017 Feb 1;144(3):365–73.

1146 78. Chavez A, Scheiman J, Vora S, Pruitt BW, Tuttle M, P R Iyer E, et al. Highly
1147 efficient Cas9-mediated transcriptional programming. Nat Methods. 2015
1148 Apr;12(4):326–8.

1149 79. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-Seq:
1150 Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled
1151 Genetic Screens. Cell. 2016 Dec 15;167(7):1853-1866.e17.

1152 80. Ding X, Singh P, Schimenti K, Tran TN, Fragoza R, Hardy J, et al. In vivo versus
1153 in silico assessment of potentially pathogenic missense variants in human
1154 reproductive genes. Proc Natl Acad Sci USA. 2023 Jul 25;120(30):e2219925120.

1155 81. Ding X, Schimenti JC. Female infertility from oocyte maturation arrest:
1156 assembling the genetic puzzle. EMBO Mol Med. 2023 Jun 7;15(6):e17729.

1157 82. Luo Y, Hartford SA, Zeng R, Southard TL, Shima N, Schimenti JC.
1158 Hypersensitivity of primordial germ cells to compromised replication-associated
1159 DNA repair involves ATM-p53-p21 signaling. PLoS Genet. 2014 Jul
1160 10;10(7):e1004471.

1161 83. Zhou Q, Wang M, Yuan Y, Wang X, Fu R, Wan H, et al. Complete Meiosis from
1162 Embryonic Stem Cell-Derived Germ Cells In Vitro. Cell Stem Cell. 2016 Mar
1163 3;18(3):330–40.

1164 84. Bock C, Reither S, Mikeska T, Paulsen M, Walter J, Lengauer T. BiQ Analyzer:
1165 visualization and quality control for DNA methylation data from bisulfite
1166 sequencing. Bioinformatics. 2005 Nov 1;21(21):4067–8.

1167 85. Read A, Gao S, Batchelor E, Luo J. Flexible CRISPR library construction using
1168 parallel oligonucleotide retrieval. Nucleic Acids Res. 2017 Jun 20;45(11):e101.

1169 86. Liu SJ, Horlbeck MA, Weissman JS, Lim DA. Genome-Scale Perturbation of
1170 Long Noncoding RNA Expression Using CRISPR Interference. Methods Mol Biol.
1171 2021;2254:323–38.

1172 87. Ding X, Fragoza R, Singh P, Zhang S, Yu H, Schimenti JC. Variants in RABL2A
1173 causing male infertility and ciliopathy. Hum Mol Genet. 2020 Dec
1174 18;29(20):3402–11.

1175 88. Ge SX, Son EW, Yao R. iDEP: an integrated web application for differential
1176 expression and pathway analysis of RNA-Seq data. BMC Bioinformatics. 2018
1177 Dec 19;19(1):534.

1178 89. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et
1179 al. Gene set enrichment analysis: a knowledge-based approach for interpreting
1180 genome-wide expression profiles. Proc Natl Acad Sci USA. 2005 Oct
1181 25;102(43):15545–50.

1182 90. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for
1183 microarray and RNA-seq data. BMC Bioinformatics. 2013 Jan 16;14:7.

1184 91. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-
1185 based method for gene-set enrichment visualization and interpretation. PLoS
1186 ONE. 2010 Nov 15;5(11):e13984.

1187 92. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al.
1188 Cytoscape: a software environment for integrated models of biomolecular
1189 interaction networks. Genome Res. 2003 Nov;13(11):2498–504.

1190 93. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, et
1191 al. Pathway enrichment analysis and visualization of omics data using g:Profiler,
1192 GSEA, Cytoscape and EnrichmentMap. Nat Protoc. 2019 Feb;14(2):482–517.

1193 94. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The
1194 STRING database in 2023: protein-protein association networks and functional
1195 enrichment analyses for any sequenced genome of interest. Nucleic Acids Res.

1196 2023 Jan 6;51(D1):D638–46.

