Reconstitution of the oocyte transcriptional network with transcription factors

https://doi.org/10.1038/s41586-020-3027-9

Received: 7 January 2020

Accepted: 28 October 2020

Published online: 16 December 2020

Check for updates

Nobuhiko Hamazaki[1]✉, Hirohisa Kyogoku[2], Hiromitsu Araki[3], Fumihito Miura[3], Chisako Horikawa[1], Norio Hamada[1,4], So Shimamoto[1], Orie Hikabe[1], Kinichi Nakashima[1], Tomoya S. Kitajima[2], Takashi Ito[3], Harry G. Leitch[5,6] & Katsuhiko Hayashi[1]✉

During female germline development, oocytes become a highly specialized cell type and form a maternal cytoplasmic store of crucial factors. Oocyte growth is triggered at the transition from primordial to primary follicle and is accompanied by dynamic changes in gene expression[1], but the gene regulatory network that controls oocyte growth remains unknown. Here we identify a set of transcription factors that are sufficient to trigger oocyte growth. By investigation of the changes in gene expression and functional screening using an in vitro mouse oocyte development system, we identified eight transcription factors, each of which was essential for the transition from primordial to primary follicle. Notably, enforced expression of these transcription factors swiftly converted pluripotent stem cells into oocyte-like cells that were competent for fertilization and subsequent cleavage. These transcription-factor-induced oocyte-like cells were formed without specification of primordial germ cells, epigenetic reprogramming or meiosis, and demonstrate that oocyte growth and lineage-specific de novo DNA methylation are separable from the preceding epigenetic reprogramming in primordial germ cells. This study identifies a core set of transcription factors for orchestrating oocyte growth, and provides an alternative source of ooplasm, which is a unique material for reproductive biology and medicine.

In mouse germline, primordial germ cells (PGCs) go through a sequential differentiation process beginning with PGC specification, followed by migration to the gonad, epigenetic reprogramming (including genome-wide DNA demethylation)[2], fate determination[3] and sex determination[4]. After sex determination, female PGCs enter meiosis, thereby becoming oocytes. Oocytes are arrested in the diplotene stage of meiotic prophase I and most are maintained in primordial follicles. Cytoplasmic expansion is triggered after activation of primordial follicles. Once oocyte growth is triggered, maternal RNAs and proteins are stored in the cytoplasm[5,6]. Meiosis resumes in full-grown oocytes, establishing metaphase II (MII) oocytes, and is not completed until after fertilization.

Oocyte differentiation therefore entails two key processes: oocyte growth and meiosis. Although concurrent, these two features are separable, as evidenced by a seminal study in which STRA8-knockout mice were shown to develop oocyte-like cells that do not enter meiosis. Several genes essential for early oocyte growth have been identified, including *Figla*[7], *Sohlh1*[8], *Sohlh2*[9], *Lhx8*[10], *Nobox*[11], *Taf4b*[12,13], *Yy1*[14] and *Tbpl2*[15]. Transcriptome analysis using oocytes lacking these genes[8,10,15–19] and identification of direct target sequences or genes[7,8,20–22] has revealed downstream gene cascades involved in early oocyte growth. In addition, a previous microarray analysis uncovered highly dynamic gene expression changes between postnatal day 2 (P2) oocytes in primordial follicles and P6 oocytes in the primary follicles (primordial-to-primary-follicle transition (PPT))[1]. Genes enriched at PPT were involved in protein synthesis and transcription[1], suggesting a role in oocyte growth.

However, there has been limited further advance towards a comprehensive description and functional investigation of the gene regulatory network specifically orchestrating oocyte growth. Here, we identify a set of transcription factors that comprise the underlying gene regulatory network and validate these findings with functional screening. Furthermore, we successfully reconstitute the network in pluripotent stem cells, thereby generating oocyte-like cells that are competent for fertilization and subsequent cleavage divisions.

Characterization of PPT in oogenesis

To identify the gene regulatory network, we used a recently established culture system that recapitulated female germline differentiation using embryonic stem (ES) cells[23]. In this culture system, ES cells are differentiated into PGC-like cells (PGCLCs) and then undergo oogenesis to give rise to MII oocytes in the presence of supporting gonadal somatic cells. Using female germline cells in this culture system and their in vivo counterparts, we mapped the trajectory of the female germline cycle by RNA sequencing (RNA-seq) analysis (Fig. 1a, Extended Data Fig. 1a). Principal component analysis (PCA) revealed that a highly dynamic gene

[1]Department of Stem Cell Biology and Medicine, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. [2]Laboratory for Chromosome Segregation, RIKEN Center for Biosystems Dynamics Research, Kobe, Japan. [3]Department of Biochemistry, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. [4]Department of Obstetrics and Gynecology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. [5]MRC London Institute of Medical Sciences (LMS), London, UK. [6]Institute of Clinical Sciences (ICS), Faculty of Medicine, Imperial College London, London, UK. ✉e-mail: hamazaki@hgs.med.kyushu-u.ac.jp; hayashik@hgs.med.kyushu-u.ac.jp

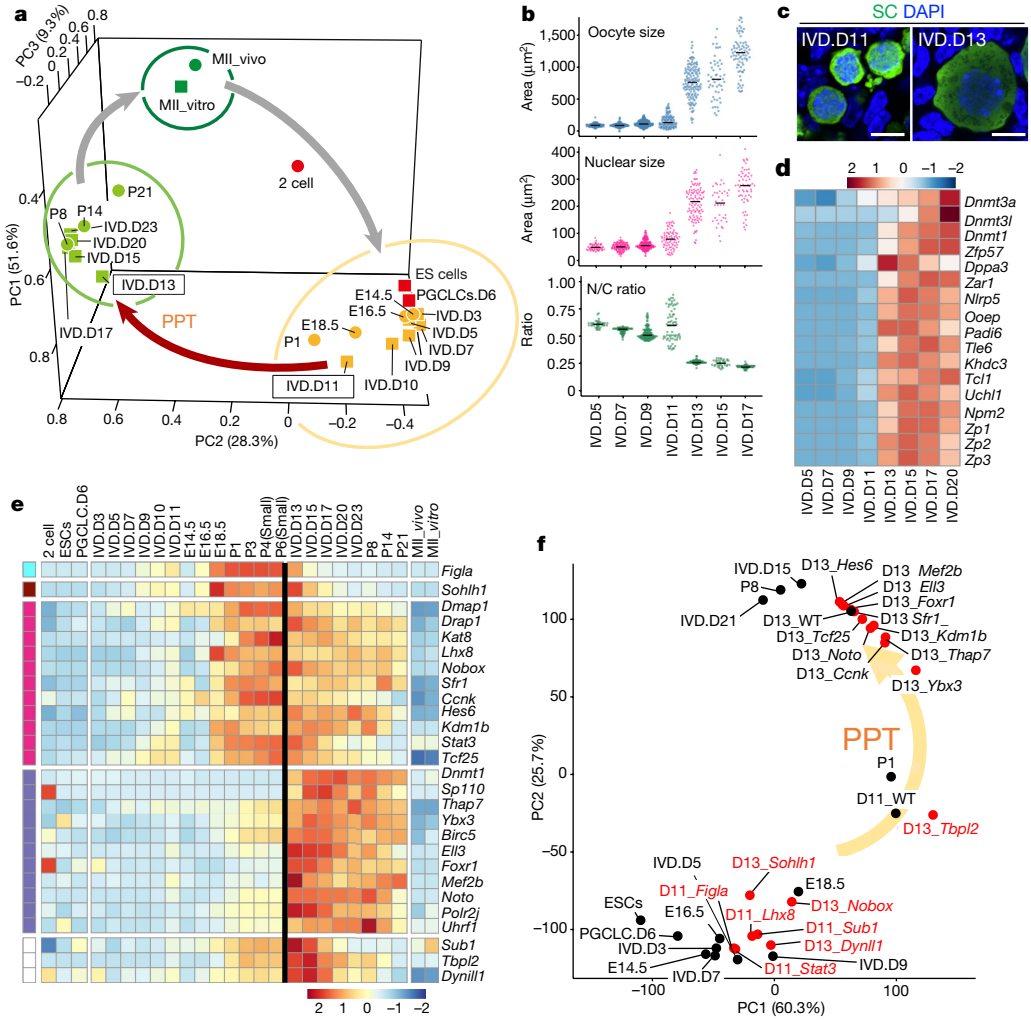


Fig. 1 | Functional screening of genes involved in the PPT. a, PCA of in vitro and in vivo oogenesis. The expression profile is based on at least biologically duplicated samples. **b**, Distribution of oocyte sizes, nuclear sizes, and the nuclear/cytosol (N/C) ratio. The values were compiled from biologically triplicated experiments. **c**, Immunostaining before and after the PPT. IVD.D11 and IVD.D13 were visualized by anti-GFP antibody (green) cross-reactive to BLIMP1–mVenus (BV) and Stella–ECFP (SC), and DAPI (blue). Scale bars, 10 μm. $n = 3$, biologically independent experiments. **d**, Heat map of gene expression dynamics of maternal factors. $n = 3$, biologically independent samples. **e**, Expression dynamics of genes from the PPT-associated modules. The thick black line shows the timing of PPT. The colour of the modules was from WGCNA (Extended Data Fig. 2a). **f**, PCA of the knockout oocytes. The transcriptomes of knockout oocytes (red) and wild-type oocytes (black) in vivo and in vitro are shown. All expression profiles are based on at least biologically duplicated samples.

expression change occurred in oocytes between in vitro differentiation culture day 11 (IVD.D11) and day 13 (IVD.D13), which corresponded to the period between P1 and P8 in vivo. As previous microarray analysis identified PPT between P2 and P6 oocytes[1], we considered the transition between IVD.D11 and IVD.D13, to represent PPT in vitro. Oocytes also exhibited a rapid increase in their cytoplasmic and nuclear volumes between IVD.D11 and IVD.D13, as observed in cytoplasmic expansion at PPT in vivo[24], along with a decrease in their nucleus/cytoplasmic ratios (Fig. 1b, c). Notably, the magnitude of gene expression changes in PPT was as large as the transition between MII oocytes and early stage preimplantation embryos, a period known as oocyte-to-embryo transition, during which zygotic genome activation occurs[25] (Fig. 1a).

During PPT, the upregulated genes included those encoding oocyte-specific structural proteins as previously reported[1]. However, we also identified genes involved in preimplantation development and DNA methylation of the maternal genome (Fig. 1d). In addition, upregulation of a subset of long terminal repeat (LTR) retrotransposon families[23] and the decline of the ratio of mean X chromosome transcripts in relation to mean transcripts from the autosomes (X/A ratio)[1] were observed at PPT (Extended Data Fig. 1b, c). As previous studies have shown that growing oocytes use TATA promoters, whereas zygotic genome activation is governed by a reciprocal switch from TATA promoter usage back to GC promoter usage[26,27], we hypothesized that the gene expression changes observed during PPT might also be coupled with a change in promoter usage. We found that GC nucleotides were enriched in the promoter region of genes highly transcribed until IVD.D11, whereas AT nucleotides were enriched in that of genes highly transcribed from IVD.D13 (Extended Data Fig. 1d). The AT nucleotides were specifically enriched at 20–30 base pairs (bp) upstream from the transcriptional start sites and included TATA-box motifs (Extended Data Fig. 1e–g). These results demonstrate that promoter usage was shifted from GC promoter to TATA promoter usage upon PPT.

Functional screening of PPT genes

To identify key transcription factors responsible for PPT, we applied weighted gene co-expression network analysis (WGCNA) to our RNA-seq data and identified 34 modules (Extended Data Fig. 2a).

Among these modules, we focused on five modules (MEcyan, MEdarkred, MEmagenta, MEwhite and MEpurple) that contained genes that were specifically expressed around PPT. From these modules, we extracted 27 transcription-related genes that fulfilled the Gene Ontology (GO) term 'regulation of transcription, DNA-templated (GO:0006355)' (Fig. 1e), and then investigated the functional requirement of these genes for PPT by loss-of-function analysis using ES cells that contained the BLIMP1–mVenus and Stella–ECFP (BVSC) reporter genes[23] (Extended Data Fig. 2b, c). Of the total 26 gene-knockout ES cell lines, 11 differentiated normally beyond IVD.D13, 8 were arrested between IVD.D11 and IVD.D13, and 7 were arrested at an earlier stage than IVD.D3 (Extended Data Fig. 2c–e). Transcriptome analysis of these knockout oocytes, except for those arrested before IVD3, revealed that the eight knockout-oocytes (*Figla*, *Sohlh1*, *Lhx8*, *Nobox*, *Stat3*, *Tbpl2*, *Dynll1* or *Sub1*-knockout oocytes) were arrested before or around PPT (Fig. 1f). These eight knockout oocytes did not show the representative features of PPT, such as a decrease in the X/A ratio, upregulation of specific retrotransposons and propensity for TATA promoter usage, with the single exception that *Dynll1*-knockout oocytes showed a decreased X/A ratio (Extended Data Fig. 3a–c). Analysis of reciprocal gene expression in each line of knockout oocytes revealed that the eight genes had a mutual effect on expression, and an imputed transcriptional network illustrated that *Lhx8*, *Sohlh1*, *Nobox* and *Tbpl2* formed a core network to which *Stat3*, *Dynll1*, *Sub1* and *Figla* were tightly connected (Extended Data Fig. 3d, e).

