

1   ***In vitro* reconstitution of epigenetic reprogramming in the human germ  
2   line**

3

4

5

6   Yusuke Murase,<sup>1,2,#</sup> Ryuta Yokogawa,<sup>1,2,#</sup> Yukihiro Yabuta,<sup>1,2</sup> Masahiro Nagano,<sup>1,2</sup>  
7   Yoshitaka Katou,<sup>1,2</sup> Manami Mizuyama,<sup>1,2</sup> Ayaka Kitamura,<sup>1,2</sup> Pimpitcha  
8   Puangsricharoen,<sup>1,2</sup> Chika Yamashiro,<sup>2</sup> Bo Hu,<sup>1,2</sup> Ken Mizuta,<sup>1,2</sup> Kosuke Ogata,<sup>3</sup> Yasushi  
9   Ishihama,<sup>3</sup> and Mitinori Saitou<sup>1,2,4</sup>

10

11

12

13   <sup>1</sup>Institute for the Advanced Study of Human Biology (ASHBi), Kyoto University,  
14   Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan.

15   <sup>2</sup>Department of Anatomy and Cell Biology, Graduate School of Medicine, Kyoto  
16   University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan.

17   <sup>3</sup>Department of Molecular Systems BioAnalysis, Graduate School of Pharmaceutical  
18   Sciences, Kyoto University, Kyoto 606-8501, Japan.

19   <sup>4</sup>Center for iPS Cell Research and Application (CiRA), Kyoto University, 53 Kawahara-  
20   cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan.

21   <sup>#</sup>These authors contributed equally to this work.

22

23

24   *Running title: Human germline epigenetic reprogramming in vitro*

25   *Key words: epigenetic reprogramming; human germ cells; human primordial germ cell-  
26   like cells; in vitro gametogenesis; bone morphogenetic protein; DNA demethylation;  
27   TET1*

28

29

30   *\*Correspondence should be addressed to:*

31   Mitinori Saitou, M.D., Ph.D.

32   E-mail: [saitou@anat2.med.kyoto-u.ac.jp](mailto:saitou@anat2.med.kyoto-u.ac.jp)

33   Tel: +81-75-753-4335; Fax: +81-75-751-7286 (MS)

34

35

36

37

38 **SUMMARY**

39 Epigenetic reprogramming resets parental epigenetic memories and differentiates  
40 primordial germ cells (PGCs) into mitotic pro-spermatogonia or oogonia, ensuring  
41 sexually dimorphic germ-cell development for totipotency<sup>1</sup>. *In vitro* reconstitution  
42 of epigenetic reprogramming in humans remains a fundamental challenge. Here,  
43 we establish a robust strategy for inducing epigenetic reprogramming and  
44 differentiation of pluripotent stem cell (PSC)-derived human PGC-like cells  
45 (hPGCLCs) into mitotic pro-spermatogonia or oogonia, coupled with their extensive  
46 amplification ( $\sim >10^{10}$ -fold). Strikingly, bone morphogenetic protein (BMP)  
47 signalling is a key driver of these processes: BMP-driven hPGCLC differentiation is  
48 accompanied by an attenuation of the mitogen-activated protein  
49 kinase/extracellular-regulated kinase (MAPK/ERK) pathway and both *de novo* and  
50 maintenance DNA methyltransferase (DNMT) activities, likely promoting  
51 replication-coupled, passive DNA demethylation. On the other hand, hPGCLCs  
52 deficient in tens-eleven translocation (TET) 1, an active DNA demethylase abundant  
53 in human germ cells<sup>2,3</sup>, differentiate into extraembryonic cells, including amnion,  
54 with de-repression of key genes bearing bivalent promoters; these cells fail to fully  
55 activate genes vital for spermatogenesis and oogenesis, with their promoters  
56 remaining methylated. Our study elucidates the framework of epigenetic  
57 reprogramming in humans, making a fundamental advance in human biology, and  
58 through the generation of abundant mitotic pro-spermatogonia and oogonia-like  
59 cells, represents a milestone for human *in vitro* gametogenesis (IVG) research and  
60 its potential translation into reproductive medicine.

61

62 Germ cells give rise to totipotency and ensure heredity and evolution. Human  
63 primordial germ cells (hPGCs) are thought to be specified at around embryonic day (E)  
64 12–16 [2 weeks post-fertilisation (2 wpf)] in the amnion or posterior epiblast of early  
65 post-implantation embryos<sup>4–6</sup>. They undergo migration through the yolk-sac and  
66 hindgut endoderm, eventually colonizing genital ridges from around 5–6 wpf<sup>7</sup>. During  
67 this period, hPGCs initiate epigenetic reprogramming, resetting parental epigenetic  
68 memories through genome-wide DNA demethylation [5-methylcytosine (5mC)  
69 demethylation] and histone modification remodelling<sup>3,8–10</sup>. By around 7–8 wpf, hPGCs  
70 show a completion of the reprogramming and differentiate into either mitotic pro-  
71 spermatogonia or oogonia, which are precursors for spermatogonia or oocyte  
72 differentiation, respectively (Fig. 1a)<sup>3,8–13</sup>.

73

74 *In vitro* gametogenesis (IVG) from pluripotent stem cells (PSCs) provides a framework  
75 for clarifying the mechanism of germ-cell development<sup>14</sup>. Accordingly, human PSCs  
76 (hPSCs) have been induced into human PGC-like cells (hPGCLCs)<sup>15–17</sup>, which, upon  
77 aggregation culture with mouse embryonic testicular or ovarian somatic cells [xenogeneic  
78 reconstituted testes or ovaries (xrTestes or Ovaries)], undergo epigenetic reprogramming

79 and differentiate into either pro-spermatogonia or oogonia-like cells, respectively<sup>17-19</sup>.  
80 However, the xrTestis/Ovary systems for hPGCLC differentiation are low in efficiency  
81 and bear limited capacity for scaling and experimental control. While hPGCLCs can be  
82 co-cultured with human hindgut organoids for differentiation, this also achieves only  
83 limited differentiation<sup>20</sup>. Thus, to explore the mechanism for hPGCLC differentiation  
84 and accelerate human IVG, a more robust methodology would be required. Here, we  
85 embarked on establishing a system for signalling molecule-driven hPGCLC  
86 differentiation.

87

### 88 Screening of signalling for hPGCLC differentiation

89 hPGCLCs cultured on m220 feeder cells under reported conditions propagate as early  
90 PGCs, but tend to de-differentiate, and therefore require cell sorting for passage  
91 (Extended Data Fig. 1a)<sup>17</sup>. We first sought to determine a condition that would  
92 minimize the de-differentiation. hiPSCs bearing *BLIMP1 (PRDM1)-tdTomato (BT)* and  
93 *TFAP2C-EGFP (AG)* alleles [585B1 *BTAG* (XY): M1-*BTAG*] (Fig. 1b) were induced into  
94 incipient mesoderm-like cells (iMeLCs) and then into BT-positive<sup>(+)</sup> AG<sup>+</sup> hPGCLCs for  
95 6 days (d6 hPGCLCs), and the resulting d6 hPGCLCs were cultured and passaged every  
96 ~10 days by cell sorting using flow cytometry (Extended Data Fig. 1a)<sup>17</sup>. Upon passage,  
97 the culture consisted of three cell populations: BT<sup>+</sup>AG<sup>+</sup> hPGCLCs, BT-negative<sup>(-)</sup>  
98 AG<sup>-</sup>/TRA-1-85<sup>+</sup> [a human-specific antigen<sup>21</sup>] de-differentiated cells, and TRA-1-85<sup>-</sup>  
99 feeders (Extended Data Fig. 1a). The ratio of hPGCLCs to de-differentiated cells was  
100 estimated as the ratio of BT<sup>+</sup>AG<sup>+</sup> cells to forward scatter-high cells (the hPGCLC  
101 enrichment score) (Extended Data Fig. 1a). Using this criterion, we evaluated the  
102 impact of inhibiting WNT, NODAL, and BMP signals on hPGCLC de-differentiation.  
103 Inhibiting NODAL or BMP signalling resulted in a decrease of the enrichment score,  
104 whereas inhibiting WNT signalling with IWR1<sup>22</sup> increased the enrichment score in a  
105 dose-dependent manner (Extended Data Fig. 1b, c). We then examined the impact of  
106 basal media and found that advanced RPMI (advRPMI) achieved a large increase in the  
107 enrichment score (Extended Data Fig. 1d). Accordingly, hPGCLCs cultured with IWR1  
108 and advRPMI exhibited vigorous propagation with high enrichment scores and could be  
109 passaged without sorting at least three times (Extended Data Fig. 1e).

110

111 During hPGC(LC) differentiation, a set of genes for spermatogenesis and oogenesis are  
112 up-regulated coupled with genome-wide DNA demethylation<sup>3,8,9,18,19</sup>. We determined  
113 the genes that showed both progressive up-regulation during hPGCLC-to-oogonia-like  
114 cell differentiation and also a ≥50% reduction in the level of promoter 5mC from hiPSCs  
115 to oogonia-like cells [epigenetic reprogramming-activated (ER) genes]<sup>19</sup> (Fig. 1c and  
116 Extended Data Fig. 1f, g). Because ER genes such as *GTSF1*, *PRAME*, and *MEG3* are  
117 up-regulated early during hPGCLC differentiation, we screened signals that induce  
118 simultaneous up-regulation of *GTSF1*, *PRAME*, and *MEG3*. hPGCLCs were cultured  
119 with IWR1 and advRPMI, together with a panel of relevant cytokines and chemicals, and

120 the expression of the three genes was examined by quantitative reverse transcriptase PCR  
121 (qRT-PCR) on culture day (c) 22 (Fig. 1d). Interestingly, BMP ligands, but not the other  
122 cytokines or chemicals, up-regulated the three genes (Fig. 1e). We reasoned that signals  
123 inducing hPGC differentiation should be active in tissues during hPGC migration,  
124 including the hypoblast, yolk sac, and hindgut endoderm<sup>7</sup>. Indeed, re-analysis of the  
125 published single-cell RNA sequence (scRNA-seq) data for human development<sup>23-25</sup>  
126 revealed that the relevant tissues expressed *BMPs* (Extended Data Fig. 2a-e).

127

128 We cultured hPGCLCs with progressively increasing doses of BMP2. We found that  
129 the hPGCLCs exhibited robust growth and high enrichment scores with low BMP2 doses  
130 (up to ~25 ng/ml), while their expansion was attenuated with high BMP2 (Extended Data  
131 Fig. 2f, g). We then cultured hPGCLCs with 25 ng/ml of BMP2 or without BMP2 until  
132 c55. hPGCLCs cultured with BMP2 continued to propagate robustly with high  
133 enrichment scores, and immunofluorescence (IF) analysis revealed that a fraction of them  
134 expressed DAZL, a key ER gene product (Extended Data Fig. 2h). In contrast,  
135 hPGCLCs cultured without BMP2 showed de-differentiation after ~c32, and de-  
136 differentiated cells predominated in the culture at c55, with the remaining BT<sup>+</sup>AG<sup>+</sup> cells  
137 barely expressing DAZL (Extended Data Fig. 2h). These findings suggest that BMP  
138 signalling stabilizes the germ-cell fate and may promote epigenetic reprogramming and  
139 differentiation of hPGC(LC)s into mitotic pro-spermatogonia or oogonia.

140

#### 141 **BMP signalling promotes hPGCLC differentiation**

142 *DAZL* and *DDX4* are two key genes signifying the differentiation of hPGCs into mitotic  
143 pro-spermatogonia/oogonia, with *DAZL* being expressed earlier and at a higher level than  
144 *DDX4*<sup>3,8,9,13,19,26</sup> (Fig. 1a, c). We therefore established hiPSC lines bearing AG and  
145 *DAZL-tdTomato* (DT) or *DDX4/human YASA homologue-tdTomato* (VT) alleles [585B1  
146 *AGDT/VT*(XY): M1-*AGDT/VT*; NCLCN/1390G3 *AGVT*(XX): F1/F2-*AGVT*<sup>19</sup>] (Fig. 1b,  
147 Extended Data Fig. 3a-g).

148

149 First, we cultured hPGCLCs from M1-*AGDT* with 25 ng/ml of BMP2, and after the first  
150 passage, with three different doses of BMP2 (25, 100, or 200 ng/ml). The increases of  
151 BMP2 dosage attenuated the AG<sup>+</sup> cell expansion but accelerated the emergence of DT<sup>+</sup>  
152 cells (Extended Data Fig. 4a, b). We chose to culture hPGCLCs from this line first with  
153 25 ng/ml of BMP2 and thereafter with 100 ng/ml of BMP2 with a passage every ~10 days.  
154 Figure 2a-c shows a representative result: hPGCLCs expanded stably and up-regulated  
155 DT from ~c32; thereafter, the DT<sup>+</sup> cell ratio was progressively increased, and nearly all  
156 the cells became DT<sup>+</sup> at c140, with an overall expansion of ~10<sup>10</sup>-fold (see the “**Cell**  
157 **number count of hPGCLC-derived cells**” section of the **Methods**). In contrast,  
158 hPGCLCs cultured without BMP2 showed substantial degrees of de-differentiation after  
159 ~c32, with the AG<sup>+</sup> cells barely expressing DT at c42 (Extended Data Fig. 4c, d).  
160 During BMP-driven differentiation, both AG<sup>+</sup>DT<sup>-</sup> and DT<sup>+</sup> cells exhibited a spindle

shape with a large, ovoid nucleus, formed loosely packed colonies with no clear AG<sup>+</sup>DT<sup>-</sup>/DT<sup>+</sup> cell segregation, and were karyotypically normal (Fig. 2d, e). We found that under the same condition, hPGCLCs from M1-*AGVT* (Fig. 1b) expanded stably, up-regulating VT from ~c32, and nearly all the cells became VT<sup>+</sup> at c92, with an overall expansion of  $\sim >3 \times 10^7$ -fold (Fig. 2b, c, Extended Data Fig. 4e, f). Similarly, hPGCLCs from the 1383D6 (XY) line with no reporters (M2) (Fig. 1b)<sup>17</sup> showed a good expansion, and ~55% and ~92% of hPGCLC-derived cells expressed DDX4 at c54 and c86, respectively, with an overall expansion of  $\sim 10^5$ -fold (Extended Data Fig. 4g–i). An orthogonal validation of DT and VT reporters revealed that DT and VT positivity are a robust quantitative indicator for DAZL and DDX4 expression, respectively, while ~one-third of DT<sup>-</sup> cells at a late stage exhibited low/middle-level DAZL expression, potentially due to a sporadic transcriptional/post-transcriptional silencing of the *DT* allele (Supplementary Figure 1 and Discussion).

174

175 Next, we cultured hPGCLCs from F1-*AGVT*. They showed a de-differentiation  
176 propensity with 25 ng/ml of BMP2, but were propagated with high enrichment scores  
177 when the BMP2 concentration was higher than 50 ng/ml (Extended Data Fig. 4j, k).  
178 Based on these results, we cultured them first with 50 ng/ml of BMP2 and thereafter with  
179 100 ng/ml of BMP2 with a passage every ~10–15 days (Fig. 2a–c). hPGCLC-derived  
180 cells expanded stably with up-regulation of VT from ~c40, and  $\geq \sim 90\%$  of the cells  
181 became VT<sup>+</sup> after c129, with an overall expansion of  $\geq \sim 10^{10}$ -fold at c141. In contrast,  
182 hPGCLCs cultured without BMP2 showed poor expansion and de-differentiated  
183 substantially as early as ~c12, with the AG<sup>+</sup> cells barely expressing VT at c46 (Extended  
184 Data Fig. 4l, m). During BMP-driven differentiation, the female AG<sup>+</sup>VT<sup>-</sup>/VT<sup>+</sup> cells  
185 showed morphologies similar both to each other and to those of the male cells and were  
186 karyotypically normal (Fig. 2d, e). Under the same condition, hPGCLCs from F2-  
187 *AGVT* (Fig. 1b) exhibited a substantially slower yet stable expansion, and up-regulated  
188 VT as early as ~c26; at c98 nearly all the cells were VT<sup>+</sup>, with an overall expansion of  
189  $\sim >3 \times 10^3$ -fold (Fig. 2b, c, Extended Data Fig. 4n, o). Note that late in culture (~c82), we  
190 observed an emergence of AG<sup>-</sup>DT<sup>+</sup>/VT<sup>+</sup> cells, which thereafter constituted a small  
191 fraction (~5–10%) of the hPGCLC-derived cells (Fig. 2a, c, Extended Data Fig. 4e, f, n,  
192 o). These findings demonstrate that BMP signalling reproducibly promotes the  
193 differentiation of hPGCLCs from at least four independent hiPSC lines into the  
194 DAZL/DDX4<sup>+</sup> state, although their propagation/differentiation dynamics show a degree  
195 of line-dependent heterogeneity.

196

197 Both male DT<sup>+</sup> cells and female VT<sup>+</sup> cells expressed key ER genes at a high level (Fig.  
198 2f, Supplementary Figure 1 and Discussion), and compared to hiPSCs/hPGCLCs,  
199 exhibited substantially reduced genomic 5mC levels (Fig. 2g). Importantly, hPGCLC-  
200 derived cells could be stored frozen and were re-expandable after thawing (Extended Data  
201 Fig. 4p). They also propagated robustly and differentiated into the DT<sup>+</sup>/VT<sup>+</sup> state under

202 a serum-free condition (Extended Data Fig. 4q, r). While hPGCLC-derived cells  
203 occasionally differentiated into an AG<sup>-</sup>DT<sup>-</sup>/VT<sup>-</sup> state, the sorted AG<sup>+</sup> cells exhibited a  
204 robust propagation and differentiation into DT<sup>+</sup>/VT<sup>+</sup> cells (Fig. 2a). Collectively, these  
205 findings support the notion that BMP signalling stabilizes the germ-cell fate, promoting  
206 epigenetic reprogramming, differentiation of hPGCLCs into mitotic pro-spermatogonia  
207 or oogonia, and robust self-renewal of the resulting pro-spermatogonia or oogonia.  
208

## 209 Transcriptome dynamics

210 To determine the dynamics of the transcriptome, we performed RNA-seq analysis  
211 (Supplementary Tables 1, 2). In response to BMP2, hPGCLC-derived cells maintained  
212 pluripotency/early PGC genes, and up-regulated ER genes and some of the 13 previously  
213 reported genes that show up-regulation in gonadal germ cells<sup>26</sup>, but did not up-regulate  
214 meiosis genes (Fig. 3a, Extended Data Fig. 5a). The expression profiles of these genes  
215 were similar to those during oogonia-like cell differentiation in xrOvaries<sup>19</sup>, and to those  
216 during mitotic pro-spermatogonia/oogonia differentiation *in vivo*<sup>3</sup> (Fig. 3a, Extended  
217 Data Fig. 5a). Male, but not female, hPGCLC-derived cells expressed genes encoded  
218 on the Y chromosome (Fig. 3a, Extended Data Fig. 5a). hPGCLC-derived cells  
219 expressed *CDH5* and *DMRT1*, the genes used as markers for human germ cells from the  
220 migration stage onward<sup>27</sup> (Extended Data Fig. 5b). Principal component analysis  
221 (PCA) revealed that BMP signalling enhanced the transcriptome maturation of hPGCLC-  
222 derived cells, which was similar to that of oogonia-like cells in xrOvaries (Fig. 3b).  
223 Notably, the transcriptome of c107 AG<sup>-</sup>VT<sup>+</sup> cells was closer to that of ag120 AG<sup>-</sup>VT<sup>+</sup>  
224 cells in xrOvaries, which show a similarity to the retinoic acid (RA)-responsive, pre-  
225 leptotene state of the first meiotic prophase<sup>19</sup> (Fig. 3b). On the other hand, we noted  
226 that BMP-driven and xrOvary-based hPGCLC differentiations were distinct with PC3,  
227 with genes up-regulated for BMP-driven differentiation enriched with gene ontology  
228 (GO) terms such as “angiogenesis” and included *ID1/2/3*, *GATA2/3*, *MSX2*, and *HAND1*,  
229 while those up-regulated for xrOvary-based differentiation were enriched for “brain  
230 development” and included *KHDC3L*, *SMARCA2*, and *DPPA5* (Extended Data Fig. 5c).  
231

232 We identified highly variable genes (HVGs) among hPGCLC-derived cells cultured with  
233 or without BMP2 (Extended Data Fig. 5d and Supplementary Table 3), which were  
234 classified into 7 clusters showing BMP-driven up- or down-regulation (Extended Data  
235 Fig. 5e, f). The BMP up-regulated genes included ER genes and were enriched for  
236 “BMP signalling pathway/negative regulation of transcription from RNA polymerase II  
237 promoter,” and “inactivation of MAPK activity” (Extended Data Fig. 5g, h). The BMP  
238 down-regulated genes were enriched for “positive regulation of transcription from RNA  
239 polymerase II promoter” and “cellular response to FGF stimulus/positive regulation of  
240 MAPK/ERK1/2 cascade” (Extended Data Fig. 5g, h). Accordingly, Western blot  
241 analysis revealed that hPGCLC-derived cells cultured with BMP2 exhibited a reduction  
242 in the phosphorylated MAPK/ERK levels (Extended Data Fig. 5i, j). Thus, BMP

243 signalling promotes up-regulation of ER genes and, directly or indirectly, leads to an  
244 attenuation of the MAPK/ERK pathways.

245

246 Next, we performed scRNA-seq analysis of the female hPGCLC culture (F1-*AGVT*) and  
247 compared the results to the oogonia/fetal meiotic oocyte development <sup>28,29</sup>  
248 (Supplementary Table 2). The analysed cells were classified into 10 clusters and were  
249 annotated as very early mitotic (VEM) 1/2 (1 and 2 represent different cell-cycle states),  
250 early mitotic (EM) 1/2, mitotic (M) 1/2/3, pre-leptotene/leptotene (PLL1/2), and  
251 zygotene/pachytene/diplotene (ZPD) (Fig. 3c and Extended Data Fig. 6a). The VEM  
252 cells formed a distinct cluster and consisted nearly exclusively of AG<sup>+</sup>VT<sup>-</sup> cells at c11.  
253 EM cells consisted primarily of AG<sup>+</sup>VT<sup>-</sup> cells at c56 and of a small number of cells at  
254 7–10 wpf that began to up-regulate oogonia markers. M was the major cell population  
255 and consisted of AG<sup>+</sup>VT<sup>+</sup> cells at c56–117 and oogonia at 7–16 wpf *in vivo*, representing  
256 the self-renewing oogonia. PLL consisted of AG<sup>-</sup>VT<sup>+</sup> cells at c86/117 and cells at 9–16  
257 wpf. ZPD consisted of cells at 13/16 wpf *in vivo* (Fig. 3c–e, Extended Data Fig. 6b, c).  
258 The emergence of cells categorized as PLL during BMP-driven hPGCLC differentiation  
259 agrees with the finding that c107 AG<sup>-</sup>VT<sup>+</sup> cells were closer to RA-responsive cells (Fig.  
260 3b).

261

262 For M, *in vivo* and *in vitro* cells contributed to each subcluster (M1: G1/S; M2: S/G2/M;  
263 M3: G2/M) in an equivalent manner (Fig. 3c, d, Extended Data Fig. 6b, c) and the DEGs  
264 between *in vivo* and *in vitro* cells were small in number, with those up-regulated *in vivo*  
265 enriched for “positive regulation of double-strand break repair” and those up-regulated *in*  
266 *vitro* enriched for “angiogenesis” and “negative regulation of transcription from RNA  
267 polymerase II promoter” (Extended Data Fig. 6d–f, Supplementary Table 4). In contrast,  
268 for PLL, *in vivo* cells contributed only to PLL1 (S/G2/M/meiotic) and although *in vitro*  
269 cells also contributed to PLL1, they were plotted in its periphery adjacent to PLL2 and  
270 were the exclusive source for PLL2 (mostly G1) (Fig. 3c, d, Extended Data Fig. 6b, c).  
271 The PLL1 signature genes included *ZGLP1*, *STRA8*, *REC8*, and *SMC1B*, and exhibited  
272 strong up-regulation in PLL1 *in vivo* cells, but showed only moderate elevation in PLL1  
273 or 2 *in vitro* cells (Extended Data Fig. 6g, h). We identified DEGs between PLL1 *in*  
274 *vivo* and PLL1 or 2 *in vitro* cells and found that those up-regulated in PLL1 *in vivo* cells  
275 were enriched for “meiotic nuclear division,” whereas those up-regulated in PLL1 or 2 *in*  
276 *vitro* cells were enriched for “regulation of cell differentiation,” “cell development,” and  
277 “regulation of signal transduction” (Extended Data Fig. 6d–h, Supplementary Table 4).  
278 Collectively, these results indicate that BMP-driven hPGCLC differentiation  
279 recapitulates the transcriptome dynamics of hPGC differentiation into mitotic pro-  
280 spermatogonia/oogonia and their robust propagation. On the other hand, the continued  
281 culture results in differentiation of aberrant cells as a minor population; these cells show  
282 low expression of genes for meiotic entry and ectopic up-regulation of developmental  
283 regulators.

284

285 **DNA methylome reprogramming**

286 As part of the process of epigenetic reprogramming, hPGCs erase their genome-wide  
287 5mC level to ~5–10% over a period of ~5–7 weeks and differentiate into either mitotic  
288 pro-spermatogonia or oogonia (Fig. 1a) <sup>3,8,9</sup>. Consequently, both mitotic pro-  
289 spermatogonia and oogonia erase their imprints and oogonia reactivate inactive X  
290 chromosomes (X-chromosome reactivation: XCR) <sup>3,8,9</sup>. To examine epigenetic  
291 reprogramming during BMP-driven hPGCLC differentiation, we performed genome-  
292 wide DNA methylation profiling (Supplementary Tables 2, 5). We also performed long-  
293 read whole-genome sequencing of the female lines, reconstructed the X-chromosome  
294 allelic sequences, and determined the active and inactive X chromosomes (Xa and Xi,  
295 respectively) <sup>30</sup> (Supplementary Table 12). Both male and female hiPSCs showed  
296 autosome-wide 5mC levels of ~85%, which, upon BMP-driven hPGCLC differentiation,  
297 decreased progressively, reaching as low as ~10%, a level equivalent to that in mitotic  
298 pro-spermatogonia/oogonia and lower than that in xrOvary-based oogonia-like cells (Fig.  
299 4a, Extended Data Fig. 7a–c, Supplementary Table 5) <sup>3,19</sup>. DNA demethylation during  
300 BMP-driven differentiation occurred more slowly than that *in vivo*, but had similar or  
301 faster dynamics than that during xrOvary-based differentiation (Fig. 4a, Extended Data  
302 Fig. 7 a–c). DNA demethylation occurred on all unique elements (Fig. 4a), including  
303 promoters of the 13 genes up-regulated in gonadal germ cells <sup>26</sup>, genes for “meiotic cell  
304 cycle,” and differentially methylated regions (DMRs) of the imprinted genes, with nearly  
305 all imprints erased (except *PEG3*, *IGF2R*, *ZFAT*), irrespective of aberrant imprints in  
306 parental hiPSCs (Fig. 4b, Extended Data Fig. 8a). The “escapees” that evaded DNA  
307 demethylation consisted primarily of evolutionarily young retrotransposons, highly  
308 overlapping with those in mitotic pro-spermatogonia/oogonia and xrOvary-based  
309 oogonia-like cells (Fig. 4c–e, Extended Data Fig. 8b, Supplementary Table 6), suggesting  
310 that DNA demethylation occurs via a common mechanism during BMP-driven, xrOvary-  
311 based, and *in vivo* PGC(LC) differentiation. Additionally, paired-end sequencing  
312 identified centromeric/telomeric satellites as major escapees (Fig. 4d, e, Extended Data  
313 Fig. 8b–d). In this context, we note that a previously reported hPGCLC induction <sup>31</sup> and  
314 previously reported hPGCLC culture <sup>32</sup> exhibited highly methylated profiles (Extended  
315 Data Fig. 8e).

316

317 Xa exhibited chromosome-wide DNA demethylation in a manner similar to autosomes  
318 (Fig. 4a, Extended Data Fig. 7b, c). In contrast, Xi showed distinct demethylation  
319 profiles: Xi in F1-*AGVT* hiPSCs had 5mC levels of ~84%, which were reduced to ~26%  
320 in AG<sup>+</sup>VT<sup>+</sup> cells at c86, but thereafter did not undergo further reduction, remaining at  
321 ~24% at c127 (Fig. 4a, Extended Data Fig. 7b, c, Supplementary Table 5). Promoters  
322 and non-promoter CpG islands (CGIs) were highly methylated in the iPSCs (~87 and  
323 ~93%, respectively), and a majority of them showed demethylation resistance, retaining  
324 ~44% and ~65% 5mC levels on average, respectively, in AG<sup>+</sup>VT<sup>+</sup> cells at c86/c127 (Fig.

325 4a, Extended Data Fig. 7b, c, Supplementary Table 5). Xi in F2-AGVThiPSCs exhibited  
326 similar demethylation resistance during both BMP-driven and xrOvary-based hPGCLC  
327 differentiation (Fig. 4a, Extended Data Fig. 7b, c, Supplementary Table 5) <sup>30</sup>. Among  
328 the escapees, Alu elements displayed greater demethylation resistance (Extended Data  
329 Fig. 8b). Thus, as in human germ cells, BMP-driven hPGCLC differentiation results in  
330 comprehensive autosome- and Xa-wide DNA demethylation, while Xi of the hiPSC lines  
331 we used were resistant to demethylation.

332

### 333 ER gene regulation and X-chromosome reactivation

334 To investigate the mechanisms associated with ER gene regulation, we first re-defined  
335 ER genes according to whether they satisfy the criteria in both xrOvary-based and BMP-  
336 driven hPGCLC differentiation (Extended Data Fig. 8f-i, 9a, b). Only minimal  
337 differences were observed in promoter-wide 5mC levels (~2%) between c82 DT<sup>-</sup> and  
338 DT<sup>+</sup> cells, and between c72 VT<sup>-</sup> and VT<sup>+</sup> cells (Fig. 4a, Supplementary Table 5). In  
339 contrast, the promoter 5mC level of *DAZL* decreased from ~42% to ~12%, and a ≥4-fold  
340 increase in *DAZL* expression was observed between c82 DT<sup>-</sup> and DT<sup>+</sup> cells (Extended  
341 Data Fig. 9c, d). Similarly, the overall promoter 5mC levels of ER genes were greater  
342 in c82 DT<sup>-</sup> than in DT<sup>+</sup> cells, while the overall ER gene expression was higher in DT<sup>+</sup>  
343 cells (Extended Data Fig. 9d). The relationship between ER-gene promoter-5mC levels  
344 and expression was similar in c72 VT<sup>-</sup> and VT<sup>+</sup> cells, although the correlation was not as  
345 strong (Extended Data Fig. 9c, d). This is presumably because *DDX4* is a late ER gene  
346 and the promoters of most ER genes were substantially demethylated when *DDX4* was  
347 up-regulated. Thus, ER-gene expression may occur, at least in part, in a coordinated  
348 manner in response to 5mC demethylation of the ER-gene promoter, though single-cell  
349 analysis would be necessary to come to any definitive conclusion.

350

351 Next, we examined the extent of XCR. We classified informative X-linked genes into  
352 four classes based on their promoter 5mC levels in hiPSCs: class 1 genes with high  
353 (≥50%) 5mC on both Xa and Xi, class 2 genes with low (< 50%) 5mC on Xa and high  
354 5mC on Xi, a class 3 gene with high 5mC on Xa and low 5mC on Xi (*XIST*), and class 4  
355 genes with low 5mC on both Xa and Xi (Extended Data Fig. 9e, Supplementary Table 7).  
356 In hiPSCs, most class 1/2 genes were expressed exclusively from Xa, while class 4 genes  
357 were biallelic and overlapped with X-chromosome inactivation (XCI) escapees <sup>33</sup>.  
358 While *XIST* (class 3) was most likely expressed from Xi <sup>30</sup>, we could not find informative  
359 single nucleotide polymorphisms (SNPs) that discriminate *XIST* expression from parental  
360 alleles by means of 3-prime RNA-seq analysis <sup>34</sup> (Extended Data Fig. 9e, f). During  
361 BMP-driven hPGCLC differentiation, XCR was relatively limited, with the majority of  
362 class 1 and 2 genes remaining expressed mainly from Xa; however, the genes that erased  
363 their promoter 5mC on Xi became biallelically expressed (Extended Data Fig. 9e, f).  
364 Accordingly, allele-usage analysis based on scRNA-seq data revealed that genes  
365 expressed predominantly from Xa in VEM cells exhibited a lower level of XCR in

366 EM/M/PLL cells, with ~25% of the transcripts derived from Xi (Extended Data Fig. 9g).

367

368 We then analysed X-chromosome dosage compensation. Bulk RNA-seq data revealed  
369 that in both male and female hiPSCs, the X chromosome:autosome (X:A) ratio of gene-  
370 expression levels was ~0.9, indicating up-regulation of Xa (X-chromosome up-  
371 regulation: XCU)<sup>35,36</sup> (Extended Data Fig. 9h). During BMP-driven hPGCLC  
372 differentiation, the X:A ratio in males decreased progressively, reaching around ~0.7 in  
373 mitotic pro-spermatogonia-like cells, whereas that in females also decreased, but  
374 plateaued at around ~0.8 in oogonia-like cells (Extended Data Fig. 9h). These unique  
375 dynamics are reminiscent of those observed during mouse, monkey, and human germ-  
376 cell development<sup>28,29</sup>. Collectively, these results led us to conclude that XCR, which  
377 correlates with promoter demethylation, partially proceeds under the current culture,  
378 while X-chromosome dosage compensation operates in a broadly normal manner.