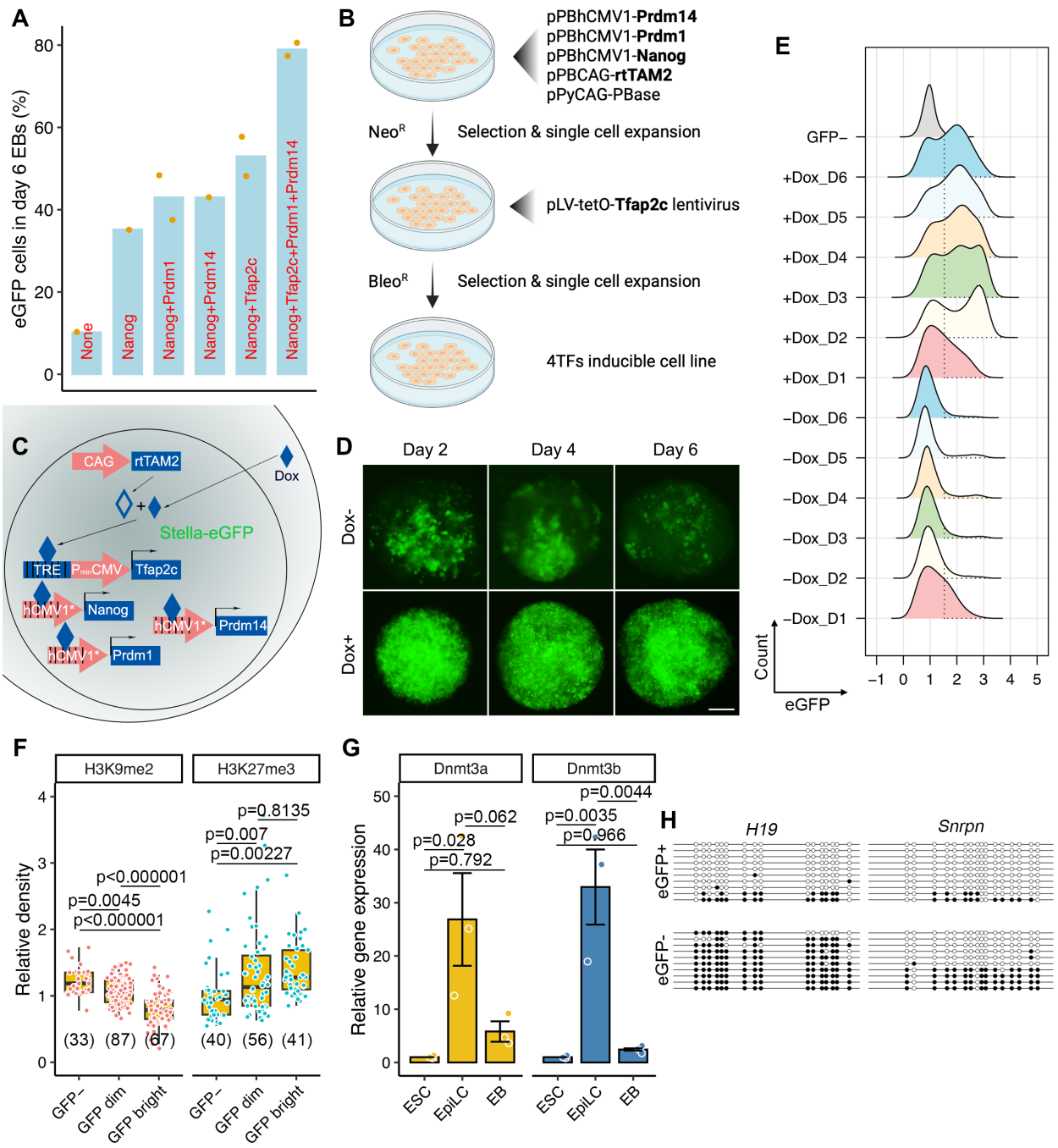

Figure 1. Construction of the 4TF-induced PGCLC differentiation platform. A. Percentages of day 6 PGCLCs (*Stella-eGFP*⁺ cells) differentiated from Dox-inducible ESCs harboring distinct combinations of TFs. Each individual clonal line is represented by a dot. **B.** The process for constructing 4TF-inducible ESCs. The 4TFs and rtTA are in bold. **C.** Schema illustrating the 4TF-induced PGCLC differentiation from ESCs. **D.** Representative EBs with eGFP signal in the presence or absence of Dox during the 6-day period. Scale bar in D represents 100 μm. **E.** FACS patterns of *Stella-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 *Stella-eGFP*⁺ and *Stella-eGFP*⁻ 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.