PPT induction by transcription factors

Next we tested whether these eight genes (hereafter referred to as PPT8) were sufficient to drive PPT and induce competence for oocyte growth. We overexpressed PPT8 in BVSCNCh-ES cells, which consisted of BVSC ES cells with the *mCherry* gene inserted into the *Npm2* locus (Extended Data Fig. 4a). Consistent with evidence that *Npm2* is specifically expressed at later oogenesis stages in vivo[28] and from IVD.D13 in vitro (Fig. 1d), NPM2–mCherry expression became visible in the nucleus of BVSCNCh-ES-cell-derived oocytes from 15 days of IVD culture (Extended Data Fig. 4b, c). Overexpression of PPT8 was controlled by the Shield1 degradation system, in which PPT8 proteins become stable after the addition of Shield1 ligand (Fig. 2a). In suspension culture with Shield1, BVSCNCh-ES cells containing PPT8 (PPT8-BVSCNCh-ES cells) immediately started to express Stella–CFP, but not BLIMP1–mVenus or NPM2–mCherry, and were deemed to cease proliferating, as the aggregations decreased in size (Fig. 2b, Extended Data Fig. 5a). These phenotypes were transgene-dependent, because no Stella–CFP expression was observed in the parental BVSCNCh-ES cells. Weak Stella–CFP expression was detected in the aggregations of PPT8-BVSCNCh-ES cells without Shield1, possibly owing to leaky expression of PPT8, but these aggregations did not decrease in size (Extended Data Fig. 5a, b). A similar phenotype was also observed in PPT8-BVSCNCh-ES cells cultured in self-renewal conditions: Stella–CFP became visible on day 1 of culture with Shield1, and the cells were no longer proliferative (Extended Data Fig. 5c). The PPT8-BVSCNCh-ES cells in aggregations with Shield1 increased in both cellular and nuclear size (Extended Data Fig. 5d) and some of the Stella–CFP-positive cells expressed DDX4 (also known as MVH), a conserved germ cell-specific marker (Extended Data Fig. 5e). At 25 days of culture, NPM2–mCherry became visible in Stella–CFP-positive cells (Fig. 2c). However, these oocyte-like cells showed irregular morphology, which in some cases included an unusual cavity in the cytoplasm.

Because follicular somatic cells are crucial for oocyte growth, we co-cultured PPT8-BVSCNCh-ES cells and E12.5 female gonadal somatic cells with Shield1, which mimicked previously described reconstituted ovaries[23] (Fig. 2d). Notably, PPT8-BVSCNCh-ES-cell-derived oocyte-like cells grew uniformly and became positive for both Stella–CFP and NPM2–mCherry, with the formation of follicle structures and expression of GDF9, an oocyte-secreted factor that is important for folliculogenesis[29], and several layers of granulosa cells (Fig. 2e–g). Notably, NPM2–mCherry was detectable from eight days of culture, which was earlier than observed in PPT8-BVSCNCh-ES cells without somatic cells (Fig. 2c), and even earlier than seen in reconstituted ovaries using PGCLCs (Extended Data Fig. 4b). PPT8-BVSCNCh-ES cells in reconstituted ovaries exhibited a rapid increase in size (Extended Data Fig. 5f). Without Shield1, parental BVSCNCh-ES cells or PPT8-BVSCNCh-ES cells proliferated extensively in reconstituted ovaries and did not form follicle-like structures (Extended Data Fig. 5g).

The transcriptomes of PPT8-BVSCNCh-ES cells with Shield1 alone or with Shield1 and gonadal somatic cells were shifted to that of P1 oocytes in vivo on day 5 of culture (Fig. 2h), which suggests that the initial transformation process was independent of the somatic cells. By contrast, the further shift of the transcriptomes of PPT8-BVSCNCh-ES cells without somatic cells after five days of culture was slower than those of PPT8-BVSCNCh-ES cells with somatic cells, which suggests that the somatic cells accelerated the PPT process. Both PPT8-BVSCNCh-ES cells with somatic cells at 14 days of culture and PPT8-BVSCNCh-ES cells without somatic cells at 25 days of culture exhibited transcriptional features beyond PPT, such as the expression of maternal factor genes and specific retrotransposons, a decrease in the X/A ratio, and a propensity for TATA promoter usage (Extended Data Fig. 5h–k). These results demonstrate that the expression of PPT8 was sufficient to induce PPT and the competence for oocyte growth directly in ES cells. These induced oocyte-like cells, hereafter termed directly induced oocyte-like cells (DIOLs), could grow to form secondary follicle structures when combined with gonadal somatic cells.

To identify a minimum set of factors sufficient for DIOL induction, we generated 66 ES cell lines, containing different combinations of the PPT8 transgenes (Extended Data Fig. 6a–d). We found that DIOLs were induced from ES cell lines containing *Nobox*, *Figla*, *Tbpl2* and *Lhx8* ('NFTL') transgenes and no DIOLs were induced from ES cell lines that lacked any of these genes. However, some ES cell lines containing NFTL transgenes showed few or no DIOL inductions (Extended Data Fig. 6d, e), possibly owing to an insufficient level and/or an inappropriate balance of the expression levels of these genes, whereas all PPT8-ES cell lines showed a robust induction of DIOLs. Therefore, although NFTL is the minimum necessary set, we used PPT8-ES cells in subsequent experiments.

To assess whether PPT8 are sufficient to induce DIOLs from somatic cells, mouse embryonic fibroblast (MEFs) were obtained from PPT8-BVSCNCh-ES cells (Extended Data Fig. 7a). However, no DIOLs were induced from the MEFs (Extended Data Fig. 7b). This suggests that a pluripotent state might be required for PPT to induce oocyte features. In keeping with this notion, DIOLs were consistently induced from induced pluripotent stem (iPS) cells reprogrammed from adult tail fibroblasts (Extended Data Fig. 7c, d). These results indicate that production of DIOLs from somatic cells is feasible through an iPS cell intermediate, emphasizing the reciprocal link between the germline cycle and the pluripotent state[30].

Oocyte growth separable from PGC fate

The trajectory of the transcriptomes in the first five days of culture suggested that DIOLs were induced from ES cells without passing through early germ cell differentiation processes before oocyte growth. The transcriptome analysis revealed that genes essential for PGC specification, such as *Prdm1* (also known as *Blimp1*), *Prdm14* and *Tfap2c*, were not expressed during DIOL induction (Extended Data Fig. 8a). Moreover, DIOLs were successfully induced from *Prdm1*-knockout PPT8-BVSCNCh-ES cells, which are unable to specify PGCs[31] (Extended Data Fig. 8b, c). These results demonstrate that DIOL induction occurred directly from ES cells without transition through a PGC intermediate.

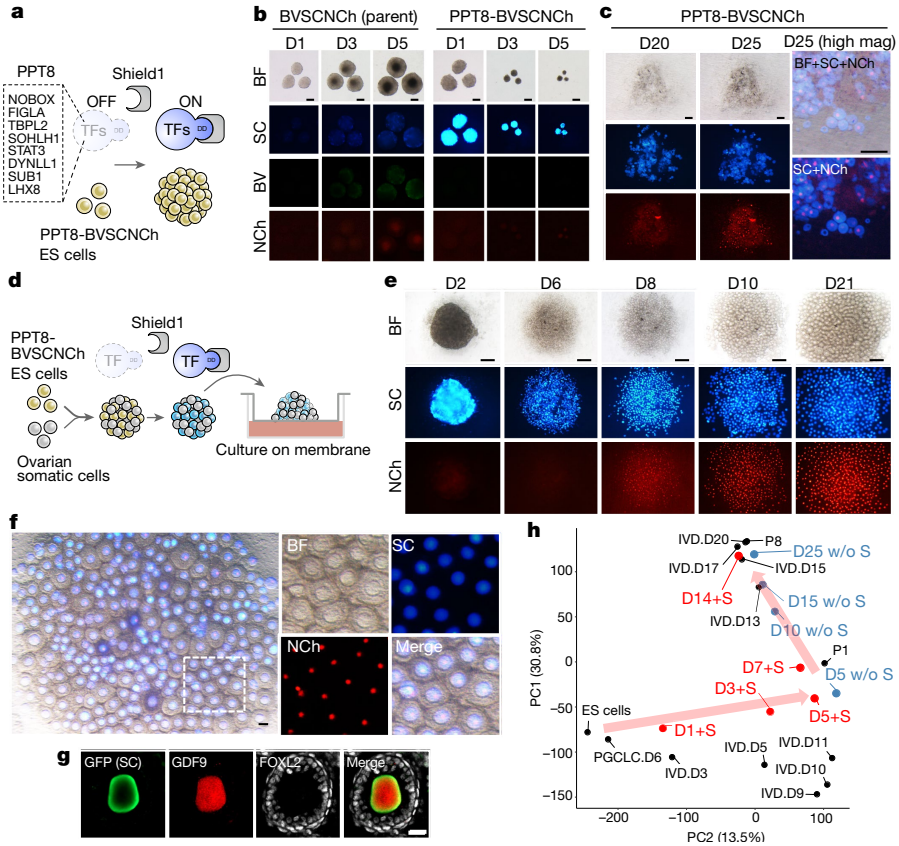


Fig. 2 | Reconstitution of PPT in ES cells. a, A schematic protocol of the overexpression of Shield1-induced PPT-associated genes. The Shield1 ligand protects the destabilization domain (DD)-tagged PPT8 proteins against endogenous proteasomes in BVSCNCh-ES cells. TFs, transcription factors. **b**, Expression of reporter genes in ES cells with PPT8 overexpression. Representative images of BVSCNCh-ES cells (parent) or PPT8-BVSCNCh ES cells at the days indicated are shown. NCh, NPM2–mCherry. Scale bars, 200 µm. **c**, Long-term culture of PPT8-BVSCNCh ES cells. Representative images of PPT8-BVSCNCh ES cells cultured at the days indicated are shown. NPM2–mCherry was clearly visible at 25 days of culture. Scale bars, 50 µm. Results in **b** and **c** were reproducible in experiments repeated more than five times. **d**, A schematic protocol of culture of PPT8-BVSCNCh ES cells with ovarian somatic cells. **e**, Formation of oocyte-like cells. Representative images of the aggregates at the days indicated are shown. NPM2–mCherry was detected from 8 days of culture. Scale bars, 200 µm. **f**, Follicle-like structure in the aggregate at 25 days of culture. Scale bar, 50 µm. Results in **e** and **f** were reproducible in experiments repeated more than ten times. **g**, Expression of marker genes in the follicle structure. Immunostaining results of the follicle structure at day 14 of culture are shown. A similar result was observed in all 26 follicles tested. Scale bar, 20 µm. **h**, PCA plot of trajectories of DIOL formation. Transcriptomes of DIOL without somatic cells (blue; days w/o S) and with somatic cells (red; days +S) are shown. The expression profile is based on biologically duplicated samples.

After specification, PGCs undergo epigenetic reprogramming, during which DNA methylation markedly decreases to 2.9% in the genome of E16.5 oocytes, and this level of DNA methylation is maintained until the non-growing oocyte stage (approximately 2%)[32–34]. Notably, methylome analysis of DIOLs at 5 days of culture with gonadal somatic cells (DIOL.D5) revealed that around 27% of CpGs in the DIOL.D5 genome were methylated, as observed in the parental PPT8-BVSCNCh-ES cell line (Extended Data Fig. 9a). The pattern of DNA methylation in DIOL.D5 mirrored that in the parental ES cell line, whereas that in non-growing oocytes in vivo showed extensive demethylation in the genome (Fig. 3a, Extended Data Fig. 9b, c). Despite a trend of severe loss of DNA methylation at differentially methylated regions (DMRs) of imprinting loci in female ES cells[35], we found that DMRs in the *H19* and *Rasgrf1* loci remained at approximately 50% methylation levels in the PPT8-ES cells and in DIOL.D5, whereas these DMRs are demethylated in non-growing oocytes in vivo (Fig. 3b). These results demonstrated that no genome-wide DNA demethylation occurred during DIOL induction.

Despite their highly methylated genome, we observed growth of DIOLs alongside the proliferation of granulosa cells under an in vitro growth (IVG) condition[23] (Fig. 3c). Transzonal projections were formed between DIOLs and the surrounding granulosa cells (Extended Data Fig. 9d). DIOLs became full-grown (fgDIOLs) with large germinal vesicles at 11 days of culture. Notably, under an in vitro maturation condition, fgDIOLs underwent germinal vesicle breakdown and bore a polar body (Fig. 3c), demonstrating that DIOLs had a potential to reach a stage morphologically similar to that of MII oocytes, which therefore we designated MII-DIOLs. During oocyte growth in vivo, a progressive gain of de novo DNA methylation is accomplished in the genome of oocytes[32–34]. Therefore, we evaluated de novo DNA methylation in fgDIOLs. Methylome analysis revealed that the pattern at high-methylation regions in fgDIOLs closely resembled that in full-grown oocytes in vivo (fgOocytes) (Fig. 3d, Extended Data Fig. 9e). By contrast, in low-methylation regions the pattern was more similar to that of DIOL.D5 than fgOocytes (Fig. 3d). The pattern of de novo DNA methylation at gene bodies, which is characteristic during oocyte growth[36], was correlated between fgDIOLs and fgOocytes ($R^2 = 0.738$); however, the absolute levels of de novo DNA methylation at these regions was lower in fgDIOLs (Fig. 3e). These results suggest that de novo DNA methylation was appropriately added to the genome during maturation of fgDIOLs, but on a background of increased global DNA methylation carried over from ES cells and DIOL.D5. Analysis of DMRs of the maternally methylated imprinting loci (*Igf2r*, *Impact*, *Mest* and *Snrpn*) revealed that