379

### 380 **BMP signalling and DNMT activities**

381 We next explored the relationship between BMP signalling and DNA methylation  
382 reprogramming. We observed that during BMP-driven hPGCLC specification and  
383 differentiation, hPGCLCs and their progeny down-regulated the expression of *de novo*  
384 DNA methyltransferases (DNMTs) (Fig. 3a, Extended Data Fig. 9i). The genome-wide  
385 CpA methylation, a readout for *de novo* DNMT activity<sup>37</sup>, was reduced in c32 cells  
386 cultured with BMP2 compared to their precursors and counterparts cultured without  
387 BMP2 [Effect sizes (Cohen's d-values) > 0.2], and the reduced level persisted in c82/c122  
388 cells (Extended Data Fig. 7b, 9i). Furthermore, hPGCLCs and their progeny down-  
389 regulated *UHRF1*, an essential co-factor for maintenance DNMT<sup>38,39</sup>, and in response to  
390 BMP2, UHRF1 appeared to partially translocate from the nucleus to the cytoplasm (Fig.  
391 3a, Extended Data Fig. 9i, k). Compared to c32 cells without BMP2 (~63%), those with  
392 BMP2 had substantially lower autosome-wide 5mC levels (~45%) (Fig. 4a).  
393 Remarkably, while hPGCLC-derived cells cultured without BMP2 were estimated to  
394 reduce their 5mC levels by only ~1.0% per cell cycle by c32, those cultured with BMP2  
395 did so by ~4.9% (Extended Data Fig. 9l). hPGCLC-derived cells from other iPSC lines  
396 cultured with BMP2 reduced their 5mC at a similar rate, with those from F2-AGVT  
397 exhibiting a ~7.6% reduction in 5mC levels (Extended Data Fig. 9l).

398

399 We analysed the characteristics of DNA demethylation dynamics across genomic bins  
400 (10-kb bins) bearing different 5mC levels ([0-20], (20-40], (40-60], (60-80], (80-100] %  
401 of 5mC levels). The DNA demethylation ratios during the differentiation from hiPSCs  
402 to hPGCLCs and from hPGCLCs to c32 cells without BMP2 [% mC original cells – %  
403 mC cultured cells / % mC original cells] were relatively small and similar among the  
404 genomic bins bearing different 5mC levels in the original cell types (Extended Data Fig.  
405 9m). In contrast, the DNA demethylation ratios during the differentiation from  
406 hPGCLCs to c32 cells with BMP2 and subsequent differentiation were larger and varied

407 across genomic bins, with the bins bearing higher and lower 5mC levels in hPGCLCs  
408 showing lower and higher demethylation ratios, respectively (Extended Data Fig. 9m).  
409 The genomic bins showing DNA demethylation resistance corresponded to DNA  
410 demethylation escapees (Extended Data Fig. 9n). Taken together, these findings suggest  
411 that BMP-driven hPGCLC specification and differentiation are accompanied by an  
412 attenuation of both *de novo* and maintenance DNMT activities, promoting a replication-  
413 coupled passive genome-wide DNA demethylation, which occurs in a heterogeneous  
414 manner depending on the properties of the genomic regions.

415

#### 416 ***TET1*-deficient hPGCLCs aberrantly differentiate into extraembryonic cells**

417 The TET proteins, TET1/2/3, are active DNA demethylases and transcriptional repressors  
418 for developmental genes<sup>2,40,41</sup>. *TET1* has been shown to be abundantly expressed  
419 during hPGC(LC) differentiation<sup>3,17,19</sup> (Fig. 3a). To explore the role of *TET1*, we  
420 generated *TET1* knockout (KO) hiPSCs (M1-BTAG *TET1*<sup>-/-</sup> KO#1 and KO#2) (Extended  
421 Data Fig. 10a–e). Both *TET1* KO hiPSC lines were induced into BT<sup>+</sup>AG<sup>+</sup> cells in an  
422 apparently normal manner (Extended Data Fig. 10f). Upon culture with BMP2, they  
423 showed a robust expansion, but after c12, their expansion was attenuated with a  
424 substantial reduction of the enrichment score, indicating an aberrant differentiation  
425 (Extended Data Fig. 10g). On the other hand, *TET1* KO BT<sup>+</sup>AG<sup>+</sup> cells could be cultured  
426 at least until c42 (Extended Data Fig. 10g).

427

428 RNA-seq analysis revealed that *TET1* KO cells underwent transcriptome changes similar  
429 to those of wild-type cells until c12, but subsequently, they did not show proper  
430 maturation, retaining immature properties and displaying a diverted trajectory from wild-  
431 type cells (Fig. 5a, Extended Data Fig. 10h, Supplementary Table 2). No compensatory  
432 up-regulation of *TET2* and *TET3* in *TET1* KO cells was observed (Extended Data Fig.  
433 10h). The number of DEGs increased with the progression of *TET1* KO-cell culture,  
434 with higher numbers of genes up-regulated in *TET1* KO cells (Extended Data Fig. 10i).  
435 The DEGs were classified into distinct clusters, with clusters 1–8 and 9–13 representing  
436 up- and down-regulated genes in *TET1* KO cells, respectively, with the former including  
437 genes for “positive regulation of transcription from RNA polymerase II promoter” (e.g.,  
438 *HAND1/2*, *CEBPA/D*, *DLX1*, *GATA6*, *TBX7/3*, and *MEIS3*), “BMP signalling pathway”  
439 (e.g., *LEFTY1/2*, *NODAL*, and *BMP4/7*), and “protein phosphorylation” (e.g.,  
440 *MAP3K5/14*), and the latter including genes for “fertilization/male meiosis” (e.g.,  
441 *SPATA22*, *TDRD12*, *MEIOB*, and *BRDT*) (Extended Data Fig. 10i–k, Supplementary  
442 Table 8). Indeed, ER genes also failed to show proper up-regulation in *TET1* KO cells  
443 (Extended Data Fig. 10l).

444

445 To determine the overall fates of the *TET1* KO cells, we performed scRNA-seq analysis  
446 of the whole culture at c18, when the *TET1* KO cells showed substantial degrees of  
447 aberrant differentiation (Extended Data Fig. 10g). Analysed cells were classified into

448 two major groups, with group 1 cells representing propagating hPGCLCs (clusters 1–3:  
449 PGC markers<sup>+</sup>; predominantly wild-type cells) and group 2 cells representing aberrantly  
450 differentiating cells (clusters 4–11: PGC markers<sup>low/−</sup>; predominantly *TET1* KO cells)  
451 (Fig. 5b, c, Extended Data Fig. 11a, b). Correlation analysis using HVGs revealed that  
452 clusters 3, 8, and 7 had a relatively high correlation between group 1 and 2 cells (Extended  
453 Data Fig. 11c). Accordingly, partition-based graph abstraction (PAGA) analysis<sup>42</sup>  
454 indicated a trajectory from cluster 3 to cluster 8 between group 1 and 2 cells (Extended  
455 Data Fig. 11d). Cluster 3 cells ectopically up-regulated transcription factors (TFs) such  
456 as *GATA3* and *HAN1*, with a subset of them repressing PGC/pluripotency genes, while  
457 cluster 8 and 7 cells more fully repressed PGC/pluripotency genes, with cluster 7 cells  
458 strongly up-regulating amnion (AM) and trophectoderm (TE) markers<sup>43,44</sup> (Fig. 5b, c).  
459 Cells of clusters 4–6 continued to express AM/TE markers, while cluster 9 cells acquired  
460 an endothelium-like profile, and cluster 10 and 11 cells were of unclear identity (Fig. 5b,  
461 c). We performed an integrated analysis of our datasets against those of PSC-based  
462 human development models and a human gastrula *in vivo*<sup>45,46</sup>, which revealed that cells  
463 of clusters 7/8 and clusters 4–6 were co-segregated with/predicted to be AM/TE-like and  
464 (extra-embryonic) mesoderm-like cells, respectively (Fig. 5d, e, Extended Data Fig.  
465 11e–h). Thus, during hPGCLC specification/differentiation, *TET1* plays a key role in  
466 repressing a diverse set of genes, including critical TFs for AM/TE development, thereby  
467 safeguarding hPGCLCs from aberrant differentiation into an extraembryonic cell fate.  
468

#### 469 ***TET1* deficiency leads to de-repression of bivalent genes and hypermethylation of 470 regulatory elements**

471 To understand how *TET1* regulates BMP-driven hPGCLC differentiation, we first  
472 evaluated the impact of *TET1* KO on the DNA methylome. At c12, before which wild-  
473 type and *TET1* KO cells exhibited an equivalent expansion and relatively similar  
474 transcriptomes (Fig. 5a, Extended Data Fig. 10f, g), *TET1* KO cells showed higher 5mC  
475 levels in all genomic elements, with genome-wide 5mC levels ~4% higher than those of  
476 wild-type cells (Fig. 6a, b, Supplementary Table 5), and DMRs were enriched in  
477 promoters, CGIs, CTCF binding sites, and coding sequences (CDSs) (Fig. 6c). We  
478 classified all open chromatin [promoters and enhancers (non-promoter open sites)] during  
479 hPGCLC induction into three major categories (active, bivalent, and poised) based on  
480 their combinatorial epigenetic states<sup>47</sup> (Fig. 6d), and examined the 5mC-level differences  
481 of each category as well as of silent regions (i.e., un-open promoters and enhancers that  
482 were not categorized into active, bivalent, or poised) between wild-type and *TET1* KO  
483 cells at c12. In *TET1* KO cells, all promoter/enhancer classes showed elevated 5mC  
484 levels, with poised promoters and bivalent/poised enhancers exhibiting especially high  
485 5mC elevations (Fig. 6e). We found that ER genes were enriched in poised promoters  
486 and that they (particularly early ER genes) exhibited highly elevated 5mC promoter levels,  
487 while imprint DMRs showed 5mC-level elevation slightly higher than the genome-wide  
488 average (Fig. 6f).

489  
490 Regarding expression, as a general trend, genes with bivalent promoters/enhancers  
491 showed significant up-regulation in *TET1* KO cells (Fig. 6g). Accordingly, among the  
492 DEGs at c12 (fold-change  $\geq 2$ ), genes with bivalent promoters/enhancers, including  
493 *SOX11*, *MSX1*, *HAND1*, *CDX2*, and *ARNT2*, were highly over-represented among up-  
494 regulated genes, while ER genes, as described earlier, displayed a trend of down-  
495 regulation and were highly over-represented in down-regulated DEGs (Fig. 6h, Extended  
496 Data Fig. 10l). The hyper-methylation of bivalent promoters associated with significant  
497 up-regulation of transcription might have been due to impaired recruitment of Polycomb  
498 repressive complexes (PRCs), as previously shown in mouse embryonic stem cells  
499 (mESCs)<sup>48,49</sup>. Among the DEGs at c12, the up-regulated genes with bivalent (the  
500 largest set) and poised promoters were highly enriched with the targets of TET1 in hESCs  
501 (Fig. 6i)<sup>50</sup>, suggesting that as demonstrated in hESCs and neuronal precursors<sup>50</sup>, key  
502 targets of TET1 may be shared in hPGCLCs and hESCs.  
503  
504 We next examined the impact of the DNA demethylase function of TET1 on intergenic  
505 regions from which all open sites during hPGCLC induction were subtracted, i.e., the  
506 “background” genome (~55% of the genome). At c12, the 5mC levels of the  
507 background genome were similar between *TET1* KO and wild-type cells (wild-type:  
508 ~74%; *TET1* KO: ~76%), whereas those of open sites were substantially different  
509 (intergenic: wild-type, ~66%; *TET1* KO, ~73%; genic: wild-type, ~67%; *TET1* KO,  
510 ~77%) (Fig. 6j, Supplementary Table 5), suggesting that during hPGCLC  
511 specification/differentiation, the impact of TET1 on the background genome might be  
512 relatively minor.  
513  
514 By c42, *TET1* KO cells showed impaired expansion (wild-type:  $3.7 \times 10^3$ ; *TET1* KO#1:  
515  $1.6 \times 10^3$ ; *TET1* KO#2:  $4.5 \times 10^2$ ) (Extended Data Fig. 10g) and their transcriptomes were  
516 both immature and aberrant (Fig. 5a, Extended Data Fig. 10h–l), precluding direct  
517 evaluation of the primary function of TET1. Nevertheless, *TET1* KO cells failed to  
518 properly promote DNA demethylation in all genomic elements, and the extent of  
519 demethylation in wild-type and *TET1* KO cells appeared to reflect, at least in part, their  
520 expansion rates (Fig. 6a). Furthermore, despite substantial genome-wide 5mC-level  
521 differences, DMRs between wild-type and *TET1* KO cells were enriched in promoters,  
522 CGIs, and CDSs (Extended Data Fig. 12a). Interestingly, the 5mC-level differences in  
523 promoters/enhancers of bivalent genes were relatively small, while those of poised genes  
524 were large, and notably, the 5mC-level differences of ER-gene promoters were the largest,  
525 while those of imprint DMRs were less than the genome-wide average (Extended Data  
526 Fig. 12b, c).  
527  
528 Regarding expression, genes with bivalent promoters/enhancers showed an overall trend  
529 of significant up-regulation (Extended Data Fig. 12d). Accordingly, among the DEGs

530 at c42, such genes, which included *LEFTY2*, *MEIS1*, *TCF7L1*, *NKX2-5*, and *TBXT*, were  
531 highly over-represented in up-regulated genes (Extended Data Fig. 12e). As described  
532 earlier, the vast majority of ER genes were down-regulated, with a significant enrichment  
533 in down-regulated DEGs (Extended Data Fig. 10l, 12e). On the other hand, interestingly,  
534 genes with poised promoters/enancers, while exhibiting failure of demethylation, did  
535 not show a general trend of down-regulation (Extended Data Fig. 10l, 12e). Among the  
536 DEGs at c42, the up-regulated genes with bivalent promoters were large in number and  
537 highly enriched with the targets of TET1 in hESCs (Extended Data Fig. 12f)<sup>50</sup>.  
538 Collectively, these findings indicate that DNA methylation reprogramming in human  
539 germ cells results, at least in part, from a replication-coupled, passive mechanism. In  
540 this process, TET1 functions as a transcriptional repressor for bivalent genes, including  
541 key TFs for the extraembryonic cell fates. Furthermore, it appears to contribute  
542 primarily to the demethylation of the regulatory elements, especially the poised and ER-  
543 gene promoters, and such demethylation is correlated with the up-regulation of ER genes,  
544 but not genes with poised promoters in general.

545

## 546 Discussion

547 We have shown that hPGCLCs cultured with BMP2 propagate stably, displaying reduced  
548 levels of MAPK/ERK signalling and both *de novo* and maintenance DNMT  
549 activities/machineries (Extended Data Fig. 5g–j, 9i–k), and thereby likely promoting  
550 replication-coupled, passive genome-wide DNA demethylation, which leads to mitotic  
551 pro-spermatogonia/oogonia-like cell differentiation. These findings create a paradigm  
552 for hallmark events in human germ-cell development (Extended Data Fig. 12g). The  
553 finding that BMP signalling stabilizes the germ-cell fate is reminiscent of a known role  
554 of BMP signalling to sustain the self-renewal of mESCs via a blockade of their  
555 differentiation<sup>51</sup>. Given that repression of MAPK/ERK signalling plays a key role in  
556 the induction of naïve pluripotency that accompanies repression of DNMT activities and  
557 DNA methylome reprogramming<sup>52–54</sup>, a similar mechanism might operate during BMP-  
558 driven hPGCLC differentiation. The precise mechanism of action of BMP signalling  
559 and whether it is involved in epigenetic reprogramming of PGCs in other species remains  
560 unclear and warrants further investigation. A recent report induced DAZL<sup>+</sup> PGCLCs  
561 using retinoic acid, Activin A, and overexpression of *SOX17* and *DMRT1*<sup>27</sup>. Although  
562 a number of genes associated with mitotic pro-spermatogonia and oogonia were up-  
563 regulated, these cells did not express *DDX4* and remained highly methylated (the  
564 genome-wide 5mC levels were ~76–79%)<sup>27</sup>. This suggests that they are a distinct entity  
565 from the mitotic pro-spermatogonia/oogonia-like cells generated in this study.

566

567 Recent studies have shown that *TET1*-deficient hPSCs can differentiate into hPGCLCs in  
568 an apparently normal manner<sup>55,56</sup> and when cultured on mouse embryonic feeders, show  
569 a reduced propensity for hPGCLC differentiation<sup>56</sup>. Here, we have shown that *TET1*-

570 deficient hiPSCs differentiate into BT<sup>+</sup>AG<sup>+</sup> cells with a number of DEGs, and intriguingly,  
571 TET1-deficient BT<sup>+</sup>AG<sup>+</sup> cells differentiated aberrantly into extraembryonic cells,  
572 including AM/TE-like cells, with de-repression of relevant genes, including critical TFs  
573 with bivalent promoters (Fig. 5, 6). The role of TET proteins in the repression of  
574 bivalent genes in cooperation with PRCs is consistent with previous findings in other  
575 contexts<sup>48,50,57,58</sup>. We have also shown that TET1-deficient hPGCLC-derived cells fail  
576 to up-regulate ER genes, with their promoters remaining methylated (Fig. 6). The  
577 consensus view of the role of *Tet1* in mice is that *Tet1* is dispensable for genome-wide  
578 DNA demethylation *per se*, but is critical in maintaining demethylation of key genes for  
579 spermatogenesis and oogenesis and of imprint DMRs<sup>59-61</sup>. In order to make a more  
580 precise comparison of the function of TET1 in epigenetic reprogramming and germ-cell  
581 development between humans and mice, it would be necessary to evaluate the impact of  
582 specifically removing the enzymatic function of TET proteins, while leaving their  
583 transcriptional repressor activity intact.

584

585 Mouse germ cells reduce their 5mCs from ~75% to ~5% over a week with an expansion  
586 of ~2<sup>9</sup>-fold (from ~40 to ~25,000 cells)<sup>62-64</sup>, diminishing their 5mCs by ~8% per cell  
587 cycle on average. In contrast, human germ cells reduce their 5mCs from ~80% to ~5%  
588 over ~5 weeks with an expansion of ~2<sup>10</sup>-fold (from ~40 to ~40,000 cells)<sup>3,8,9,65</sup>,  
589 decreasing their 5mCs by ~7.5% per cell cycle on average. Thus, mouse and human  
590 germ cells reduce 5mCs at a similar rate per cell cycle, and accordingly, the difference in  
591 time scale for genome-wide DNA demethylation between mice and humans might simply  
592 reflect the difference in the doubling time between mouse (~0.8 days on average) and  
593 human germ cells (~3.5 days on average). This would further support the notion of a  
594 replication-dependent, passive mechanism, which is non-species specific, as the primary  
595 mechanism underlying genome-wide DNA demethylation. Unlike mPGCLCs, which  
596 reduce their 5mCs by ~10% per cell cycle on average<sup>66</sup>, hPGCLC-derived cells from 3  
597 out of 4 hiPSC lines reduce their 5mCs at a lower rate (~3–5%) per cell cycle (Extended  
598 Data Fig. 9l), indicating that the hPGCLC differentiation condition would necessitate  
599 additional optimization, including the provision of additional factors. Indeed, *UHRF1*  
600 remained at a slightly higher level during BMP-driven hPGCLC differentiation than  
601 during *in vivo* hPGC differentiation (Fig. 3a), and the former requires substantial  
602 propagation for genome-wide demethylation. Given that BMP-driven hPGCLC  
603 differentiation displays an attenuation of the MAPK/ERK signalling (Extended Data Fig.  
604 5g–j), a fine tuning of the balance between BMP and MAPK/ERK signalling could be a  
605 key for such optimization. On the other hand, hPGCLC-derived cells from F2-AGVT  
606 reduce their 5mCs by ~7.6% per cell cycle, but with a prolonged doubling time (~8.8  
607 days) (Fig. 2b, Extended Data Fig. 9l), which might be due to a variance of signalling  
608 efficiency associated with this line.

609

610 During BMP-driven hPGCLC differentiation, despite the comprehensive autosome- and

611 Xa-wide DNA methylome reprogramming, Xi—especially Xi on promoters and non-  
612 promoter CGIs—exhibited a resistance to reprogramming with limited XCR (Fig. 4a,  
613 Extended Data Fig. 7b, c, 9e–g). This was also the case for hPGCLC-derived cells in  
614 xrOvaries<sup>19,30</sup> and could result from an aberrant Xi epigenetic state, including Xi-wide  
615 H3K9me3-based repression, in female hiPSCs<sup>67</sup>. This idea was supported by a recent  
616 finding in *in vitro* oogenesis in cynomolgus monkeys (cy), in which cyPGCLCs induced  
617 from cyESCs bearing stable Xi with H3K27me3-based repression underwent more  
618 comprehensive Xi reprogramming and XCR in xrOvaries, and differentiated up to fetal  
619 oocyte-like cells at the zygotene stage of the meiotic prophase<sup>30</sup>. In the future, it will  
620 be important to perform more comprehensive characterizations of the epigenetic states,  
621 including the histone-modification states, for the BMP-driven hPGCLC differentiation  
622 and explore their significance in the properties and functions, including Xi  
623 reprogramming and differentiation competence, of hPGCLC-derived cells.  
624 Interestingly, a recent study has shown that hPGCLCs induced from epigenetically “reset”  
625 hPSCs may have an accelerated differentiation potential<sup>20</sup>; this finding warrants further  
626 investigation, including using the BMP-driven hPGCLC differentiation system.  
627

628 We noted that in the late phases of culture, a fraction (~5%) of hPGCLC-derived cells  
629 differentiate into the AG<sup>-</sup>DT<sup>+</sup>/VT<sup>+</sup> state, and the AG<sup>-</sup>VT<sup>+</sup> cells show gene expression  
630 suggestive of their potential entry into the meiotic prophase (Fig. 2a, c, 3b). On the other  
631 hand, these cells failed to acquire sufficient levels of genes for meiosis and aberrantly up-  
632 regulated genes involved in ectopic developmental/signalling pathways (Extended Data  
633 Fig. 6d–h), indicating that while BMP-driven hPGCLC differentiation recapitulates  
634 hPGC differentiation into mitotic pro-spermatogonia/oogonia and their propagation, the  
635 continued culture also results in differentiation of aberrant cells with low expression of  
636 genes for meiotic entry and ectopic up-regulation of developmental regulators. BMP  
637 signalling and its downstream effector, ZGLP1, robustly induce competent mPGCLCs  
638 into the oogenic pathway in mice<sup>68,69</sup>, and the mechanism for the oogenic differentiation  
639 may be conserved in humans. On the other hand, compared to mice, humans and  
640 monkeys exhibit a distinct transcriptomic transformation upon oogonia-to-oocyte  
641 transition<sup>13,29</sup> and may employ a distinct mechanism for the oogenic differentiation.  
642 BMP-driven hPGCLC differentiation, which does not require reaggregate formation with  
643 ovarian/other somatic cells and generates abundant numbers of oogonia with appropriate  
644 properties (Fig. 2), should serve as a robust system to identify conditions for the proper  
645 induction of human fetal oocytes and to investigate the mechanism of human fetal oocyte  
646 differentiation. In a broader context, with its capacity to provide abundant numbers  
647 (>10<sup>10</sup> cells) of both mitotic pro-spermatogonia and oogonia, the BMP-driven hPGCLC  
648 differentiation system will provide a promising foundation for future directions of human  
649 IVG research.

650

651

652      **References**

- 653      1 Tang, W. W., Kobayashi, T., Irie, N., Dietmann, S. & Surani, M. A. Specification  
654      and epigenetic programming of the human germ line. *Nat Rev Genet* **17**, 585-600,  
655      doi:10.1038/nrg.2016.88 (2016).
- 656      2 Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine  
657      in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935, doi:1170116  
658      [pii] 10.1126/science.1170116 (2009).
- 659      3 Tang, W. W. *et al.* A Unique Gene Regulatory Network Resets the Human  
660      Germline Epigenome for Development. *Cell* **161**, 1453-1467,  
661      doi:10.1016/j.cell.2015.04.053 (2015).
- 662      4 Hertig, A. T. *et al.* A thirteen-day human ovum studied histochemically. *Am J  
663      Obstet Gynecol* **76**, 1025-1040; discussion 1040-1023, doi:10.1016/0002-  
664      9378(58)90185-6 (1958).
- 665      5 Sasaki, K. *et al.* The Germ Cell Fate of Cynomolgus Monkeys Is Specified in the  
666      Nascent Amnion. *Dev Cell* **39**, 169-185, doi:10.1016/j.devcel.2016.09.007 (2016).
- 667      6 Chen, D. *et al.* Human Primordial Germ Cells Are Specified from Lineage-Primed  
668      Progenitors. *Cell reports* **29**, 4568-4582 e4565, doi:10.1016/j.celrep.2019.11.083  
669      (2019).
- 670      7 Witschi, E. Migration of germ cells of human embryos from the yolk sac to the  
671      primitive gonadal folds. *Contrib. Embryol. Carnegie Inst.* **209**, 67-80 (1948).
- 672      8 Gkountela, S. *et al.* DNA Demethylation Dynamics in the Human Prenatal  
673      Germline. *Cell* **161**, 1425-1436, doi:10.1016/j.cell.2015.05.012 (2015).
- 674      9 Guo, F. *et al.* The Transcriptome and DNA Methylome Landscapes of Human  
675      Primordial Germ Cells. *Cell* **161**, 1437-1452, doi:10.1016/j.cell.2015.05.015  
676      (2015).
- 677      10 Gruhn, W. H. *et al.* Epigenetic resetting in the human germ line entails histone  
678      modification remodeling. *Sci Adv* **9**, eade1257, doi:10.1126/sciadv.ade1257  
679      (2023).
- 680      11 Baker, T. G. A Quantitative and Cytological Study of Germ Cells in Human  
681      Ovaries. *Proc R Soc Lond B Biol Sci* **158**, 417-433 (1963).
- 682      12 Fukuda, T., Hedinger, C. & Groscurth, P. Ultrastructure of developing germ cells  
683      in the fetal human testis. *Cell Tissue Res* **161**, 55-70 (1975).
- 684      13 Li, L. *et al.* Single-Cell RNA-Seq Analysis Maps Development of Human  
685      Germline Cells and Gonadal Niche Interactions. *Cell Stem Cell* **20**, 858-873 e854,  
686      doi:10.1016/j.stem.2017.03.007 (2017).
- 687      14 Saitou, M. & Hayashi, K. Mammalian in vitro gametogenesis. *Science* **374**,  
688      eaaz6830, doi:10.1126/science.aaz6830 (2021).
- 689      15 Irie, N. *et al.* SOX17 is a critical specifier of human primordial germ cell fate.  
690      *Cell* **160**, 253-268, doi:10.1016/j.cell.2014.12.013 (2015).
- 691      16 Sasaki, K. *et al.* Robust In Vitro Induction of Human Germ Cell Fate from  
692      Pluripotent Stem Cells. *Cell Stem Cell* **17**, 178-194,

- 693       doi:10.1016/j.stem.2015.06.014 (2015).
- 694     17    Murase, Y. *et al.* Long-term expansion with germline potential of human  
695           primordial germ cell-like cells in vitro. *EMBO J* **39**, e104929,  
696           doi:10.15252/embj.2020104929 (2020).
- 697     18    Hwang, Y. S. *et al.* Reconstitution of pro-spermatogonial specification in vitro  
698           from human induced pluripotent stem cells. *Nat Commun* **11**, 5656,  
699           doi:10.1038/s41467-020-19350-3 (2020).
- 700     19    Yamashiro, C. *et al.* Generation of human oogonia from induced pluripotent stem  
701           cells in vitro. *Science* **362**, 356-360, doi:10.1126/science.aat1674 (2018).
- 702     20    Alves-Lopes, J. P. *et al.* Specification of human germ cell fate with enhanced  
703           progression capability supported by hindgut organoids. *Cell reports* **42**, 111907,  
704           doi:10.1016/j.celrep.2022.111907 (2023).
- 705     21    Williams, B. P. *et al.* Biochemical and genetic analysis of the OKa blood group  
706           antigen. *Immunogenetics* **27**, 322-329, doi:10.1007/BF00395127 (1988).
- 707     22    Lu, J. *et al.* Structure-activity relationship studies of small-molecule inhibitors of  
708           Wnt response. *Bioorg Med Chem Lett* **19**, 3825-3827,  
709           doi:10.1016/j.bmcl.2009.04.040 (2009).
- 710     23    Mole, M. A. *et al.* A single cell characterisation of human embryogenesis  
711           identifies pluripotency transitions and putative anterior hypoblast centre. *Nat  
712           Commun* **12**, 3679, doi:10.1038/s41467-021-23758-w (2021).
- 713     24    Tyser, R. C. V. *et al.* Single-cell transcriptomic characterization of a gastrulating  
714           human embryo. *Nature* **600**, 285-289, doi:10.1038/s41586-021-04158-y (2021).
- 715     25    Elmentait, R. *et al.* Single-Cell Sequencing of Developing Human Gut Reveals  
716           Transcriptional Links to Childhood Crohn's Disease. *Dev Cell* **55**, 771-783 e775,  
717           doi:10.1016/j.devcel.2020.11.010 (2020).
- 718     26    Nicholls, P. K. *et al.* Mammalian germ cells are determined after PGC  
719           colonization of the nascent gonad. *Proc Natl Acad Sci U S A* **116**, 25677-25687,  
720           doi:10.1073/pnas.1910733116 (2019).
- 721     27    Irie, N. *et al.* DMRT1 regulates human germline commitment. *Nat Cell Biol* **25**,  
722           1439-1452, doi:10.1038/s41556-023-01224-7 (2023).
- 723     28    Chitiashvili, T. *et al.* Female human primordial germ cells display X-chromosome  
724           dosage compensation despite the absence of X-inactivation. *Nat Cell Biol* **22**,  
725           1436-1446, doi:10.1038/s41556-020-00607-4 (2020).
- 726     29    Mizuta, K. *et al.* Ex vivo reconstitution of fetal oocyte development in humans  
727           and cynomolgus monkeys. *EMBO J* **41**, e110815,  
728           doi:10.15252/embj.2022110815 (2022).
- 729     30    Gyobu-Motani, S. *et al.* Induction of fetal meiotic oocytes from embryonic stem  
730           cells in cynomolgus monkeys. *EMBO J*, e112962,  
731           doi:10.15252/embj.2022112962 (2023).
- 732     31    von Meyenn, F. *et al.* Comparative Principles of DNA Methylation  
733           Reprogramming during Human and Mouse In Vitro Primordial Germ Cell

- 734 32 Specification. *Dev Cell* **39**, 104-115, doi:10.1016/j.devcel.2016.09.015 (2016).
- 735 32 Kobayashi, M. *et al.* Expanding homogeneous culture of human primordial germ  
736 cell-like cells maintaining germline features without serum or feeder layers. *Stem  
737 cell reports* **17**, 507-521, doi:10.1016/j.stemcr.2022.01.012 (2022).
- 738 33 Balaton, B. P., Cotton, A. M. & Brown, C. J. Derivation of consensus inactivation  
739 status for X-linked genes from genome-wide studies. *Biol Sex Differ* **6**, 35,  
740 doi:10.1186/s13293-015-0053-7 (2015).
- 741 34 Nakamura, T. *et al.* SC3-seq: a method for highly parallel and quantitative  
742 measurement of single-cell gene expression. *Nucleic Acids Res* **43**, e60,  
743 doi:10.1093/nar/gkv134 (2015).
- 744 35 Ohno, S. *Sex Chromosomes and Sex-Linked Genes.*, (Springer-Verlag, 1967).
- 745 36 Deng, X., Berletch, J. B., Nguyen, D. K. & Disteche, C. M. X chromosome  
746 regulation: diverse patterns in development, tissues and disease. *Nat Rev Genet*  
747 **15**, 367-378, doi:10.1038/nrg3687 (2014).
- 748 37 Ramsahoye, B. H. *et al.* Non-CpG methylation is prevalent in embryonic stem  
749 cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci U S  
750 A* **97**, 5237-5242, doi:97/10/5237 [pii] (2000).
- 751 38 Bostick, M. *et al.* UHRF1 plays a role in maintaining DNA methylation in  
752 mammalian cells. *Science* **317**, 1760-1764 (2007).
- 753 39 Sharif, J. *et al.* The SRA protein Np95 mediates epigenetic inheritance by  
754 recruiting Dnmt1 to methylated DNA. *Nature* **450**, 908-912 (2007).
- 755 40 Rasmussen, K. D. & Helin, K. Role of TET enzymes in DNA methylation,  
756 development, and cancer. *Genes Dev* **30**, 733-750, doi:10.1101/gad.276568.115  
757 (2016).
- 758 41 Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism,  
759 function and beyond. *Nat Rev Genet* **18**, 517-534, doi:10.1038/nrg.2017.33 (2017).
- 760 42 Wolf, F. A. *et al.* PAGA: graph abstraction reconciles clustering with trajectory  
761 inference through a topology preserving map of single cells. *Genome Biol* **20**, 59,  
762 doi:10.1186/s13059-019-1663-x (2019).
- 763 43 Lee, C. Q. *et al.* What Is Trophoblast? A Combination of Criteria Define Human  
764 First-Trimester Trophoblast. *Stem cell reports* **6**, 257-272,  
765 doi:10.1016/j.stemcr.2016.01.006 (2016).
- 766 44 Io, S. *et al.* Capturing human trophoblast development with naive pluripotent stem  
767 cells in vitro. *Cell Stem Cell* **28**, 1023-1039 e1013,  
768 doi:10.1016/j.stem.2021.03.013 (2021).
- 769 45 Zheng, Y. *et al.* Controlled modelling of human epiblast and amnion development  
770 using stem cells. *Nature* **573**, 421-425, doi:10.1038/s41586-019-1535-2 (2019).
- 771 46 Zhao, C. *et al.* Reprogrammed blastoids contain amnion-like cells but not  
772 trophectoderm. *bioRxiv*, doi:10.1101/2021.05.07.442980 (2021).
- 773 47 Tang, W. W. C. *et al.* Sequential enhancer state remodelling defines human  
774 germline competence and specification. *Nat Cell Biol* **24**, 448-460,

- 775 48 doi:10.1038/s41556-022-00878-z (2022).
- 776 48 Wu, H. *et al.* Dual functions of Tet1 in transcriptional regulation in mouse  
777 embryonic stem cells. *Nature* **473**, 389-393, doi:nature09934 [pii]  
778 10.1038/nature09934 (2011).
- 779 49 Lu, F., Liu, Y., Jiang, L., Yamaguchi, S. & Zhang, Y. Role of Tet proteins in  
780 enhancer activity and telomere elongation. *Genes Dev* **28**, 2103-2119,  
781 doi:10.1101/gad.248005.114 (2014).
- 782 50 Verma, N. *et al.* TET proteins safeguard bivalent promoters from de novo  
783 methylation in human embryonic stem cells. *Nat Genet* **50**, 83-95,  
784 doi:10.1038/s41588-017-0002-y (2018).
- 785 51 Ying, Q. L., Nichols, J., Chambers, I. & Smith, A. BMP induction of Id proteins  
786 suppresses differentiation and sustains embryonic stem cell self-renewal in  
787 collaboration with STAT3. *Cell* **115**, 281-292, doi:10.1016/s0092-  
788 8674(03)00847-x (2003).
- 789 52 Habibi, E. *et al.* Whole-genome bisulfite sequencing of two distinct  
790 interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell*  
791 **13**, 360-369, doi:10.1016/j.stem.2013.06.002 (2013).
- 792 53 Takashima, Y. *et al.* Resetting transcription factor control circuitry toward ground-  
793 state pluripotency in human. *Cell* **158**, 1254-1269, doi:10.1016/j.cell.2014.08.029  
794 (2014).
- 795 54 Theunissen, T. W. *et al.* Systematic identification of culture conditions for  
796 induction and maintenance of naive human pluripotency. *Cell Stem Cell* **15**, 471-  
797 487, doi:10.1016/j.stem.2014.07.002 (2014).
- 798 55 Li, Z. *et al.* The balance between NANOG and SOX17 mediated by TET proteins  
799 regulates specification of human primordial germ cell fate. *Cell Biosci* **12**, 181,  
800 doi:10.1186/s13578-022-00917-0 (2022).
- 801 56 Hsu, F. M. *et al.* TET1 facilitates specification of early human lineages including  
802 germ cells. *iScience* **26**, 107191, doi:10.1016/j.isci.2023.107191 (2023).
- 803 57 Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA  
804 methylation fidelity. *Nature* **473**, 343-348, doi:nature10066 [pii]  
805 10.1038/nature10066 (2011).
- 806 58 Charlton, J. *et al.* TETs compete with DNMT3 activity in pluripotent cells at  
807 thousands of methylated somatic enhancers. *Nat Genet* **52**, 819-827,  
808 doi:10.1038/s41588-020-0639-9 (2020).
- 809 59 Yamaguchi, S. *et al.* Tet1 controls meiosis by regulating meiotic gene expression.  
810 *Nature* **492**, 443-447, doi:10.1038/nature11709 (2012).
- 811 60 Yamaguchi, S., Shen, L., Liu, Y., Sendler, D. & Zhang, Y. Role of Tet1 in erasure  
812 of genomic imprinting. *Nature* **504**, 460-464, doi:10.1038/nature12805 (2013).
- 813 61 Hill, P. W. S. *et al.* Epigenetic reprogramming enables the transition from  
814 primordial germ cell to gonocyte. *Nature* **555**, 392-396, doi:10.1038/nature25964  
815 (2018).