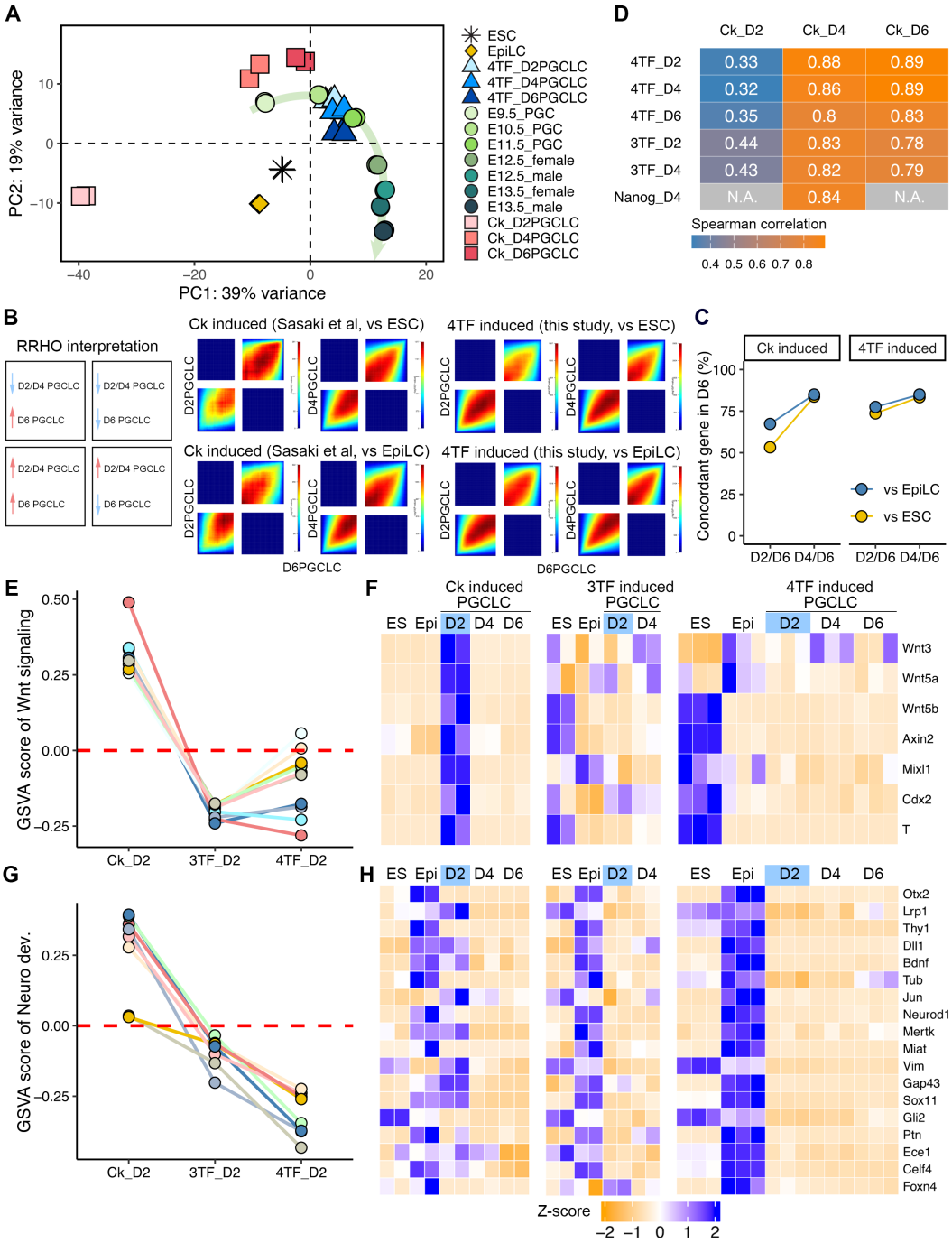

Figure 2. 4TF-induced PGCLCs share similar transcriptomes with cytokine- and 3TF-induced PGCLCs. A. PCA of indicated transcriptomes from 4TF- and cytokine (Ck)-induced PGCLCs, *in vivo* germ cells at indicated times of gestation, EpiLCs and ESCs. "D" = days of induction. PGCLCs were isolated by flow sorted for diagnostic markers. **B.** 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. **C.** Percentage of concordant genes encompassing both up and down regulated genes compared with D6PGCLCs. **D.** Spearman correlation coefficients between PGCLCs. 4TF_D2/4/6PGCLCs were compared with RNA-seq datasets of Ck_D2/4/6PGCLCs. 3TF_D2/4PGCLCs were compared with microarray datasets of Ck_D2/4/6PGCLCs, generated using the Affymetrix Mouse Genome 430 2.0 Array. Nanog_D4PGCLC was compared with microarray dataset of Ck_D4PGCLC, generated using the Illumina MouseWG-6 v2.0 expression beadchip. N.A., not available. **E.** GSVA score of Wnt signaling pathways (10 out of the 11 pathways in total were selected, with one specific to the planar cell polarity being excluded) 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. **G.** GSVA score of neurodevelopment-related pathways (see Fig. S4C) in D2PGCLC between Ck-, 3TF- and 4TF-induced conditions (Ck_D2, 3TF_D2 and 4TF_D2). **H.** Heatmap of selected genes associated with neurodevelopment.