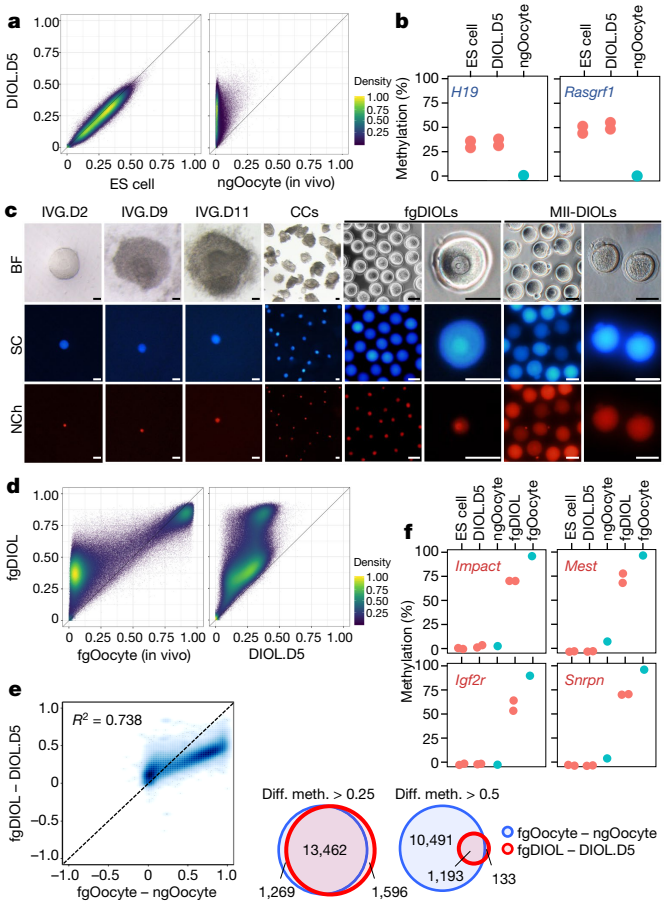


Fig. 3 | Dispensability of epigenetic reprogramming for DIOL growth and de novo methylation. a, Comparison of genome-wide patterns of DNA methylation. Each dot represents the DNA methylation level for a 10-kb window. $n = 2$, biologically independent samples. **b**, Level of DNA methylation in the imprinting loci. Graphs show the percentage of methylated CpG in the DMR region of *H19* and *Rasgrf1*. ngOocyte, non-growing oocyte. **c**, Maturation of DIOLs. Representative bright-field (BF) and fluorescence images of the follicle structure during IVG, complexes with cumulus cells (CCs), fgDIOLs and MII-DIOLs are shown. Results were reproducible in experiments repeated more than eight times. Scale bars, 50 µm. **d**, Comparison of genome-wide patterns of DNA methylation. Each dot represents the DNA methylation level for a 10-kb window. $n = 2$, biologically independent samples. **e**, Overlapped region of the gene body with de novo methylation. The *x* and *y* axis of the smooth scatter plot represent the level of de novo CpG methylation in the gene body regions in fgOocytes and fgDIOLs, respectively. Venn diagrams show the number of genes assigned to the gene body regions that are de novo CpG methylated more than 25% (left) and 50% (right). The gene bodies with de novo methylation almost overlapped between fgOocytes and fgDIOLs. The methylome profile is based on biologically duplicated samples, except in the case of ngOocytes and fgOocytes, for which single samples were used. Diff. meth., differential methylation. **f**, Level of DNA methylation in the imprinting loci. Graphs show the percentage of methylated CpG in the DMR region of maternally imprinted genes.

the appropriate pattern of de novo methylation was observed in the genome of fgDIOLs (Fig. 3f). However, the methylation level in fgDIOLs was lower than that in vivo. This could be due to the heterogeneity of DIOLs, as individual reads showed that either all or none of the CpGs were methylated (Extended Data Fig. 9f). Of note, more than half the reads in all loci showed complete DNA methylation, which suggests that a certain number of oocytes had correctly laid down maternal imprints. Together, these results demonstrate that oocyte growth represented by cytoplasmic maturation and de novo DNA methylation in the nucleus was accomplished without PGC specification or genome-wide DNA demethylation.

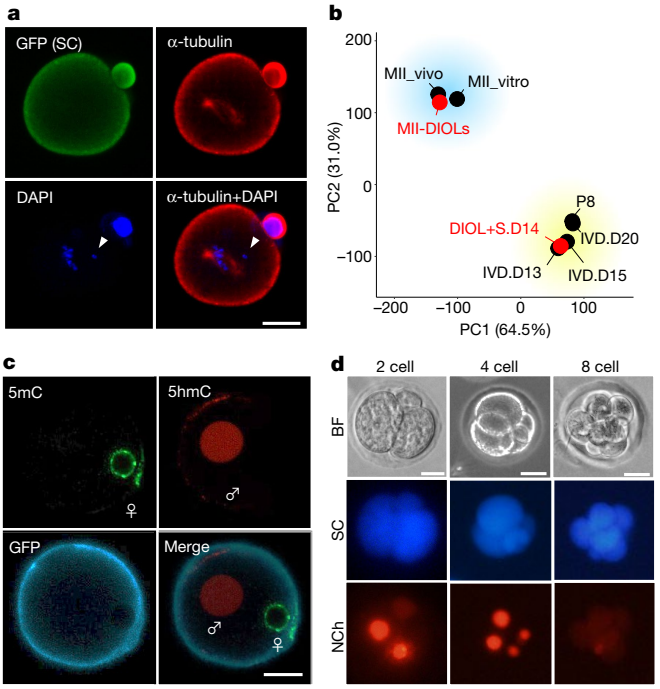


Fig. 4 | Maturation and developmental potential of DIOLs. **a**, Immunostaining analysis of spindle formation in an MII-DIOL. Stella–ECFP stained by anti-GFP antibody is a marker of the DIOL. Arrowhead indicates a mis-segregated chromosome. Spindle formation was observed in 7 out of 18 MII-DIOLs. Scale bar, 20 µm. **b**, PCA of MII-DIOLs. $n = 2$, biologically independent samples. **c**, Formation of pronuclei in a fertilized DIOL. The image shows immunostaining of maternal and paternal pronuclei that were highly methylated and hydroxymethylated, respectively. The epigenetically dimorphic pronuclei were observed in three out of five fertilized DIOLs tested. Scale bar, 20 µm. **d**, Cleavage of DIOL-derived embryos. Scale bars, 20 µm. These results were reproducible experiments repeated seven times.

Developmental potential of DIOLs

Finally, we validated the developmental potential of MII-DIOLs. Immunofluorescence analysis revealed that MII-DIOLs formed a spindle-like structure with unusual shape and location in the cytoplasm (Fig. 4a). In contrast to bivalent chromosome structure observed in oocytes in vivo after germinal vesicle breakdown, chromosomes in DIOLs exhibited univalent chromosome structures after germinal vesicle breakdown (Extended Data Fig. 10a). Live-imaging analysis showed that the chromosomes were abnormally separated in DIOLs after germinal vesicle breakdown (Supplementary Video 1), which indicates a lack of meiotic recombination in DIOLs. In line with this, genes essential for entry or progression of meiotic prophase were not upregulated during DIOL induction, except a temporal upregulation of *Sycp3* in DIOL.D5 plus somatic cells, which nevertheless resulted in aberrant protein distribution in the nuclei of DIOLs (Extended Data Fig. 10b, c). The percentage of a polar body extrusion in DIOLs (22.0%, 202 out of 917) was comparable ($P > 0.05$, Pearson's chi-square test) to that in *Stra8*-knockout oocyte-like cells (26.2%, 11 out of 42) that lacked meiosis[37] (Extended Data Fig. 10d). These observations demonstrated that DIOLs had abnormal configuration of chromosomes owing to lack of meiosis.

Notably, despite the abnormal configuration of chromosomes, the transcriptome of MII-DIOLs was comparable to that of MII oocytes derived from PGCLCs in vitro and in vivo (Fig. 4b). Genes involved in oocyte maturation, zygotic genome activation and cell cycle regulation were upregulated in MII-DIOLs (Extended Data Fig. 10e). MII-DIOLs were fertilized in vitro with wild-type sperm, forming both paternal and maternal pronuclei, which were recognized by specific antibodies

against 5-hydroxymethycytosine and 5-methylcytosine, respectively (Fig. 4c). This dimorphic epigenetic remodelling mirrors that found in normal zygotes, which suggests that key epigenetic regulatory mechanisms are established in MII-DIOLs. At 24 h after in vitro fertilization, 14.4 ± 9.7% (mean ± s.d.) of fertilized DIOLs developed to two-cell embryos (Fig. 4d). At 48 h and 72 h, 6.5 ± 5.6% and 3.6 ± 3.5% of embryos progressed to the 4- and 8-cell stages, respectively. However, further development of DIOL-derived embryos was severely compromised, probably owing to aneuploidy. Nevertheless, DIOLs had a capacity for fertilization, the formation of pronuclei with appropriate epigenetic configuration, and subsequent cleavage divisions. These results show that reconstitution of the gene regulatory network by defined transcription factors was sufficient to confer the competence of oocyte growth on ES cells.

Discussion

In this study, we identified a set of transcription factors that drive PPT and can convert pluripotent stem cells into DIOLs. The DIOL system indicates that DNA demethylation is not a prerequisite to activate the gene-regulatory network driving oocyte growth. As such, our results complement recent studies that suggest that the principle role of epigenetic reprogramming is to activate the meiotic program[38,39]. We therefore revealed the distinctive control mechanisms regulating meiotic entry versus oocyte growth. In addition, our observation that de novo DNA methylation during oocyte growth is independent from epigenetic reprogramming in PGCs provides insight into the epigenetic control in oocytes and is also of importance in itself, because de novo DNA methylation is crucial for conferring developmental competence to the oocyte nucleus[40,41]. Possible roles of transcription factors in DIOL induction are further discussed in Supplementary Information.

The oocyte cytoplasm is a unique material that is highly valuable for applications in assisted reproductive technology, such as mitochondrial replacement therapy, and somatic cell nuclear transfer technology. However, the bottleneck of these applications is cytoplasts obtained from oocyte donation, which requires an invasive process with associated risk factors. This study raises the possibility that DIOLs could provide an alternative source of the cytoplast. As DIOLs can be produced from pluripotent stem cells by transgenes, which are eventually discarded with the nucleus, the cytoplasm of DIOLs will be a powerful tool to promote our understanding of the mechanisms that underlie reprogramming in somatic cell nuclear transfer and of their potential application in assisted reproductive technologies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-3027-9.

1. Pan, H., O'Brien, M. J., Wigglesworth, K., Eppig, J. J. & Schultz, R. M. Transcript profiling during mouse oocyte development and the effect of gonadotropin priming and development in vitro. *Dev. Biol.* **286**, 493–506 (2005).
2. Saitou, M. & Yamaji, M. Primordial germ cells in mice. *Cold Spring Harb. Perspect. Biol.* **4**, a008375 (2012).
3. Nicholls, P. K. et al. Mammalian germ cells are determined after PGC colonization of the nascent gonad. *Proc. Natl Acad. Sci. USA* **116**, 25677–25687 (2019).
4. McLaren, A. & Southee, D. Entry of mouse embryonic germ cells into meiosis. *Dev. Biol.* **187**, 107–113 (1997).
5. Schultz, R. M., Letourneau, G. E. & Wassarman, P. M. Program of early development in the mammal: changes in the patterns and absolute rates of tubulin and total protein synthesis during oocyte growth in the mouse. *Dev. Biol.* **73**, 120–133 (1979).
6. Sternlicht, A. L. & Schultz, R. M. Biochemical studies of mammalian oogenesis: kinetics of accumulation of total and poly(A)-containing RNA during growth of the mouse oocyte. *J. Exp. Zool.* **215**, 191–200 (1981).
7. Soyal, S. M., Amleh, A. & Dean, J. FIGalpha, a germ cell-specific transcription factor required for ovarian follicle formation. *Development* **127**, 4645–4654 (2000).
8. Pangas, S. A. et al. Oogenesis requires germ cell-specific transcriptional regulators Sohlh1 and Lhx8. *Proc. Natl Acad. Sci. USA* **103**, 8090–8095 (2006).
9. Choi, Y., Yuan, D. & Rajkovic, A. Germ cell-specific transcriptional regulator sohlh2 is essential for early mouse folliculogenesis and oocyte-specific gene expression. *Biol. Reprod.* **79**, 1176–1182 (2008).
10. Choi, Y., Ballow, D. J., Xin, Y. & Rajkovic, A. Lim homeobox gene, lhx8, is essential for mouse oocyte differentiation and survival. *Biol. Reprod.* **79**, 442–449 (2008).
11. Rajkovic, A., Pangas, S. A., Ballow, D., Suzumori, N. & Matzuk, M. M. NOBOX deficiency disrupts early folliculogenesis and oocyte-specific gene expression. *Science* **305**, 1157–1159 (2004).
12. Falender, A. E., Shimada, M., Lo, Y. K. & Richards, J. S. TAF4b, a TBP associated factor, is required for oocyte development and function. *Dev. Biol.* **288**, 405–419 (2005).
13. Grive, K. J., Seymour, K. A., Mehta, R. & Freiman, R. N. TAF4b promotes mouse primordial follicle assembly and oocyte survival. *Dev. Biol.* **392**, 42–51 (2014).
14. Griffith, G. J. et al. Yin-yang1 is required in the mammalian oocyte for follicle expansion. *Biol. Reprod.* **84**, 654–663 (2011).
15. Gazdag, E. et al. TBP2 is essential for germ cell development by regulating transcription and chromatin condensation in the oocyte. *Genes Dev.* **23**, 2210–2223 (2009).
16. Choi, Y. et al. Microarray analyses of newborn mouse ovaries lacking Nobox. *Biol. Reprod.* **77**, 312–319 (2007).
17. Joshi, S., Davies, H., Sims, L. P., Levy, S. E. & Dean, J. Ovarian gene expression in the absence of FIGLA, an oocyte-specific transcription factor. *BMC Dev. Biol.* **7**, 67 (2007).
18. Shin, Y. H. et al. Transcription factors SOHLH1 and SOHLH2 coordinate oocyte differentiation without affecting meiosis I. *J. Clin. Invest.* **127**, 2106–2117 (2017).
19. Wang, Z., Liu, C. Y., Zhao, Y. & Dean, J. FIGLA, LHX8 and SOHLH1 transcription factor networks regulate mouse oocyte growth and differentiation. *Nucleic Acids Res.* **48**, 3525–3541 (2020).
20. Choi, Y. & Rajkovic, A. Characterization of NOBOX DNA binding specificity and its regulation of Gdf9 and Pou5f1 promoters. *J. Biol. Chem.* **281**, 35747–35756 (2006).
21. Choi, M. et al. The oocyte-specific transcription factor, Nobox, regulates the expression of Pad6, a peptidylarginine deiminase in the oocyte. *FEBS Lett.* **584**, 3629–3634 (2010).
22. Park, M. et al. Identification and characterization of LHX8 DNA binding elements. *Dev. Reprod.* **16**, 379–384 (2012).
23. Hikabe, O. et al. Reconstitution in vitro of the entire cycle of the mouse female germ line. *Nature* **539**, 299–303 (2016).
24. Da Silva-Buttkus, P. et al. Effect of cell shape and packing density on granulosa cell proliferation and formation of multiple layers during early follicle development in the ovary. *J. Cell Sci.* **121**, 3890–3900 (2008).
25. Schultz, R. M., Stein, P. & Svoboda, P. The oocyte-to-embryo transition in mouse: past, present, and future. *Biol. Reprod.* **99**, 160–174 (2018).
26. Ram, P. T. & Schultz, R. M. Reporter gene expression in G2 of the 1-cell mouse embryo. *Dev. Biol.* **156**, 552–556 (1993).
27. Davis, W., Jr & Schultz, R. M. Developmental change in TATA-box utilization during preimplantation mouse development. *Dev. Biol.* **218**, 275–283 (2000).
28. Burns, K. H. et al. Roles of NPM2 in chromatin and nucleolar organization in oocytes and embryos. *Science* **300**, 633–636 (2003).
29. Dong, J. et al. Growth differentiation factor-9 is required during early ovarian folliculogenesis. *Nature* **383**, 531–535 (1996).
30. Leitch, H. G. & Smith, A. The mammalian germline as a pluripotency cycle. *Development* **140**, 2495–2501 (2013).
31. Zhang, J. et al. OTX2 restricts entry to the mouse germline. *Nature* **562**, 595–599 (2018).
32. Smallwood, S. A. et al. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat. Genet.* **43**, 811–814 (2011).
33. Kobayashi, H. et al. High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Res.* **23**, 616–627 (2013).
34. Shirane, K. et al. Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. *PLoS Genet.* **9**, e1003439 (2013).
35. Yagi, M. et al. Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation. *Nature* **548**, 224–227 (2017).
36. Veselovska, L. et al. Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biol.* **16**, 209 (2015).
37. Dokshin, G. A., Baltus, A. E., Eppig, J. J. & Page, D. C. Oocyte differentiation is genetically dissociable from meiosis in mice. *Nat. Genet.* **45**, 877–883 (2013).
38. Yamaguchi, S. et al. Tet1 controls meiosis by regulating meiotic gene expression. *Nature* **492**, 443–447 (2012).
39. Hill, P. W. S. et al. Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. *Nature* **555**, 392–396 (2018).
40. Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B. & Bestor, T. H. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294**, 2536–2539 (2001).
41. Kaneda, M. et al. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**, 900–903 (2004).