- 816 62 Tam, P. P. & Snow, M. H. Proliferation and migration of primordial germ cells  
817 during compensatory growth in mouse embryos. *J Embryol Exp Morphol* **64**, 133-  
818 147 (1981).
- 819 63 Seisenberger, S. *et al.* The dynamics of genome-wide DNA methylation  
820 reprogramming in mouse primordial germ cells. *Mol Cell* **48**, 849-862,  
821 doi:10.1016/j.molcel.2012.11.001 (2012).
- 822 64 Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M. & Saitou, M. Replication-  
823 coupled passive DNA demethylation for the erasure of genome imprints in mice.  
824 *EMBO J* **32**, 340-353, doi:10.1038/emboj.2012.331 (2013).
- 825 65 Mamsen, L. S., Lutterodt, M. C., Andersen, E. W., Byskov, A. G. & Andersen, C.  
826 Y. Germ cell numbers in human embryonic and fetal gonads during the first two  
827 trimesters of pregnancy: analysis of six published studies. *Hum Reprod* **26**, 2140-  
828 2145, doi:10.1093/humrep/der149 (2011).
- 829 66 Ohta, H. *et al.* In vitro expansion of mouse primordial germ cell-like cells  
830 recapitulates an epigenetic blank slate. *EMBO J* **36**, 1888-1907,  
831 doi:10.15252/embj.201695862 (2017).
- 832 67 Yokobayashi, S. *et al.* Inherent genomic properties underlie the epigenomic  
833 heterogeneity of human induced pluripotent stem cells. *Cell reports* **37**, 109909,  
834 doi:10.1016/j.celrep.2021.109909 (2021).
- 835 68 Miyauchi, H. *et al.* Bone morphogenetic protein and retinoic acid synergistically  
836 specify female germ-cell fate in mice. *EMBO J* **36**, 3100-3119,  
837 doi:10.15252/embj.201796875 (2017).
- 838 69 Nagaoka, S., I. *et al.* ZGLP1 is a determinant for the oogenic fate in mice. *Science*  
839 **367**, doi:10.1126/science.aaw4115 (2020).
- 840 70 Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA  
841 velocity to transient cell states through dynamical modeling. *Nat Biotechnol* **38**,  
842 1408-1414, doi:10.1038/s41587-020-0591-3 (2020).
- 843 71 Okae, H. *et al.* Genome-wide analysis of DNA methylation dynamics during early  
844 human development. *PLoS Genet* **10**, e1004868,  
845 doi:10.1371/journal.pgen.1004868 (2014).
- 846  
847  
848  
849

850 **Figure Legends**

851 **Fig. 1. Exploration of the signaling for epigenetic reprogramming and hPGCLC**  
852 **differentiation.**

853 **a**, Scheme of human germ-cell development. Differentiation stages, key markers, and  
854 stages covered by this study (yellow) are shown. PGCs: primordial germ cells; wpf:  
855 weeks post-fertilisation.

856 **b**, Summary of the acronyms used in this study.

857 **c**, Expression (top) and promoter methylation (bottom) dynamics of epigenetic  
858 reprogramming-activated genes (ER genes) during hPGCLC differentiation in xrOvaries.  
859 Top eight ER genes in the expression level at ag35, and *DAZL* and *DDX4* are annotated.

860 **d**, Scheme for the screening of cytokines/chemicals that induce ER gene up-regulation.

861 **e**, Expression of *PRDM1*, *GTSF1*, *PRAME*, and *MEG3* measured by qRT-PCR at culture  
862 day (c) 22 with the indicated cytokines/chemicals. For each gene,  $\Delta Ct$  from the average  
863 Ct values of two housekeeping genes, *RPLP0* and *PPIA* (set as 0), were calculated and  
864 plotted for two independent experiments. Mean values are shown as a red bar. \*, \*\*:  
865 Not detected or  $\Delta Ct < -10$  in one or two replicates, respectively. ag77: expression  
866 values in hPGCLC-derived cell at ag77 in xrOvaries<sup>19</sup>. Statistical significances were  
867 determined by Welch's *t*-test for the none (n=2) vs. BMPs (n=6) comparison. *p*-values:  
868 0.42 for *PRDM1*, 0.012 for *GTSF1*, 0.14 for *PRAME*, and 0.0038 for *MEG3*.

869

870 **Fig. 2. BMP signaling promotes epigenetic reprogramming and hPGCLC**  
871 **differentiation.**

872 **a**, Flow cytometric analysis of the expression of AG and DT or VT during BMP-driven  
873 M1-*AGDT* or F1-*AGVT* hPGCLC differentiation on the indicated culture days.  
874 Percentages of the cells in each gate are shown.

875 **b**, Growth curve (left) and doubling time (right) of hPGCLC-derived cells induced from  
876 the indicated hiPSC lines. The number of hPGCLC-derived cells was calculated as the  
877 sum of reporter<sup>+</sup> cells or EpCAM<sup>+</sup>/ITGA6<sup>+</sup> cells (for M2). For the doubling time, each  
878 dot represents a doubling time for one passage interval and the red bar represents the  
879 average of all passage intervals. \*: Passaged by flow cytometry. Color coding is as  
880 indicated.

881 **c**, Proportion of cells with the indicated reporter expression during BMP-driven M1-  
882 *AGDT* or F1-*AGVT* hPGCLC differentiation on the indicated culture days.

883 **d**, Relief contrast and fluorescence (DT and AG) images of M1-*AGDT* and F1-*AGVT*  
884 hPGCLC-derived cells at the indicated culture days. Bar, 200  $\mu$ m.

885 **e**, Karyotype (left; percentage of cells with 46 or other chromosome numbers; right:  
886 chromosome spreads) of M1-*AGDT* and F1-*AGVT* hPGCLC-derived cells at the indicated  
887 culture days.

888 **f**, Expression of the indicated genes in the indicated cells (two biological replicates)  
889 measured by qRT-PCR. Quantification was as in Fig. 1d.

890 **g**, Dot blot analysis of the genomic 5mC level in the indicated cells (two biological

891 replicates for hiPSCs and hPGCLCs).

892

893 **Fig. 3. Transcriptome dynamics during BMP-driven hPGCLC differentiation.**

894 **a**, Heatmap showing the expression levels of the indicated genes in the indicated cell  
895 types (see Supplementary Table 2 for full sample information). Color coding is as  
896 indicated. NA: not applicable.

897 **b**, PCA of transcriptomes during hPGCLC induction and BMP-driven or xrOvary-based  
898 hPGCLC differentiation. The left and right panels are color coded with reporter  
899 expression and culture days, respectively, as indicated. The dotted area in the right panel  
900 is magnified to clarify the difference of transcriptome progression between cultures with  
901 and without BMP2.

902 **c**, UMAP and Louvain clustering of scRNA-seq data of female germ cells at 7–16 weeks  
903 post fertilization (wpf) *in vivo*<sup>28,29</sup> and F1-AGVT hPGCLC-derived cells at c11/56/86/117  
904 *in vitro*. Cell-type (left) and cell-cycle (right) annotations are shown.

905 **d–f**, UMAP plot as in (c), with the annotation of *in vivo* and *in vitro* samples (d), with  
906 potential developmental trajectories of *in vivo* and *in vitro* cell types analyzed by RNA  
907 velocity<sup>70</sup> (e), and with expression levels of the indicated genes (f).

908

909 **Fig. 4. DNA methylome reprogramming during BMP-driven hPGCLC  
910 differentiation.**

911 **a**, Violin plots of the average 5mC levels on the indicated genomic loci in the indicated  
912 cell types (see Supplementary Table 2 for full sample information). Bars represent the  
913 average values. In (a–e), the DNA methylome data for human spermatozoa, oocytes,  
914 and blastocysts are from<sup>71</sup> and those for human male and female germ cells at 9 wpf are  
915 from<sup>3</sup>. H/I/LCP: high/intermediate/low CpG promoter.

916 **b**, Heatmap of the 5mC levels of the imprint DMRs in the indicated samples. Color  
917 coding is as indicated.

918 **c**, Escapee numbers common or specific between/in M1-AGDT c122 cells and M1-AGVT  
919 c82 cells (left), among/in the union of M1-AGDT c122/M1-AGVT c82 cells, M2 c76 cells,  
920 and *in vivo* male germ cells at 9 wpf (middle), and among/in F1-AGVT c127 cells, F2-  
921 AGVT c68 cells, ag 120 cells, and *in vivo* female germ cells at 9 wpf (right). Color  
922 coding is as indicated.

923 **d, e**, Venn diagram showing the relationships of the DNA demethylation escapees among  
924 the indicated samples, and composition of the escapees in the indicated samples (d: male  
925 samples; e; female samples, with autosomes and X chromosomes separately indicated).  
926 Color coding is as indicated.

927

928 **Fig. 5. TET1 protects hPGCLCs from differentiation into extraembryonic cells.**

929 **a**, PCA (left: PC1/2; right: PC1/2/3) of the transcriptomes of BMP-driven wild-type and  
930 *TET1* KO hPGCLC differentiation (see Supplementary Table 2 for full sample  
931 information). Color coding is as indicated.

932 **b**, UMAP and Louvain clustering of scRNA-seq data of wild-type and *TET1* KO  
933 hPGCLC culture at c18 (top left), with the annotation of the genotype, or with the  
934 expression levels of the indicated genes. Color code is as indicated.  
935 **c**, Genotype composition (top) and expression profiles of key lineage markers and cell-  
936 type annotation (bottom) of each cluster. AMLC: amnion-like cells; EXMLC: extra-  
937 embryonic mesoderm-like cells. Color coding is as indicated.  
938 **d**, (left) Integrated UMAP plots and Louvain clustering of scRNA-seq data in (**b**) with  
939 those of a PSC-based model of early human post-implantation development<sup>45</sup>. (right)  
940 Distributions of data in this study and the study of Zheng et al.  
941 **e**, Cell-type composition of each cluster. Annotations are based on Zheng et al. and the  
942 results in panel (**c**). The actual proportions of major cell types are shown within the  
943 histogram. The full information is provided in Source Data Fig. 5.  
944

945 **Fig. 6. De-repression of bivalent genes and hypermethylation of regulatory elements**  
946 **in *TET1* deficient cells.**

947 **a**, Violin plots of the average 5mC levels on the indicated genomic loci in wild-type and  
948 *TET1* KO hPGCLC-derived cells at c12 and c42. Bars represent the average values.  
949 **b**, Scatter plot of 5mC levels across all 2-kb bins in wild-type and *TET1* KO hPGCLC-  
950 derived cells at c12. The numbers of bins with higher 5mC levels ( $\geq 30\%$ ) in wild-type  
951 (10,920) and *TET1* KO (238) hPGCLC-derived cells are indicated. For the KO lines,  
952 the average values of KO#1 and KO#2 were used.  
953 **c**, Odds ratio and *q*-value of the enrichment of the 2-kb bins with higher 5mC levels in  
954 wild-type cells compared to *TET1* KO cells at c12 in the ENCODE genomic annotation  
955 categories.  
956 **d**, Two dimensional UMAP embedding of all open sites (ATAC-seq peaks) during  
957 hPGCLC induction based on epigenetic signals of relevant cell types using public data<sup>47</sup>,  
958 with labels derived from semi-supervised HDBSCAN (hierarchical density-based spatial  
959 clustering of applications with noise). The open sites were colored according to the  
960 labels (top, left) or signal intensities of relevant histone modifications (bottom). The  
961 averaged signal intensities of relevant histone modifications in each label (cluster) are  
962 shown (top, left).  
963 **e–g**, Violin plots for the 5mC-level (%) (**e**, **f**) and the expression-level ( $\log_2$  fold-change)  
964 (**g**) differences between wild-type and *TET1* KO hPGCLC-derived cells at c12 on the  
965 indicated elements. Promoters, enhancers (non-promoter open sites), and their labels are  
966 based on the data for d4 hPGCLCs<sup>47</sup>. Silent promoters: promoters that did not overlap  
967 with open sites; silent enhancers: enhancers categorized neither into active, bivalent, nor  
968 poised. Number of each element is indicated. In (**g**), \*\*\*:  $p < 0.001$  (*t*-test adjusted by  
969 Bonferroni correction).  
970 **h**, Odds ratio of the enrichment of genes with indicated promoters defined in d4  
971 hPGCLCs<sup>47</sup> and ER genes (for down-regulated genes) in genes up- (left) or down- (right)  
972 regulated in *TET1* KO hPGCLC-derived cells at c12. Number of each gene class is

973 indicated.

974 **i**, Odds ratio of the c12 up-regulated genes bound by TET in hESCs<sup>50</sup> in each category  
975 of promoters. The odds ratio was calculated relative to the background ratio of all genes  
976 bound by TET in each respective promoter category. Number of each gene class is  
977 indicated.

978 **j**, Violin plots for the 5mC-level differences (%) between wild-type and *TET1* KO  
979 hPGCLC-derived cells at c12 on the indicated elements. All open sites during hPGCLC  
980 induction<sup>47</sup> were defined as regulatory elements (RE). NRE: non-RE regions.  
981 Intergenic NRE are defined as “background” genome. Number of each element is  
982 indicated. \*\*\*:  $p < 2.2 \times 10^{-16}$  (Wilcoxon rank-sum test).

983

984

985

986

987 **Methods**

988 **hiPSC culture**

989 All experiments using hiPSCs were approved by the Institutional Review Board of Kyoto  
990 University and were performed according to the guidelines of the Ministry of Education,  
991 Culture, Sports, Science, and Technology (MEXT) of Japan. The hiPSC lines used were  
992 585B1, 1383D6, 1390G3, and NCLCN, and their derivatives<sup>16,72,73</sup>. NCLCN (derived  
993 from cord blood cells) was purchased from XCell Science and licensed for academic use.  
994 All hiPSC cultures were performed in a humidified incubator with 5% CO<sub>2</sub> at 37°C. All  
995 hiPSC lines used in this study were maintained in StemFit AK02N or AK03N  
996 (Ajinomoto) on a plate coated with iMATRIX-511 (Nippi, 892014). For a passage,  
997 hiPSC colonies were dissociated into single cells by treating with 0.5 × Tryple select [1:1  
998 mixture of Tryple select (Gibco, 12563-011) and PBS (−) containing 0.5 mM EDTA  
999 (Nacalai Tesque, 06894-14)]. Dissociated cells were plated with culture media  
1000 containing 10 µM of Y-27632 (TOCRIS, 1254), and the medium was replaced with a  
1001 fresh one without Y27632 the next day. Pictures of hiPSC colonies were taken with a  
1002 CKX41 inverted microscope (OLYMPUS) equipped with a DS-Fi2 camera (Nikon).

1003

1004 **hPGCLC induction**

1005 All iMeLC/hPGCLC induction experiments were performed in a humidified incubator  
1006 with 5% CO<sub>2</sub> at 37°C. hiPSCs were induced to iMeLCs and then into hPGCLCs as  
1007 described previously<sup>16</sup>. To induce iMeLCs, 1.0–2.0 × 10<sup>5</sup> cells/well of hiPSCs were  
1008 plated with GK15 medium [15% knockout serum replacement (KSR) (Gibco, 10828-028),  
1009 1% MEM non-essential amino acids solution (NEAA) (Gibco, 11140-050), 1% penicillin  
1010 & streptomycin, 2 mM L-glutamine, 2 mM sodium pyruvate (Gibco, 11360-070) and 0.1  
1011 mM 2-mercaptoethanol in GMEM (Gibco, 11710-035)] supplemented with 3 µM  
1012 CHIR99021 (TOCRIS, 4423), 50 ng/mL activin A (PeproTech, AF-120-14) and 10 µM  
1013 Y-27632 on a fibronectin (Millipore, FC010)-coated 12-well plate (the induction time was  
1014 44 h for M1-BTAG and M1-AGDT hiPSCs, 48 h for M1-AGVT, M2, and F2-AGVT  
1015 hiPSCs, and 54 h for F1-AGVT hiPSCs). To induce hPGCLCs, 3.0–6.0 × 10<sup>3</sup> cells/well  
1016 of iMeLCs were plated with GK15 medium supplemented with 200 ng/mL BMP4 (R&D  
1017 SYSTEMS, 314-BP), 100 ng/mL SCF (R&D SYSTEMS, 255-SC), 10 ng/mL LIF (Merck  
1018 Millipore, LIF1010), 50 ng/mL EGF (PeproTech, AF-100-15) and 10 µM Y-27632 in a  
1019 v-bottom 96-well plate (Greiner, 651970). At 6 or 8 days after hPGCLC induction,  
1020 iMeLC aggregates were subjected to FACS sorting [see the Fluorescence-activated cell  
1021 sorting (FACS) section] to isolate hPGCLCs. Pictures of iMeLC aggregates under the  
1022 hPGCLC induction condition were taken with a M205C microscope (Leica) equipped  
1023 with a DP72 camera (OLYMPUS).

1024

1025 **hPGCLC expansion/differentiation culture**

1026 All hPGCLC expansion/differentiation cultures were performed in a humidified incubator  
1027 with 5% CO<sub>2</sub> at 37°C and maintained in the expansion culture medium [aRK15: advanced

1028 RPMI (Gibco, 12633-012) supplemented with 15% knockout serum replacement (KSR)  
1029 (Gibco, 10828-128), 1× GlutaMAX (Gibco, 35050061), 1% penicillin & streptomycin  
1030 (Gibco, 15140-122) and 0.1 mM 2-mercaptoethanol (Gibco, 21985-023)], 2.5% FBS  
1031 (NICHIREI, 175012), 100 ng/mL hSCF (R&D Systems, 255-SC), 10  $\mu$ M forskolin  
1032 (SIGMA, F3917), 20 ng/mL hbFGF (Wako, 064-04541) and 1.5  $\mu$ M IWR1 (SIGMA,  
1033 I0161)] unless otherwise specified. For the male hPGCLC differentiation, 25 ng/mL  
1034 BMP2 was supplemented until the first passage. For the female hPGCLC differentiation,  
1035 50 ng/mL BMP2 was supplemented until the first passage. After the first passage, both  
1036 male and female hPGCLCs were cultured with 100ng/mL BMP2. Only when cells were  
1037 plated upon passage, 10  $\mu$ M Y27632 was added. For the cytokine/chemical screening  
1038 (Fig. 1d), the expansion culture medium was supplemented with the following  
1039 cytokines/chemicals in the indicated amounts: DMSO (Sigma, D8418), VO-OHpic (10  
1040  $\mu$ M; Selleckchem, S8174), PS48 (10  $\mu$ M; Sigma, P0022), MHY1485 (5  $\mu$ M; Sigma,  
1041 SML0810), PD0325901 (1  $\mu$ M; StemGent, 04-0006), PD173074 (1  $\mu$ M; StemGent, 04-  
1042 0008), CHIR99021 (1  $\mu$ M; TOCRIS, 4423), RA (1  $\mu$ M; Sigma, R2625), BMS 493 (1  $\mu$ M;  
1043 Sigma, B6688), WIN18,499 (1  $\mu$ M; Cayman Chemical, 14018), SAG (1  $\mu$ M; TOCRIS,  
1044 4366), PGD2 (100 ng/mL; Cayman Chemical, 12010), FGF4 (100 ng/mL; R&D Systems,  
1045 235-F4-025), FGF9 (100 ng/mL; R&D Systems, 7399-F9), FGF18 (100 ng/mL; R&D  
1046 Systems, 8988-F18), Wnt5a (100 ng/mL; R&D Systems, 645-WN), BMP2 (25 ng/mL;  
1047 R&D Systems, 355-BM-01M/CF), BMP4 (25 ng/mL; R&D Systems, 314BP01M),  
1048 BMP7 (100 ng/mL; R&D Systems, 5666-BP-010), Nodal (50 ng/mL; R&D Systems,  
1049 1315-ND), InhibinA (10 ng/mL; R&D Systems, 8346-IN), NOGGIN (50 ng/mL;  
1050 Proteintech, HZ-1118), Pleiotrophin (100 ng/mL; R&D SYSTEMS, 252-PL-050),  
1051 Midkine (100 ng/mL; R&D Systems, 9760-MD), Osteopontin (100 ng/mL; R&D  
1052 Systems, 441-OP), SDF-1a (100 ng/mL; R&D SYSTEMS, 460-SD), IGF2 (100 ng/mL;  
1053 R&D Systems, 792-MG), Desert Hedgehog (50 ng/mL; R&D Systems, 733-DH), and  
1054 Jagged1 (100 ng/mL; R&D Systems, 599-JG). Details are summarized in Supplementary  
1055 Table 9. For the basal medium comparison (Extended Data Fig. 1d), aRK15 was  
1056 replaced with the indicated basal media listed in Supplementary Table 9. At the  
1057 beginning of the expansion culture, hPGCLCs were directly plated on the Mitomycin c-  
1058 treated m220-5 feeder cells by FACS using the automated cell deposition unit (ACDU)  
1059 sorting mode. Alternatively, hPGCLCs were sorted into FACS buffer [0.1%BSA-  
1060 PBS(−) containing 10  $\mu$ M Y27632]. Then, the sorted hPGCLCs were collected by  
1061 centrifugation at 200g for 7 min and were plated with the expansion culture medium.  
1062 hPGCLCs were identified as described in the “**Fluorescence-activated cell sorting**  
1063 (**FACS**)” section.

1064

1065 A step-by-step protocol for the passage of hPGCLC expansion/differentiation culture  
1066 [modified from <sup>17</sup>] is as follows:

1067

1068 1. Preparation of the culture plate and m220-5 feeder cells

- 1069           1.1. Coat a flat-bottom cell culture plate (i.e., Falcon, 353072) with 0.1%  
1070            gelatin solution at room temperature for at least 1 h.  
1071           1.2. Plate MMC-treated (4 µg/ml, 2 h) m220-5 feeder cells on a gelatin-  
1072            coated cell culture plate. Typically,  $6.25 \times 10^4$  cells are plated in one  
1073            well of a 96-well plate with 1 ml of DMEM containing 10% FBS, 1%  
1074            penicillin and streptomycin, and 1x GlutaMAX.  
1075           1.3. Incubate m220-5 feeder cells in a humidified incubator maintained at 5%  
1076            CO<sub>2</sub> and 37°C until use.  
1077        2. Cell harvesting and re-plating  
1078           2.1. Wash cultured cells with PBS(–) once.  
1079           2.2. Treat cultured cells with a 1:4 mixture of 0.5% trypsin–EDTA and PBS  
1080           (–) at 37°C for 5 min. Alternatively, Accumax (AM105, Innovative  
1081           Cell Technologies) can be used to dissociate the cultured cells.  
1082           2.3. Neutralize the trypsin by adding advanced RPMI containing 10 % FBS,  
1083           0.1 mg/mL DNase1 and 10 uM Y27632. Alternatively, the expansion  
1084           culture medium can be used for the neutralization.  
1085           2.4. Repeat the pipetting of the cell suspension to dissociate it into single  
1086           cells.  
1087           2.5. If needed, harvested cells were subjected to immunolabeling (see the  
1088           **“Fluorescence-activated cell sorting (FACS)” section**).  
1089           2.6. Plate the desired volume of cell suspension with culture medium  
1090           supplemented with 10 µM Y27632 on the m220-5 feeder cells.  
1091           Typically, the collected cell suspension is passaged at a dilution ratio of  
1092           around 1:5.  
1093           2.7. Incubate cells in a humidified incubator maintained at 5% CO<sub>2</sub> and 37°C.  
1094           2.8. Remaining cell suspension was subjected to FACS after adding DAPI at  
1095           the final concentration of 0.1 µg/mL. To measure the growth of AG<sup>+</sup>  
1096           cells, DAPI<sup>–</sup> live cells were used for analysis.  
1097        3. Maintenance of the expansion culture  
1098           3.1. Replace the medium with a new one on the second day.  
1099           3.2. Replace half the medium with a new one every other day until the next  
1100           passage on around day 10 after cell plating.  
1101

1102       **Establishment of the reporter-knockin iPSC lines**

1103       To generate iPSC lines bearing *TFAP2C-p2A-eGFP*, *DAZL-p2A-tdTomato* and *DDX4-*  
1104       *p2A-tdTomato* reporters, the *p2A-eGFP/p2A-tdTomato* sequences were replaced with the  
1105       termination codon of the targeted gene by TALEN-assisted homologous recombination.  
1106       First, homologous regions surrounding the termination codon of human *DAZL/DDX4*  
1107       were cloned into the pCR2.1 vector (pCR2.1-DAZL/pCR2.1-DDX4) (TOPO TA Cloning;  
1108       Thermo Fisher Scientific, 450641). Next, the *p2A-tdTomato* sequences concatenated  
1109       with the PGK promoter-driven Neomycin resistance gene flanked by loxP sites were

1110 subcloned into the pCR2.1-DAZL/pCR2.1-DDX4 vector by using a GeneArt seamless  
1111 cloning and assembly kit (Thermo Fisher Scientific, A13288). Finally, the MC1  
1112 promoter-driven DT-A gene was isolated by XbaI/ApaI or XhoI/ApaI digestion and then  
1113 subcloned into the region downstream of the homologous regions by enzymatic ligation.  
1114

1115 To assemble TALEN constructs targeting the human *DAZL/DDX4* gene locus, a  
1116 GoldenGate TALEN and TAL Effector kit (Addgene, #1000000016) was employed. The  
1117 amino acid sequences of repeat variable di-residues in each TALEN were designed as  
1118 follows: *DAZL* 5' end [NG NG NN NI NG HD HD NG HD HD NG NN NN HD NG NG  
1119 NI NG HD]; *DAZL* 3' end [NI NG NG HD NI NI NI NI HD HD NI NN HD NI NI HD  
1120 NG NG]; *DDX4* 5'end [HD HD HD NI NI NG HD HD NI NN NG NI NN NI NG NN  
1121 NI NG]; and *DDX4* 3'end [NN NI NI NN NN NI NG NN NG NG NG NG NN NN HD  
1122 NG].  
1123

1124 5 µg of the donor vector and 2.5 µg of each TALEN plasmid were introduced into  $1.0 \times$   
1125  $10^6$  hiPSCs (585B1#17, *TFAP2C-p2A-eGFP* with the PGK promoter-driven puromycin  
1126 resistance gene flanked by loxP sites)<sup>16</sup> by electroporation using a NEPA21 type II  
1127 electroporator (Nepagene). 150 µg/mL of G418 (2 to 4 days after the electroporation)  
1128 and 0.2 µg/mL puromycin (4 days after the electroporation) were added to eliminate  
1129 untargeted cells. About 2 weeks after the electroporation, single iPSC colonies were  
1130 isolated by manually picking them up using disposable plastic pipette tips. The  
1131 zygosities of the targeted locus and random integrations of isolated clones were assessed  
1132 by PCR using genomic DNA samples. Isolated clones bearing reporter genes knocked-  
1133 in in a heterozygous manner were subsequently transfected with Cre recombinase  
1134 expression plasmids. Finally, the excision of loxP-flanked drug-resistance genes in  
1135 isolated clones was verified by PCR using genomic DNA samples. DNA sequences of  
1136 the oligonucleotides used were listed in Supplementary Table 10.  
1137

1138 M1-*AGDT*, M1-*AGVT* and F1-*AGVT* iPSC lines were confirmed to be karyotypically  
1139 normal (46, XY for M1 and 46, XX for F1, respectively) by G-banding analysis  
1140 performed by Nihon Gene Research Laboratories (Miyagi, Japan).  
1141

#### 1142 Establishment of *TET1* KO human iPS cell lines

1143 *TET1* KO human iPS cell lines were established as previously described<sup>74</sup>. M1-*BTAG*  
1144 was employed as a parental line. The DNA oligos used for vector construction are listed  
1145 in Supplementary Table 10. One million BTAG iPSCs were transfected with 3.0 µg each  
1146 of two vectors respectively encoding guide RNAs targeting the human *TET1* gene and  
1147 mCherry-tagged Cas9 nickase in 100 µL OptiMEM (Gibco, 31985-062) using the  
1148 NEPA21 type II electroporator. Two days after the transfection, the transfected BTAG  
1149 iPSCs highly expressing mCherry were subjected to single cell sorting by FACS using  
1150 the ACDU sorting mode. The sorted cells were directly plated on 96-well plates pre-

1151 filled with AK02N supplemented with 10 µM of Y27632 and 0.25 µg/cm<sup>2</sup> iMATRIX.  
1152 Three days after the single cell sorting, the culture medium was replaced with AK02N.  
1153 At 12–14 days after single cell sorting, the viable cells were cryopreserved in STEM-  
1154 CELLBANKER cryopreservation solution (ZENOAQ, CB045) or snap-frozen in liquid  
1155 nitrogen for genotyping PCR.

1156

#### 1157 **Genotype analysis of *TET1* KO human iPSC cell lines**

1158 Each cell pellet was resuspended in the lysis buffer [3 µL of 10× KOD Plus Neo Buffer  
1159 (Toyobo, KOD-4B), 3 µL of 5% Nonidet P40 (Nacalai Tesque, 23640-94), 1 µL of 20  
1160 mg/mL proteinase K, and 23 µL of water]. All cell suspensions were incubated at 55°C  
1161 for 2 h, followed by 95°C for 5 min. Alternatively, genomic DNA was extracted from  
1162 the iPSC pellet by using NucleoSpin Tissue (Macherey-Nagel, U0952S) according to the  
1163 manufacturer's instruction. 1 µL of the cell lysate or genomic DNA was subjected to  
1164 PCR amplification using KOD -Plus- Neo (Toyobo, KOD-401) with primers whose  
1165 sequences were as follows: forward, TTTTAAAGAGTGATCTTGAGGATAAGTGA;  
1166 reverse, ATCACTCTGGCTATTGTAAATGTAT. After the PCR product purification,  
1167 amplicons were cloned into pGEM®-T Easy Vector (Promega, A1360). The vectors  
1168 were transformed into homemade competent cells of the DH5 $\alpha$  strain. Transformed cells  
1169 were plated on an LB plate coated with X-gal and IPTG for blue–white selection and  
1170 were incubated at 37°C until colony formation. White colonies were directly used as a  
1171 template for the PCR amplification using KOD FX Neo (Toyobo, KFX-201) with primers  
1172 whose sequences were as follows: forward, CAGGAAACAGCTATGAC; reverse,  
1173 GTAAAACGACGGCCAG. In order to identify the indels of each clone, PCR products  
1174 were subjected to Sanger sequence analysis by Eurofins Genomics (Tokyo, Japan).

1175

#### 1176 **Cell number count of hPGCLC-derived cells**

1177 To calculate the expansion of hPGCLC-derived cells, we used the sum of reporter/surface  
1178 marker<sup>+</sup> cells, i.e., the sum of BT<sup>+</sup>AG<sup>+</sup> cells for M1-*BTAG*, the sum of AG<sup>+</sup>DT<sup>-</sup>/VT<sup>-</sup>,  
1179 AG<sup>+</sup>DT<sup>+</sup>/VT<sup>+</sup>, and AG<sup>-</sup>DT<sup>+</sup>/VT<sup>+</sup> cells for M1-*AGDT/VT* and F1/F2-*AGVT*, and the sum  
1180 of EpCAM<sup>+</sup>/ITGB6<sup>+</sup> cells for M2.