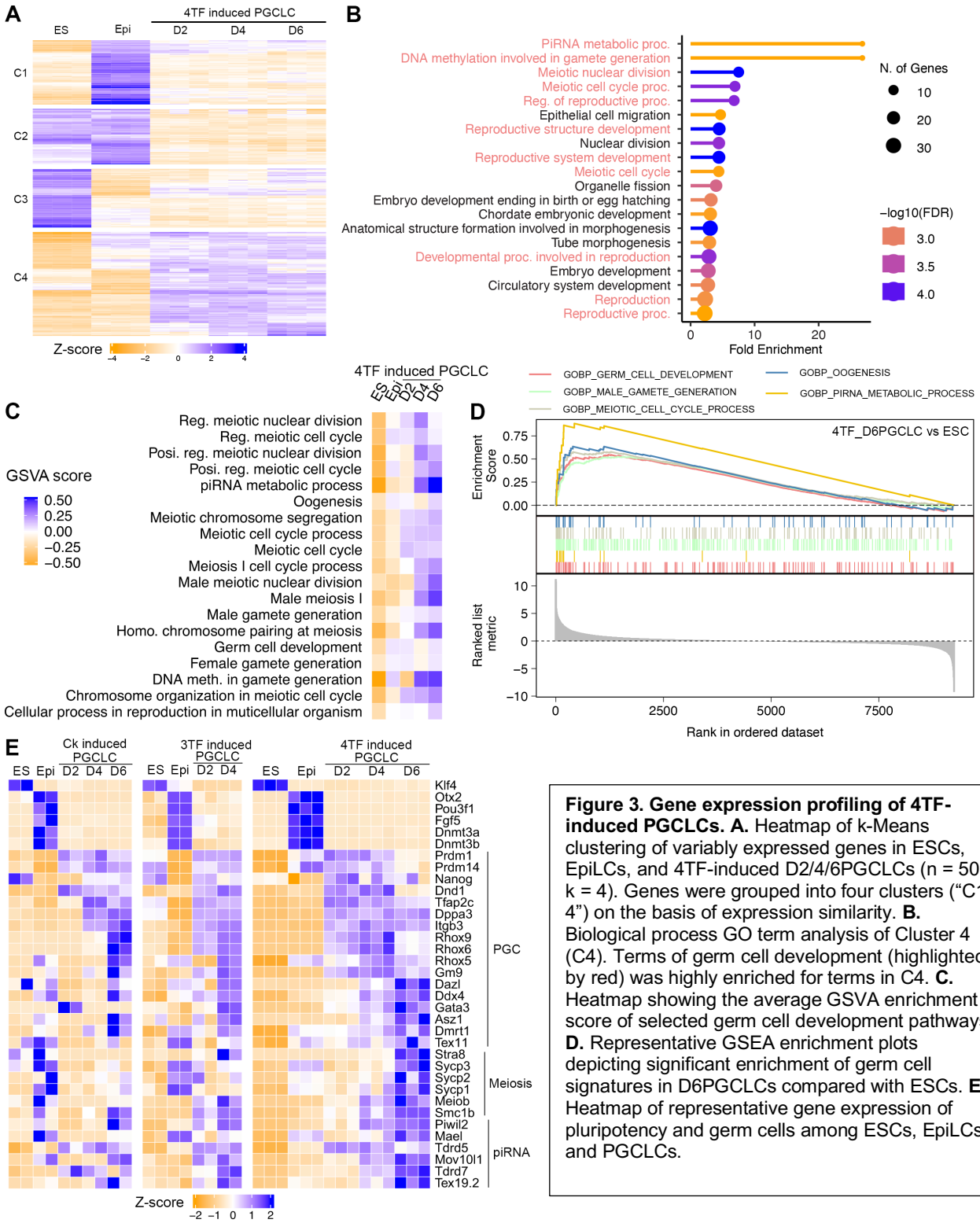

Figure 3. Gene expression profiling of 4TF-induced PGCLCs. A. Heatmap of k-Means clustering of variably expressed genes in ESCs, EpiLCs, and 4TF-induced D2/4/6PGCLCs (n = 500, k = 4). Genes were grouped into four clusters ("C1-4") on the basis of expression similarity. **B.** Biological process GO term analysis of Cluster 4 (C4). Terms of germ cell development (highlighted by red) was highly enriched for terms in C4. **C.** Heatmap 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.