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment unless stated otherwise.

Animals and cells

All animal experiments were approved by a relevant committee (Kyushu University, approved numbers A28-109-4 and 26-74). All animal experiments followed all relevant guidelines and regulations. Female ICR mice for collection of blastocysts and gonadal somatic cells were purchased from Japan SLC. BVSC H18 ES cells[23] were used as wild-type controls. iPS cells were obtained from tail-tip fibroblasts of 10-week-old female BVSC mice (129svj × C57BL/6) by introducing retroviral vectors containing *Pou5f1*, *Sox2*, *Klf4* and *Myc*[23]. All ES cell lines were cultured in 2i+LIF with a reduced concentration of PD0325901[42].

Collection of germ cells, oocytes and DIOLs

Ovaries and reconstituted ovaries at each developmental stage were treated with CTK solution (0.1 mg ml^{-1} collagenase IV, 0.25% trypsin, 20% KSR and 1 mM $CaCl_2$ in PBS) for 30 min at 37 °C, then transferred into Accutase (Nacalai Tesque) and incubated at 37 °C for 15 min. The IVD.D11, IVD.D13 or DIOLs except DIOL.D14 plus somatic cells were isolated by FACSAria II (BD Bioscience) and the IVD.D21 or DIOL.D14 plus somatic cells were manually collected under a stereotype microscope. For FACS sorting, the cells were dissociated by gentle pipetting and filtered by using a 70-μm-pored nylon mesh to remove cell clumps. FSC-A and SSC-A scatter plot were visualized by FACS_Diva v.6.1.3 and FlowJo v.10.6.1 and used to separate cell events from debris and/or dead cells. For manual collection, oocytes were picked-up using a glass capillary.

Transcriptome analysis

For transcriptome analyses, P8, P14 and P21 oocytes, IVD.D23 oocytes, and oocytes from each knockout line of ES cells and DIOLs were transferred into a 1.5-ml DNA LoBind tube (Eppendorf). For each sample, poly(A) RNAs were extracted from more than 300 P8 oocytes, 300 P14 oocytes, 100 P21 oocytes, 400 IVD.D23, 200 oocytes from each KO ES cell line, 300 DIOL.D1–D7 plus somatic cells, 150 DIOL.D14 plus somatic cells, 3,000 DIOL.D5, 500 DIOL.D10, 400 DIOL.D15, 75 DIOL.D25, and 30 MII-DIOLs by using a Dynabead mRNA DIRECT Micro Purification Kit (Thermo). Biologically duplicated (knockout oocytes of each knockout ES cell line, and DIOLs) or triplicated (P8, P14 and P21 oocytes, and IVD.D23 oocytes) samples were prepared in each stage. Directional RNA-seq libraries were prepared as previously described[43]. HiSeq 2500 (Illumina) or Next-Seq was used to perform single-end sequencing according to the manufacturer's instructions. The RNA-seq data of the two-cell embryo were downloaded by DRA002400[43]. RNA-seq data of in vivo and in vitro female germ cells were downloaded from GSE128305[44]. Obtained reads were then mapped to the mouse mm10 genome using STAR v.2.6.0. The reads mapped onto each feature were counted by using featureCounts of the subread package v1.6.0[45]. The principal component analysis was performed using R software with FactoMineR v.2.0[46] based on TPM values. For identification of differentially expressed genes (DEGs), we performed an exact test for the negative binomial distribution using the edgeR library v.3.28.0[47]. The DEGs were defined as genes exhibiting more than fourfold difference between the samples (FDR < 0.001 and $\log_2$CPM >5 as the mean of the expression level in the group). The DAVID database (https://david.ncifcrf.gov/) or clusterProfiler v.3.14.3 was used for gene ontology analysis[48,49]. For analysis of the motif enrichment, MEME v.4.12.0[50] was used with the following parameters: -mod zoops -nmotifs 3 -minw 6 -maxw 30 -dna -revcomp -nostatus. Heat maps were generated by the pheatmap package v.1.0.12 (https://CRAN.R-project.org/package=pheatmap).

Establishment of knockout ES cells

For each guide RNA, DNA oligonucleotides were annealed and then inserted into BbsI-digested pX330 (Addgene 42230). For transfection, the pX330 and a vector expressing the puromycin-resistance gene were co-transfected into BVSC H18 ES cells by Lipofectamine 2000. Puromycin (1 μg ml^{-1}) of was added to the 2i+LIF medium at 1 day after transfection, and the culture was continued for an additional 2 days. At 3 days of culture after the transfection, ES cell colonies were dissociated and re-seeded on MEF-coated 6-cm culture dishes. After culture for 5 days, the colonies were picked up and cultured on MEF and gelatin-coated 96-well plates. DNA was extracted from the cells on gelatin-coated 96-well plates and used to determine the genotype by PCR. The primers were designed for regions flanking the target site. All DNA oligonucleotides used in this study are listed in Supplementary Table 1. Amplified DNA fragments were analysed by the indel detection by amplicon analysis (IDAA) method as previously described[51,52]. For knockout of *Nobox*, whole exons were removed by two gRNAs.

Establishment of BVSCNCh-ES cells

For the construction of the knock-in vector for NPM2–mCherry ES cells, the 5′ and 3′ regions of *Npm2* were amplified from the H18 ES cell genome and inserted into a vector containing mCherry, thymidine kinase and the *loxP*-Neo cassette. The knock-in vector was transfected into BVSC H18 ES cells with Lipofectamine 2000. G418 (400 ng ml^{-1}) of was added to the culture medium at 1 day after transfection, and the condition was continued for 5 days. At 5 days, ES cell colonies were picked up and spread onto MEF- and gelatin-coated 96-well plates. DNA was extracted from the cells on gelatin-coated 96-well plates and used to determine the genotype by PCR. All primers used in this study are listed in Supplementary Table 1. Uncropped images of the gels are shown in Supplementary Information.

Establishment of Shield1-inducible transgenic ES cells

For construction of the pPB-CAG-DD vector, the CAG promoter and destabilized domain (DD) were cloned from the CAG-DD-hTFAP2C plasmid[53] and inserted into the PiggyBAC vector used previously[44]. For construction of the Shield1-inducible cDNA overexpression system, all cDNAs were amplified from the cDNA of mouse P8 ovaries by PCR using KOD Fx Neo or KOD Plus Neo DNA polymerase (Toyobo) and cloned into a pPB-CAG-DD vector with an In-Fusion HD cloning kit (Takara Bio) according to the manufacturer's protocol. The sequences of all fragments were confirmed by Sanger sequencing. These pPB-CAG-DD plasmids and pCMV-hyPBase[54] (Wellcome Trust Sanger Institute) were co-transfected to BVSCNCh-ES cells by using Lipofectamine 2000 (Thermo). After 24 h, ES cells were cultured with 1 μg ml^{-1} puromycin (InvivoGen). After 3 days, ES cells were re-spread onto MEFs. At 5 days of culture, single colonies were picked up and seeded on MEFs. To remove MEFs, we passaged these ES cells under feeder-free conditions at least three times. MEF-removed ES cells were subjected to qPCR analysis for estimation of the copy numbers of each transgene. Primers amplifying both the endogenous and exogenous *Kit* promoter, and those amplifying the *Il2* gene locus were used as a reference. qPCR was performed by using PowerSYBR Green PCR Master Mix (Thermo) on a CFX384 Real-Time PCR Detection System (BioRad), and the relative copy number of the transgene was calculated according to the following formula: $\Delta C_t = C_t$ value of *Il2* locus $- C_t$ value of each transgene.

Weighted gene co-expression network analysis

To avoid noise, the lowly expressed genes (less than 50 reads per kilobase of transcript, per million mapped reads (RPKM) in all stages) were excluded from downstream analysis. Using transcriptome gene expression data, genes were assigned to co-expression modules that were given arbitrary colour names by WGCNA v.1.61[55,56]. For this WGCNA, the transcriptome data of P3, P4 small and P6 small oocytes[44] were

added. For the construction of co-expression modules, we set the following options: power = 12, TOMType = 'signed', minModuleSize = 20, reassignThreshold = 0, mergeCutHeight = 0.12, numericLabels = TRUE, pamRespectsDendro = FALSE, saveTOMs = F, maxBlockSize = 24000, verbose = 3, and networkType = 'signed hybrid'. We chose five candidate modules whose genes were specifically expressed around PPT (MEcyan, MEdarkred, MEmagenta, MEwhite and MEpurple). For the extraction of candidate PPT-associated transcription factors, we selected genes belonging to the GO term, GO:0006351 'transcription, DNA-templated'. We selected genes that were specifically and highly expressed in IVD.D13 (RPKM in IVD.D13 > 50, s.d. > 20 for purple, magenta, and white; RPKM in IVD.D13 > 200, s.d. > 20 for MEcyan and MEdarkred).

Analysis of the X chromosome number

The DNA fragments amplified by PCR from the X chromosome of the C57BL/6 strain were sensitive to SfaN1 digestion, whereas those from the 129+Ter/svj strain were insensitive due to the SNP difference. The amplified DNA fragments were 1,064 bp in length, and those derived from the C57BL/6 strain were digested into 845 bp and 219 bp fragments. PCR products were electrophoresed on an agarose gel to identify the pair of X chromosomes in the ES cells containing a C57BL/6 × 129+Ter/svj background.

PGCLC induction

All cultures in this study were performed under a normoxic condition (5% CO_2 and air) at 37 °C. PGCLC induction was performed as previously described[57]. PGCLCs were purified by FACSAria II or Fusion (BD Bioscience) and aggregated with E12.5 female gonadal somatic cells in a low-binding U-bottom 96-well plate (NUNC) for 2 days of culture in GK15 medium supplemented with 1 μM retinoic acid. To strictly remove residual PGCs from dissociated gonadal cells, both SSEA1 and CD31 antibodies (Miltenyi Biotech) were used according to the manufacturer's instructions. Approximately 5,000 PGCLCs were cultured with 75,000 gonadal somatic cells to produce one reconstituted ovary.

IVD culture of PGCLCs

IVD culture was performed as previously described[23]. Reconstituted ovaries (5,000 PGCLCs and 75,000 gonadal somatic cells) were placed on Transwell-COL membranes soaked in αMEM-based IVDi medium: αMEM supplemented with 2% FCS, 150 μM ascorbic acid (Sigma), 1 × Glutamax, 1 × penicillin/streptomycin and 55 μM 2-mercaptoethanol (Life Technologies). At 4 days of culture, the culture medium was changed to a StemPro-34-based IVDi medium: StemPro-34 SFM (Life Technologies) supplemented with 10% FCS, 150 μM ascorbic acid, 1 × Glutamax, 1 × penicillin/streptomycin, and 55 μM 2-mercaptoethanol. From 7 to 10 days of culture, 500 nM ICI182780 was added to the StemPro-34-based IVDi medium.