1181

#### 1182 **Preparing samples for mass spectrometry by in-gel digestion**

1183 About  $2 \times 10^6$  M1-*BTAG* iPSCs or *TET1*-KO iPSCs were lysed in RIPA buffer (Santa  
1184 Cruz, SC-24948). After 30 min incubation at 4°C on the rotator, the cell lysate was  
1185 subjected to sonication. Then, the cell lysate was centrifuged at 14,000 rpm at 4°C for  
1186 10 min, and the supernatant was used for the experiment. Total protein concentration  
1187 was measured with a Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, 23227).  
1188 Cell lysate containing 12 µg protein was mixed with 4 × Laemmli sample buffer (Bio-  
1189 Rad, 1610747) supplemented with 2-mercaptoethanol, and boiled at 95°C for 5 min.  
1190 After incubation on ice, the cell lysate was separated by 8% SDS-PAGE in MOPS buffer  
1191 at 100V for 20 min. Gel pieces containing separated proteins of around 260 kDa and

1192 160 kDa were harvested into microtubes independently. The gel pieces were  
1193 sequentially treated as follows: destaining with 200  $\mu$ L of 50 mM ammonium bicarbonate  
1194 (ABC) (Fujifilm Wako, 017-02875)/50% acetonitrile (ACN) for 30 min; dehydration with  
1195 100% ACN; reduction with 500  $\mu$ L of 10 mM dithiothreitol (DTT) (Fujifilm Wako, 045-  
1196 08974)/50 mM ABC for 30 min; alkylation with 50 mM iodoacetamide (IAA) (Fujifilm  
1197 Wako, 095-02151)/50 mM ABC for 30 min in the dark; washing with 200  $\mu$ L of 0.5%  
1198 acetic acid/50% methanol twice; equilibration with 50 mM ABC; dehydration with 100%  
1199 ACN; protein digestion with 10  $\mu$ L of trypsin solution (10 ng/ $\mu$ L in 50 mM ABC) for 5  
1200 min; and additional incubation with 50  $\mu$ L of 50 mM ABC buffer at 37°C overnight.  
1201 Trypsinization was terminated by adding 5  $\mu$ L of 10% TFA. The supernatant was  
1202 subjected to additional extraction with 50% ACN/0.1% TFA and 80% ACN/0.1% TFA.  
1203 Finally, the supernatant was dried and suspended in 0.1% TFA. The resulting samples  
1204 were desalted using SDB-XC StageTips<sup>75,76</sup>.

1205

#### 1206 **LC/MS/MS**

1207 LC/MS/MS for in-gel digested samples was performed with an Orbitrap Fusion Lumos  
1208 (Thermo Fisher Scientific) linked with an Ultimate 3000 pump (Thermo Fisher Scientific)  
1209 and HTC-PAL autosampler (CTC Analytics). Samples were subjected to a needle  
1210 column pulled in-house (150 mm length, 100  $\mu$ m ID, 6  $\mu$ m needle opening) house-packed  
1211 with Reprosil-Pur 120 C18-AQ 1.9  $\mu$ m reversed-phase material (Dr. Maisch GmbH),  
1212 using a 60 min gradient of 5–40% B at a flow rate of 500 nL/min (mobile phase A: 0.5%  
1213 acetic acid; mobile phase B: 0.5% acetic acid/80% acetonitrile). Orbitrap MS1 spectra  
1214 were collected at a resolution of 120,000. Data-dependent ion trap (IT) MS2 scans were  
1215 collected in the Top Speed mode using a cycle time of 1.5 sec between Full MS scans.

1216

#### 1217 **LC/MS/MS data analysis**

1218 Acquired data files were processed using FragPipe v20.0 (MSFragger v3.8, Philosopher  
1219 v5.0.0, IonQuant v1.9.8) for protein identification and quantitation<sup>77-79</sup>. Peptides and  
1220 proteins were identified by automated database searching against the human  
1221 UniprotKB/Swissprot database (accessed on 2021/07/15) with settings of strict trypsin  
1222 (C-terminal of K and R) specificity and the allowance for up to 2 missed cleavages.  
1223 Carbamidomethylation of cysteine (C, +57.021465) was set as a fixed modification.  
1224 Oxidation of methionine (M, +15.9949) and acetylation of protein N-terminus (Protein  
1225 N-term, +42.0106) were set as variable modifications. False discovery rates of peptide  
1226 and protein levels were controlled to 1% using PeptideProphet and ProteinProphet<sup>80</sup>.  
1227 Peptide quantification was performed using IonQuant.

1228

#### 1229 **Fluorescence-activated cell sorting (FACS)**

1230 hPGCLCs were isolated from the aggregates of iMeLCs as described previously<sup>17</sup>.  
1231 Briefly, iMeLC aggregates were treated with a 1:1 mixture of PBS(–) and 0.5% Trypsin-  
1232 EDTA (Gibco, 15400-054). After the neutralization with DMEM (Gibco, 10313-021)

1233 containing 10% FBS, 1% penicillin & streptomycin, 2 mM L-glutamine, 10  $\mu$ M Y-27632,  
1234 and 0.1 mg/mL DNase I (Sigma, DN25), the cell suspensions were subjected to  
1235 centrifugation at 200g for 5 min. The cells were then resuspended in FACS buffer and  
1236 sorted using a FACS Aria III. hPGCLCs were gated as BT<sup>+</sup>AG<sup>+</sup> cells (M1-BTAG line)  
1237 or AG<sup>+</sup> cells (M1-AGDT, M1-AGVT, F1-AGVT, and F2-AGVT lines) or EpCAM<sup>+</sup>ITGA6<sup>+</sup>  
1238 cells (M2) on hPGCLC induction day 6 or day 8.

1239

1240 To analyze the surface antigen expression, trypsinized cells were resuspended in FACS  
1241 buffer containing antibodies [BV421-conjugated anti TRA-1-85 antibody (BD  
1242 Bioscience, 563302) or BV421-conjugated isotype control antibody (BD Bioscience,  
1243 562438), PE- or BV421-conjugated anti ITGA6/CD49f antibody (BioLegend,  
1244 313611/313624), APC-conjugated anti EpCAM/CD326 antibody (BioLegend, 324208)]  
1245 and were incubated on ice for 30 min in the dark. After washing with PBS (-) and  
1246 pelleting, the cells were resuspended in FACS buffer and were subjected to FACS analysis.  
1247

#### 1248 **Immunofluorescence (IF) analysis**

1249 IF analysis was performed as described previously <sup>17</sup>. The expanded PGCLCs were  
1250 cultured in a  $\mu$ -Slide 8-Well chamber slide (ibidi, 80826) for several days. After washing  
1251 with PBS (-) once, the cells were fixed in 4% paraformaldehyde for 15 min at room  
1252 temperature and washed with PBS (-) three times. The fixed cells were incubated in  
1253 blocking buffer [PBS (-) with 10% normal donkey serum (Jackson ImmunoResearch,  
1254 017-000-121), 3% BSA (Sigma, A4503), and 0.1% Triton-X 100 (Nacalai Tesque, 35501-  
1255 02)]. Then, the cells were incubated in 0.5  $\times$  blocking buffer [a 1:1 mixture of blocking  
1256 buffer and PBS (-)] containing primary antibodies for 30 minutes at room temperature.  
1257 The dilution rates for the primary antibody reactions were as follows: rat anti-GFP  
1258 antibody (Nacalai Tesque, 04404-84, 1:250), goat anti-tdTomato antibody (Origene,  
1259 AB8181-200, 1:200), mouse anti-UHRF1 antibody (Milipore, MABE308, 1:200), mouse  
1260 anti-AP2 $\gamma$ /TFAP2C (1/100; Santa Cruz, sc-12762), mouse anti-DAZL antibody (Santa  
1261 Cruz, sc-390929, 1:100), rabbit anti-DDX4 antibody (abcam, ab13840, 1:250). Then the  
1262 cells were washed with PBS (-) three times and incubated in 0.5  $\times$  blocking buffer  
1263 containing 1  $\mu$ g/mL of DAPI and secondary antibodies [Alexa Fluor 488 donkey anti-rat  
1264 IgG (Invitrogen, A21208, 1:800); Alexa Fluor 568 donkey anti-goat IgG (Invitrogen,  
1265 A11057, 1:800); Alexa Fluor 647 donkey anti-mouse IgG (Invitrogen, A31571, 1:800);  
1266 and Alexa Fluor 647 donkey anti-rabbit IgG (Invitrogen, A31573, 1:800)] for 1 h at room  
1267 temperature. The cells were washed with PBS (-) three times and were mounted in ibidi  
1268 mounting media (ibidi, 50001). Images were acquired by a confocal microscopy system  
1269 (IX81-FV1000, Olympus) or fluorescence microscope (BZX810, Keyence). For the  
1270 line-plot analysis (Extended Data Fig. 9k), the signal intensities of the lines placed across  
1271 randomly picked 10 cells were measured with “plot profile” function of the Fiji software  
1272 <sup>81</sup>. The outlines of the nucleus and cytoplasm were determined manually based on the  
1273 visual inspection of the DAPI and AG signals, respectively. The UHRF1 signals were

1274 normalized to set the maximum value as 1. To plot the signals from the various lengths  
1275 of the nucleus and cytoplasm in an integrated manner, the lengths of the nucleus and  
1276 cytoplasm were normalized to the range from 0 to 1. The curve fitting analysis was  
1277 performed by the geom\_smooth function with "method = "gam"" option from the ggplot2  
1278 package in R.

1279

### 1280 **Automated image analysis**

1281 Immunofluorescence images were quantified by CellProfiler 4.2.1 automated image-  
1282 processing software <sup>82</sup>. For analyzing UHRF1 subcellular localization (Extended Data  
1283 Fig. 9k), first, nuclei were identified based on the DAPI channel signal using the  
1284 IdentifyPrimaryObjects module. Second, hPGCLC-derived cells were identified based  
1285 on the GFP channel signal using the IdentifySecondaryObjects module. Third,  
1286 cytoplasms were identified by subtracting the nuclei from the cells with the  
1287 IdentifyTertiaryObjects module. Finally, the intensity of the nuclei and cytoplasms was  
1288 calculated using the MeasureObjectIntensity module. Mean intensity values of each cell  
1289 were used for calculating the ratio (cytoplasm over nucleus). For the validation of  
1290 fluorescence reporter expression, nuclei and cells were identified as described above.  
1291 The intensity from the nuclei (TFAP2C, GFP) or cells (DAZL, DDX4, GFP, tdTomato)  
1292 was calculated using the MeasureObjectIntensity module. Mean intensity values were  
1293 used for analysis.

1294

### 1295 **Chromosome counting**

1296 hiPSCs were incubated with 0.1 µg/mL demecolcine (WAKO, 045-18761) for 4 h. After  
1297 the incubation, iPSCs were dissociated into single cells with 0.5 × Tryple select and then  
1298 centrifuged at 160g for 5 min.

1299

1300 Expanding hPGCLCs were incubated with 0.1 µg/mL demecolcine overnight. To isolate  
1301 expanded hPGCLCs, AG<sup>+</sup> cells were sorted using a FACS Aria III (BD) from the whole  
1302 cell suspension harvested from the expansion culture. To prevent the resumption of cell  
1303 division, cells were sorted into CELLOTION solution (ZENOAQ, CB051) containing 0.1  
1304 µg/mL demecolcine and centrifuged at 200g for 7 min.

1305

1306 After centrifugation, cells were treated with the hypotonic buffer (Genial Helix,  
1307 GGSJL006B) for 30 min at 37°C and then treated cells were placed on ice for 3 min.  
1308 Treated cells were then fixed in a fixative solution [methanol:acetic acid = 3:1] and spread  
1309 on slide glasses. Chromosome spreads were stained with 0.1 µg/mL DAPI for 10 min at  
1310 room temperature. Images were taken by a confocal microscopy system (IX81-FV1000,  
1311 Olympus).

1312

### 1313 **DNA dot blot**

1314 Genomic DNA was extracted from cell pellets using NucleoSpin Tissue (MACHEREY-

NAGEL, U0952S) by following the manufacturer's instructions. Eluted DNA was diluted to 5 ng/ $\mu$ L (for 5mC detection) or 25 ng/ $\mu$ L (for 5hmC detection). 5  $\mu$ L of 0.5N NaOH was added to 20  $\mu$ L of the DNA. Then, the mixture was incubated at 99°C for 5 min. After the incubation, 2.5  $\mu$ L of 6.6M ammonium acetate was added to the mixture. Finally, 2.75  $\mu$ L of the mixture, which is equivalent to 50 ng of genomic DNA, was spotted onto a nitrocellulose membrane pre-soaked in 20  $\times$  SSC. The membrane was baked for 2 h at 80°C and then incubated in the blocking buffer [PBST (PBS containing 0.1% Tween20) containing 10% skim milk (BD, 232100) and 1% BSA (Sigma, A4503)] for 1 h at room temperature. Then, the membrane was incubated with rabbit anti-5hmC antibody (Active Motif, 39069) at 1:5000 dilution or rabbit anti-5mC antibody (Sigma, SAB5600040) at 1:1000 dilution in the blocking buffer overnight at 4°C. The membrane was washed in PBST for 5 min three times, and then incubated with anti-rabbit IgG goat IgG conjugated with HRP (Sigma, A6154) at 1:8000 dilution in the blocking buffer for 1 h at room temperature. The membrane was washed in PBST for 5 min three times, and then the membrane was subjected to chemiluminescence detection with Chemi-Lumi One Super (Nacalai Tesque, 02230-30) or Chemi-Lumi One Ultra (Nacalai Tesque, 11644-24). The chemiluminescence signal was detected by Fusion solo S (VILBER LOURMAT).

1332

### 1333 Western blot analysis

About  $2.0 \times 10^4$  BT<sup>+</sup>AG<sup>+</sup> cells were collected into a tube containing Cellotion by FACS at c33. After centrifugation, cells were lysed in 4 $\times$  Laemmli sample buffer supplemented with 2-mercaptoethanol, and boiled at 95°C for 5 min. After incubation on ice, the cell lysates were separated by 10% SDS-PAGE (~ $6.0 \times 10^3$  cells/lane) in MOPS buffer at 200 V for 25 min. Separated proteins were transferred onto the PVDF membrane, and the membrane was treated with 5% skim milk in TBS containing 0.1% Tween20 (TBST) for 1 h at room temperature. After blocking, the membrane was incubated with anti-p44/42 MAPK(ERK1/2) antibody (1:1000, CST, 4695) or anti-phospho p44/42 MAPK(ERK1/2) antibody (1:2000, CST, 4730) overnight at 4°C in 5%BSA-TBST. Then, the membrane was washed three times in TBST and subjected to secondary antibody reaction with anti-rabbit IgG goat IgG conjugated with HRP. Immunoblotting for  $\alpha$ TUBLIN was performed on the same membrane after antibody stripping by using WB Stripping Solution (Nacalai Tesque, 05364-55). Then the membrane was treated as described above except for 1-h primary antibody reaction in 5% skim milk in TBST (anti- $\alpha$ TUBLIN antibody, 1:8000, Sigma, T9026). After washing with TBST three times, chemiluminescence was detected with Chemi-Lumi One Super (for ERK1/2 and  $\alpha$ TUBLIN, Nacalai Tesque, 02230-30) or Immunostar LD (for phosho ERK1/2, Fujifilm, 296-69901). Images were taken using a Fusion Solo imaging system. The signal intensity was quantified by CaptAdvanced software (Vilber Lourmat). To determine the relative phosphorylation levels of ERK1/2, first, the signal intensities of ERKs or pERKs were divided by the signal intensities of  $\alpha$ TUBLIN. Then, the normalized pERK intensity was divided by the normalized ERK intensity.

1356

1357 **cDNA amplification and qPCR**

1358 For the screening of cytokines/chemicals, the whole cell lysate of the hPGCLC expansion  
1359 culture was subjected to reverse transcription and subsequent qPCR using a CellAmp™  
1360 Direct TB Green® RT-qPCR Kit (TaKaRa, 3735A) following the manufacturer's  
1361 instructions. For other experiments, total RNA was extracted from cells using a  
1362 NucleoSpin RNA XS Kit (MACHEREY-NAGEL, U0902A) according to the  
1363 manufacturer's instructions. 1 ng of extracted RNA was used for the reverse  
1364 transcription and the cDNA amplification described in the previous report<sup>34</sup>. Amplified  
1365 cDNA was used for qPCR with Power SYBR Green PCR Master Mix (Applied  
1366 Biosystems, 4367659) on a CFX384 Touch Real-Time PCR detection system (Bio-Rad  
1367 Laboratories). For the cytokines/chemicals screening, the whole cell lysate of the  
1368 hPGCLC expansion culture was subjected to reverse transcription and subsequent qPCR  
1369 using a CellAmp™ Direct TB Green® RT-qPCR Kit (TaKaRa, 3735A) following the  
1370 manufacturer's instructions. DNA sequences of the primers used for qPCR are listed in  
1371 Supplementary Table 10.

1372

1373 **Reference genomes and annotations**

1374 Genome sequences and transcript annotations for the human genome (GRCh38.p12) were  
1375 downloaded from the NCBI ftp site. Consensus sequences for retrotransposons were  
1376 downloaded from Repbase v27.08<sup>83</sup>. Repeatmasker information for GRCh38 was  
1377 downloaded from the UCSC table browser. CpG islands were from the data defined in  
1378 Illingworth et al.<sup>84</sup> in hg18 format, which were converted into GRCh38 using the binary  
1379 version of the LiftOver program (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>).  
1380 Differentially methylated regions of imprinted genes were obtained from Court et al.<sup>85</sup>.  
1381 Promoter regions of genes were defined as sequences between 900 bp upstream and 400  
1382 bp downstream of the transcription start sites (TSSs). Intergenes were defined as regions  
1383 between genes and excluding retrotransposons. Exons, introns and intergenes with a  
1384 length of between 500 bp and 10kb were used for the analysis.

1385

1386 **Bulk RNA-seq library preparation, data processing and analysis**

1387 Sequencing libraries were generated from amplified cDNA as described previously<sup>17,34,86</sup>.  
1388 Mapping and processing were performed as described in<sup>34</sup> and on the GEO depository  
1389 site (see Data availability). In brief, all reads were processed with Cutadapt v1.18<sup>87</sup> for  
1390 trimming of adaptor and poly-A sequences, then mapped onto GRCh38.p12 transcript  
1391 references using TopHat2 v2.1.1<sup>88</sup>. The resulting BAM files were processed with HTseq  
1392 v0.9.1<sup>89</sup> to calculate read counts per gene. Read counts per gene were divided by total  
1393 read counts in all chromosomal genes to estimate the RPM values for genes. For  
1394 chromosomal expression ratio analysis, the autosome [chromosome (chr.) 10]: total  
1395 autosome ratio and the chr. X: total autosome ratio (X:A ratio) was calculated according  
1396 to the method described previously<sup>90</sup>. Briefly, 75<sup>th</sup>-percentile log<sub>2</sub>(RPM+1) values of

1397 the expressed genes [the maximum log<sub>2</sub>(RPM + 1) values > 3, among the cells analyzed]  
1398 on the autosomes and chr. X except for the pseudo-autosomal regions  
1399 (<https://www.ncbi.nlm.nih.gov/grc/human>) in individual samples/cells were calculated  
1400 and used as representative expression values in the samples/cells. 12,880/532/509 genes  
1401 on total autosome/Chr.10/Chr.X were used. Gene ontology analysis was performed  
1402 using DAVID 2021<sup>91</sup>.

1403

1404 **Single cell RNA-seq (scRNA-seq) library preparation, data processing and analysis**  
1405 Whole-cell suspensions harvested from the expansion culture were pelleted and were  
1406 subjected to Cell Multiplexing Oligo labeling. Alternatively, samples for analysis were  
1407 cryopreserved as follows: whole-cell suspensions harvested from the expansion culture  
1408 were pelleted and were resuspended in Cell Banker Type 1 Plus for cryopreservation at  
1409 -80°C. For Cell Multiplexing Oligo labeling, after thawing in the water bath at 37°C if  
1410 needed, collected cells were resuspended in Cell Multiplexing Oligo solution and  
1411 incubated for 5 min at room temperature. After the incubation, cells were washed with  
1412 1% BSA (Sigma, A1519)-PBS (-) and were collected by centrifugation at 200g for 10  
1413 min. Collected cells were resuspended in 1% BSA-PBS (-) containing anti-TRA-1-85  
1414 antibody (BD Bioscience, 563302, 1:20) and placed on ice for 30 min in the dark. Then,  
1415 the cells were washed with PBS once and were resuspended in 1% BSA-PBS (-)  
1416 containing 3 µM DRAQ7 (abcam, ab109202). To isolate live human cells, TRA-1-85  
1417 (+) DRAQ7(-) cells were sorted into DMEM/F12 (Gibco, 10565-018) + 1% BSA by  
1418 using a FACS Aria III. Sorted cells were used for the subsequent library construction  
1419 with a Chromium Single Cell 3' Reagent Kit (v3.1 Chemistry Dual Index), followed by  
1420 sequencing on an Illumina NovaSeq 6000 platform. All steps were performed in  
1421 accordance with the manufacturer's instructions. Public 10X single cell RNA-seq data  
1422<sup>28,29,45</sup> were downloaded and were subjected to subsequent processing steps. Obtained  
1423 read data were processed with CellRanger v6.0.1 with default settings using the transcript  
1424 reference as described above. Read data generated from samples labeled with Cell  
1425 Multiplexing Oligo were processed with CellRanger v6.0.1 with the "multi" option.

1426

1427 *Comparison of hPGCLC-derived cells in vitro and human germ cells in vivo*

1428 The analysis of gene-barcode matrices was performed using the Seurat R package (v4.2.1)  
1429<sup>92</sup>, following the online tutorials (<https://satijalab.org/seurat/index.html>). Cells with low  
1430 quality, putative doublets/multiplets, and putative stripped nuclei were excluded from  
1431 further analysis. Doublet cells were removed using the Scrublet python package (v0.2.3)  
1432 according to the instructions reported in<sup>93</sup>. The threshold for doublet estimation was  
1433 determined based on the bimodal distribution of simulated doublet scores. Detailed  
1434 information for parameters used in the quality filtering were described in Supplementary  
1435 Table 11. For further analysis, *in vitro* datasets were merged with cells in the DDX4<sup>+</sup>  
1436 germ-cell cluster extracted from human *in vivo* 10x datasets<sup>28,29</sup>. The merged count  
1437 matrix was applied to noise reduction using the RECODE.fit\_transform function from

the screcode python package v.0.1.2<sup>94</sup>. The denoised data were normalized and log-transformed [ $\log_e(\text{size-scaled (ss) UMI} + 1)$ ] using the NormalizeData (scale.factor=100,000) function in the Seurat R package. To identify highly variable genes (HVGs), we applied the FindVariableFeatures (selection.method="vst,") function in the Seurat R package and detected 5,000 HVGs, which were expressed [ $\log_e(\text{ssUMI} + 1) > 2$ ] in at least 5 cells. Using the HVGs, batch correction was performed by the RunFastMNN function in the Seurat R package. Cell clusters were characterized and visualized using the Seurat RunUMAP, FindNeighbors and FindClusters functions with the parameter "dims=1:50". The FindClusters function identified 12 clusters. After removing three small clusters consisting of non-germ cells or putative doublets/multiplets that co-expressed the markers of other cell types, 10 clusters were defined by the expression pattern of key germ cell markers as "very early mitotic 1/2", "early mitotic 1/2", "mitotic 1/2/3", "pre-leptotene/leptotene 1/2", and "zygotene/pachytene/diplotene." To select variably expressed genes during human germ-cell development, we defined 5,000 HVGs again using human *in vivo* datasets. For GO enrichment analysis and heatmap visualization, genes with low expression levels [cluster average of  $\log_e(\text{ssUMI} + 1) \leq 1.0$ : 3,745 genes] were excluded. Cell-cycle scoring and annotation were performed using the CellCycleScoring function of the Seurat package based on the scoring strategy described in<sup>95</sup>. The human Quick GO term "Meiotic cell cycle" was used to score the meiotic prophase I. Cell-Cycle scores for the S, G2/M, and meiotic prophase I phases were calculated using the AddModuleScore function, and cells without any of these annotations were assigned to the G1 phase. RNA velocity analysis was processed using the "scvelo.pp.moments(n\_pcs = 30, n\_neighbors = 30)" and "scvelo.tl.velocity(mode = 'stochastic')" functions in the scVelo python package (v0.2.5)<sup>70</sup>. To analyze the allelic gene expression, the 10X transcriptome analysis data were remapped to the masked GRCh38.p12 reference using the "cellranger multi" pipeline in Cell Ranger (v. 6.0.1) with the following parameters: "include-introns"=true, and modified star parameters "--alignEndsType EndToEnd" and "--outSAMattributes NH HI NM MD". The BAM files resulting from the mapping were processed with SNPsSplit (v. 0.5.0) using the "--paired" option to split the BAM files into phased alleles. The BAM files were then converted to FASTQ files using the "bamtofastq" command in Cell Ranger (v. 6.0.1). The "cellranger count" pipeline was used to calculate UMI counts with default parameters except for "--include-introns" for both sets of FASTQ files, separately. UMI counts for Xa and Xi were generated using the phasing blocks defined below. Genes with low expression (total UMI counts from Xa and Xi < 3 in datasets) and those overlapping with multiple phasing blocks were excluded. Subsequently, Xa allele usage was calculated as [UMI counts from Xa] over [total UMI counts from Xa and Xi] ([total UMI counts from Xa and Xi > 2]) for each gene and cell.

1476

1477 *Analysis of BMP-driven differentiation of TET1 KO hPGCLCs*

1478 CellRanger output files were loaded by the Read10X and CreateSeuratObject functions

of Seurat package v4.1.1 in R software. RECODE was applied to mitigate the course of dimensionality of single cell RNA-seq data through the desktop application v1.1.1. Doublet cells were detected by Scrublet<sup>93</sup> with default settings and were excluded from the data. Low-quality cells were excluded by the “subset” function of the Seurat package based on the following criteria: nCount\_RNA >10000 & nCount\_RNA <100000 & percent.mt < 10. UMI counts of the cells passing the filters described above were normalized by the NormalizeData function of Seurat with the “scale.factor = 100000” option. Highly variable genes were identified by the FindVariableFeatures function with the following options: “selection.method = “vst””, “nfeatures = 2000”. Data scaling, PCA, construction of a shared nearest neighbor graph, identification of clusters, and dimensional reduction by UMAP were done by the ScaleData, RunPCA, FindNeighbors, FindClusters, and runUMAP functions, respectively. For the data of *TET1* KO hPGCLC-derived cells, the specific parameters in the FindNeighbors and FindClusters were as follows, respectively: “dims = 1:22” and “resolution = 0.5”. For the data of<sup>45</sup>, the specific parameters in FindNeighbors and FindClusters were as follows, respectively: “dims = 1:10” and “resolution = 0.4”. Data integration was performed by the IntegrateData function<sup>96</sup>.

1496

#### 1497 **Public scRNA-seq data processing and analysis**

1498 Processed scRNA-seq data of human blastocysts<sup>23</sup> was the kind gift of Dr. Tim Coorens.  
1499 The cell clusters were annotated in the original manuscript<sup>23</sup>. For the scRNA-seq data  
1500 of a human gastrula<sup>24</sup>, read data were processed with Trim\_Galore! v0.6.3 for quality  
1501 check and mapped on the human GRCh38.p12 genome with TopHat2 v2.1.1. Mapped  
1502 bam files were converted to FPKM using Cufflinks v2.2.1<sup>97</sup> with the “--compatible-hits-  
1503 norm” option and human GRCh38.p12 transcript reference. Processed scRNA-seq data  
1504 of human embryonic gut<sup>25</sup> were downloaded from CZ CELLxGENE<sup>98</sup>.

1505

#### 1506 **EM-seq library preparation, data processing and analysis**

1507 EM-seq libraries were generated and sequenced on an Illumina NovaSeq 6000 platform  
1508 as described previously<sup>30</sup>. EM-seq data for the F2-AGVT iPSCs were reported  
1509 previously<sup>30</sup> and were used for the subsequent processing steps. Paired-end reads were  
1510 processed with Trim\_Galore! v0.6.3 and mapped on the GRCh38.p12 genome using  
1511 Bismark v0.22.1<sup>99</sup> and Bowtie2 v2.3.4.1 with the “-X 2000” option. BAM files were  
1512 then processed with the deduplicate\_bismark script to remove PCR duplicates, and with  
1513 bismark\_methylation\_extractor script to count methyl- and unmethyl-cytosines per CpG  
1514 or CpA sequences. PBAT data for human sperm, oocytes, blastocysts<sup>71</sup>, human fetal  
1515 germ cells (female and male germ cells at 9 wpf)<sup>3</sup>, and hPGCLC-derived cells in  
1516 xrOvaries (ag120 AG and ag120 AGVT)<sup>19</sup> were processed with Trim\_Galore! and  
1517 Bismark with the “--pbat” option, followed by processing with the  
1518 bismark\_methylation\_extractor script as described previously<sup>19</sup>. Data from biological  
1519 replicates were averaged by following procedures and were used for the presentation

unless otherwise stated. Average CpG levels in binned loci were calculated as the average of % mC in CpG (depth  $\geq 4$ ), and only bins with  $\geq 4$  CpG were used for analysis. Average CpA levels in binned loci were calculated as [sum of mC calls] over [sum of mC and C calls] (sum of mC and C calls  $\geq 10$ ). To obtain the average methylation level of replicates, bins with a read depth greater than three in all replicates were used for the calculation.

1526

## 1527 **Methylome analysis**

### 1528 *Methylation analysis on retrotransposons*

1529 EM-seq reads were mapped on humrep of the Repbase database. Deduplication and  
1530 5mC calculation were performed as described earlier.

1531

### 1532 *Demethylation escapees*

1533 Mapped bam files for M1-*AGDT* c122 AGDT, M1-*AGVT* c82 AGVT, M2 c76, F1-*AGVT*  
1534 c127 AGVT, F2-*AGVT* c68 and ag120 AGVT, hGC Wk9M, and hGC Wk9F were  
1535 processed with MethPipe v3.4.3<sup>100</sup> to calculate hyper-methylated regions with the default  
1536 setting. Overlap of the regions was calculated using the intersect command of BEDTools  
1537<sup>101</sup>. Annotations of the regions were calculated using Homer v4.11.1<sup>102</sup>.

1538

### 1539 *Calculating demethylation rates*

1540 Demethylation rates of bins were calculated by dividing the decrease in the DNA  
1541 methylation level of each bin, which was one of the DNA methylation intervals defined  
1542 as (0, 20], (20, 40], (40, 60], (60, 80], and (80, 100], by the initial DNA methylation level  
1543 of that bin.

1544

## 1545 **Oxford Nanopore Technology (ONT) library preparation, mapping and processing**

1546 Genomic DNA from more than  $6 \times 10^6$  iPSCs was purified by using a Monarch® HMW  
1547 DNA Extraction Kit for Tissue. Library generation was performed using an Ultra-Long  
1548 Sequencing Kit (Oxford Nanopore Technology, SQK-ULK001) according to the  
1549 manufacturer's instructions. The sequencing was done using R9.4.1 flow cells. ONT  
1550 data in fast5 format were processed with megalodon v2.5.0 on the GRCh38.p12 genome  
1551 with the options “--guppy-config dna\_r9.4.1\_450bps\_sup\_prom.cfg”, “--remora-  
1552 modified-bases dna\_r9.4.1\_e8 sup 0.0.0 5mc CG 0”, and “--outputs basecalls mappings  
1553 mod\_mappings mods”. The “mapping.bam” files from two runs were sorted and merged  
1554 using samtools v1.15.1<sup>103</sup>. SNP calling was performed using Clair3 v0.1.12 with the  
1555 model of “r941\_prom\_sup\_g5014” and “--platform=ont” option. The “PASS” marked  
1556 SNPs with genotypes of “1/1” and “1/2” were extracted, and the reference fasta files for  
1557 GRCh38.p12 at these positions were replaced with ALT bases using the “bcftools  
1558 consensus” command of bcftools v1.15.1 to make F1/F2-*AGVT* custom genomic  
1559 reference. Initial fast5 files were reprocessed with megalodon on the GRCh38.p12-  
1560 F1/F2-*AGVT* genome, followed by Clair3 as described above.

1561  
1562 **Haplotype phasing for ONT data and assignment of phasing blocks to Xa and Xi**  
1563 “PASS” marked SNPs with phasing block marks in chromosome X were used for  
1564 haplotype phasing ONT, EM-seq and transcriptome data. The bam files were  
1565 subgrouped into two alleles using the “whatshap haplotag” and “whatshap split”  
1566 commands of Whatshap v1.4. Methylation calls at the CG sequence were calculated  
1567 using modbam2bed v0.5.3 with the “-e -m 5mC --cpg” option. From the careful  
1568 observation of 5mC levels in the promoter, CGI and genome-wide 10kb-binned regions,  
1569 we found that (1) 5mC levels of the promoters and the CGIs in one allele of the  
1570 chromosome Xs tended to be more highly methylated than those in the other allele; and  
1571 (2) the allele with higher 5mC levels in the promoters and the CGIs showed relatively  
1572 lower 5mC levels in the genome-wide 10kb-bin data than the other allele. Therefore,  
1573 among the phasing blocks longer than 500 kb, we assigned Xa as the allele having (1)  
1574 5mC levels in the promoters and the CGIs lower by 50% than those in the other allele;  
1575 and (2) 5mC levels in the genome-wide 10kb-bins higher than those in the other allele for  
1576 the blocks that did not contain any promoter or CGI (Supplementary Table 12).

1577  
1578 **Mapping and calculation of allelic 5mC levels for EM-seq data**  
1579 Human, GRCh38.p12- F1/F2-AGVT custom reference, as described earlier, was masked  
1580 with N at the allelic SNP positions using the “bedtools maskfasta” command of BEDTools.  
1581 EMseq reads from F1/F2-AGVT samples were mapped on this genome as described  
1582 earlier. Bam files were split by allele using SNPsplits v0.3.2 with the “--paired” and “--  
1583 bisulfite” options and masked GRCh38.p12- F1/F2-AGVT custom reference. Resulting  
1584 bam files for two alleles were processed with deduplicate\_bismark and  
1585 bismark\_methylation\_extractor to calculate 5mC levels per CpG or CpA as described  
1586 earlier. Average 5mC levels for both CpG and CpA for allelic 5mC levels within bins  
1587 were calculated as [sum of mC calls] over [sum of mC and C calls] (sum of mC and C  
1588 calls $\geq$ 10) to detect limited signals.

1589  
1590 **Mapping and calculation of allelic 5mC levels for public ONT, EM-seq and PBAT**  
1591 **data**

1592 ONT data in fast5 format for F2-AGVT (R9.4.1 flow cell, Ultra-Long Sequencing Kit)  
1593 were processed for base calling, mapping, SNP calling and haplotype phasing as  
1594 described above. Allelic 5mC levels for the F2-AGVT iPSC (EM-seq) and ag120 AGVT  
1595 cells in xrOvaries [post-bisulfite adaptor tagging sequence (PBAT)] were as described  
1596 above and previously <sup>30</sup>.