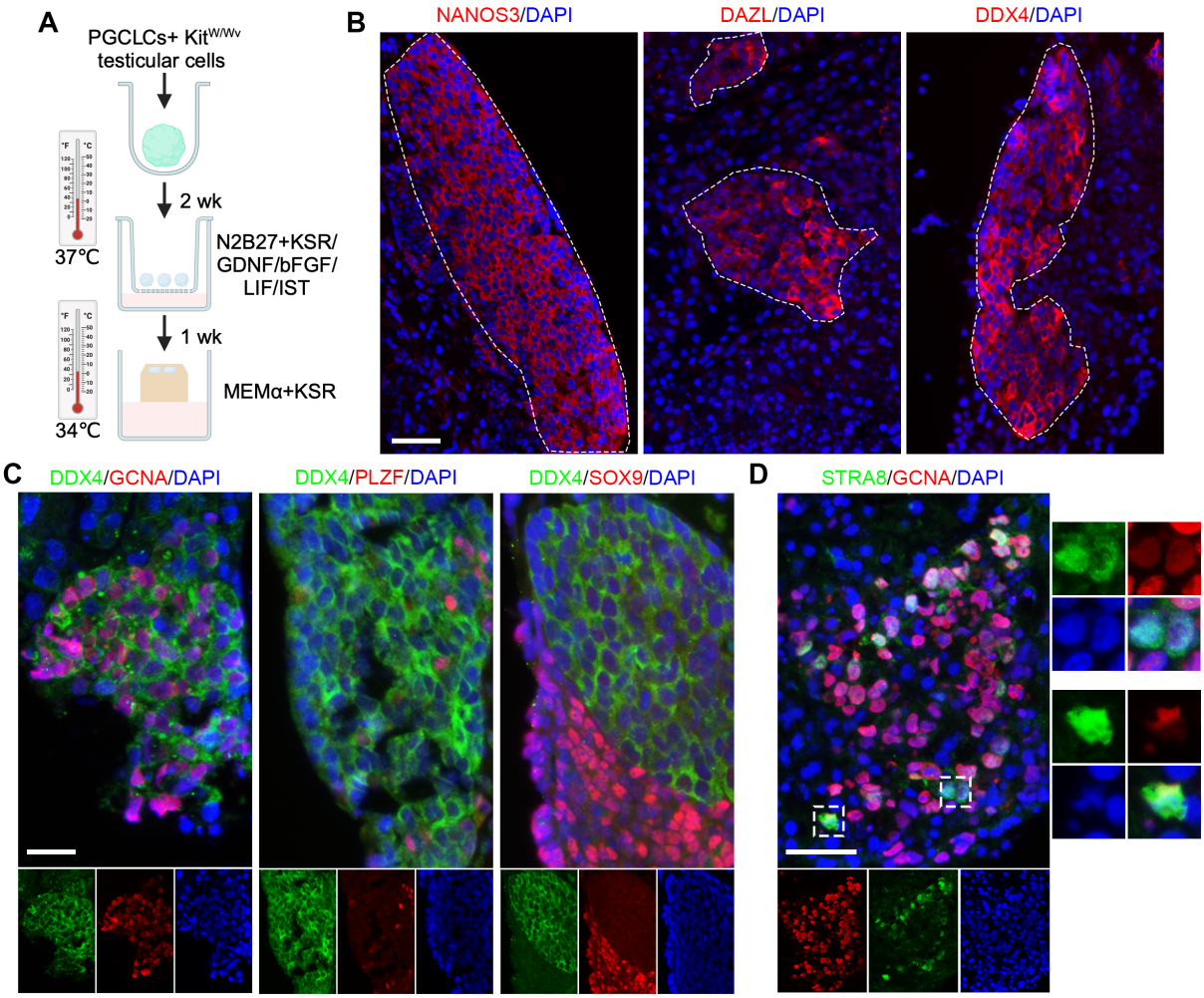

Figure 4. Development of cells past PGCLC stage. A. Further development of 4TF_D6PGCLCs by aggregation with testicular somatic cells isolated from *Kit$^{w/wv}$* mice. The cultured aggregates were subjected to IF. **B.** Clusters of germ cells labeling by NANOS3, DAZL and DDX4 were identified from aggregates cultured after 2 weeks. Dashed lines denote regions with germ cell clusters. **C.** Dual IF staining of DDX4 with GCNA, PLZF and SOX9 in the aggregates after 2-week culture. **D.** Dual IF staining of GCNA and STRA8 in the aggregates cultured for an additional week. The boxed region are magnifications of representive dual positive cells. Scale bars in B, C and D represent 50 μm.