In vitro growth and maturation of PGCLC-derived oocytes

IVG culture was performed as described previously[23]. In brief, the isolated secondary follicles on the Transwell-COL membranes were soaked in IVG medium: αMEM supplemented with 5% FCS, 2% polyvinylpyrrolidone (Sigma), 150 μM ascorbic acid, 1 × Glutamax, 1 × penicillin/streptomycin, 100 μM 2-mercaptoethanol, 55 μg ml^{-1} sodium pyruvate (Nacalai Tesque), 0.1 IU ml^{-1} follicle-stimulating hormone (Follistim; MSD), 15 ng ml^{-1} BMP15 and 15 ng ml^{-1} GDF9 (R&D Systems). At 2 days of culture, BMP15 and GDF9 were withdrawn from the medium and then follicles were incubated in 0.1% type IV collagenase (MP Biomedicals). After washing with αMEM supplemented with 5% FCS several times, the follicles were cultured in IVG-αMEM without BMP15 and GDF9. At 11 days of culture, cumulus–oocyte complexes grown on the membrane were picked up by a fine glass capillary, and transferred to in vitro maturation medium: αMEM supplemented with 5% FCS, 25 μg ml^{-1} sodium pyruvate, 1 × penicillin/streptomycin, 0.1 IU ml^{-1} follicular-stimulating hormone, 4 ng ml^{-1} EGF, and 1.2 IU ml^{-1} hCG (gonadotropin; ASKA). At 16 h of culture, swollen cumulus cells were stripped from the oocytes by treating with hyaluronidase (Sigma), and then MII oocytes were determined by the first polar body.

DIOL induction

For the short-term induction of DIOLs without somatic cells, 2,000 ES cells were transferred into a low-cell-binding U-bottom 96-well plate (NUNC) in S10 medium (StemPro-34 SFM (Life Technologies) supplemented with 10% FCS, 150 μM ascorbic acid, 1 × Glutamax, 1 × penicillin/streptomycin and 55 μM 2-mercaptoethanol) with 150 ng ml^{-1} SCF, 0.5 μM of Shield1 (Clontech), and 10 nM Y-27632 (Wako), and then cultured for 5 days. For the long-term induction of DIOLs without somatic cells, 50,000 ES cells were transferred into the low-cell-binding U-bottom 96-well plate in the same medium as used in the short-term induction. At 2 days of culture, aggregates were placed on Transwell-COL membranes (Corning) in S10 medium with 0.5 μM of Shield1 and cultured for 2–4 weeks. For induction of DIOLs with somatic cells, 50,000 ES cells and 30,000–75,000 E12.5 female gonadal somatic cells were transferred into the low-cell-binding U-bottom 96-well plate in S10 medium with 0.5 μM of Shield1. At 2 days of culture, aggregates were placed on Transwell-COL membranes and cultured for 28 days with Shield1. For IVG culture of DIOLs, individual follicles were manually isolated using a sharpened tungsten needle. The isolated DIOLs were cultured on Transwell-COL membranes in IVG medium. At 2 days of culture, BMP15 and GDF9 were withdrawn from the medium and then follicles were incubated in 0.1% type IV collagenase. After washing with αMEM supplemented with 5% FCS several times, the follicles were cultured in IVG-αMEM without BMP15 and GDF9 for 7–11 days. In vitro maturation and fertilization were performed as described above and previously[23].

Preparation of PPT8 MEFs

Mix8_6 ES cells were injected into blastocysts obtained from mating between ICR mice, and the blastocysts were transferred into the uterus of pseudopregnant females at 2.5 days post-coitum. At 11 days after the embryo transfer, chimaera embryos were minced, digested by trypsin-EDTA (Invitrogen), and then cultured in DMEM containing 10% FBS supplemented with 2 mM glutamine and 50 U ml^{-1} penicillin/streptomycin and 1 μg ml^{-1} of puromycin for 10 days to select MEFs derived from Mix8_6 ES cells having a puromycin-resistance gene. After the selection, Mix8_6 MEFs were cultured for one more passage before being used for DIOL induction.

Immunofluorescence analysis

For whole-mount analysis, reconstituted ovaries were fixed in 4% paraformaldehyde for 1 h at room temperature, washed with PBST (0.2% Tween20), soaked in blocking buffer (PBS containing 0.1% BSA and 0.3% Triton X-100) overnight at 4 °C and then incubated with primary antibodies diluted with blocking buffer overnight at 4 °C. The samples were washed with washing buffer (PBS containing 0.3% Triton X-100), incubated with secondary antibodies and DAPI overnight at 4 °C, and washed and mounted in Fluoro-KEEPER antifade reagent (Nacalai Tesque). For phalloidin staining, Cytopainter Phalloidin-iFluor 647 reagent (ab176759; Abcam) was added to the secondary antibody at a dilution of 1:500. All samples were analysed with an LSM700 or LSM800 confocal microscope (Zeiss). The antibodies used in this study are listed in Supplementary Table 2.

Live cell imaging

Live cell imaging was performed as previously described with some modifications[58]. mRNA for labelling each protein was synthesized in vitro using linearized template DNA plasmids with a mMESSAGE mMACHINE KIT (Ambion). Oocytes were microinjected with mRNAs (1.6 pl of 650 ng μl^{-1} EGFP-MAP4 and 0.6 pl of 350 ng μl^{-1} H2B-mCherry), cultured for 2.5–3.0 h, and then released to IBMX-free M2 medium for meiotic maturation. To obtain images, an LSM880 confocal microscope

equipped with a 40× C-Apochromat 1.2NA water immersion objective lens (Carl Zeiss) was controlled by a multi-position autofocus macro (https://www-ellenberg.embl.de/resources/microscopyautomation). For spindle imaging, 11 confocal *z*-sections (every 3.0 µm) of 512 × 512 pixel *xy* images covering a total volume of 53.2 × 53.2 × 33.0 µm were acquired at 5-min intervals for at least 14 h after induction of meiotic resumption.

Whole-genome bisulfite sequencing by PBAT

DIOL.D5 were FACS sorted by Stella-CFP fluorescence. fgDIOLs were collected by a fine glass capillary under a stereotype microscope. We confirmed that all fgDIOLs had Stella–CFP fluorescence under a fluorescent microscope. For preparation of each sample, DNA was extracted from 100,000 ES cells, 20,000 or 30,000 DIOL.D5, and 500 fgDIOLs using a QIAamp DNA Micro Kit (Qiagen). For each developmental stage, biologically duplicated samples were prepared. Extracted DNA was subjected to the PBAT library construction steps, as previously described[59]. Libraries were sequenced with HiSeq X Ten at Macrogen Japan. Obtained reads were mapped to the mm10 genome by BMap. PBAT data of non-growing oocytes and full-grown oocytes were downloaded from DRP000598. The Metilene package[60] was used to identify DMRs. DMRs were defined as the regions whose DNA methylation levels showed more than a 20% difference between the comparison pairs with *P* values less than 0.01 in a two-dimensional Kolmogorov–Smirnov test. For the locally estimated scatter plot smoothing regression on DNA methylation levels across chromosome 15, DNA methylation levels of 100-kb windows were used for the fitting. For analysis of methylation during the growth of oocytes and DIOLs, gene body methylation levels were defined by averaging the methylation levels of gene body CG sites covered by at least five reads in individual samples. De novo methylation was defined by subtracting the gene body methylation levels in ngOocytes and DIOL.D5 from those in fgOocytes and fgDIOLs, respectively.

Oocyte number counting and size measurement

Oocyte areas and numbers were quantified using ImageJ software v.1.52 (National Institutes of Health). In ImageJ, we binarized CFP images and defined Stella–CFP-positive regions as oocyte areas. The total area of each Stella–CFP-positive region was measured and the number of regions was counted.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The RNA-seq and methylome data have been deposited at the Gene Expression Omnibus (GEO) database under accession number GSE143218 and GSE143219, and DDBJ Sequence Read Archive (DRS001541 and DRS001547). There is no restriction on data availability. Source data are provided with this paper.

Code availability

Custom code used in this article can be accessed at https://github.com/nhamazaki/2020_DIOL.

42. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532 (2011).
43. Hamazaki, N., Uesaka, M., Nakashima, K., Agata, K. & Imamura, T. Gene activation-associated long noncoding RNAs function in mouse preimplantation development. *Development* **142**, 910–920 (2015).
44. Shimamoto, S. et al. Hypoxia induces the dormant state in oocytes through expression of *Foxo3*. *Proc. Natl Acad. Sci. USA* **116**, 12321–12326 (2019).
45. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
46. Lê, S., Josse, J. & Husson, F. FactoMineR: an R package for multivariate analysis. *J. Stat. Soft.* **25**, 1–18 (2008).
47. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
48. Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocols* **4**, 44–57 (2009).
49. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
50. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
51. Yang, Z. et al. Fast and sensitive detection of indels induced by precise gene targeting. *Nucleic Acids Res.* **43**, e59 (2015).
52. Lonowski, L. A. et al. Genome editing using FACS enrichment of nuclease-expressing cells and indel detection by amplicon analysis. *Nat. Protocols* **12**, 581–603 (2017).
53. Kobayashi, T. et al. Principles of early human development and germ cell program from conserved model systems. *Nature* **546**, 416–420 (2017).
54. Yusa, K. et al. Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. *Nature* **478**, 391–394 (2011).
55. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, 17 (2005).
56. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
57. Hayashi, K. & Saitou, M. Generation of eggs from mouse embryonic stem cells and induced pluripotent stem cells. *Nat. Protocols* **8**, 1513–1524 (2013).
58. Yoshida, S., Sakakibara, Y. & Kitajima, T. S. Live imaging of intracellular dynamics during meiotic maturation in mouse oocytes. *Methods Mol. Biol.* **1457**, 241–251 (2016).
59. Miura, F., Enomoto, Y., Dairiki, R. & Ito, T. Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res.* **40**, e136 (2012).
60. Jühling, F. et al. metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome Res.* **26**, 256–262 (2016).

Acknowledgements We thank all members of the Hayashi laboratory for their support and input. We are grateful to F. Arai, T. Matsuda and T. Ishiuchi for technical support and Y. Hayashi for comments. We thank the staff of the Research Support Center, Research Center for Human Disease Modeling, Kyushu University Graduate School of Medical Sciences for their technical assistance; we particularly thank M. Amago for support with the FACS sorting. We are grateful to Y. Hamazaki and KN international for proofreading. This study was supported in part by KAKENHI Grants-in-Aid from MEXT, Japan (numbers 17H01395, 18H05544 and 18H05545 to K.H., 15H06475, 16K18816, 16H06279 and 18K14605 to N. Hamazaki and 16H06527 to K.N.); by Management Expenses Grants of Kyushu University (K.H.); by the Advanced Computational Scientific Program of the Research Institute for Information Technology, Kyushu University; by the Uehara Memorial Foundation (K.H.); by the Takeda Science Foundation (K.H.); by a Hayashi Grant-in-Aid for Basic Medical Research (K.H.); by a JSPS Fellowship (N. Hamazaki); by the Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research) from AMED (JP19am0101103, 1804) (K.H.); by a Grant-in-Aid from The Open Philanthropy Project (K.H.); by MRC core funding (H.G.L.); and by a BBSRC grant (BB/R002703/1) (H.G.L.). H.G.L. is an Academic Clinical Lecturer and acknowledges support from the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre (BRC). The authors apologize to colleagues whose work could not be cited owing to length limitations.

Author contributions N. Hamazaki and K.H. conceived and designed the project. N. Hamazaki and C.H. performed the molecular experiments, cellular experiments and analysis of RNA-seq data. O.H., S.S. and N. Hamazaki collected RNA-seq samples. N. Hamada and N. Hamazaki performed the immunostaining. F.M., H.A., T.I. and N. Hamazaki performed PBAT library construction and analysis of data. H.K. and T.S.K. performed live imaging of oocytes. N. Hamazaki, K.N., H.G.L. and K.H. prepared the figures and wrote the manuscript, incorporating feedback from all the authors.