1597  
1598 **ChIP-seq data processing**

1599 Public ChIP-seq data <sup>47</sup> were processed as previously described <sup>104</sup>. Briefly, reads were  
1600 processed using Trim Galore v0.4.1/Cutadapt v1.9.1 to remove adaptor sequences. The  
1601 truncated reads were then aligned to (GRCh38p2) using Bowtie2 v2.3.4.1 with the “-

1602 very-sensitive” option. Reads aligned to chromosomes 1 to 22, X, and Y were converted  
1603 to the BAM format by samtools v1.7. BED files were obtained from the BAM files using  
1604 the bamtobed command of BEDTools v2.29.2. BigWig files were generated from the  
1605 BAM files using bamcoverage for raw count with the “--normalizeUsing CPM -bs 25” or  
1606 bamcompare for IP/Input command with the “--pseudocount 1 -bs 1000” option of  
1607 deepTools v3.5.0<sup>105</sup>. In both cases, the blacklist regions<sup>106</sup>  
1608 (<https://www.encodeproject.org/files/ENCF419RSJ>) were excluded. ChIP-seq read  
1609 data for TET1<sup>50</sup> were processed with Trim Galore to remove adapter and low-quality  
1610 bases and mapped onto genome reference GRCh38.p12 with bowtie2. Mapped bam  
1611 files were sorted by coordinates using samtools. Peak calling was performed using  
1612 MACS2 v2.2.7.1 with the default settings<sup>107</sup>. Peaks in the promoter regions were  
1613 annotated to the nearest TSSs by the annotatePeakInBatch function in the ChIPpeakAnno  
1614 package in R.

1615

#### 1616 ATAC-seq data processing

1617 Public ATAC-seq data<sup>47</sup> were processed as previously described<sup>104</sup>. Briefly, adaptor  
1618 sequences were trimmed from the reads using Trim Galore v0.4.1/Cutadapt v1.9.1.  
1619 These reads were aligned using Bowtie2 v2.3.4.1 to GRCm38p2 with the “--very-  
1620 sensitive -X 2000” option. The properly mapped reads with the flag (99, 147, 83 or 163)  
1621 were extracted by awk, and mitochondrial reads were excluded. Duplicated reads were  
1622 removed using the MarkDuplicates command of Picard Tools v2.18.23  
1623 (<https://broadinstitute.github.io/picard/>). These de-duplicated reads were then filtered  
1624 for high quality (MAPQ $\geq$ 30). The reads with an insert size of less than 100 bp were  
1625 extracted as nucleosome free region (NFR) reads. Bed files for downstream analysis  
1626 were generated by the “bamtobed” command of BEDTools v2.29.2 with the “-bedpe”  
1627 option. BigWig files were generated from the BAM files using bamcoverage for raw  
1628 count with the “--normalizeUsing CPM -bs 25” option of deepTools v3.5.0. The  
1629 blacklist regions were excluded as well as ChIP-seq data processing.

1630

1631 Peak calling was performed using MACS v2.1.1 with the “--nomodel --shift -100 --  
1632 extsize 200 --keep-dup all” option after shifting NFR reads with the offset by +4 bp in  
1633 the + strand and by -5 bp in the - strand.

1634

#### 1635 Overlap enrichment analysis

1636 The overlap between genomic regions and annotated intervals was examined using  
1637 Fisher’s exact tests as implemented in the fisher.test() from R library stats version 3.6.1.  
1638 Ensembl Regulatory build annotations v20180516 were sourced directly from Ensembl.  
1639

#### 1640 Epigenome-based clustering of cis-regulatory elements

1641 ATAC-seq peaks identified using MACS2 were pooled across cell types using DiffBind  
1642 v3.8.4<sup>108</sup> to identify consensus peaks with re-centered summits. Twenty 200 bp bins

surrounding peak summits were selected to capture the chromatin state  $\pm$  2 kb from accessible regions, in which the  $\log_2(\text{enrichment over input})$  values of ChIP-seq signals were extracted as input for dimension reduction through UMAP<sup>109</sup> and subsequently clustered through HDBSCAN<sup>110</sup> as implemented in cuML v23.02.00. For UMAP, correlation distances were used together with a grid search over `min_dist` of [0.0, 0.01, 0.1], `n_neighbors` of [15, 30, 50, 70, 100] and `n_components` of 2–10, and the resultant embeddings were all subjected to HDBSCAN clustering to identify epigenetically distinct clusters via visual inspection. For HDBSCAN, a grid search over `min_cluster_size` and `min_samples` over [50, 100, 200, 500, 1000] were tested. In a semi-supervised fashion, individual clusters were isolated and subjected to further sub-clustering until the embedding no longer exhibited distinct segregation of data points for any individual epigenetic signal. We assessed the epigenetic profiles within each cluster and categorized the major clusters (i.e., clusters 1, 2, and 4) as follows: “Active” clusters were characterized by high levels of H3K4me3 and H3K27ac marks. “Bivalent” clusters exhibited high levels of H3K4me3 and H3K27me3 marks. “Poised” clusters were specifically marked by high H3K4me1 levels. Promoters that overlapped with each category of open sites were considered active, bivalent, or poised promoters, while promoters that did not overlap with open sites were considered silent promoters. Active, bivalent, and poised enhancers were defined as the open sites within each cluster that did not overlap with promoters. Open sites that did not fall into any of these categories (active, bivalent, or poised) were classified as silent enhancers. To investigate enhancer-mediated gene regulation, we focused on genes with transcription start sites (TSSs) located within a proximity of 10 kilobases both upstream and downstream of the enhancers.

1667

## 1668 Statistical analysis

1669 Principal component analysis was performed by the `prcomp` function in R.  
1670 Unsupervised hierarchical clustering was performed by the `hclust` function with the  
1671 “method = “ward.D2”” option based on the Euclidean distance calculated by the `Dist`  
1672 function in the `amap` package in R. Welch’s t-test was performed by using the `t.test`  
1673 function with the “paired = TRUE” option in R (Fig. 1d) or the `T.TEST` function in Excel  
1674 (Extended Data Fig. 5g). Tukey-Kramer test (Extended Data Fig. 9k) was performed by  
1675 using the `TukeyHSD` function in R. Two-sided Dunnet’s test (Extended Data Fig. 10l)  
1676 was performed by the `SimTestDiff` function with “type = “Dunnet”” option from the  
1677 `SimComp` package in R. Paired two-sided t-test (Extended Data Fig. 10l) was performed  
1678 by the `t.test` function with the “paired = TRUE” option in R. t-test with Bonferroni  
1679 correction (Fig. 6g and Extended Data Fig. 12d) and Wilcoxon rank-sum test (Fig. 6j)  
1680 were performed by the `geom_signif` function in the `ggsignif` package in R. Cohen’s d-  
1681 values were calculated by using the `cohens_d` function from the `rstatix` package in R  
1682 (Extended Data Fig. 9m).

1683

1684 **Reporting summary**

1685 Further information on research design is available in the Nature Portfolio Reporting  
1686 Summary linked to this article.

1687

1688 **Data and code availability**

1689 The accession number for all the sequencing data generated in this study is GSE232078  
1690 (the GEO database). The R script is available on request. The raw MS data and  
1691 analysis files have been deposited in the ProteomeXchange Consortium  
1692 (<http://proteomecentral.proteomexchange.org>) via the jPOST partner repository<sup>111</sup>  
1693 (<https://jpostdb.org>) and can be accessed using the dataset identifier PXD048118.

1694

1695

1696

1697

1698      **Supplementary References**

- 1699      72      Okita, K. *et al.* An efficient nonviral method to generate integration-free human-  
1700      induced pluripotent stem cells from cord blood and peripheral blood cells. *Stem  
1701      Cells* **31**, 458-466, doi:10.1002/stem.1293 (2013).
- 1702      73      Yokobayashi, S. *et al.* Clonal variation of human induced pluripotent stem cells  
1703      for induction into the germ cell fate. *Biol Reprod* **96**, 1154-1166,  
1704      doi:10.1093/biolre/iox038 (2017).
- 1705      74      Kojima, Y. *et al.* Evolutionarily Distinctive Transcriptional and Signaling  
1706      Programs Drive Human Germ Cell Lineage Specification from Pluripotent Stem  
1707      Cells. *Cell Stem Cell* **21**, 517-532 e515, doi:10.1016/j.stem.2017.09.005 (2017).
- 1708      75      Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification,  
1709      enrichment, pre-fractionation and storage of peptides for proteomics using  
1710      StageTips. *Nature protocols* **2**, 1896-1906, doi:10.1038/nprot.2007.261 (2007).
- 1711      76      Ogata, K. & Ishihama, Y. CoolTip: Low-Temperature Solid-Phase Extraction  
1712      Microcolumn for Capturing Hydrophilic Peptides and Phosphopeptides. *Mol Cell  
1713      Proteomics* **20**, 100170, doi:10.1016/j.mcpro.2021.100170 (2021).
- 1714      77      Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii,  
1715      A. I. MSFragger: ultrafast and comprehensive peptide identification in mass  
1716      spectrometry-based proteomics. *Nat Methods* **14**, 513-520,  
1717      doi:10.1038/nmeth.4256 (2017).
- 1718      78      da Veiga Leprevost, F. *et al.* Philosopher: a versatile toolkit for shotgun  
1719      proteomics data analysis. *Nat Methods* **17**, 869-870, doi:10.1038/s41592-020-  
1720      0912-y (2020).
- 1721      79      Yu, F. *et al.* Fast Quantitative Analysis of timsTOF PASEF Data with MSFragger  
1722      and IonQuant. *Mol Cell Proteomics* **19**, 1575-1585,  
1723      doi:10.1074/mcp.TIR120.002048 (2020).
- 1724      80      Nesvizhskii, A. I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for  
1725      identifying proteins by tandem mass spectrometry. *Anal Chem* **75**, 4646-4658,  
1726      doi:10.1021/ac0341261 (2003).
- 1727      81      Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis.  
1728      *Nat Methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).
- 1729      82      Stirling, D. R. *et al.* CellProfiler 4: improvements in speed, utility and usability.  
1730      *BMC Bioinformatics* **22**, 433, doi:10.1186/s12859-021-04344-9 (2021).
- 1731      83      Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive  
1732      elements in eukaryotic genomes. *Mob DNA* **6**, 11, doi:10.1186/s13100-015-0041-  
1733      9 (2015).
- 1734      84      Illingworth, R. S. *et al.* Orphan CpG islands identify numerous conserved  
1735      promoters in the mammalian genome. *PLoS Genet* **6**, e1001134,  
1736      doi:10.1371/journal.pgen.1001134 (2010).
- 1737      85      Court, F. *et al.* Genome-wide parent-of-origin DNA methylation analysis reveals  
1738      the intricacies of human imprinting and suggests a germline methylation-

- 1739 independent mechanism of establishment. *Genome Res* **24**, 554-569,  
1740 doi:10.1101/gr.164913.113 (2014).
- 1741 86 Ishikura, Y. *et al.* In Vitro Derivation and Propagation of Spermatogonial Stem  
1742 Cell Activity from Mouse Pluripotent Stem Cells. *Cell reports* **17**, 2789-2804,  
1743 doi:10.1016/j.celrep.2016.11.026 (2016).
- 1744 87 Martin, M. Cutadapt removes adapter sequences from high-throughput  
1745 sequencing reads. *EMBnet.journal* **17**, 10 (2011).
- 1746 88 Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of  
1747 insertions, deletions and gene fusions. *Genome Biol* **14**, R36, doi:10.1186/gb-  
1748 2013-14-4-r36 (2013).
- 1749 89 Anders, S., Pyl, P. T. & Huber, W. HTSeq-a Python framework to work with high-  
1750 throughput sequencing data. *Bioinformatics* **31**, 166-169,  
1751 doi:10.1093/bioinformatics/btu638 (2015).
- 1752 90 Okamoto, I. *et al.* The X chromosome dosage compensation program during the  
1753 development of cynomolgus monkeys. *Science* **374**, eabd8887,  
1754 doi:10.1126/science.abd8887 (2021).
- 1755 91 Sherman, B. T. *et al.* DAVID: a web server for functional enrichment analysis and  
1756 functional annotation of gene lists (2021 update). *Nucleic Acids Res* **50**, W216-  
1757 W221, doi:10.1093/nar/gkac194 (2022).
- 1758 92 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-  
1759 3587 e3529, doi:10.1016/j.cell.2021.04.048 (2021).
- 1760 93 Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification  
1761 of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281-291 e289,  
1762 doi:10.1016/j.cels.2018.11.005 (2019).
- 1763 94 Imoto, Y. *et al.* Resolution of the curse of dimensionality in single-cell RNA  
1764 sequencing data analysis. *Life Sci Alliance* **5**, doi:10.26508/lsa.202201591 (2022).
- 1765 95 Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by  
1766 single-cell RNA-seq. *Science* **352**, 189-196, doi:10.1126/science.aad0501 (2016).
- 1767 96 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-  
1768 1902 e1821, doi:10.1016/j.cell.2019.05.031 (2019).
- 1769 97 Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-  
1770 seq experiments with TopHat and Cufflinks. *Nature protocols* **7**, 562-578,  
1771 doi:10.1038/nprot.2012.016 (2012).
- 1772 98 Program, C. S.-C. B. CZ CELL×GENE Discover: A single-cell data platform for  
1773 scalable exploration, analysis and modeling of aggregated data. *bioRxiv*,  
1774 doi:<https://doi.org/10.1101/2023.10.30.563174> (2023).
- 1775 99 Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller  
1776 for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-1572,  
1777 doi:10.1093/bioinformatics/btr167 (2011).
- 1778 100 Song, Q. *et al.* A reference methylome database and analysis pipeline to facilitate  
1779 integrative and comparative epigenomics. *PLoS One* **8**, e81148,

- 1780 101 doi:10.1371/journal.pone.0081148 (2013).
- 1781 101 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing  
1782 genomic features. *Bioinformatics* **26**, 841-842,  
1783 doi:10.1093/bioinformatics/btq033 (2010).
- 1784 102 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors  
1785 prime cis-regulatory elements required for macrophage and B cell identities. *Mol  
1786 Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).
- 1787 103 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics*  
1788 **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
- 1789 104 Nagano, M. *et al.* Nucleome programming is required for the foundation of  
1790 totipotency in mammalian germline development. *EMBO J* **41**, e110600,  
1791 doi:10.15252/embj.2022110600 (2022).
- 1792 105 Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing  
1793 data analysis. *Nucleic Acids Res* **44**, W160-165, doi:10.1093/nar/gkw257 (2016).
- 1794 106 Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist:  
1795 Identification of Problematic Regions of the Genome. *Scientific reports* **9**, 9354,  
1796 doi:10.1038/s41598-019-45839-z (2019).
- 1797 107 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**,  
1798 R137, doi:10.1186/gb-2008-9-9-r137 (2008).
- 1799 108 Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with  
1800 clinical outcome in breast cancer. *Nature* **481**, 389-393, doi:10.1038/nature10730  
1801 (2012).
- 1802 109 McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold  
1803 Approximation and Projection. *Journal of Open Source Software* **3**, 861 (2018).
- 1804 110 Campello, R., Moulavi, D. & Sander, J. in *Advances in Knowledge Discovery and  
1805 Data Mining* 160-172 (Springer, 2013).
- 1806 111 Okuda, S. *et al.* jPOSTrepo: an international standard data repository for  
1807 proteomes. *Nucleic Acids Res* **45**, D1107-D1111, doi:10.1093/nar/gkw1080  
1808 (2017).
- 1809
- 1810
- 1811

1812 **Acknowledgements**

1813 We thank the members of our laboratory for their helpful input on this study, and Y. Nagai,  
1814 N. Konishi, E. Tsutsumi, and M. Kawasaki of the Saitou Laboratory for their technical  
1815 assistance. We are grateful to Spyros Goulas for the critical review of the manuscript  
1816 and Single-Cell Genome Information Analysis Core (SignAC) in ASHBi for the  
1817 RNA/EM/long-read sequence analyses. We thank the XCell Science and Healios for  
1818 the NCLCN hiPSCs. We dedicate this work to T. Mori. This work was supported by  
1819 Grants-in-Aid for Specially Promoted Research from JSPS (17H06098, 22H04920), a  
1820 Grant from HFSP (RGP0057/2018), and Grants from the Open Philanthropy Project  
1821 (2018-193685) to M.S.

1822

1823

1824 **Author contributions**

1825 Y.M. and M.S. conceived the project, and Y.M., R.Y., and M.S. designed the experiments.  
1826 Y.M. and R.Y. conducted BMP-driven hPGCLC differentiation and analyzed the data.  
1827 M.M., A.K., P.P., C.Y., K.M., K.O., and Y.I. assisted with the experiments, and Y.Y., Y.K.,  
1828 B.H., and M.N. assisted with the data analysis. Y.M., R.Y., Y.Y., Y.K., B.H., M.N., and  
1829 M.S. wrote the manuscript.

1830

1831

1832 **Competing interests**

1833 M.S., Y.M. and R.Y. are inventors on patent applications relating to induction of germ  
1834 cells from pluripotent stem cells filed by Kyoto University.

1835

1836

1837

1838

1839 **Extended Data Figure Legends**

1840 **Extended Data Fig. 1. Stable expansion of hPGCLCs.**

1841 **a**, Scheme for hPGCLC expansion culture (left)<sup>17</sup> and flow cytometric plot for BTAG  
1842 expression of the hPGCLC culture and for forward and side scatter (FSC and SSC) of the  
1843 non-BTAG cells (middle). The P1 cells in the middle panel are TRA-1-85<sup>+</sup>, i.e., de-  
1844 differentiated hPGCLC-derived cells, whereas a majority of the P2 cells are TRA-1-85<sup>-</sup>,  
1845 i.e., m220 feeders (right). Accordingly, the enrichment score is defined as log<sub>2</sub> (the  
1846 number of BT<sup>+</sup>AG<sup>+</sup> cells/the number of cells in the P1 gate) (right). hPGCLCs were  
1847 cultured as in<sup>17</sup>. See Fig. 1a for the summary of acronyms used in this study.

1848 **b-d**, hPGCLC expansion and the enrichment score of the hPGCLC culture with IWR1,  
1849 A83-01, and LDN193189 at culture day (c) 10 and 20 (**b**), with different doses of IWR1  
1850 at c10, 20, and 30 (**c**), and with different basal media (**d**). The passages were performed  
1851 using flow cytometry. The color coding is as indicated. hPGCLCs were cultured as in  
1852<sup>17</sup> with or without indicated chemicals. The numbers of experiments were as follows;  
1853 n=1 for (**b**) and (**d**), n=3 for (**c**).

1854 **e**, hPGCLC expansion and the enrichment score of the hPGCLC culture with IWR1 (1.5  
1855 μm) in DMEM or advanced RPMI at c12 and 22 (top), and FACS plots for BTAG  
1856 expression and FSC/SSC of the non-BTAG cells of the hPGCLC culture with IWR1 (1.5  
1857 μm) in DMEM or advanced RPMI at c22 (bottom). The passages were performed with  
1858 dilution. The color coding is as indicated. Note that there were nearly no de-  
1859 differentiated cells in the P1 gate in the culture with advanced RPMI.

1860 **f**, Principal component analysis (PCA) of transcriptomes of key cell types during  
1861 hPGCLC induction and hPGCLC differentiation in xrOvaries<sup>19</sup> (top) and the  
1862 identification of genes making significant contributions [radius of standard deviations  
1863 (SDs) ≥ 3] to scaled PC1 and PC2 loadings (bottom). Genes expressed in at least one  
1864 sample [ $\log_2(\text{RPM}+1) \geq 4$ ] were used for PCA.

1865 **g**, (left) Unsupervised hierarchical clustering (UHC) of the genes selected in (**f**) based on  
1866 their expression dynamics, and (right) promoter methylation dynamics of the genes in the  
1867 five clusters in (left) during hPGCLC induction and hPGCLC differentiation in xrOvaries  
1868<sup>19</sup>. Among the cluster 2 genes, those showing promoter 5mC-level reduction from  
1869 hiPSCs to oogonia-like cells by ≥ 50% are defined as epigenetic reprogramming-activated  
1870 genes (ER genes).

1871

1872 **Extended Data Fig. 2. Exploration of the signaling for epigenetic reprogramming**  
1873 **and hPGCLC differentiation.**

1874 **a**, Expression of key lineage markers and BMP ligands in single cells of cultured human  
1875 embryos (~E11)<sup>23</sup> visualized by Uniform manifold approximation and projection  
1876 (UMAP) and Louvain clustering. Color coding is as indicated. STB:  
1877 syncytiotrophoblast; CTB: cytotrophoblast; Epi: epiblast; Hyp: hypoblast.

1878 **b**, Expression of BMP2 in various cell types in a gastrulating human embryo at ~E16<sup>24</sup>.

1879 **c, d**, Unsupervised hierarchical clustering (UHC) (**c**) and cell-type annotation (**c, d**) based

on key marker expression of endoderm cells in a gastrulating human embryo at ~E16 in (b)<sup>24</sup> and expression of BMP ligands in each cell type (d).  
e, Expression of BMP ligands in cells composing human embryonic gut at week 6.1<sup>25</sup>. Note that BMP ligands are expressed at a high level in colonic (i.e., hindgut) epithelium and mucosal mesoderm. FLC: fibroblasts; SMC: smooth muscle cells.  
f, g, Representative FACS plots for BTAG expression (f, top) and for FSC/SSC fluorescence of the BT<sup>-</sup>AG<sup>-</sup> cells (f, bottom), and hPGCLC fold-change (g, left) and the enrichment scores (g, right) of the hPGCLC culture with various concentrations of BMP2 with IWR1 (1.5 μM) in advanced RPMI at c22 (The numbers of experiments were as follows: n = 2 for BMP2 5 ng/mL; n=3 for BMP2 0, 10–200 ng/mL). The passages were performed with dilution. Note that there were nearly no de-differentiated cells in the P1 gate under all conditions.  
h, Immunofluorescence (IF) analysis of the expression of GFP (*TFAP2C-EGFP*: AG), tdTomato (*BLIMPI-tdTomato*: BT), and DAZL in hPGCLC-derived cells cultured without (top) or with (bottom) BMP2 (25 ng/ml) at c55. ~19% (5/26) of BT<sup>+</sup>AG<sup>+</sup> cells were DAZL<sup>+</sup> in the culture with BMP2, whereas no DAZL<sup>+</sup> cells were found in the culture without BMP2. Bar, 50 μm.

### Extended Data Fig. 3. Generation of fluorescent reporters for hPGCLC differentiation.

a, (top) Schematic illustrations of the human *TFAP2C* locus with knock-in of the 2A-*EGFP* and *PGK-Puro* cassette<sup>16</sup> and the same locus with the excision of *PGK-Puro* by Cre-recombinase. (bottom) Schematic illustrations of the human *DAZL* locus, the *DAZL* targeting vector for knocking in the 2A-*tdTomato* and *PGK-Neo* cassette, the knocked-in locus, and the knocked-in locus with the excision of *PGK-Neo* by Cre-recombinase. Positions of the primer pairs for the screening by PCR of the genotypes are shown. Black boxes indicate the exons.  
b, Screening by PCR of the targeted alleles for *DAZL-2A-tdTomato* (DT) and *TFAP2C-2A-EGFP* (AG), and of random integration of the targeting vectors. Targeted: bands for the targeted allele; wild-type: bands for the wild-type allele; arrowheads: random integration of the targeting vectors. The 585B1-AGDT #7453 line (M1-AGDT) was selected for subsequent experiments.  
c, (top) Schematic illustrations of the human *TFAP2C* locus, the *TFAP2C*-targeting vector for knocking in the 2A-*EGFP* and *PGK-Puro* cassette, the knocked-in locus, and the knocked-in locus with the excision of *PGK-Puro* by Cre-recombinase. (bottom) Schematic illustrations of the human *DDX4* (human VASA homolog) locus, the *DDX4* targeting vector for knocking in the 2A-*tdTomato* and *PGK-Neo* cassette, the knocked-in locus, and the knocked-in locus with the excision of *PGK-Puro* by Cre-recombinase. Positions of the primer pairs for the screening by PCR of the genotypes are shown. Black boxes indicate the exons.  
d, e, Screening by PCR of the targeted alleles for *TFAP2C-2A-EGFP* (AG) and *DDX4*

(human VASA homolog)-*2A-tdTomato* (*VT*), and of random integration of the targeting vectors. Targeted: bands for the targeted allele; wild-type: bands for the wild-type allele; arrowheads: random integration of the targeting vectors. The 585B1-AGVT #1375 line (*M1-AGVT*) (**d**) and the NCLCN-AGVT #26-1 line (*F1-AGVT*) (**e**) were selected for subsequent experiments.

**f**, Representative result for the G-band analysis of *M1-AGDT*, *M1-AGVT*, and *F1-AGVT* bearing normal karyotypes (46, XY or 46, XX).

**g**, Bright-field and fluorescence [AG and DT or VT] images and flow cytometric plots for AGDT or AGVT expression of the iMeLC aggregates induced for hPGCLCs for 6 days from the *M1-AGDT* (left), *M1-AGVT* (middle), and *F1-AGVT* (right) lines. Bar, 200  $\mu$ m.

#### Extended Data Fig. 4. BMP signaling promotes hPGCLC differentiation.

**a, b**, Growth curve (**a**) and proportion of cells with the indicated fluorescence-marker expression at c42 and c52 (**b**) during BMP-driven *M1-AGDT* hPGCLC differentiation with varying concentrations of BMP2 as indicated.

**c, d**, Flow cytometric plots for AGDT expression at the indicated culture days (**c**) and growth curve (**d**) of *M1-AGDT* hPGCLC-derived cells cultured without BMP2.

**e, f**, Flow cytometric plots for AGVT expression (**e**) and proportion of cells with the indicated fluorescence-marker expression (**f**) at the indicated culture days during BMP-driven *M1-AGVT* hPGCLC differentiation.

**g–i**, Flow cytometric plots for EpCAM and ITGA6 expression (**g**), IF analysis of TFAP2C and DDX4 expression (**h**), and proportion of DDX4 $^{+}$  cells (**i**) at the indicated culture days during BMP-driven *M2* hPGCLC differentiation. In (**i**), the numbers of experiments (N) and of cells analyzed (n), and typical images for the positivity of DDX4 staining are shown. Bar, 200  $\mu$ m.

**j, k**, Growth curve and enrichment scores (**j**) and flow cytometric plots for AGVT expression at c43 (**k**) during BMP-driven *F1-AGVT* hPGCLC differentiation with 25 ng/ml or 50 ng/ml of BMP2.

**l, m**, Flow cytometric plots for AGVT expression at the indicated culture days (**l**) and growth curve (**m**) of *F1-AGVT* hPGCLC-derived cells cultured without BMP2.

**n, o**, Flow cytometric plots for AGVT expression (**n**) and proportion of cells with the indicated fluorescence-marker expression (**o**) at the indicated culture days during BMP-driven *F2-AGVT* hPGCLC differentiation.

**p**, Relief contrast and fluorescence (VT and AG) images of *F1-AGVT* hPGCLC-derived cells frozen at c64 and thawed and cultured for an additional 24 days. Bar, 200  $\mu$ m.

**q, r**, Growth curve (**q**) and proportion of cells with the indicated fluorescence-marker expression (**r**) during BMP-driven *M1-AGDT* hPGCLC differentiation with or without FBS.

#### Extended Data Fig. 5. Identification of distinctive transcriptional processes driven

1962 by BMP signaling.

1963 **a**, Heatmap showing the expression levels of 13 previously reported genes that show up-  
1964 regulation in gonadal germ cells (*DAZL* and *DDX4* are excluded)<sup>26</sup> (top), and the unique  
1965 genes on the Y chromosome (bottom) in the indicated cell types (see Supplementary Table  
1966 2 for full sample information). Color coding is as indicated. NA: not applicable.

1967 **b**, Expression dynamics of *CDH5* and *DMRT1*, the genes used as markers for human  
1968 germ cells from the migration stage onward<sup>27</sup>, during hPGCLC induction and BMP-  
1969 driven hPGCLC differentiation. The average (bar) and replicate (circles) values are  
1970 shown (see Supplementary Table 2 for full sample information). The data for iPSCs,  
1971 iMeLCs were with M1-*BTAG*, and the data for hPGCLCs were with the M1-*BTAG*, M1-  
1972 *AGDT*, and F1-*AGVT* lines. Color coding is as indicated.

1973 **c**, PC1–PC3 plane of the PCA of transcriptomes during hPGCLC induction and BMP-  
1974 driven and xrOvary-based hPGCLC differentiation in Fig. 3b (left, top), and the GO  
1975 enrichments with *p* values of genes contributing to the negative [standard deviation (SD)  
1976 < -2: BMP-up-regulated genes] and positive [SD > 2: xrOvary-up-regulated genes]  
1977 scores of PC3 (left, bottom; right). Color coding is as indicated.

1978 **d**, PCA of M1-*AGDT* hPGCLC-derived AD<sup>+</sup>DT<sup>-</sup> cells cultured with or without BMP2.  
1979 The color coding is as indicated. Genes expressed in at least one sample [ $\log_2(\text{RPM}+1)$   
1980  $\geq 4$ ] were used for PCA.

1981 **e**, UHC of highly variable genes (top 1,000 genes with high coefficient of variance) in  
1982 (**d**) based on their expression dynamics.

1983 **f**, Box plots of the expression dynamics of the 7 gene clusters in (**e**) during hPGCLC  
1984 culture with or without BMP2. The 7 gene clusters are classified into those showing  
1985 progressive up- (clusters 4, 5, 7) or down- (clusters 1, 2, 3, 6) regulation during BMP-  
1986 driven hPGCLC differentiation, and the numbers of the genes in each cluster are shown.

1987 **g**, Gene ontology (GO) enrichments and representative genes in up- (clusters 4, 5, 7) (left)  
1988 and down- (clusters 1, 2, 3, 6) (right) regulated genes. The color coding is as indicated.

1989 **h**, Expression dynamics of *DUSP4* and *DUSP6* (GO:0070373~negative regulation of  
1990 ERK1 and ERK2 cascade), and *INSR* and *SHC2* (GO:0043410~positive regulation of  
1991 MAPK cascade) during hPGCLC induction and BMP-driven hPGCLC differentiation.  
1992 The average (bar) and replicate (circles) values are shown (see Supplementary Table 2  
1993 for full sample information). The data for iPSCs, iMeLCs were with M1-*BTAG*, and the  
1994 data for hPGCLCs were with the M1-*BTAG*, M1-*AGDT*, and F1-*AGVT* lines.

1995 **i**, Western blot analysis of the levels of phosphorylated or total ERK1 and 2 in M1-*AGDT*  
1996 hPGCLC-derived cells at c33 cultured with or without BMP2. Three independent  
1997 cultures were analyzed for two biological replicates.  $\alpha$ TUBLIN were used for the  
1998 loading control. For the gel source data, see Supplementary Figure 2. pERK:  
1999 phosphorylated ERK.

2000 **j**, Quantification of pERK1 and 2 levels normalized by  $\alpha$ TUBLIN in M1-*AGDT*  
2001 hPGCLC-derived cells at c33 cultured with or without BMP2 in (**h**). The average fold-  
2002 differences of the Western blot signals for pERK1 and pERK2 were ~4.5-fold and ~2.9-

2003 fold (Expt. 1) and ~4.3-fold and ~3.9-fold (Expt. 2), respectively. *p* values with Welch's  
2004 *t*-test are shown. Data from two independent experiments with three biological  
2005 replicates were shown in (i) and (j).

2006

2007 **Extended Data Fig. 6. scRNA-seq analysis of BMP-driven female hPGCLC**  
2008 **differentiation.**

2009 **a**, Heatmap showing the expression levels of key genes in oogonia/fetal oocytes *in vivo*  
2010 (left) and F1-AGVT hPGCLC-derived cells *in vitro* (right) classified into 10 clusters in  
2011 Fig. 3c. The actual expression levels [ $\log_2(\text{normalized read counts}+1)$ ] are provided in  
2012 Source Data Extended Data Fig. 6. The color coding is as indicated.

2013 **b, c**, Proportion of the 10 clusters in Fig. 3c in the indicated samples (**b**) and of the  
2014 indicated samples in each cluster (**c**). The actual percentages of major clusters  
2015 (**b**)/culture days/weeks post-fertilization (wpf) (**c**) are shown within the histogram. The  
2016 full information is provided in Source Data Extended Data Fig. 6. The color coding is  
2017 as indicated.

2018 **d–f**, The numbers of differentially expressed genes (DEGs) between *in vivo* and *in vitro*  
2019 cell types in the EM, M, and PLL clusters (**d**), volcano plots for the comparisons in the  
2020 M and PLL clusters (**e**), and the GO enrichments with *p* values of DEGs in the M and  
2021 PLL clusters (**f**).

2022 **g**, Heatmap showing the expression levels of PLL1 (left) or PLL2 (right) signature genes  
2023 (top 50 genes highly expressed in PLL1 or 2 relative to all other clusters) in the indicated  
2024 samples. The color coding is as indicated.

2025 **h**, GO enrichments with *p* values of DEGs between PLL1 *in vivo* and *in vitro* cells (top)  
2026 and between PLL1 *in vivo* and PLL2 *in vitro* cells (bottom).

2027

2028 **Extended Data Fig. 7. DNA methylome reprogramming during BMP-driven**  
2029 **hPGCLC differentiation.**

2030 **a**, Scatter-plot comparisons (contour representation) of the 5mC levels (genome-wide 2-  
2031 kb bins), combined with histogram representation (top and right of the scatter plots),  
2032 between the indicated cell types. Note that genome-wide 5mC profiles of F1 and F2-  
2033 AGVT hiPSCs measured by EM-seq are highly similar to those of F2-AGVT hiPSCs  
2034 measured by whole genome bisulfite sequence (WGBS)<sup>30</sup>.

2035 **b**, Heatmap of the 5mC [CpG (top) or CpA (bottom)] levels on chromosome 1 (left) and  
2036 chromosome X (right) in the indicated samples. For chromosome X (right), data were  
2037 generated using the reads overlapping with allelic SNPs. mb, megabases. The color  
2038 coding is as indicated. N.D.: bins without enough CpGs (4) with read depth  $\geq 4$  in CpG  
2039 or bins without enough mC + C calls ( $\geq 10$ ) in CpA.