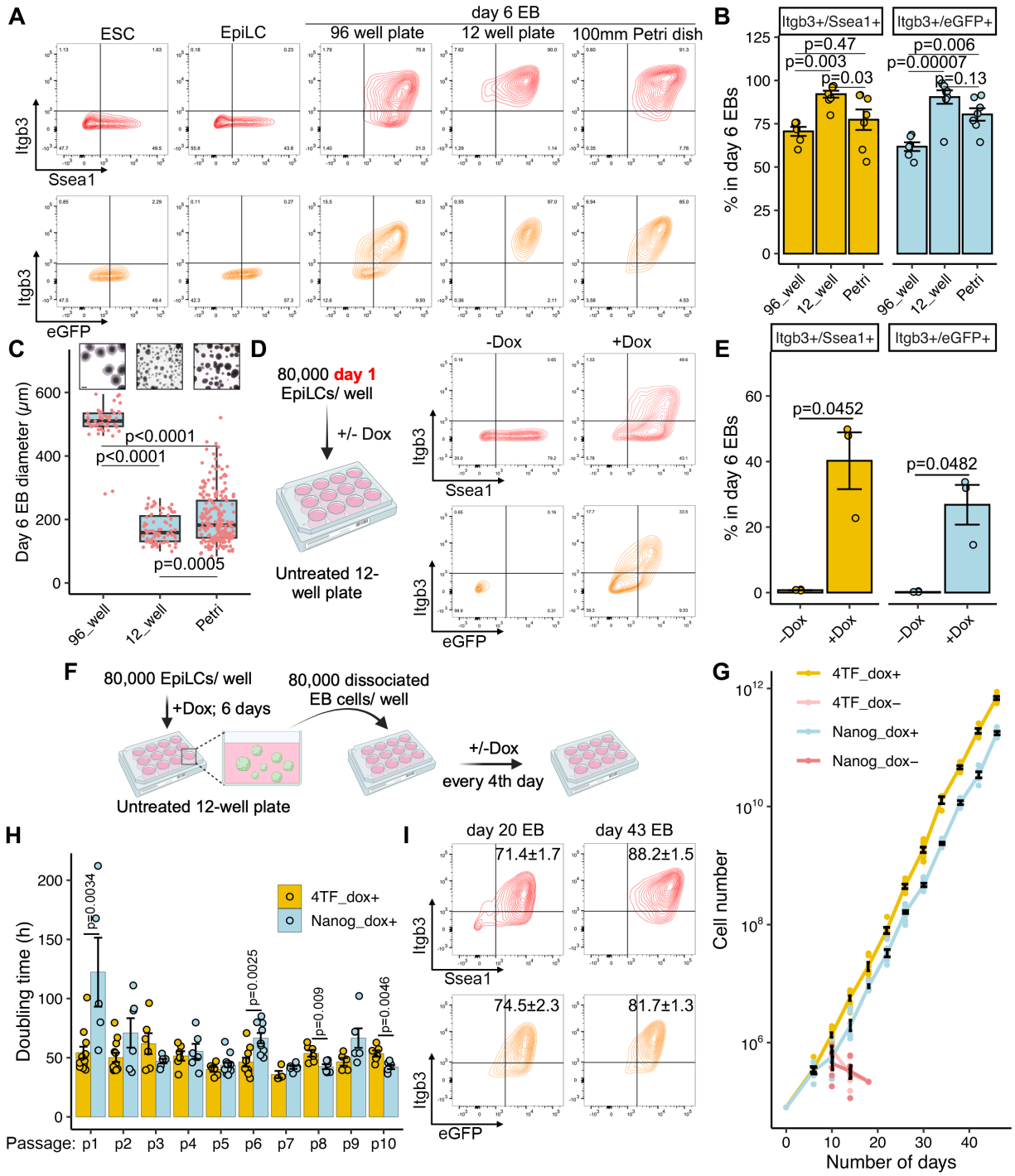

Figure 5. Scale-up and long-term differentiation of 4TF-induced PGCLCs. A. Representative FACS pattern of Itgb3$^+$/Ssea1$^+$ and Itgb3$^+$/eGFP$^+$ cells in ESCs, EpiLCs and day 6 EBs from U bottom 96-well plates, 12-well non-treated plates and 100mm bacteriological petri dishes. eGFP, Stella-eGFP. **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.** Induction of 4TF-induced PGCLCs from day 1 EpiLCs. **E.** Quantification of 4TF-induced PGCLC populations in day 6 EBs in the presence or absence of Dox. Day 1 EpiLCs were used for EB formation in non-treated 12-well plates. **F.** Schema depicting the process of EB passage. The dissociated EB cells were passaged every 4 days into each well of non-treated 12-well plates. **G.** Growth curves of long-term cultured EBs induced by 4TFs, Nanog or not. **H.** Doubling time of passaged EBs in 4TF- and Nanog-induced PGCLC differentiation systems. **I.** FACS pattern of Itgb3$^+$/Ssea1$^+$ and Itgb3$^+$/eGFP$^+$ cells induced by 4TFs in day 20 and 43 EBs cultured in 12-well non-treated plates. Data in B, C, E, G and H are represented as the mean ± SEM. Data in B and C were analyzed using one-way ANOVA with Tukey's *post hoc* test. Data in E were analyzed using a two-tailed paired *t* test. Data in H were analyzed using a two-tailed unpaired *t* test. Scale bar in C represents 200 μm.