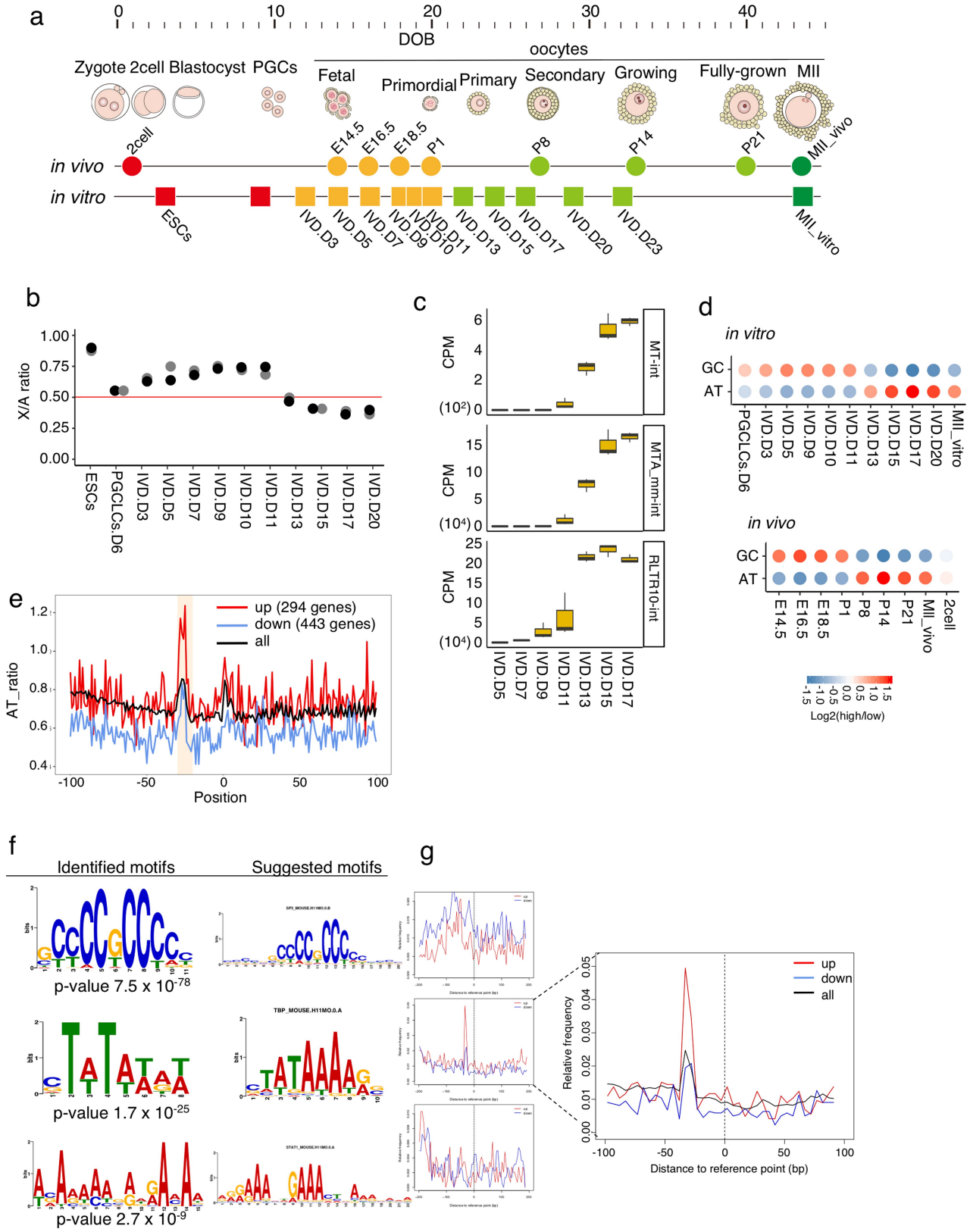
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-3027-9.

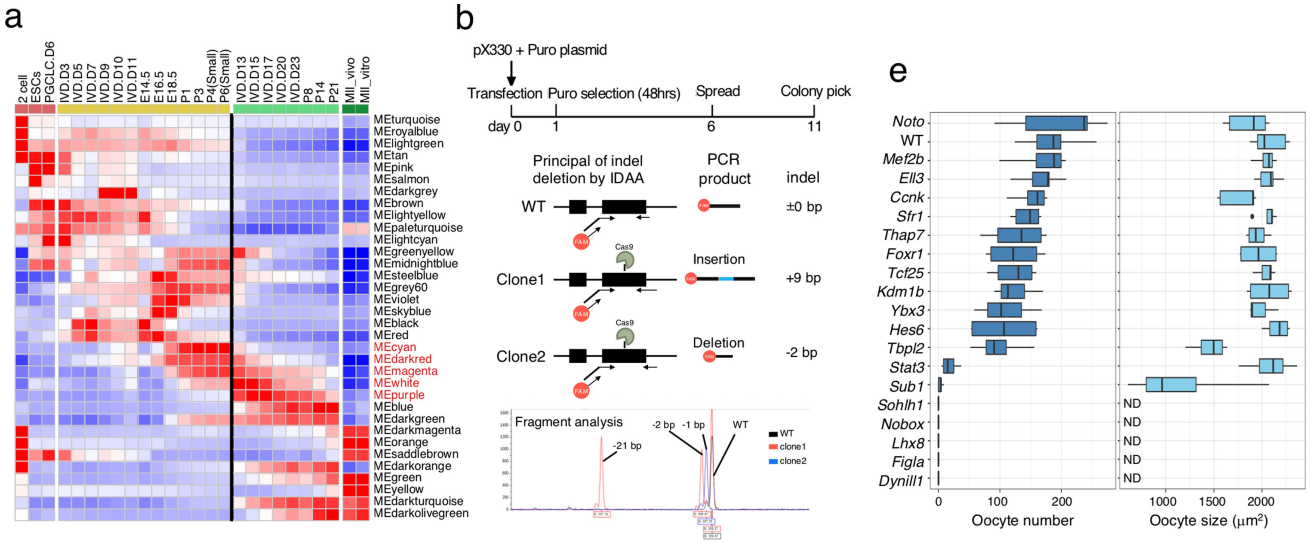
Correspondence and requests for materials should be addressed to N.H. or K.H.

Peer review information *Nature* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.



Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Transcriptional signatures in PPT. a, Schematic diagram of the transcriptome analysis. Each stage of the female germ line in vivo (circles) and in vitro (rectangles) was subjected to analysis. The top scale shows days of differentiation. The colour of the points corresponds to that in the PCA plot in Fig. 1a. DOB, day of birth. **b**, Dynamics of the X/A ratio. The ratio was determined by dividing the transcripts of X chromosomes by the average of the transcripts of autosomes. The putative half-active value (0.5) is shown by a red line. The expression profile is based on at least biologically duplicated samples. **c**, Expression dynamics of LTR families in oogenesis. The centre of the box plot is the median and the box correspond to the interquartile range (25th and 75th for bottom and top edge of the box, respectively), the distance between the first and third quartiles, the whiskers extend no more than 1.5 times the interquartile range. The expression profile shown is based on biologically triplicated samples. **d**, The ratio of expression levels of the AT-high-genes/AT-low-genes and GC-high/GC-low genes. Genes expressing any of the stages of the female germ line were classified into groups by the nucleotide compositions of their core promoters (−100 to −1 nt). High genes and low genes possessed higher AT or GC numbers compared to the (median +10) values and lower AT or GC numbers compared to the (median −10) values, respectively. **e**, Enrichment of the AT-rich sequences around the transcription start sites (TSSs). Shown are the AT ratios at −100 to +100 bp of the TSSs of genes whose expressions were up- or downregulated more than fourfold between IVD.D11 and IVD.D13. Local enrichment was observed at 20–30 bp upstream of the TSSs of these genes (orange box). Note that the TATA-boxes are known to be enriched at −20 to −30 bp of the TSS. **f**, Motif analysis of the sequence around the TSSs. Shown are de novo searched motifs within −40 bp to 0 bp of the upregulated genes (left) and similar motifs suggested by TOMTOM (right). *P* values were computed by a two-sided Fisher's exact test for enrichment of the motif sequences. For the correction of multiple comparisons, *P* values are multiplied by the number of candidate motifs tested. **g**, Enrichment of the motifs. Shown are the distributions of each motif shown in **f** at −200 to +200 bp of the TSSs of the up- or downregulated genes (left) and the magnified view of the enrichment of TATA-box-like motifs at −100 bp to +100 bp of the TSSs (right).



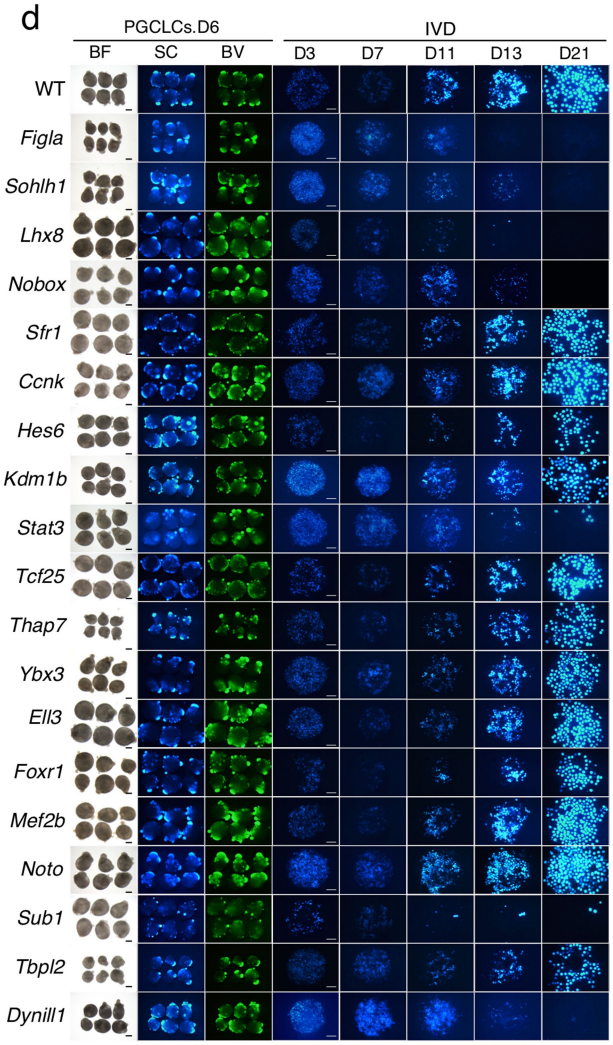
c

CDS mutation by Cas9

Genes	Analyzed lines	Mutants	PGCLC	IVD .D3	IVD .D11	IVD .D13	IVD .D21
Nobox	16	2	√	√	√	√ (small)	-
Lhx8	6	4	√	√	√	-	
Dnmt1	10	3	-				
Uhrf1	10	6	-				
Kdm1b	15	4	√	√	√	√	√
Kat8	41	0					
Drap1	26	0					
Hes6	4	3	√	√	√	√	√
Tbpl2	4	2	√	√	√	√	√ (small)
Birc5	26	0					
Thap7	4	3	√	√	√	√	√
Ybx3	4	3	√	√	√	√	√
Noto	20	2	√	√	√	√	√
Ell3	4	2	√	√	√	√	√
Foxr1	4	3	√	√	√	√	√
Mef2b	4	2	√	√	√	√	√
Sp110	34	0					
Ccnk	7	1	√	√	√	√	√
Dynll1	26	0					
Polr2j	37	0					
Sfr1	7	1	√	√	√	√	√
Stat3	27	0					
Sub1	22	3	√	√	-		
Tcf25	7	3	√	√	√	√	√
Dmap1	29	0					
Sohlh1	8	2	√	√	√	√ (small)	-
Figla	8	3	√	√	-		

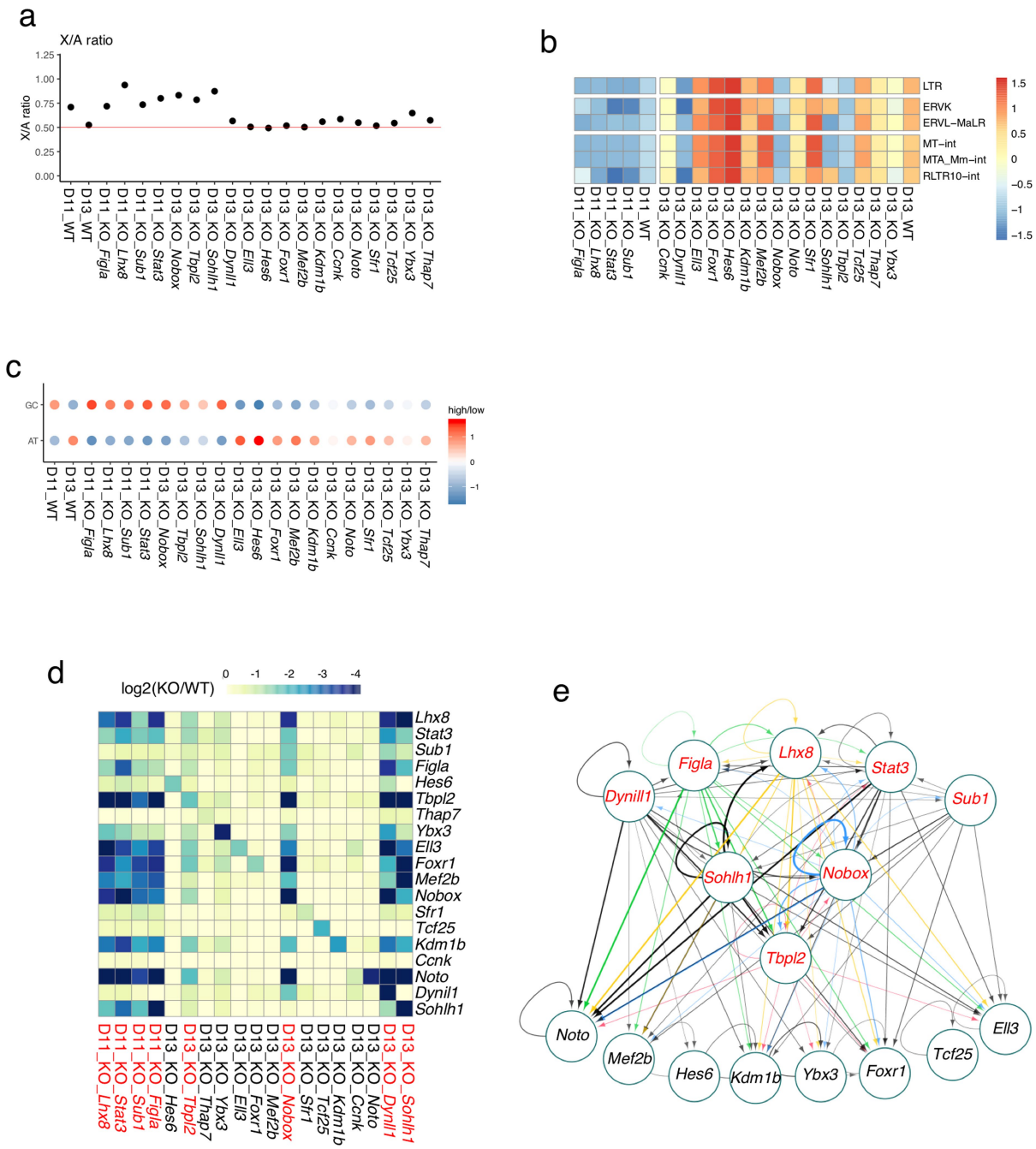
Exon deletion with dox-inducible cDNA rescue