2040 **c**, PCA of the indicated samples using the 5mC levels on the autosome-wide (left) or Xa-  
2041 and Xi-wide (right) 2-kb bins (top) and promoters (bottom). The color coding is as  
2042 indicated.

2043

2044 **Extended Data Fig. 8. DNA methylome reprogramming and identification of core**  
2045 **ER genes during BMP-driven hPGCLC differentiation.**

2046 **a**, Violin plots of the promoter 5mC-level dynamics of 13 previously reported genes that  
2047 show up-regulation in gonadal germ cells<sup>26</sup> (left) and genes included in the GO term  
2048 “meiotic cell cycle” (GO: 0051321) in the indicated cells during BMP-driven hPGCLC  
2049 differentiation and in *in vivo* germ cells<sup>3</sup>. All relevant promoters are classified into  
2050 H/I/LCP (high/intermediate/low CpG promoter) and plotted.

2051 **b**, Violin plots of the average 5mC levels on the indicated repeat elements in the indicated  
2052 cell types (see Supplementary Table 2 for full sample information). Bars represent the  
2053 average values. The DNA methylome data for human spermatozoa, oocytes, and  
2054 blastocysts are from<sup>71</sup> and those for human male and female germ cells at 9 wpc are from  
2055<sup>3</sup>.

2056 **c**, Genome coverage (%) by EM-seq with paired-end sequencing (this study), EM-seq  
2057 with computationally simulated single-end sequencing, whole genome bisulfite sequence  
2058 (WGBS) with single-end 101 bp sequencing<sup>19</sup>, and WGBS with single-end 50 bp  
2059 sequencing<sup>3</sup>.

2060 **d**, Annotation of differentially covered regions between paired-end and single-end  
2061 sequencing in **(c)**. Color coding is as indicated.

2062 **e**, (left) Violin plots of the average 5mC levels (genome-wide 2 kb bins) in the indicated  
2063 cell types. Bars represent the average values. (right) Scatter-plot comparisons  
2064 (contour representation) of the 5mC levels (genome-wide 2-kb bins), combined with  
2065 histogram representation (top and right of the scatter plots), between the indicated cell  
2066 types. Note that DAZL<sup>+</sup> PGCLCs by Irie et al.<sup>27</sup> are highly methylated (~76%) and that  
2067 hPGCLCs by von Meyenn et al.<sup>31</sup> and cultured hPGCLCs by Kobayashi et al.<sup>32</sup> remain  
2068 methylated (the average 5mC levels of 57.9% and 61.4%, respectively) and show a  
2069 methylome similar to that in M1-AGDT hPGCLC-derived cells at c32 cultured without  
2070 BMP2 (AG B-).

2071 **f**, PCA of transcriptomes of key cell types during hPGCLC induction and BMP-driven  
2072 hPGCLC differentiation (top) and the identification of the genes with significant  
2073 contributions [radius of standard deviations (SDs)  $\geq 3$ ] to scaled PC1 and PC2 loadings  
2074 (bottom). Genes expressed in at least one sample [ $\log_2(\text{RPM}+1) \geq 4$ ] were used for PCA.

2075 **g, h**, UHC of the genes selected in **(f)** based on their expression dynamics **(g)**, and  
2076 promoter methylation dynamics of the genes in the five clusters in **(g)** **(h)** during hPGCLC  
2077 induction and BMP-driven hPGCLC differentiation. Among the cluster 3 genes, those  
2078 showing promoter 5mC-level reduction from hiPSCs to oogonia-like cells by  $\geq 50\%$  are  
2079 defined as epigenetic reprogramming-activated genes (ER genes), which are classified  
2080 into early and late ER genes based on their expression dynamics.

2081 **i**, Venn diagram showing the overlap of ER genes defined for xrOvaried-based (Fig. 1b,  
2082 and Extended Data Fig. 1f, g) and BMP-driven (Extended Data Fig. 8f–h) hPGCLC  
2083 differentiation.

2084

2085 **Extended Data Fig. 9. ER gene regulation and XCR.**  
2086 **a**, Expression dynamics of core ER genes (early: yellow; late: red) (Extended Data Fig.  
2087 8f–i) during BMP-driven M1-AGDT (top) and F1-AGVT (bottom) hPGCLC  
2088 differentiation.  
2089 **b**, Box plots showing the expression of ER genes in *in vitro* and *in vivo* EM, M, and PLL  
2090 cells in Fig. 3c.  
2091 **c**, 5mC-level tracks of *DAZL* (top) and *DDX4* (bottom) loci in the indicated cell types.  
2092 Green bars represent the promoters [+400 bp and –900 bp of the transcription start sites  
2093 (TSSs)], and their 5mC levels are indicated.  
2094 **d**, Scatter-plot representations of the relationship between promoter-5mC-level  
2095 differences and expression-level differences for early (yellow, left) and late (red, right)  
2096 ER genes between c82 AG<sup>+</sup>DT<sup>−</sup> and DT<sup>+</sup> cells (top) and between c72 AG<sup>+</sup>VT<sup>−</sup> and VT<sup>+</sup>  
2097 cells (bottom). Regression lines are indicated.  
2098 **e**, Heatmap of the promoter 5mC (left) and expression (right) level dynamics of the X-  
2099 linked genes during BMP-driven F1-AGVT hPGCLC differentiation. The Xa and Xi  
2100 allelic data were generated using the reads overlapping allelic SNPs. The genes were  
2101 classified according to their promoter 5mC levels on the Xa and Xi alleles in hiPSCs:  
2102 class 1 genes with high ( $\geq 50\%$ ) 5mC on both Xa and Xi (16 genes), class 2 genes with  
2103 low ( $< 50\%$ ) 5mC on Xa and high 5mC on Xi (40 genes), a class 3 gene with high 5mC  
2104 on Xa and low 5mC on Xi (*XIST*), and class 4 genes with low 5mC on both Xa and Xi (3  
2105 genes) (Supplementary Table 7). The color coding is as indicated. N.D., promoters  
2106 with insufficient read depths. Note that there were no informative single nucleotide  
2107 polymorphisms (SNPs) that discriminate *XIST* expression from parental alleles with the  
2108 3-prime RNA-seq analysis<sup>34</sup>.  
2109 **f**, Expression dynamics from the Xa and Xi alleles during BMP-driven F1-AGVT  
2110 hPGCLC differentiation. (left) Proportions of the expression from the Xa allele in the  
2111 three gene classes in (e) are plotted, with individual values plotted as diamonds and their  
2112 averages shown as colored lines. The distributions of the Xa ratio of all genes are shown  
2113 as violin plots. Data points at 100% are dispersed within the range of 5% for better  
2114 visualization. Raw data are available in (Supplementary Table 7). (right) Proportions  
2115 of the expression from the Xa allele of the class 1 and 2 genes are box-plotted, with genes  
2116 retaining high ( $\geq 50\%$ ) and low ( $\leq 25\%$ ) 5mC levels in c117/118 AG<sup>+</sup>VT<sup>+</sup> cells colored  
2117 blue and green, respectively.  
2118 **g**, Xa allele usage of genes expressed similarly from Xa and Xi in VEM cells at c11 (%  
2119 Xa usage < 90%; 8 genes) (top) or those expressed predominantly from Xa in VEM cells  
2120 at c11 (% Xa usage  $\geq 90\%$ ; 34 genes) (bottom) in the indicated cell types. Xa: active X  
2121 chromosome. VEM, M, and PLL are defined in Fig. 3c.  
2122 **h**, Dynamics of the X chromosome:autosome ratio (X:A ratio) of gene-expression levels  
2123 (top) and *XIST* expression (bottom) during BMP-driven M1-AGDT (left) and F1-AGVT  
2124 (right) hPGCLC differentiation, based on the bulk RNA-seq data. The ratios of the  
2125 75%-tile expression values of the genes from the chromosome X or chromosome 10,

relative to those of all genes are plotted in the log<sub>2</sub> (left) or linear (right) scale.

**i**, Absolute expression-level fold-changes of *UHRF1*, *DNMT3A*, and *DNMT3B* during BMP-driven hPGCLC specification and differentiation. The data in Fig. 3a are used and the value in one replicate in hiPSCs is set as one. The numbers of biological replicates are: n = 3 for hiPSCs; n = 2 for other cells. The red circles present the average.

**j**, Violin plots for the methylated CpA levels (genome-wide 10-kb bin) in the indicated cell types. Effect sizes (Cohen's d-values) are shown. \*: 0.2 < d < 0.5; \*\*: 0.5 < d < 0.8. The d-values for comparisons with hiPSCs and hPGCLCs were > 0.8, except for the comparison with c32 AG BMP(−). The d-values for comparisons among BMP-driven hPGCLC differentiation samples were < 0.2.

**k**, IF analysis of the expression and subcellular localization of UHRF1 co-stained with GFP (*TFAP2C-EGFP*: AG) and DAPI in the indicated cell types (left, top), and normalized UHRF1 signal intensities across the nucleus and cytoplasm (magenta lines) of randomly chosen 10 cells (left, bottom) and their curve fitting representation by Generalized additive model with 95% confidence intervals highlighted in gray (right, top). The outlines of the nucleus (nuc) and cytoplasm (cyto) were determined based on the visual inspection of the DAPI and AG signals (dotted lines), respectively. In (left, top), AG appeared to be enriched in the nucleus, but the reason was unclear. (right, bottom) Quantification of the nuclear/cytoplasmic ratio of UHRF1 by an automated image analysis was shown in. The numbers of cells measured in each sample (n) were indicated. \*: Tukey-Kramer test,  $p < 5 \times 10^{-6}$ . ns: no statistical significance. Bar, 10 μm.

**l**, Doubling times, 5mC demethylation levels, and 5mC demethylation rates per cell division in the indicated culture periods.

**m**, 5mC demethylation ratios of genomic bins bearing different 5mC levels in the originated cell types during the indicated cell-type transitions. Pie charts indicate the proportion of each bin in the originated cell types. The color coding is as indicated.

**n**, A pie chart showing overlap of the bins bearing ≥ 80% 5mC levels in c32 cells cultured with BMP2 with the DNA demethylation escapees in human germ cells *in vivo*.

**Extended Data Fig. 10. Generation of *TET1* knockout hiPSCs and analysis of BMP-driven *TET1*-knockout hPGCLC differentiation.**

**a**, Scheme of the human *TET1* locus, with the illustration of PAM (protospacer adjacent motif) and guide RNA sequences in the exon 6. Black boxes indicate the exons.

**b**, Sequences of the targeted loci in two *TET1* knockout (KO) cell lines [*TET1* KO#1 and # 2 (M1-BTAG *TET1*<sup>−/−</sup> #142 and #2725)].

**c**, Dot-blot analysis of genomic 5hmC levels in wild-type and *TET1* KO hiPSCs.

**d**, Karyotype of *TET1* KO#1 and #2 hiPSCs (top: chromosome spreads; bottom: percentage of cells with 46 or other chromosome numbers). Bar, 10 μm.

**e**, Mass spectrometric analysis [log<sub>2</sub>(signal intensities)] for *TET1* and its truncated protein potentially derived from the *TET1* KO allele in wild-type and *TET1* KO cells. Peptides

2167 from the full-length (top), but not the truncated (bottom), TET1 were detected from the  
2168 wild-type cells in two independent experiments (red and blue bars), whereas neither form  
2169 was detected from the *TET1* KO cells.

2170 **f**, Induction of hPGCLCs from wild-type (M1-BTAG) and *TET1* KO#1 and #2 hiPSCs.  
2171 Photomicrographs of hiPSCs and iMeLC aggregates induced for hPGCLCs for 6 days  
2172 (bright-field and fluorescence images for AG and BT) (left), their flow cytometric plots  
2173 for AG and BT expression (middle), and percentages of BT<sup>+</sup>AG<sup>+</sup> cells (4 independent  
2174 experiments) (right) from each genotype are shown. Bar, 500 μm.

2175 **g**, Growth curves of BT<sup>+</sup>AG<sup>+</sup> cells and enrichment scores during BMP-driven (~c12: 25  
2176 ng/ml; c12~: 100 ng/ml) wild-type and *TET1* KO#1 and #2 hPGCLC differentiation.  
2177 The numbers of experiments were as follows; n = 5 for c12–c32, n = 2 for c42. The  
2178 color coding is as indicated.

2179 **h**, UHC of the transcriptomes during hPGCLC induction and BMP-driven hPGCLC  
2180 differentiation from wild-type and *TET1* KO hiPSCs, with the expression levels of key  
2181 genes indicated. The color coding is as indicated.

2182 **i–k**, The numbers of the differentially expressed genes (DEGs) [ $\log_2(\text{RPM}+1) \geq 3$ , fold  
2183 change  $\geq 2$ ] between wild-type and *TET1* KO cells (up- or down-regulated in *TET1* KO  
2184 cells) (**i**), UHC of the DEGs (**j**), and the GO enrichments and representative genes in the  
2185 indicated DEG clusters (**k**). DEGs were defined using average expression values of  
2186 biological replicates. The DEG numbers were unions of two comparisons (i.e., wild-  
2187 type vs KO#1 and vs KO#2). Core ER genes were highlighted in red in (**j**).

2188 **l**, Box plots for the expression dynamics of ER genes (n = 42) during hPGCLC induction  
2189 and BMP-driven hPGCLC differentiation from wild-type and *TET1* KO hiPSCs. *p*-  
2190 values of Two-sided Dunnet's test (except c42) or paired two-sided t-test (c42) were  
2191 shown.

2192

2193 **Extended Data Fig. 11. TET1 protects hPGCLCs from differentiation into**  
2194 **extraembryonic cells.**

2195 **a, b**, Proportion of wild-type and *TET1* KO cells (**a**) and cell-cycle phases (**b**) in the 11  
2196 clusters in Fig. 5b. The actual proportions of major clusters (**a**)/cell-cycle phases (**b**) are  
2197 shown within the histogram. The full information is provided in Source Data Extended  
2198 Data Fig. 11. The color coding is as indicated.

2199 **c**, Correlations among clusters in Fig. 5b based on expression of the top 2,000 highly  
2200 variable genes (HVGs). Spearman's rank correlation coefficient was calculated for  
2201 analysis.

2202 **d**, Partition-based graph abstraction (PAGA) analysis of the relationships of the clusters  
2203 in Fig. 5b.

2204 **e–g**, UMAP plots and Louvain clustering of scRNA-seq data of a PSC-based model of  
2205 early human post-implantation development<sup>45</sup> (**e**), the expression of key lineage markers  
2206 in the 7 clusters in (**e**) (**f**), and the annotation of the 7 clusters based on their gene  
2207 expression (**g**). AMLC: amnion-like cells; MeLC: mesoderm-like cells. The color

2208 coding is as indicated.

2209 **h**, Prediction of the cell types of the clusters in Fig. 5b using the prediction tool by Zhao  
2210 et al<sup>46</sup>. The color coding is as indicated. PriS: primitive streak; ExE\_Mes: extra-  
2211 embryonic mesoderm; Epi: epiblast; PriS\_Amnion; primitive streak\_amnion; Mes:  
2212 mesoderm; TE: trophectoderm; NoSigHts: no significant hits.

2213

2214 **Extended Data Fig. 12. De-repression of bivalent genes and hypermethylation of**  
2215 **regulatory elements in *TET1* deficient cells.**

2216 **a**, Odds ratio and *q*-value of the enrichment of the 2-kb bins with higher 5mC levels in  
2217 *TET1* KO cells compared to wild-type cells at c42 in the ENCODE genomic annotation  
2218 categories.

2219 **b**, Violin plots for the 5mC-level differences (%) in the promoters and enhancers between  
2220 wild-type and *TET1* KO hPGCLC-derived cells at c42. Promoters, enhancers (non-  
2221 promoter open sites), and their labels are based on the data for d4 hPGCLCs<sup>47</sup>. Silent  
2222 promoters: promoters that did not overlap with open sites; silent enhancers: enhancers  
2223 categorized neither into active, bivalent, nor poised.

2224 **c**, Violin plots for the 5mC-level differences (%) in the indicated elements between wild-  
2225 type and *TET1* KO hPGCLC-derived cells at c42.

2226 **d**, Violin plots for the expression-level differences ( $\log_2$  fold-change) of genes bearing  
2227 the indicated promoters (top) and enhancers (bottom) between wild-type and *TET1* KO  
2228 hPGCLC-derived cells at c42. \*\*\*:  $p < 0.001$  (*t*-test adjusted by Bonferroni correction).

2229 **e**, Odds ratio of the enrichment of genes with indicated promoters defined in d4 hPGCLCs  
2230<sup>47</sup> and ER genes (for down-regulated genes) in genes up- (left) or down- (right) regulated  
2231 in *TET1* KO hPGCLC-derived cells at c42. Number of each gene class is indicated.

2232 **f**, Odds ratio of the c42 up-regulated genes bound by TET in hESCs<sup>50</sup> in each category  
2233 of promoters. The odds ratio was calculated relative to the background ratio of all genes  
2234 bound by TET in each respective promoter category. Number of each gene class is  
2235 indicated.

2236 **g**, A summery scheme of the present work.

2237

2238

2239      **Supplementary Tables**, see the separate Excel documents  
2240      **Supplementary Table 1.** Expression matrix of bulk RNA-seq.  
2241  
2242      **Supplementary Table 2.** Overview of bulk RNA-seq, scRNA-seq, EM-seq, ONT long-read-  
2243      seq samples used in this study.  
2244  
2245      **Supplementary Table 3.** List of HVGs in AG<sup>+</sup>DT<sup>-</sup> cells cultured with or without BMP2, of  
2246      DEGs between BMP-driven and xrOvary-based hPGCLC differentiation, and associated  
2247      GO terms.  
2248  
2249      **Supplementary Table 4.** Lists of DEGs between germ cells *in vivo* and *in vitro* in EM, M or  
2250      PLL stages, and associated GO terms.  
2251  
2252      **Supplementary Table 5.** Statistic values of EM-seq analysis.  
2253  
2254      **Supplementary Table 6.** Composition of DNA demethylation escapees.  
2255  
2256      **Supplementary Table 7.** Allelic expression and promoter DNA methylation of X-linked genes,  
2257      and demethylation escapees.  
2258  
2259      **Supplementary Table 8.** List of DEGs between wild-type and *TET1* KO cells, and associated  
2260      GO terms.  
2261  
2262      **Supplementary Table 9.** Information of cytokines, chemicals, and culture media.  
2263  
2264      **Supplementary Table 10.** DNA oligonucleotides used in this study.  
2265  
2266      **Supplementary Table 11.** Information of parameters for quality filtering.  
2267  
2268      **Supplementary Table 12.** Haplotype phasing of chromosome X.  
2269  
2270  
2271

Fig.1 Murase et al.

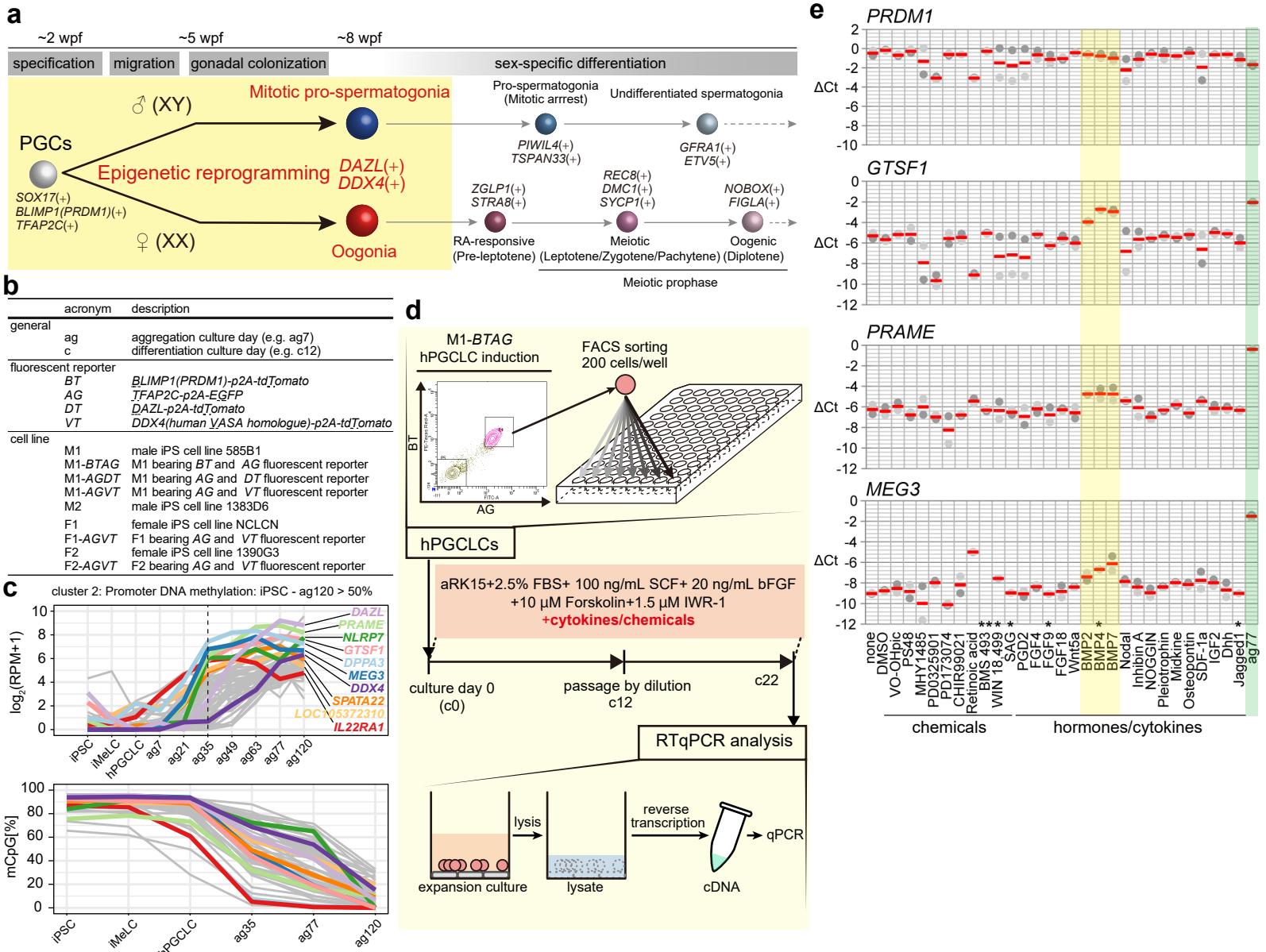
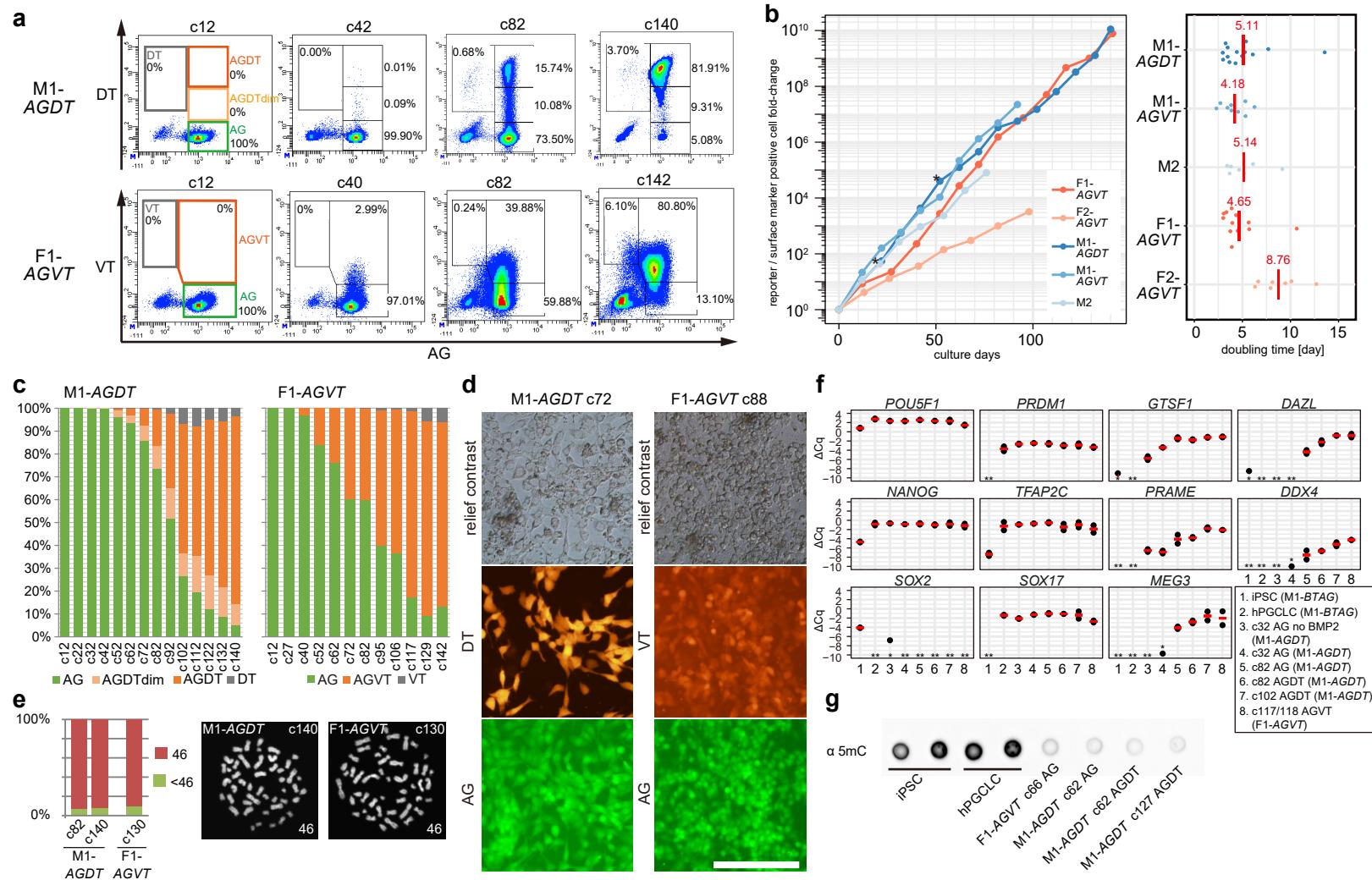
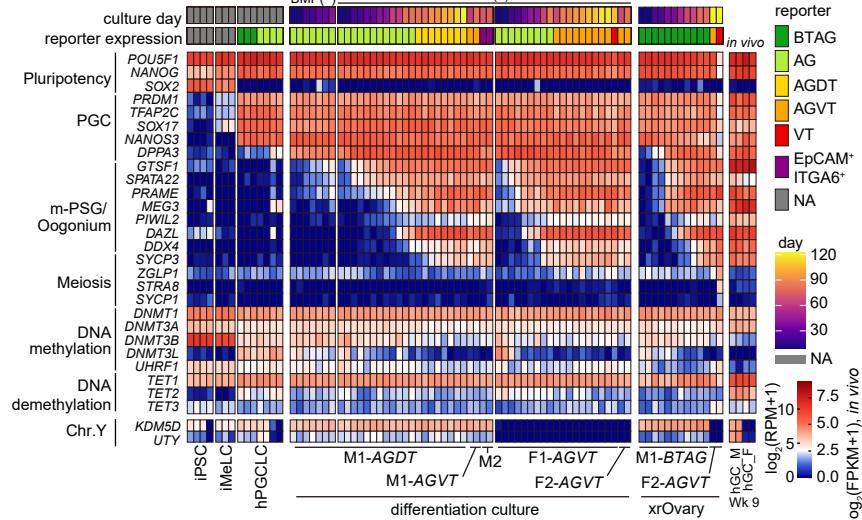


Fig. 2 Murase et al.

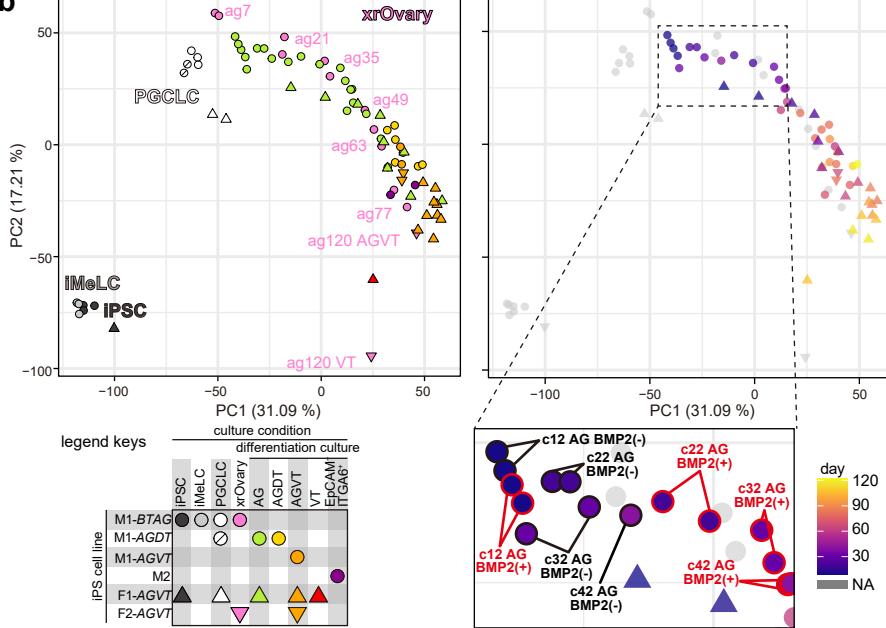


**Fig. 3 Murase et al.**

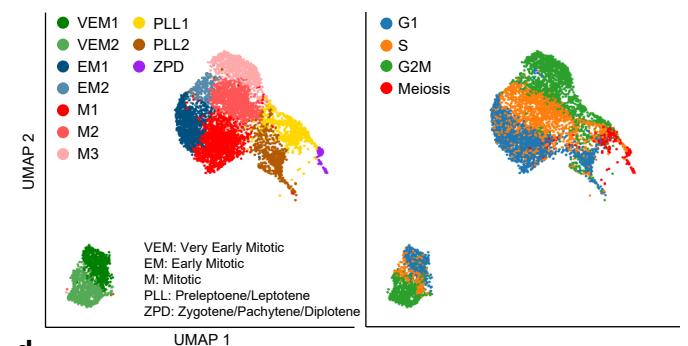
**a**



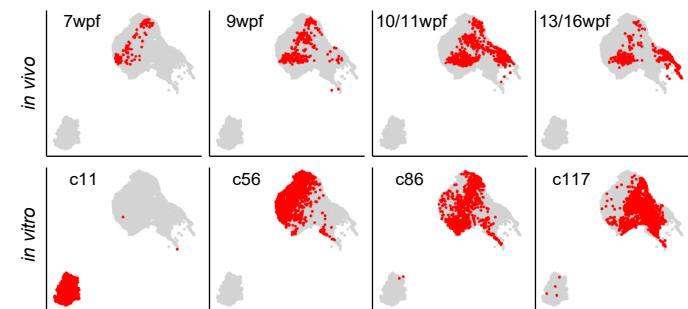
**b**



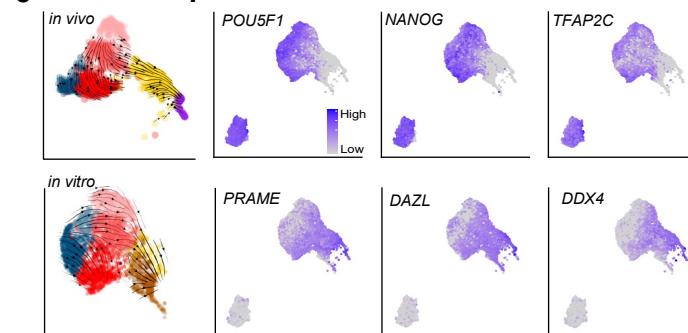
**c**



**d**



**e**



**f**



Fig. 4 Murase et.al.

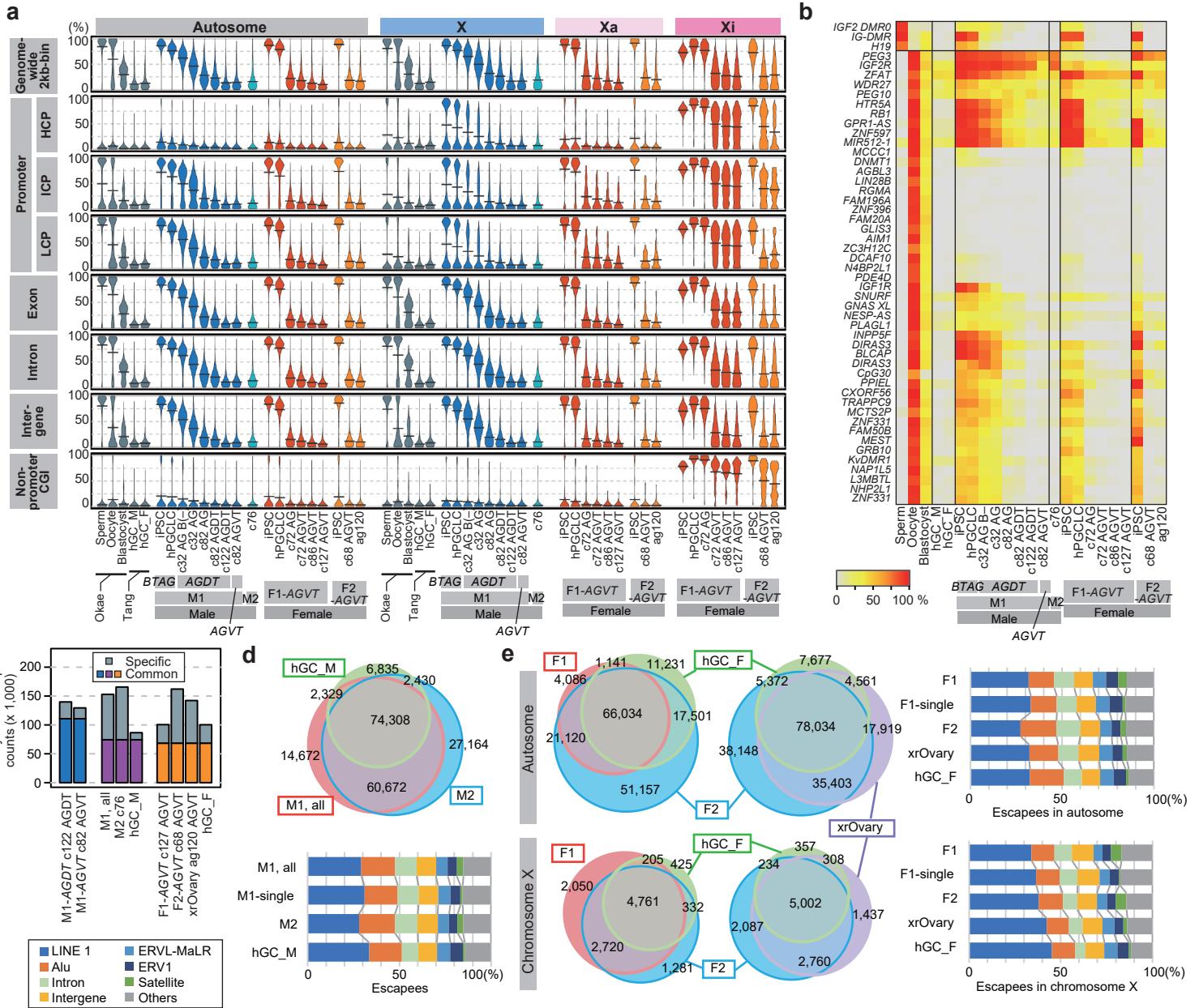


Fig. 5 Murase et al.