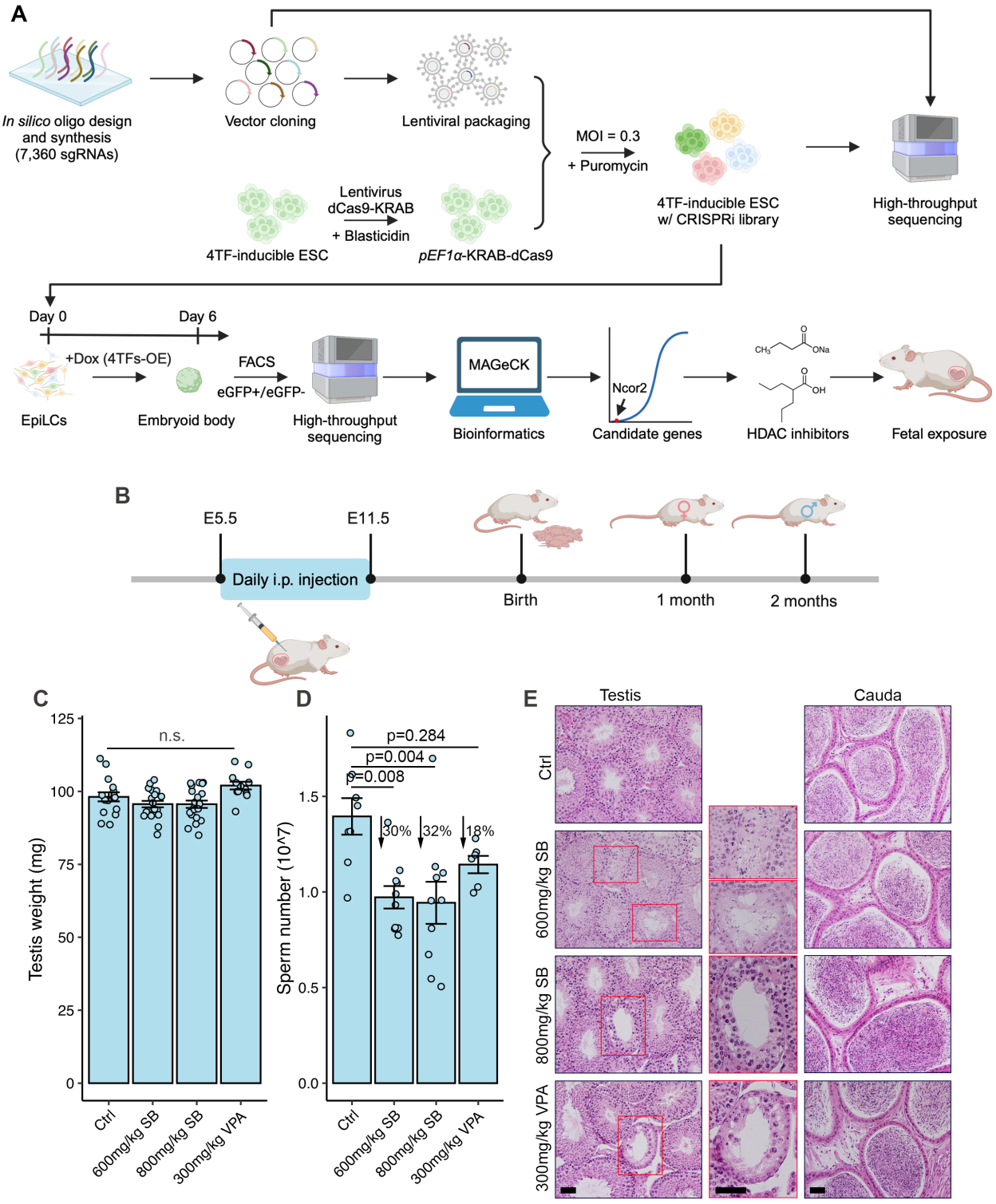

Figure 6. Application of 4TF-inducible system for high throughput screening epigenetic regulators. A. Schema of CRISPRi library generation and validation is shown on the top. The uniformity of sgRNA library representation was tested by sequencing of batch-cloned plasmids and lentivirus-infected ESCs (colored cell clusters). On the bottom is a schema of the PGCLC differentiation protocol and CRISPRi screening process. Day 6 EBs were sorted into Stella-eGFP$^+$ and Stella-eGFP$^-$ cells to generate count representatives for downstream analysis. Candidate gene *Ncor2* and HDAC inhibitors SB and VPA were selected for validation. OE, overexpression. **B.** Experimental schema showing SB and VPA treatment on pregnant mice and analysis their pups. i.p., intraperitoneal. **C.** Testis weight of 2-month-old pups. n.s., no significant difference. **D.** Sperm counts. **E.** Histological analyses of 2-month-old testes and cauda epididymis. The boxed region are magnifications of atrophic seminiferous tubules. Data in C and D are represented as the mean ± SEM and were analyzed using one-way ANOVA with Tukey's *post hoc* test. Scale bars in E represent 50 μm.