Genes	Analyzed lines	Candidate lines by PCR	Seq-validated lines	PGCLC	IVD .D3	IVD .D11	IVD .D13	IVD .D21
Kat8	32	2	2	-				
Birc5	32	3	1	-				
Sp110	32	1	0					
Dynll1	32	8	1	√	√	√	√	-
Polr2j	32	4	3	√	-			
Drap1	32	7	2	√	-			
Stat3	32	9	2	√	√	√	-	
Dmap1	32	9	1	-				

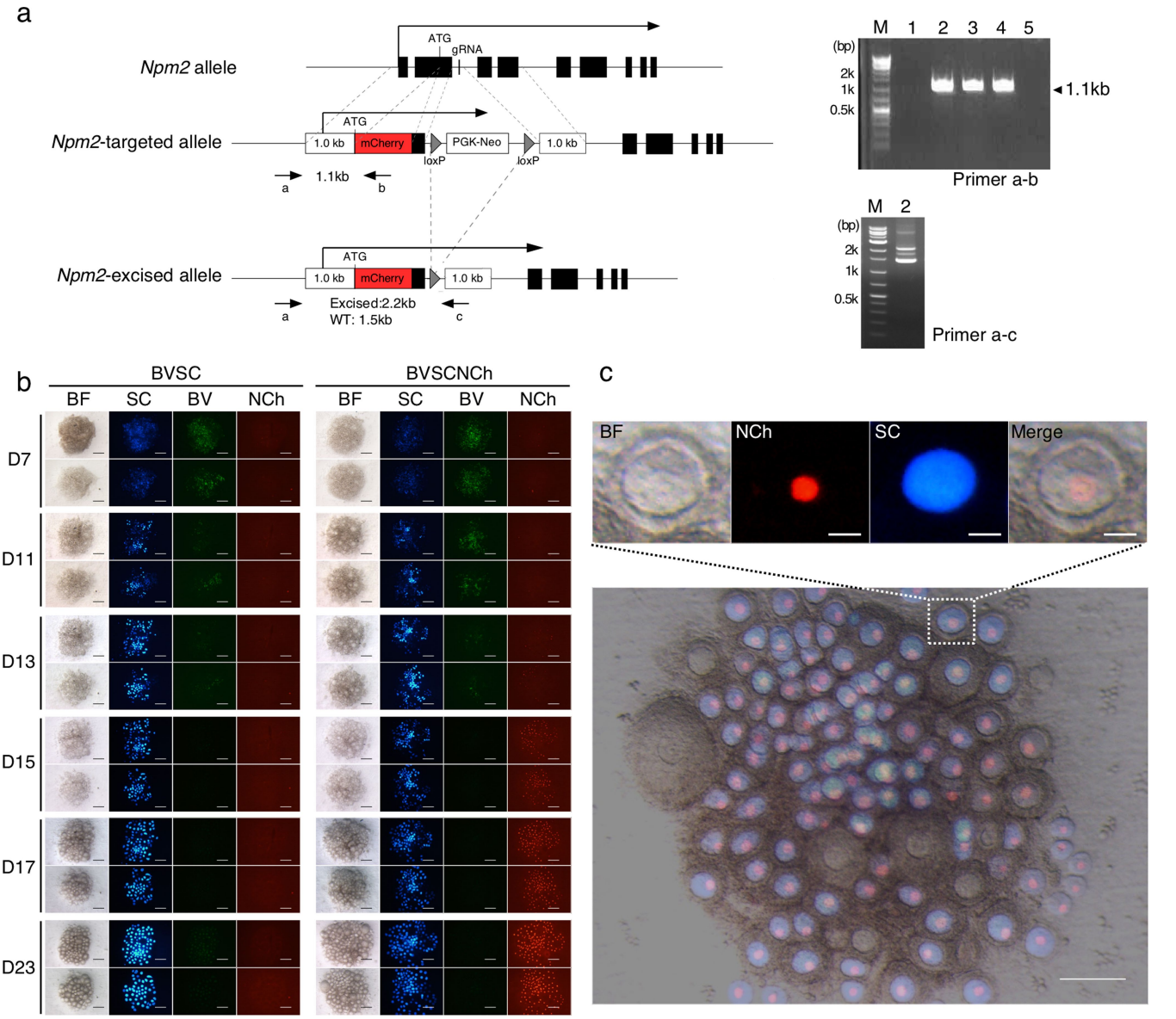


Extended Data Fig. 2 | See next page for caption.

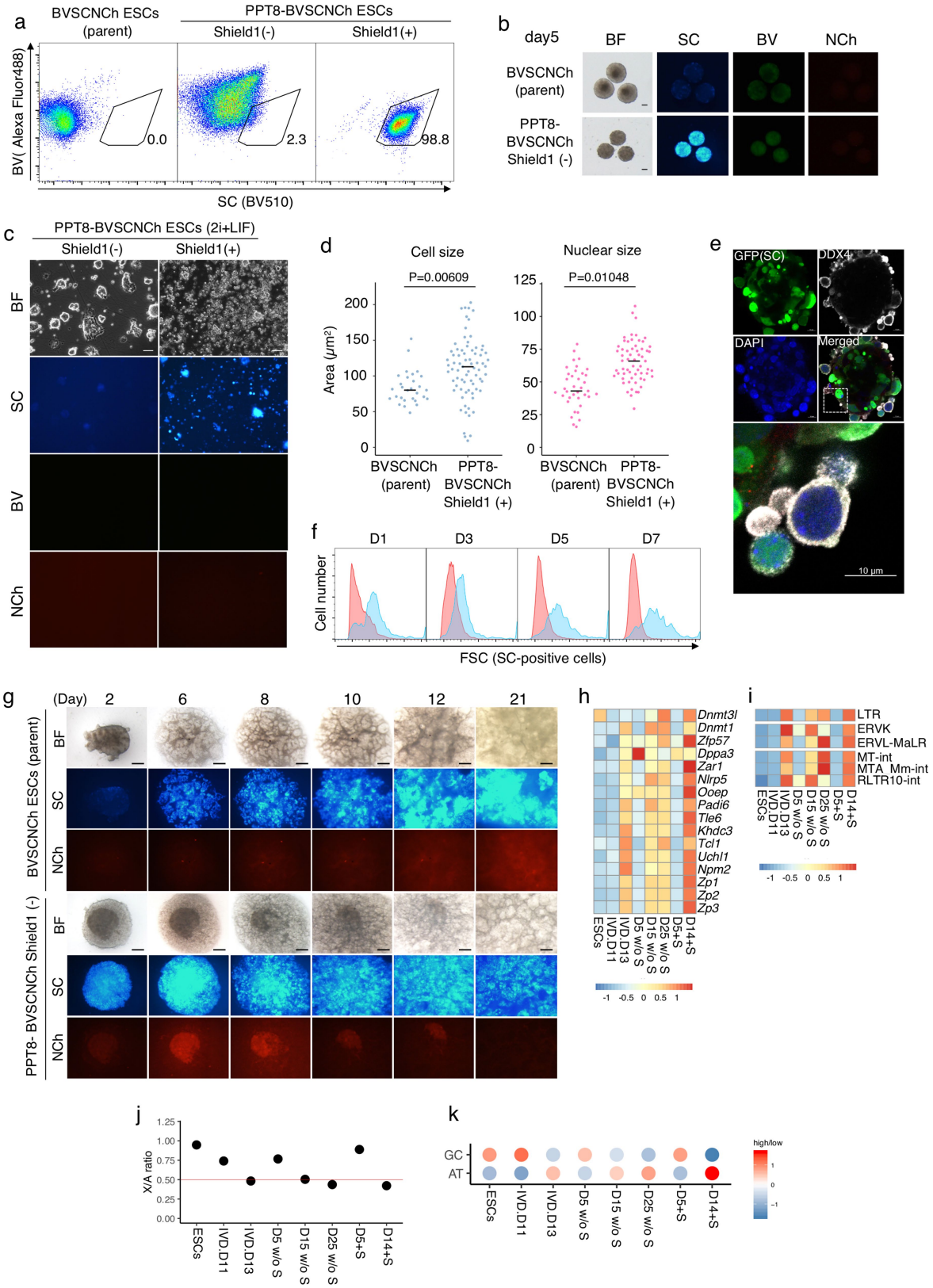
Extended Data Fig. 2 | Establishment and functional validation of knockout ES cell lines. a, Heat map obtained by average linkage hierarchical clustering. The colour names at the right indicate the module assignment determined by WGCNA. Putative PPT-associated modules are written in red. Colour bars on the top of the heatmap indicate the groups in Extended Data Fig. 1a. **b**, Experimental scheme for the establishment of knockout ES lines. Shown are time course of cell culture for the establishment of knockout ES lines (top), the principal of indel detection by amplicon analysis (IDAA)[51] (middle) and a representative IDAA profile of the *Tbpl2* locus (bottom). For cell culture procedures, see Methods. For the amplicon analysis, amplicons derived from alleles with either deletion or insertion were analysed in a fragment analyser. The fragment analyser shows the size of the amplicons in *x* axis and their frequency in *y* axis. The sequences of the alleles in ES cells selected by the IDAA profiles were confirmed by Sanger sequencing. **c**, Summary of oocyte induction from knockout ES lines. The table shows results from the Cas9-mediated knockout lines. The number of ES cell lines tested for the IDAA profile (analysed lines) and knockout lines tested for oocyte induction (mutants) are shown. Among the 27 genes targeted, 19 genes were successfully disrupted. KO alleles of these lines were confirmed by Sanger sequencing. The bottom table shows eight genes (*Kat8*, *Birc5*, *Sp110*, *Dynll1*, *Polr2j*, *Drap1*, *Stat3* and *Dmap1*) for which mutants could not obtained by the Cas9 system. These genes were rescued by the dox (Dox)-inducible expression vector. Shown are the numbers of ES cell lines tested for the IDAA profile (analysed lines), positive for the IDAA profile (candidate lines by PCR), and validated by Sanger sequence (seq-validated lines). *Sp110* knockout ES cells could not be established in this study. In these ES cell lines, the gene expression was rescued by addition of Dox until after PGCLC differentiation, and then Dox was removed from the medium at IVD culture, by which 7 out of 8 genes were successfully disrupted in oocytes. Checkmarks indicate successful differentiation into the stage indicated at the top of the table. Pink boxes indicate the stage where differentiation was not observed. **d**, Oocyte derivation from knockout ES cell lines. The disrupted genes are shown at the left. Results of PGCLCs at 6 days of induction (PGCLCs. D6) and IVD at the days indicated are shown. BF, bright field. Scale bars, 200 µm. $n = 3$, biologically independent experiments. **e**, Quantification of oocyte formation in knockout lines. Box plots are as in Extended Data Fig. 1c. The oocyte numbers and sizes were measured by Stella-CFP signals. The values were compiled from biologically triplicated experiments.



Extended Data Fig. 3 | Transcriptional properties of knockout oocytes. a, X/A ratio in knockout oocytes. The ratio was determined as shown in Extended Data Fig. 1b. **b**, Expression dynamics of LTRs in knockout oocytes. **c**, Promoter usage of knockout oocytes. Shown are the ratios of expression levels of the AT-high-genes/AT-low-genes and GC-high/GC-low genes. Genes are classified as shown in Extended Data Fig. 1d. Expression profiles shown in **a**–**c** are based on biologically duplicated samples. **d**, Reciprocal gene expression in each line of knockout oocytes. Shown are heat map of the expression of PPT-associated genes in the knockout oocytes. Differences of gene expression in the knockout oocytes at each stage compared to the wild-type are shown. Knockout oocytes arrested before PPT are highlighted in red. **e**, Imputed transcriptional network of PPT-associated genes from the RNA-seq data of KO-oocytes. Arrows indicate positive regulations. Line widths indicate the strength of the regulations. Arrow colours indicate the source genes of the arrows. Genes associated with the arrest of knockout oocytes before PPT are highlighted in red.