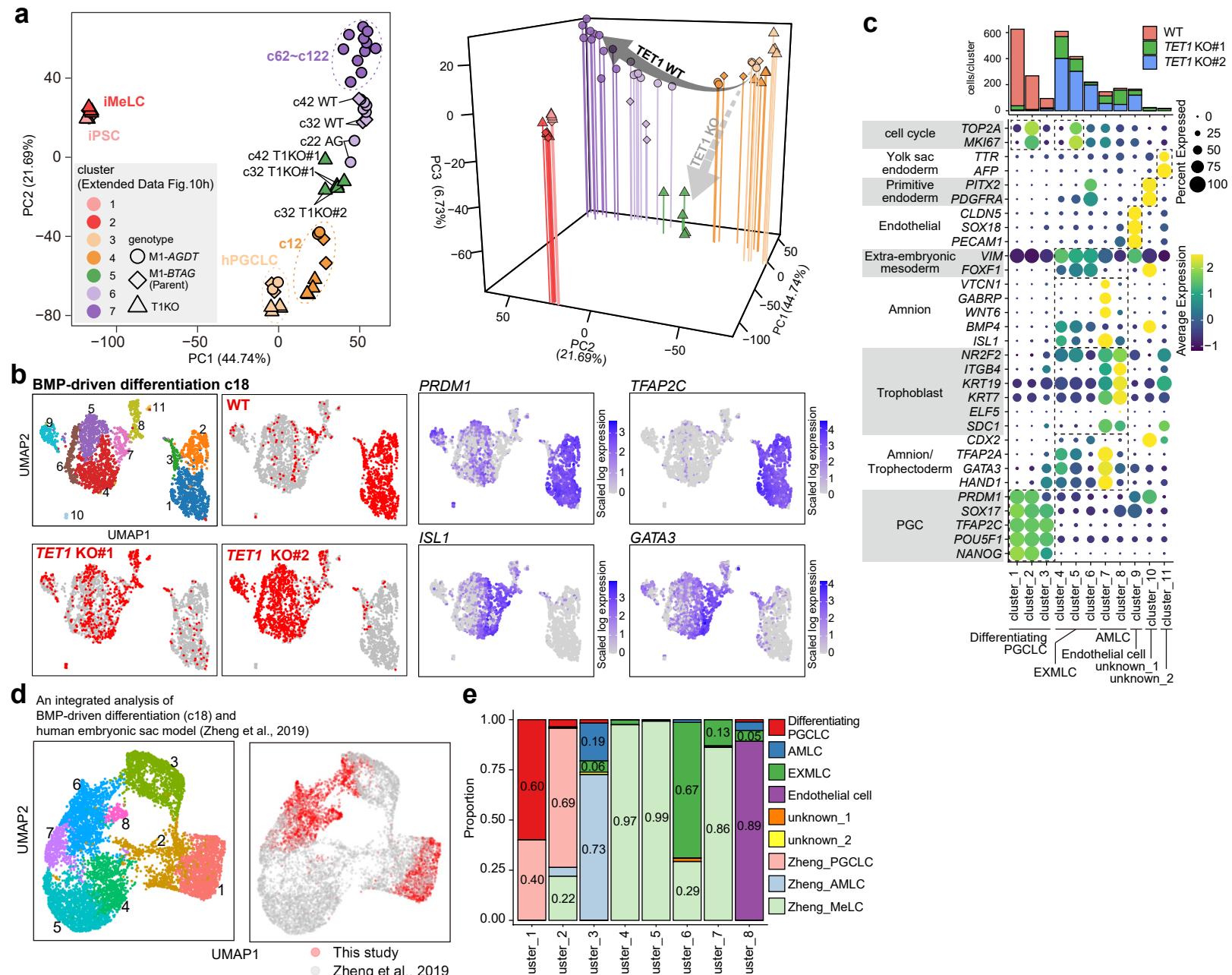
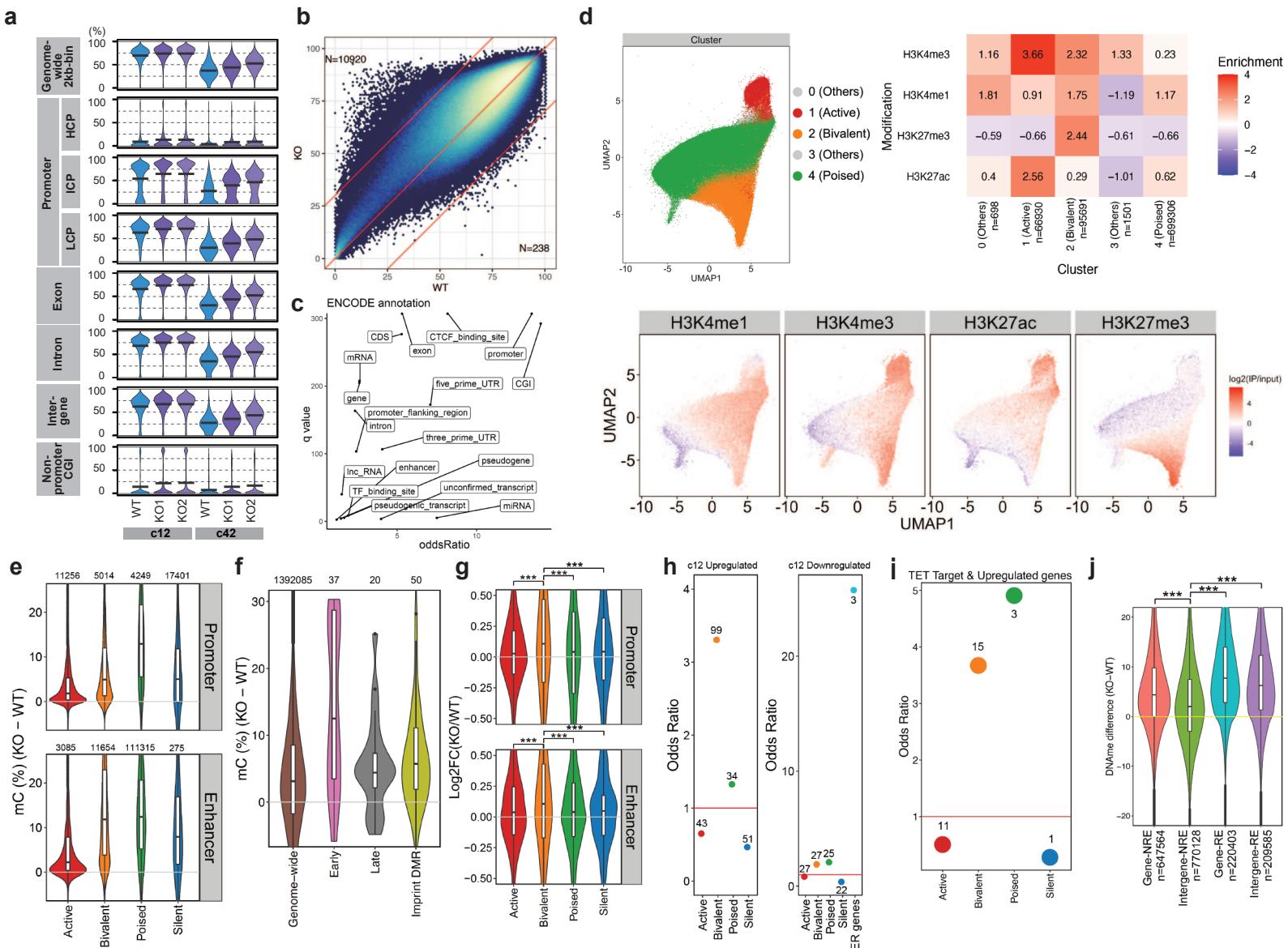
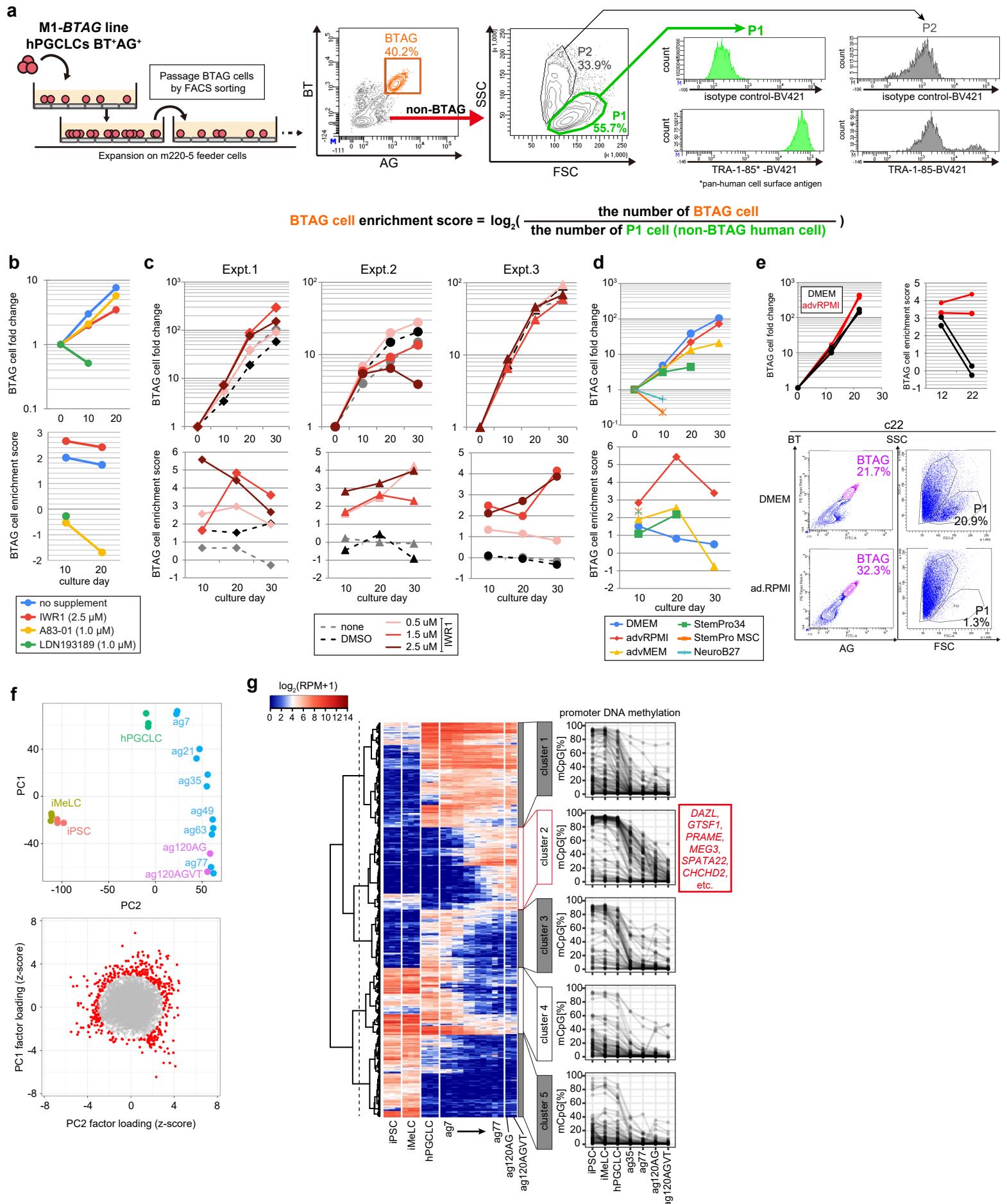


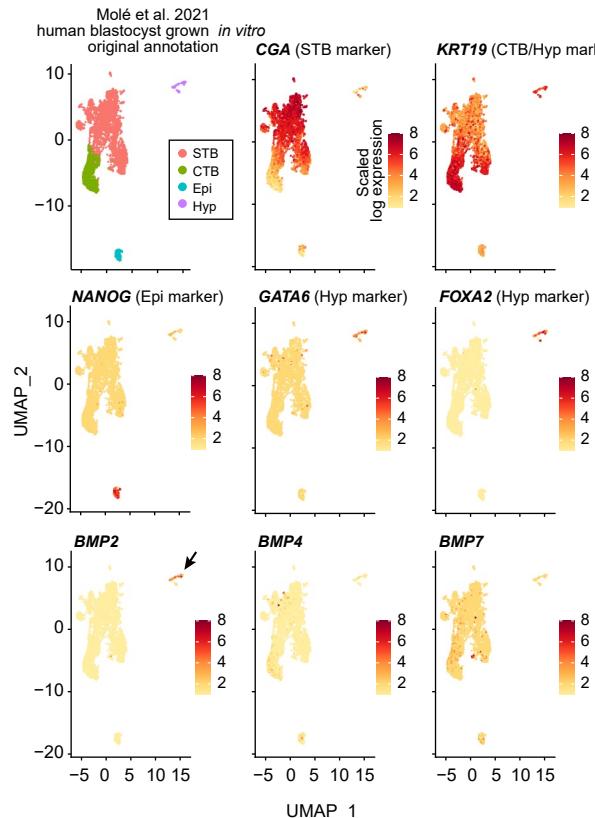
Fig. 6 Murase et.al.



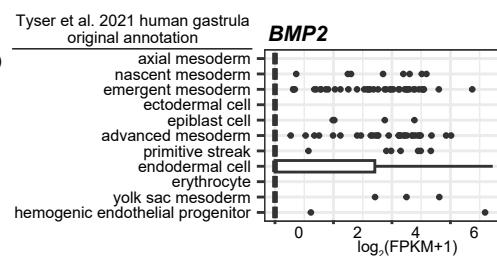


# Extended Data Fig.2 Murase et al.

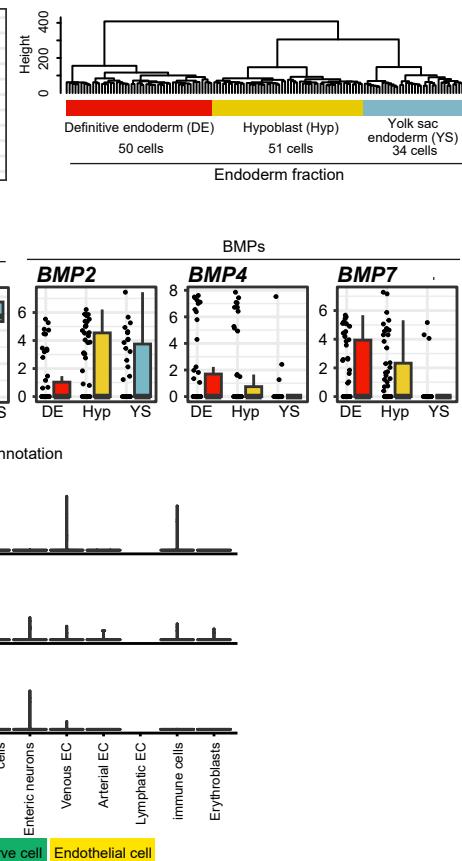
**a**



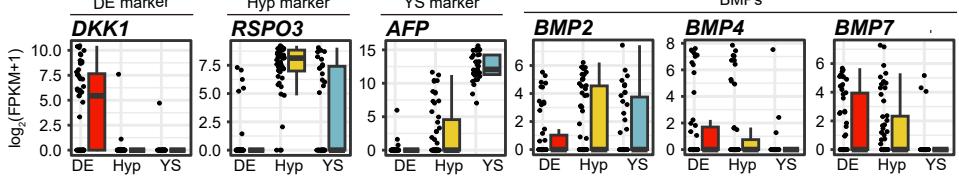
**b**



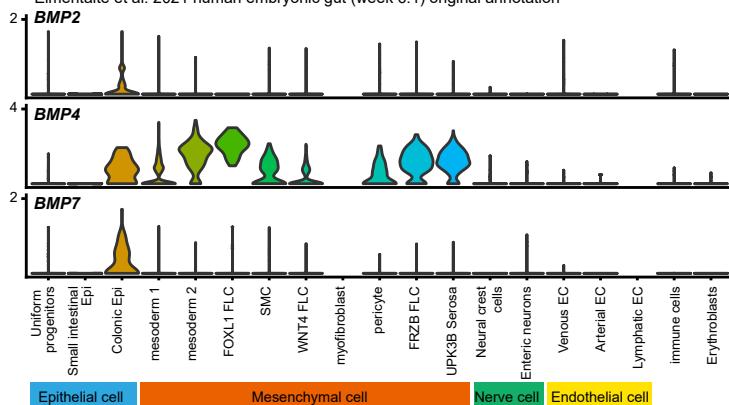
**c**



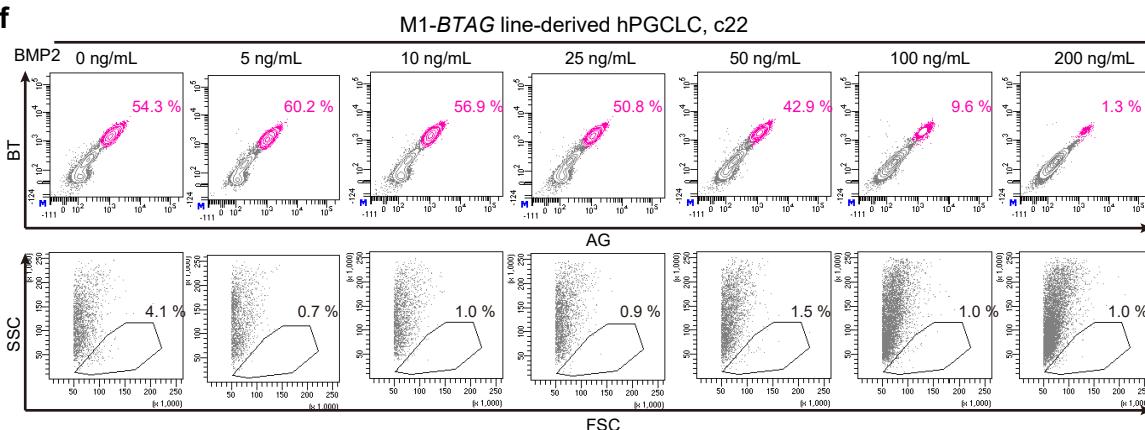
**d**



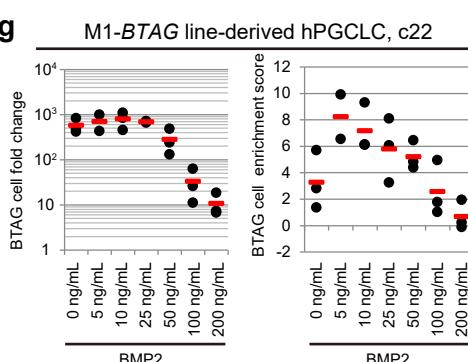
**e**



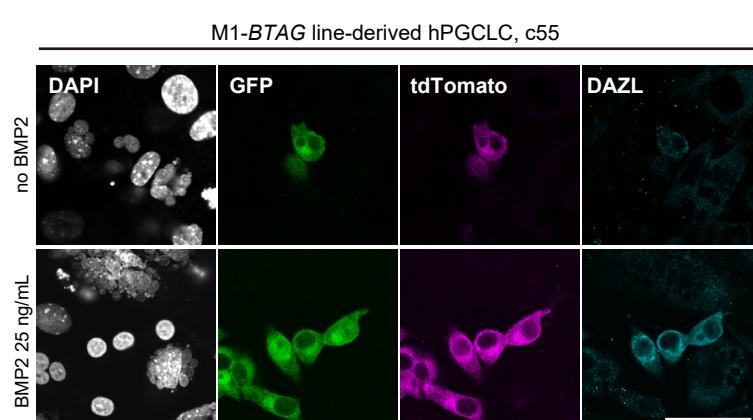
**f**



**g**

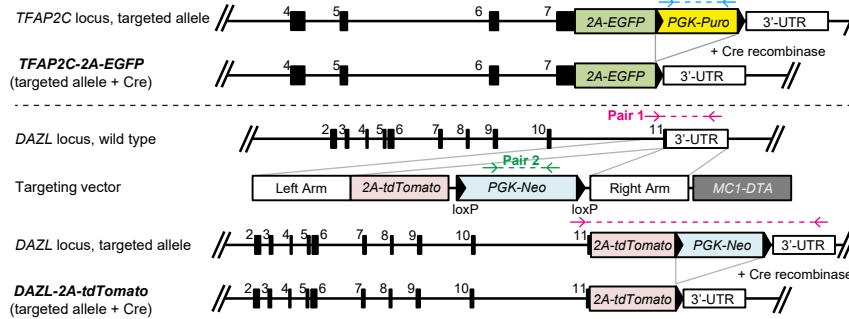


**h**



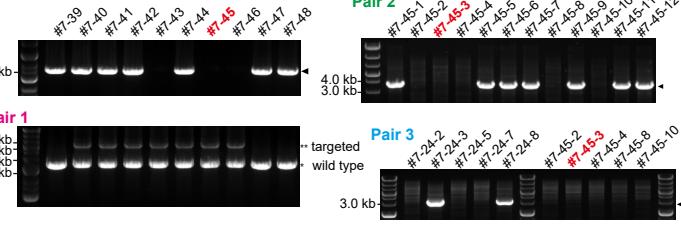
# Extended Data Fig.3 Murase et al.

## a AGDT reporter construction

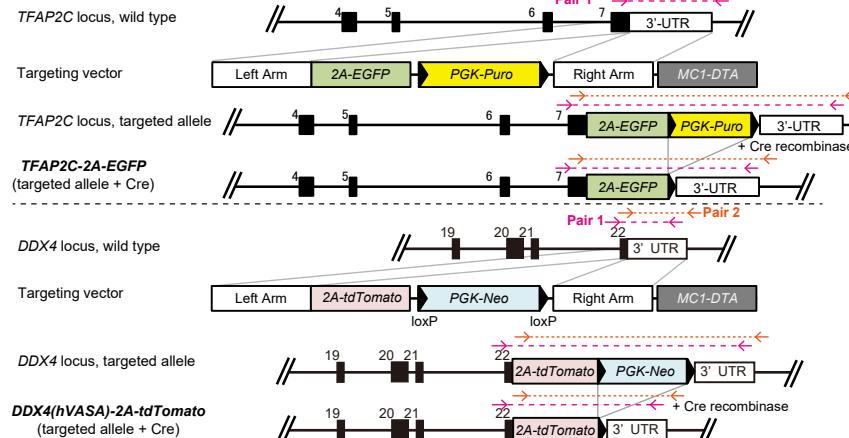


## b M1-AGDT establishment

### Random integration

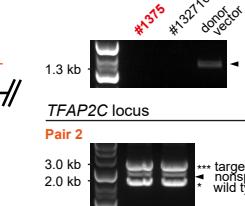


## c AGVT reporter construction

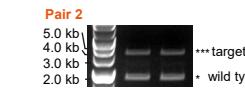


## d M1-AGVT establishment

### Random integration

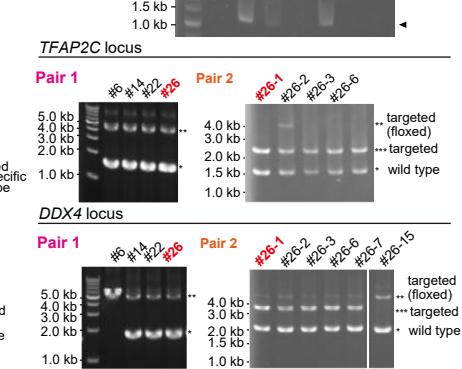


### DDX4 locus



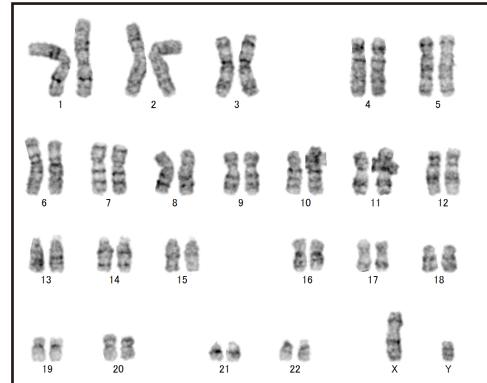
## e F1-AGVT establishment

### Random integration

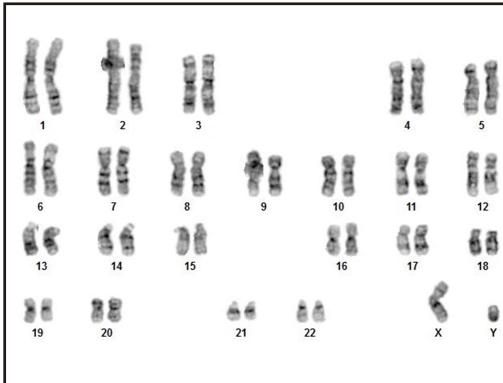


## f

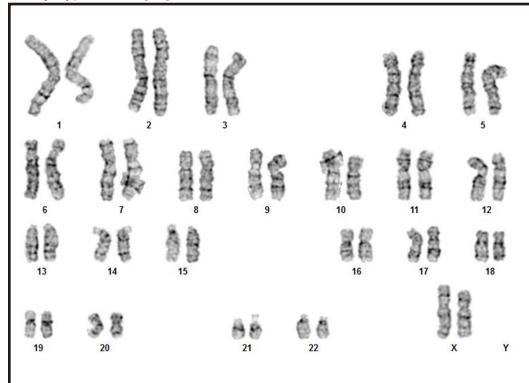
M1-AGDT (585B1 AGDT #7453)  
Karyotype: 46,XY[20]



M1-AGVT (585B1 AGVT #1375)  
Karyotype: 46,XY[20]

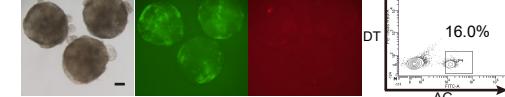


F1-AGVT (NCLCN AGVT #261)  
Karyotype: 46,XX[20]

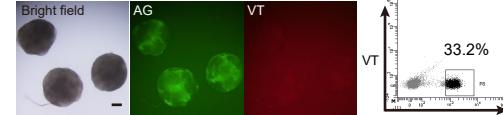


## g

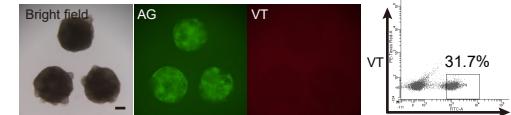
### M1-AGDT

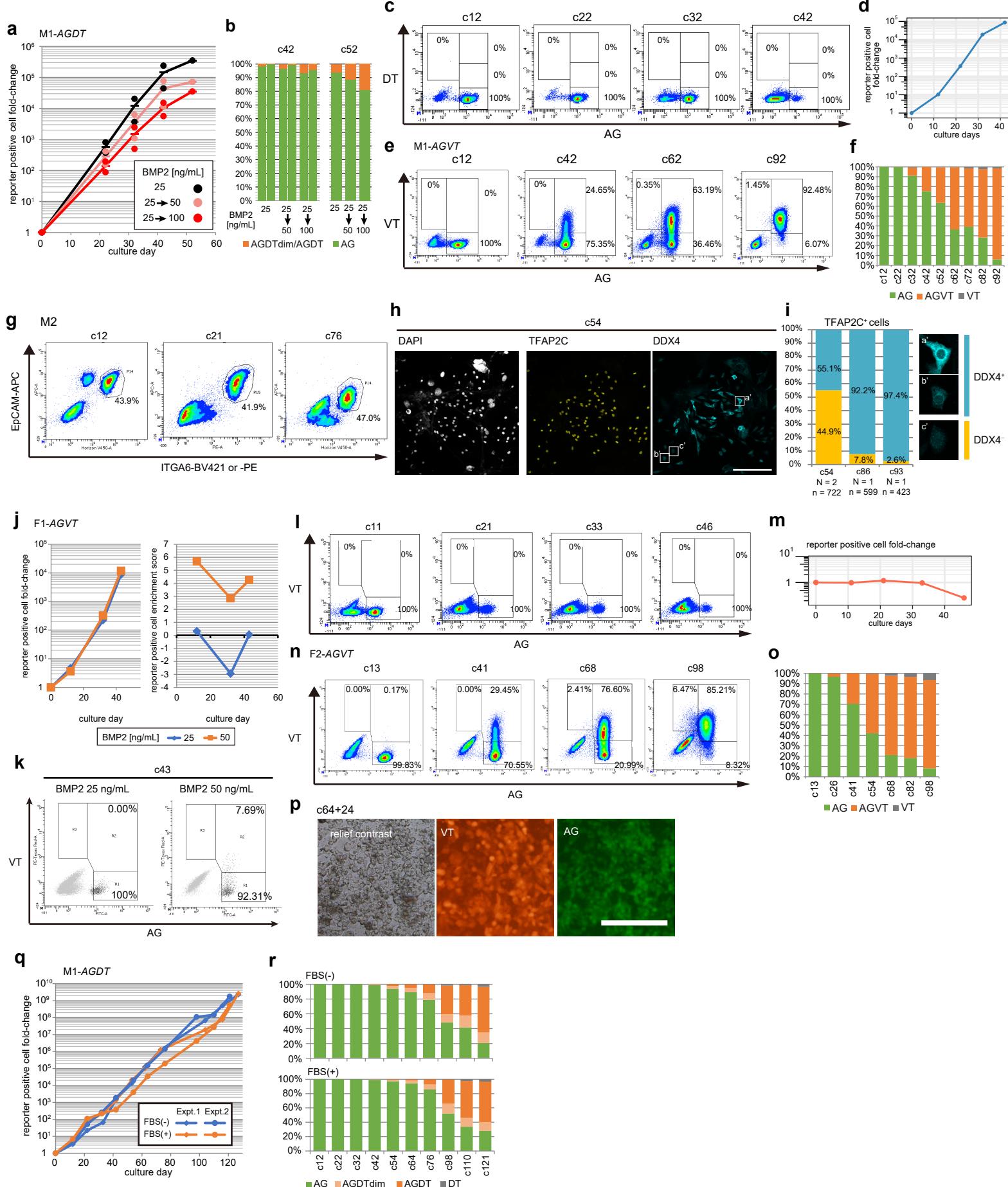


### M1-AGVT



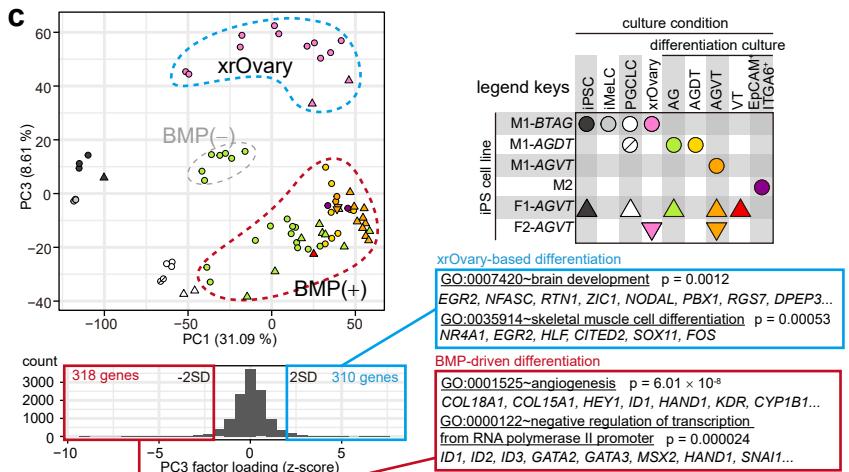
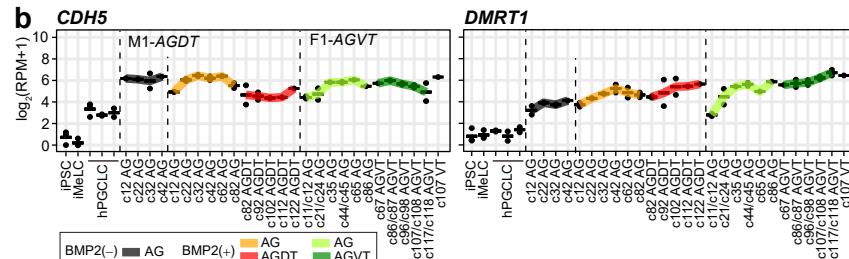
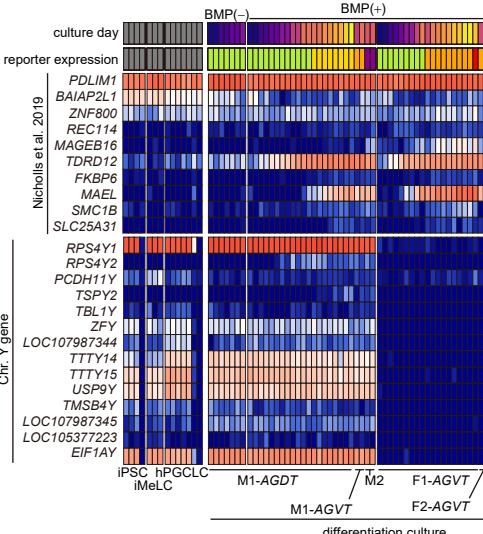
### F1-AGVT