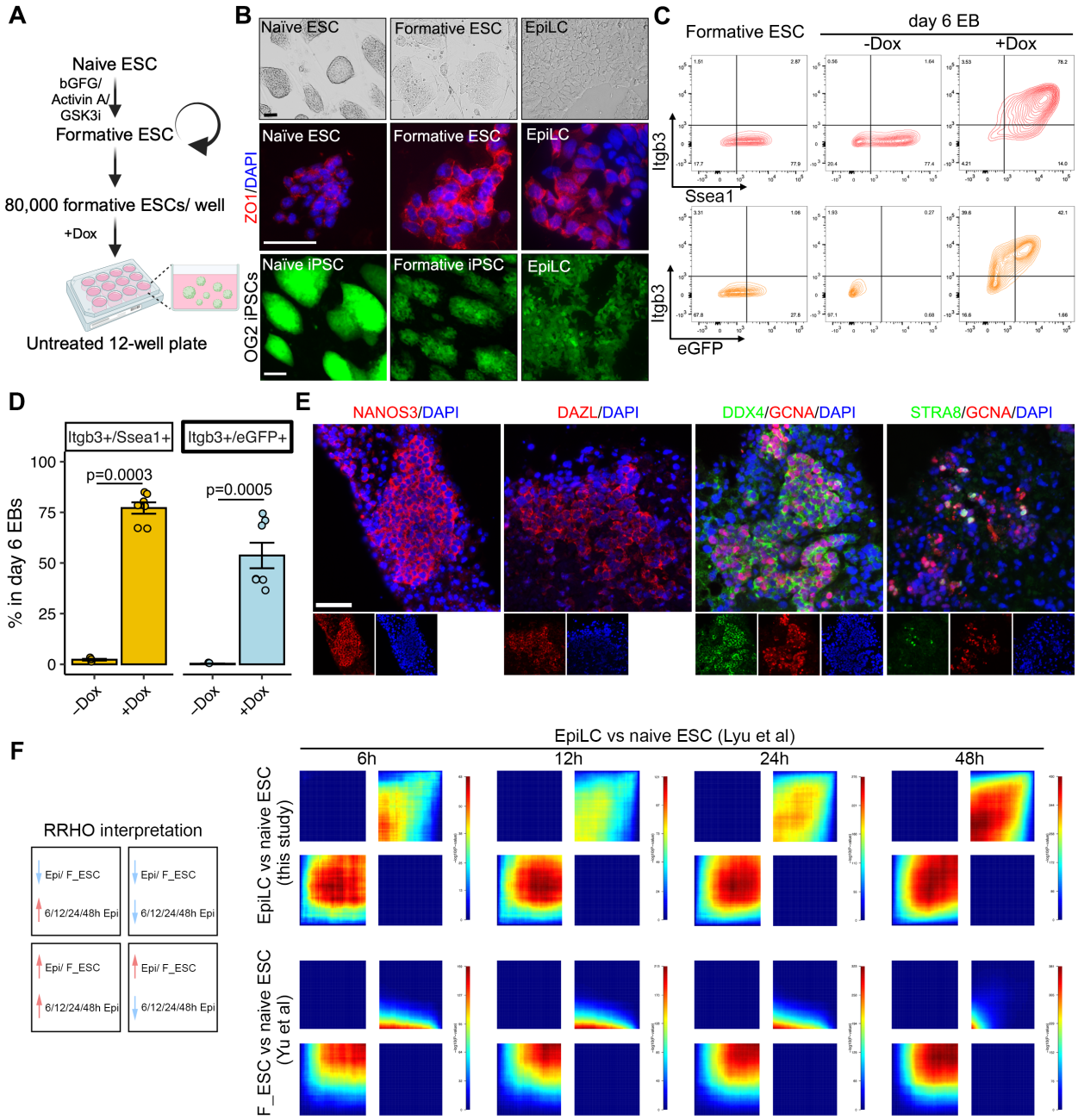

Figure 7. Direct differentiation of PGCLCs from formative 4TF-inducible ESCs. A. Schema for inducing PGCLCs from formative ESCs. **B.** Characterization of formative ESC derived from naïve ESCs. Top panel: representative morphologies of naïve ESCs, formative ESCs and EpiLCs. Middle panel: IF staining of tight junction protein ZO1 on naïve ESCs, formative ESCs and EpiLCs. Bottom panel: GFP expression in naïve and formative OG2 (Oct4-DE-GFP) iPSCs, and EpiLCs differentiated from naïve OG2 iPSCs. **C.** Representative FACS pattern of Itgb3^{+}/Ssea1^{+} and Itgb3^{+}/Stella-eGFP^{+} cells in formative ESCs and day 6 EBs with or without Dox treatment in 12-well non-treated plates. **D.** Quantification of PGCLC subpopulations after 4TF induction. **E.** IF staining of germ cell markers in cultured D6PGCLCs (*Stella-eGFP*$^{+}$) differentiated from formative ESCs aggregated with testicular cells from *Kit*$^{w/wv}$ mice. **F.** RRHO analysis of transcriptomes from 4TF formative ESCs and EpiLCs compared to those from a published study (55). Sampled time points during differentiation are indicated (6h, 12h, 24h, and 48h). Data in D are represented as the mean ± SEM and analyzed using a two-tailed paired *t* test. Scale bars in top and bottom panels of B represent 100μm, and in middle panel of B and E represent 50 μm.