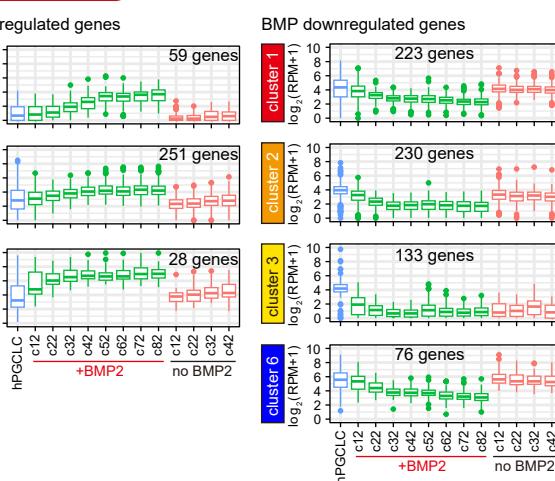
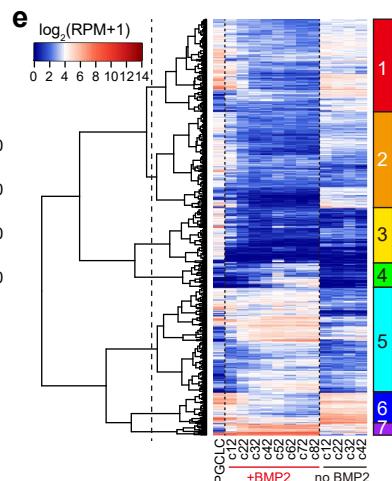
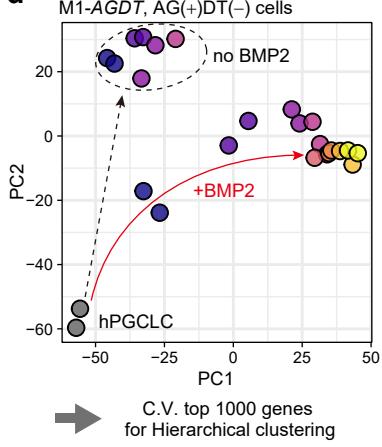


# Extended Data Fig 5, Murase et al.

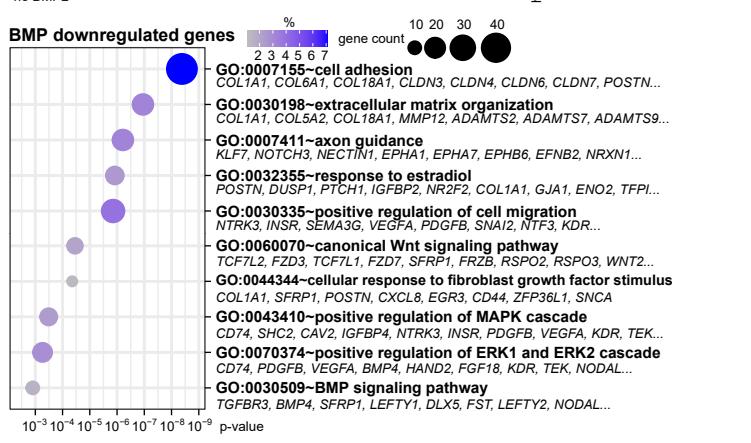
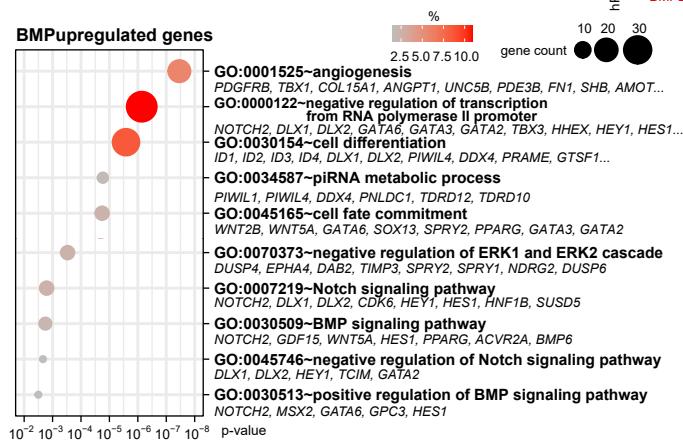
**a**



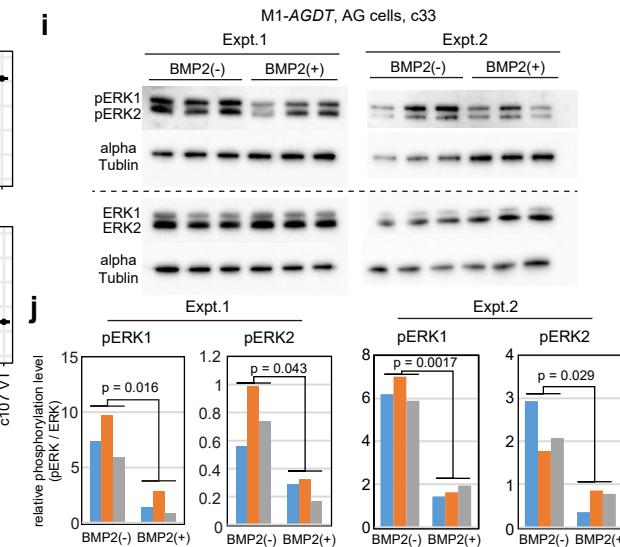
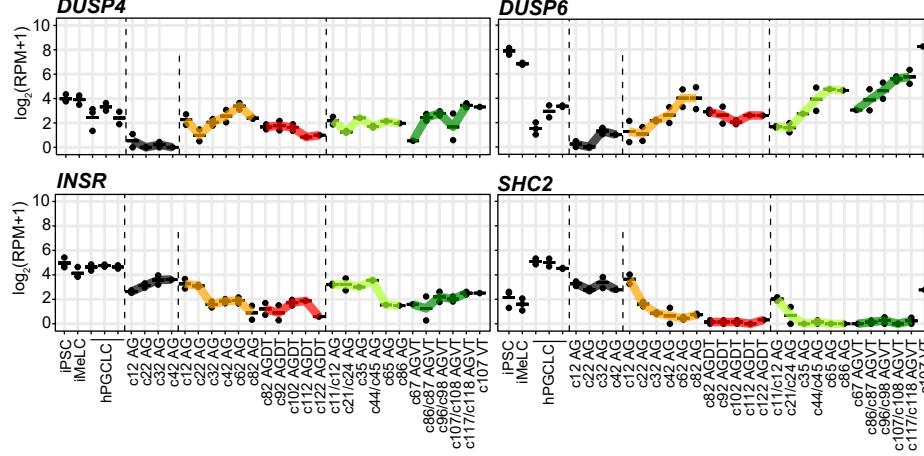
**d**



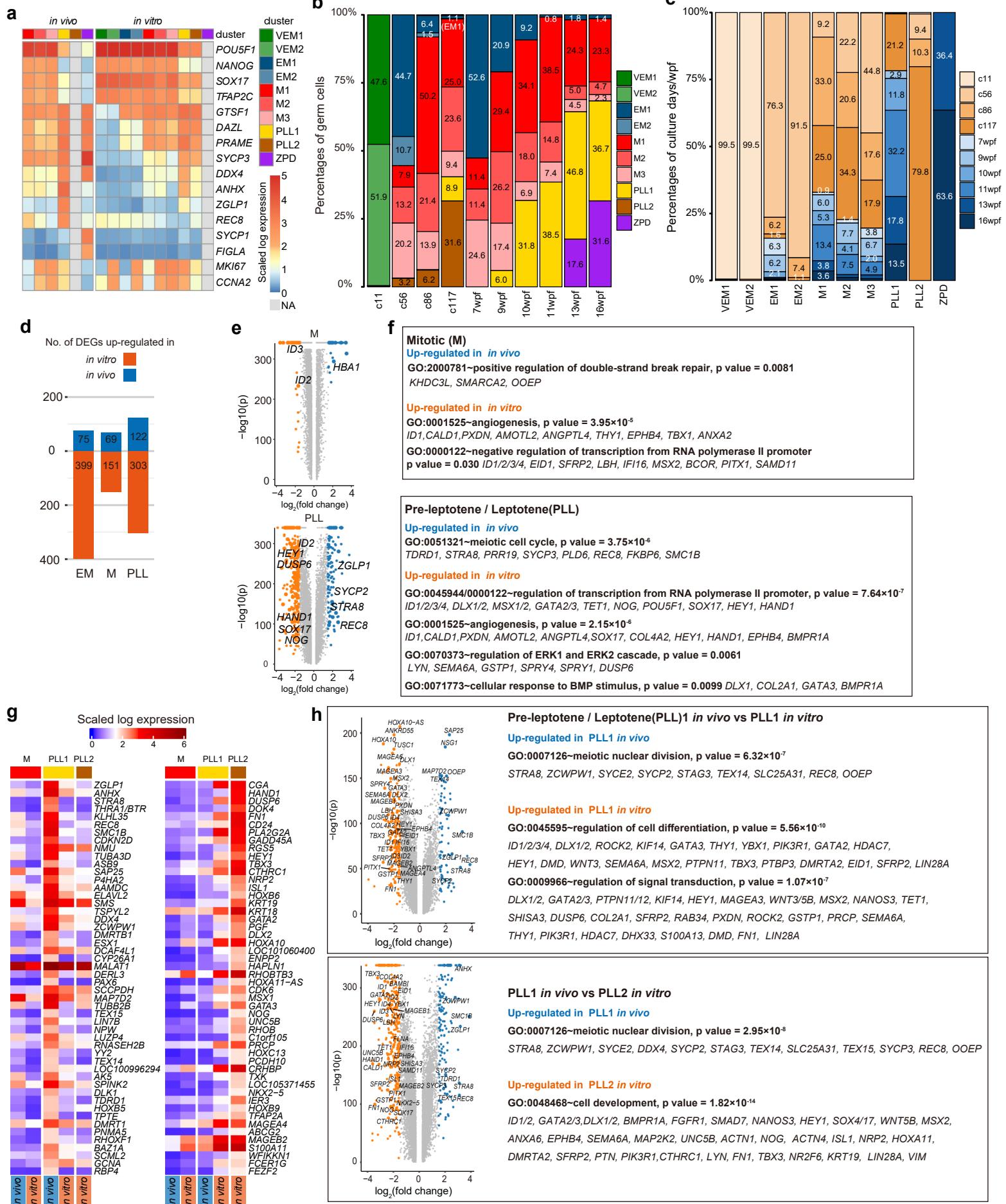
**g**



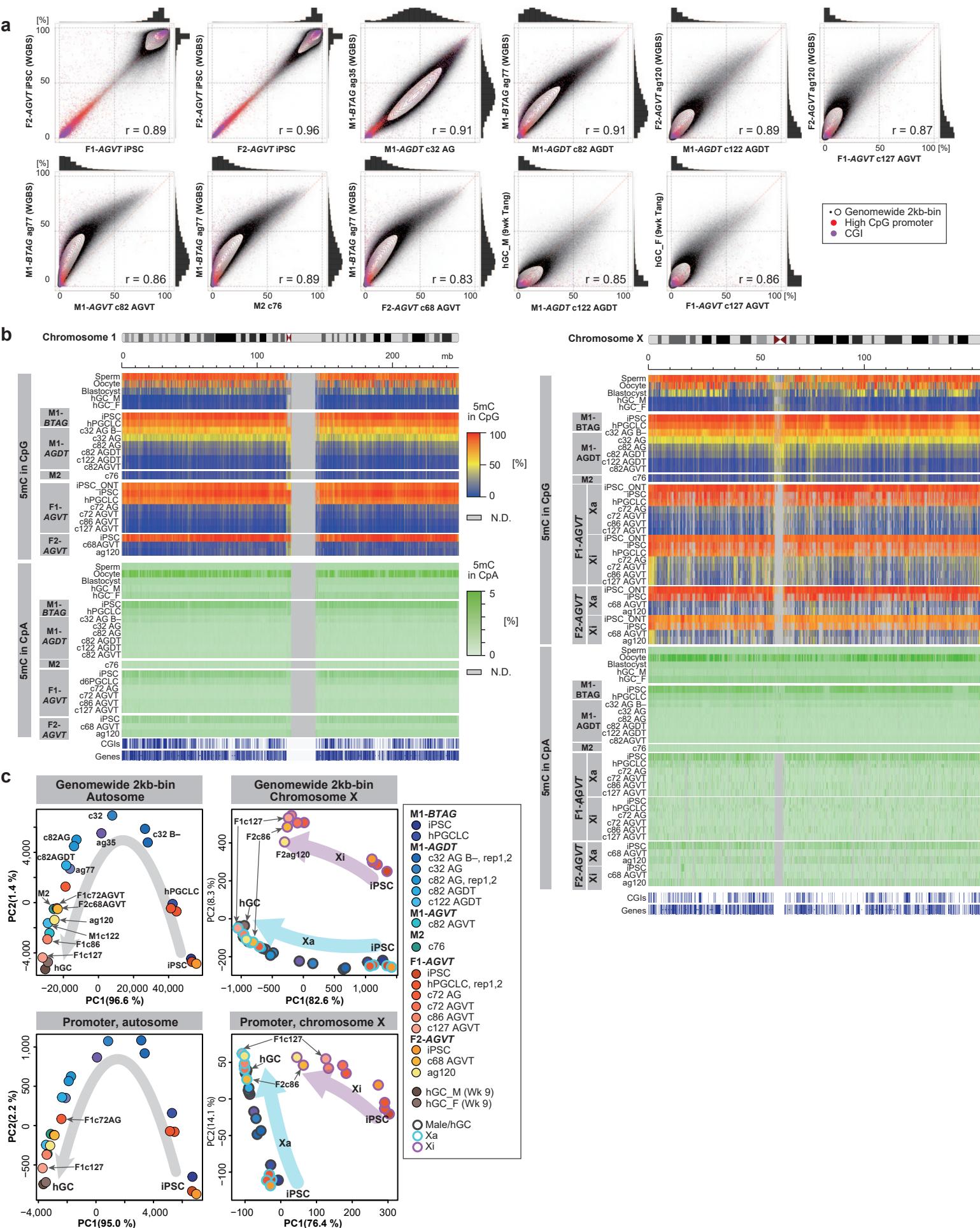
**h**



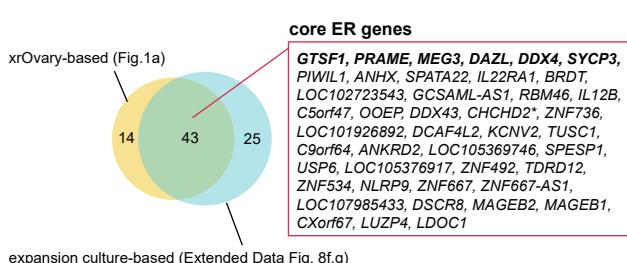
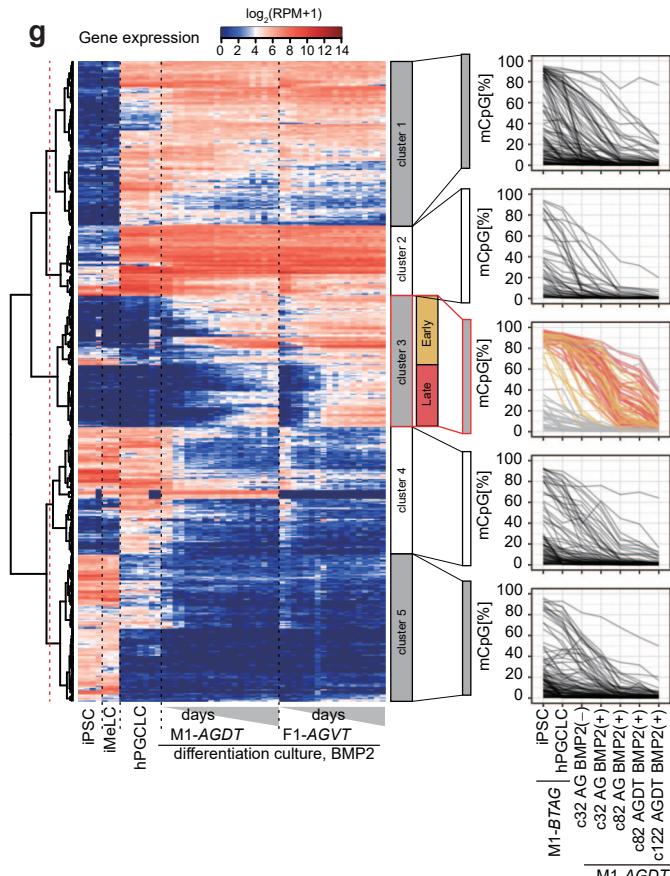
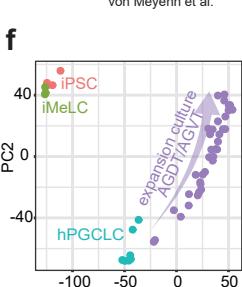
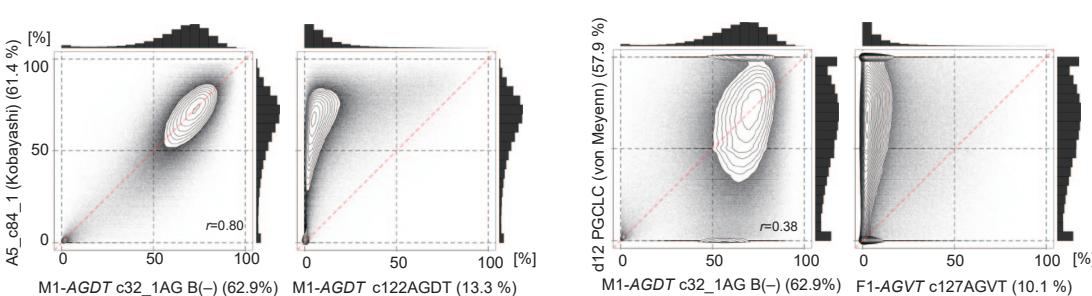
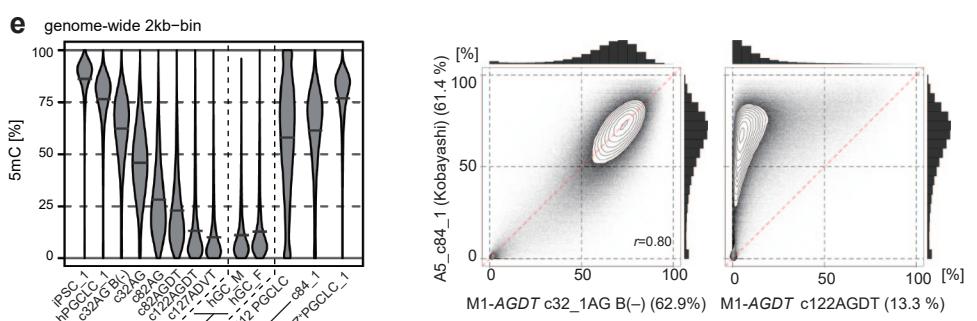
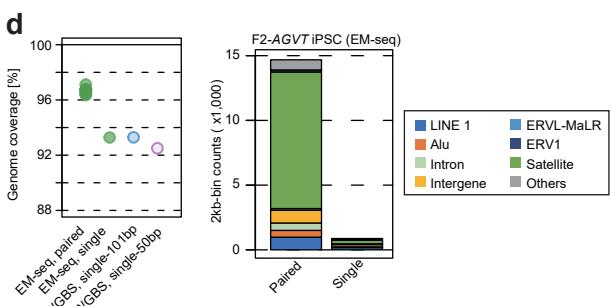
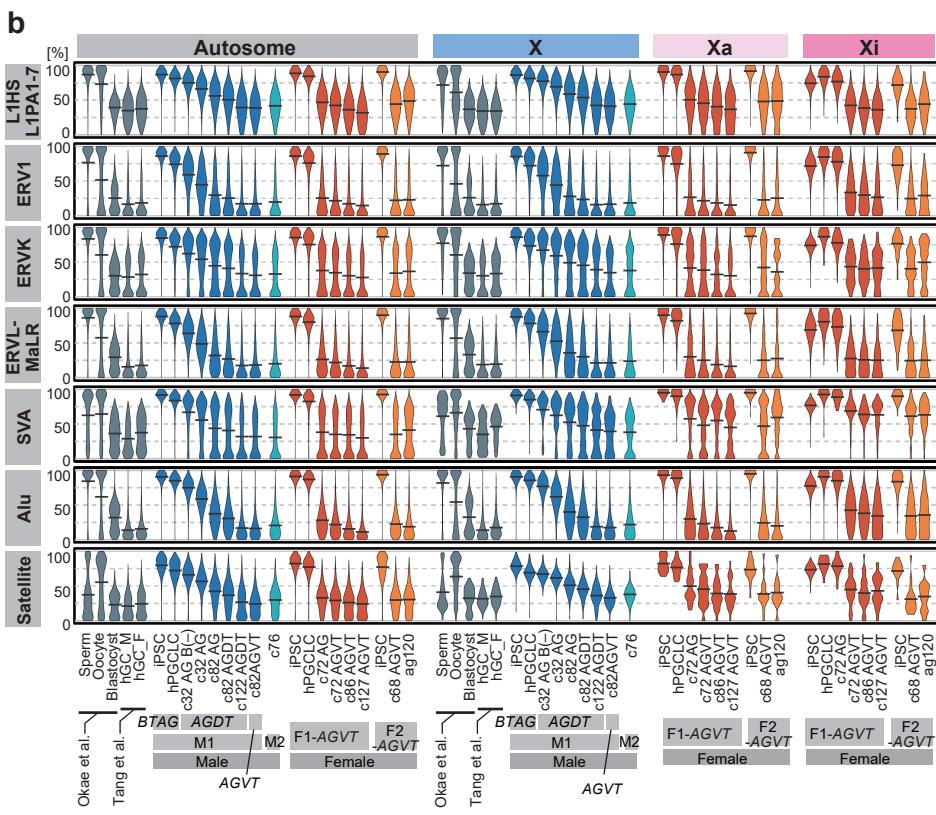
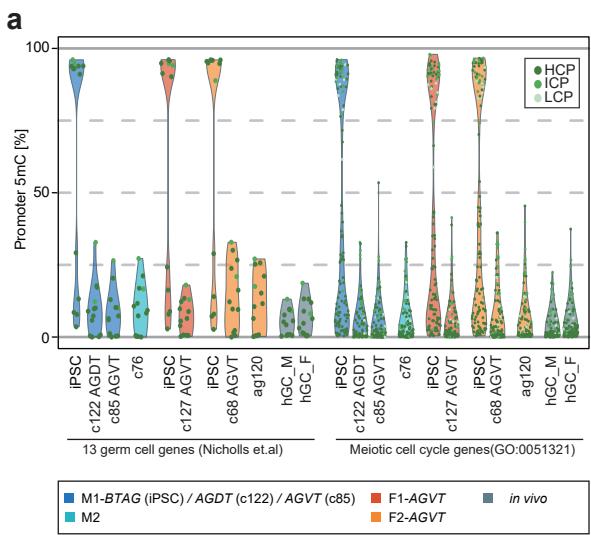
Extended Data Fig.6 Murase et al.



# Extended Data Fig. 7 Murase et.al.

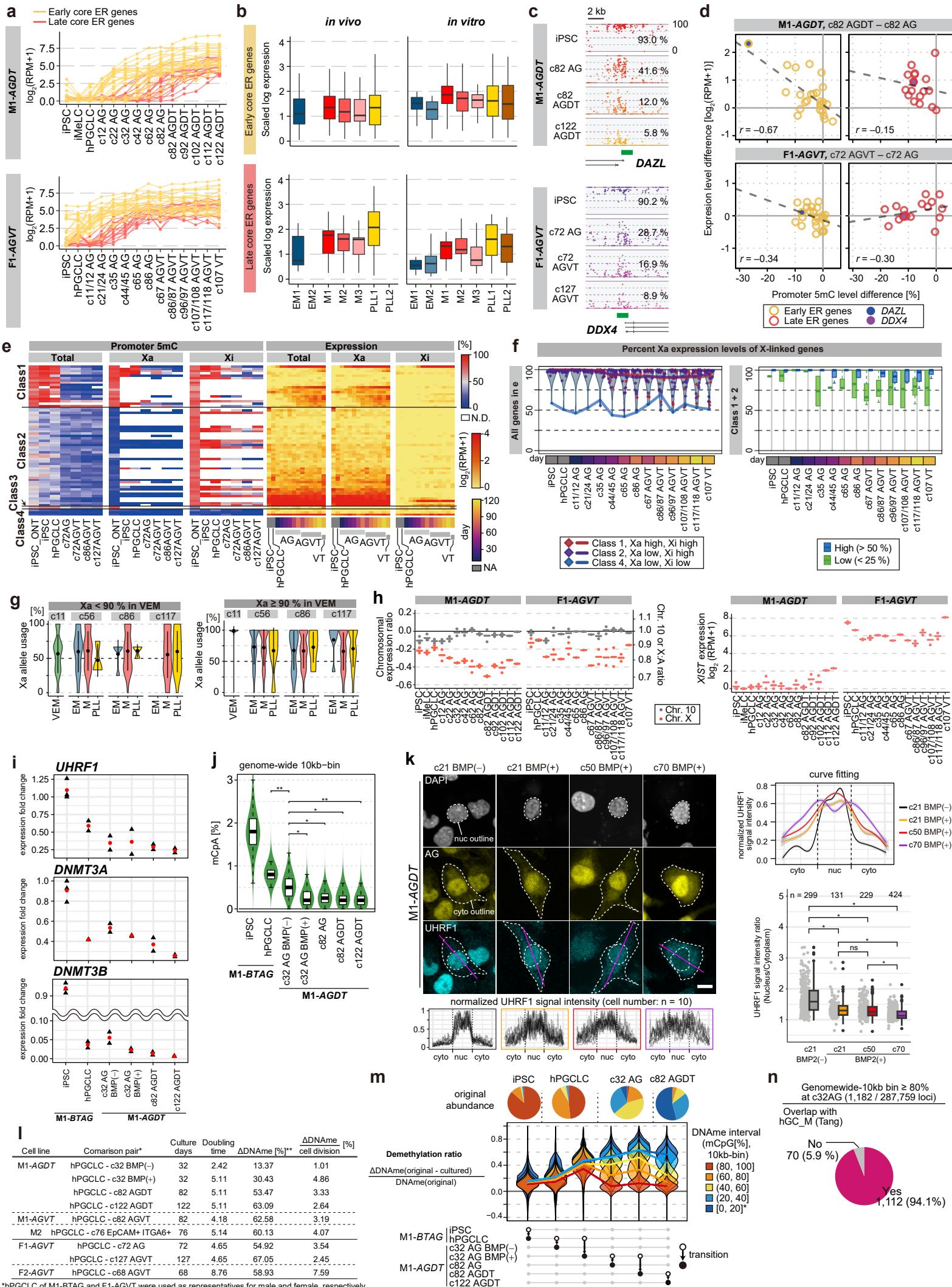


# Extended Data Fig. 8 Murase et.al.



\*Note that CHCHD2 was excluded from the following analysis due to the persistent expression in F1-AGVT iPSCs and their derivatives.

# Extended Data Fig. 9 Murase et al.



\*hPGCLC of M1-BTAG and F1-AGVT were used as representatives for male and female, respectively.

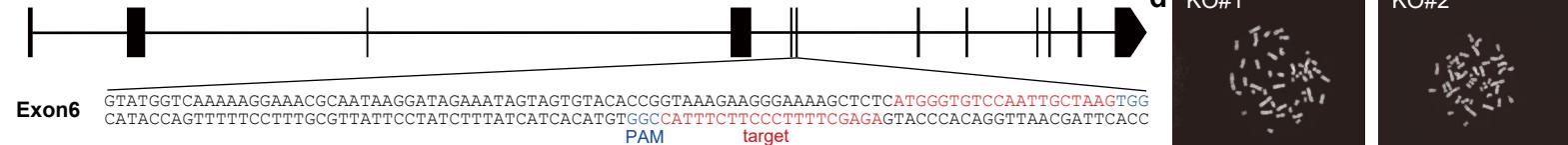
\*\*decreases from hPGCLC (genome-wide 2kb-bin)

\*Bins having zero value were excluded from the demethylation ratio calculation.

# Extended Data Fig. 10 Murase et al.

**a**

human *TET1*



**b**

KO#1 (585B1-BTAG-TET1<sup>-/-</sup> #142)

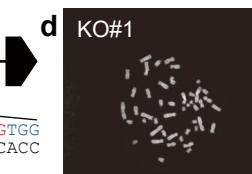
AGTGTACAccgttaaaaggaaaaagcttc-----a-tgggtgtccaaattgctaagtggTAAG Genomic sequence  
AGTGTACACCGTAAAGAAGGGAAAAGCTCTCCTGGGAAAGCTCTCATTGGGTGTCCTAAGTGGGTAAG 14bp insertion  
AGTGTACACCGTAAAGAAGGGAAAAGCGCTC-----A-TGG-TGTCCAATTGCTAAGTGGGTAAG 1bp deletion

KO#2 (585B1-BTAG-TET1<sup>-/-</sup> #2725)

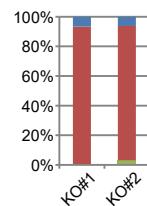
AGTGTACAccgttaaaaggaaaaagcttcCatgggtgtccaaattgctaagtggTAAG Genomic sequence  
AGTGTACACCGTAAAGA-----GCTAAGTGGGTAAG 28bp deletion  
AGTGTACACCGTAAAGAAGGGAAAAGGGAAAA-----GCTAAGTGGGTAAG 20bp deletion

**c**

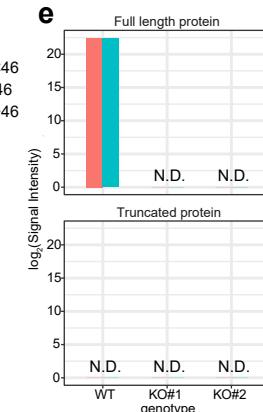
parent  
KO#1  
KO#2  
α 5hmC



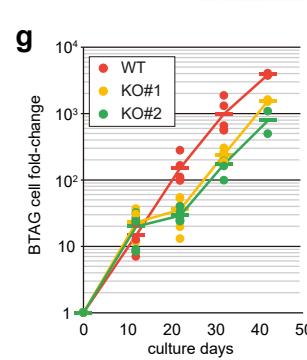
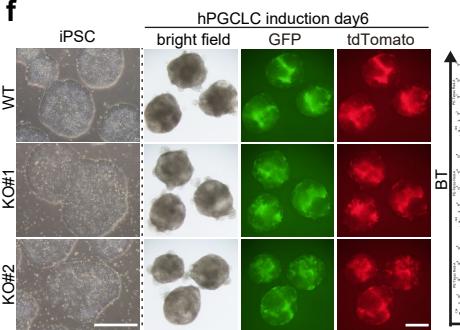
**d**



**e**



**f**



**k**

**cluster 1/2**  
GO:0032956~regulation of actin cytoskeleton organization

p value = 0.0075

ARHGAP18, GRHL3, FZD10, PAK3

GO:0009972~cytidine deaminase, p value = 0.0012

APOBEC3C, APOBEC3F, APOBEC3G

**cluster 3/4**

GO:0045944~positive regulation of transcription from RNA polymerase II promoter, p value = 2.39×10<sup>-10</sup>

HAND1, CEBPA, DLX1, CEBPD, GATA6, TBXT, MEIS3, TBX3...

GO:0007155~cell adhesion, 7.41×10<sup>-5</sup>

POSTN, ITGB5, ITGA3, ITGA2, CXCL12, COL5A1, ADAM9...

**cluster 5-8**

GO:0007155~cell adhesion, p value = 8.99×10<sup>-17</sup>

COL18A1, COL16A1, COL12A1, ITGA1, IGFBP7, CTNNA3, HES1...

GO:0011525~angiogenesis, p value = 3.30×10<sup>-12</sup>

SHC1, EPAS1, PDGFRB, ANGPT2, HAND2, HOXB3, TEK...

GO:0030509~BMP signaling pathway, p value = 1.70×10<sup>-7</sup>

LEFTY1, WNT5A, BMP7, BMP4, HES1, LEFTY2, NODAL, ENG...

GO:0006468~protein phosphorylation, p value = 0.0035

DAPK3, RIPK4, WNK2, COQ8B, JAK2, TRIB2, MAP3K14, MAP3K5...

**cluster 9**

GO:1903765~negative regulation of potassium ion export across plasma membrane, p value = 0.0051

KCN3, KCNE5

GO:0048599~oocyte development, p value = 0.029

MEIOC, AURKC

**cluster 10**

GO:0007399~nervous system development, p value = 0.0069

OLFM1, CHRNA3, ZEB1, CRMP1, GFRA1, DCLK1

**cluster 11/12**

GO:0009566~fertilization, p value = 0.013

SPATA22, SPESP1, TDRD12, MEIOB

GO:0043046~DNA methylation involved in gamete generation, p value = 0.015

KDM1B, TDRD12, CTCFL...

GO:0007140~male meiosis, p value = 0.023

TDRD12, MEIOB, BRDT

GO:0007288~sperm exosome assembly, p value = 0.028

SPEF2, NEURL1, MNS1

**cluster 13**

GO:0006357~regulation of transcription from RNA polymerase II promoter, p value = 4.77×10<sup>-6</sup>

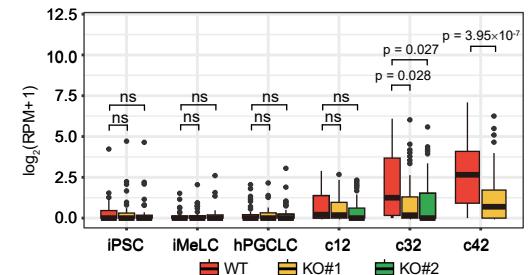
ZNF594, ZNF792, ZNF670, ZBTB49, FOXO3, TBX22, SOX13...

**i**

The number of DEGs in each cell type

	iPSC	iMeLC	hPGCLC	c12	c32	c42
upregulated	41	19	273	226	608	1219
downregulated	61	69	155	101	159	416

IPSC  
KO#1  
KO#2  
iMeLC  
KO#1  
KO#2  
hPGCLC  
KO#1  
KO#2  
c12  
KO#1  
KO#2  
c32  
KO#1  
KO#2  
c42  
KO#1  
KO#2



# Extended Data Fig.11 Murase et al.