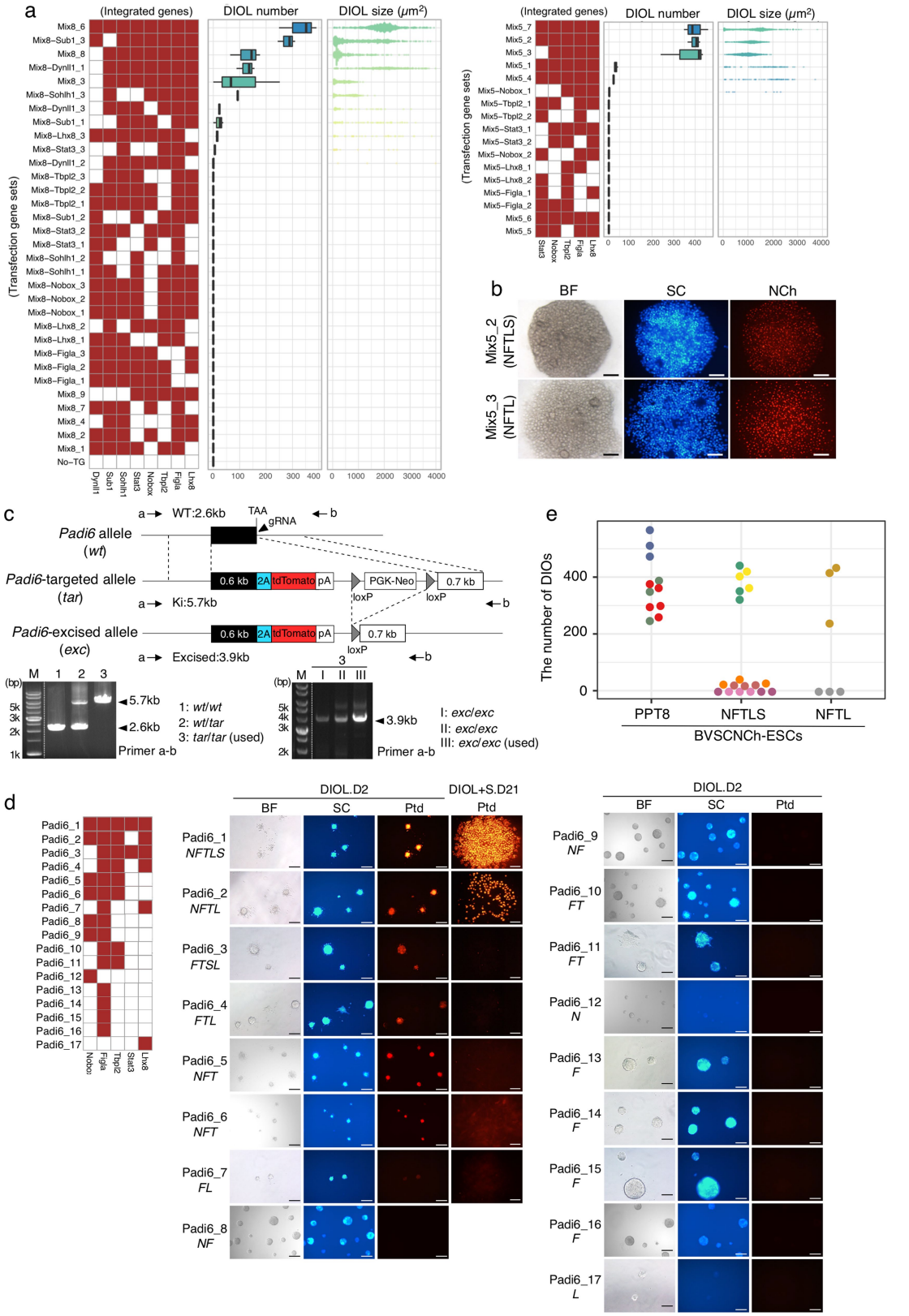


Extended Data Fig. 4 | Establishment of BVSCNCh-ES cells. a, Knock-in of mCherry into the *Npm2* locus. As shown in the schematic diagram (left), *mCherry* was inserted into the ATG of the endogenous *Npm2* gene. Primers for genotyping (arrows) and the expected size of the amplicons are shown. The right images show the results of PCR using the primers and samples numbered. M, size marker. For gel source data, see Supplementary Fig. 1. **b**, mCherry expression in oocytes from BVSCNCh-ES cells. Oocytes were induced by culturing PGCLCs from BVSC H18 ES cells and BVSCNCh-ES cells aggregated with ovarian somatic cells. Note that NPM2-mCherry signals were detected at day 13 in nuclei of oocytes derived from BVSCNCh-ES cells. Scale bar, 100 µm. $n = 9$, biologically independent experiments. **c**, High magnification oocytes from BVSCNCh-ES cells. Shown are a merged view of oocytes derived from BVSCNCh-ES cells at 23 days of culture in the aggregate and magnified views of the secondary follicle. Scale bars, 20 µm (top), and 100 µm (bottom). $n = 9$, biologically independent experiments.



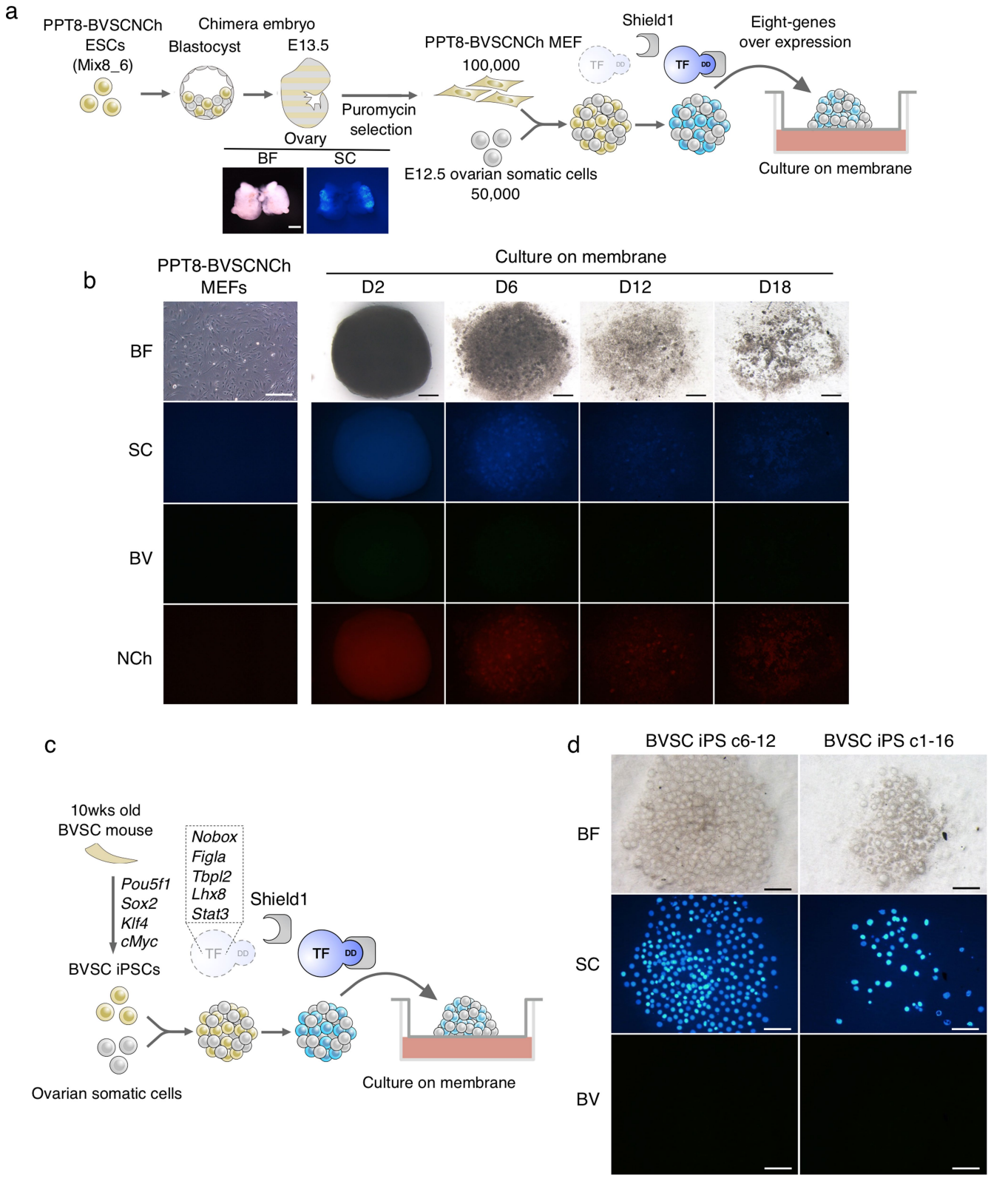
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Requirement of PPT8 for oocyte-like cell induction from ES cells. a, FACS analysis of BVSCNCh+PPT8 ES cells. Note that over 98% of BVSCNCh+PPT8 ES cells expressed Stella-ECFP at day 5 of Shield1-inducible overexpression of PPT8. $n = 3$, biologically independent experiments. For the gating strategy, see Supplementary Fig. 2. **b**, Reporter gene expression in BVSCNCh+PPT8 ES cells without Shield1. Scale bars, 100 μm. $n = 8$, biologically independent experiments. **c**, Overexpression of PPT8 in the ground state. BVSCNCh+PPT8 ES cells were cultured in 2i+LIF with Shield1. Note that Stella-ECFP, but not BLIMP1-mVenus or NPM2-mCherry, was clearly detectable at 1 day of the induction. Scale bars, 100 μm. $n = 12$, biologically independent experiments. **d**, Change in cell size and nuclear size upon overexpression of PPT8. Both the cell and nuclear sizes of BVSCNCh+PPT8 ES cells were increased at day 5 of the induction. *P* values were determined by two-sided Student's *t*-test. **e**, DDX4 expression in BVSCNCh+PPT8 ES cells. At day 5 of the induction, DDX4 expression was detectable in some Stella-ECFP-positive cells. The dashed box in the merged image is enlarged at right. **f**, Dynamics of size of oocyte-like cells. FACS analysis shows the forward scatter of Stella-ECFP-positive cells (blue) and somatic cells (red) in the aggregates at the days indicated. For the gating strategy, see Supplementary Fig. 2. **g**, PPT8-dependent follicle formation. Shown are cultures of parental BVSCNCh-ES cells with ovarian somatic cells (top) and BVSCNCh+PPT8 ES cells with ovarian somatic cells without addition of Shield1 (bottom). Results in **a**–**g** were reproducible in experiments repeated three times. Scale bars, 200 μm. **h**, Heat map of maternal gene expression in oocyte-like cells. **i**, Expression dynamics of LTRs in oocyte-like cells. **j**, X/A ratio in oocyte-like cells. The ratio was determined as shown in Extended Data Fig. 1b. **k**, Promoter usage in oocyte-like cells. Shown are the ratios of expression levels of the AT-high-genes/AT-low-genes and GC-high/GC-low genes. Genes are classified as shown in Extended Data Fig. 1d. The sample names in **h**–**k** correspond to those in Fig. 2h. Expression profiles shown in **h**–**k** are based on biologically duplicated samples.

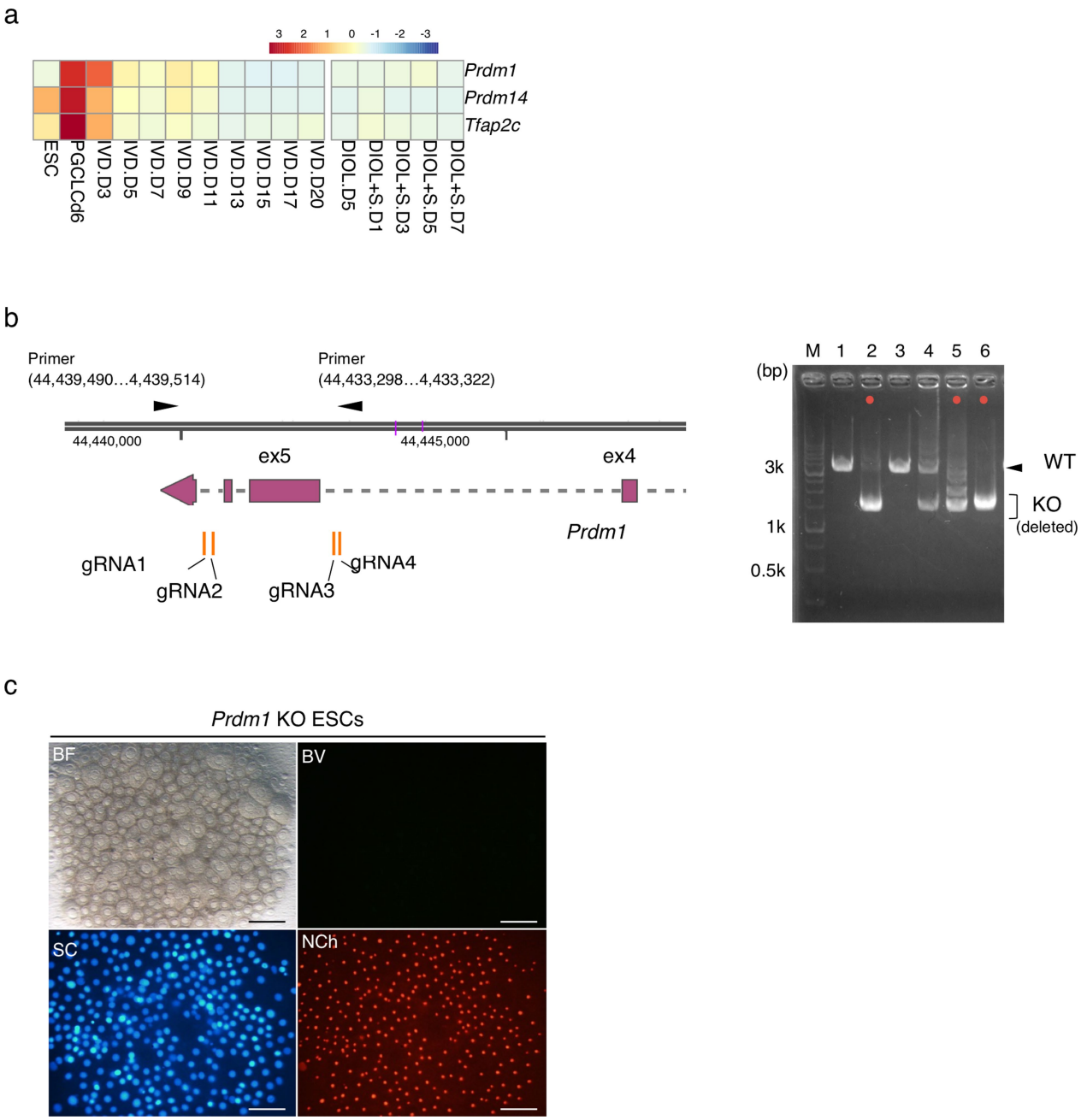


Extended Data Fig. 6 | See next page for caption.

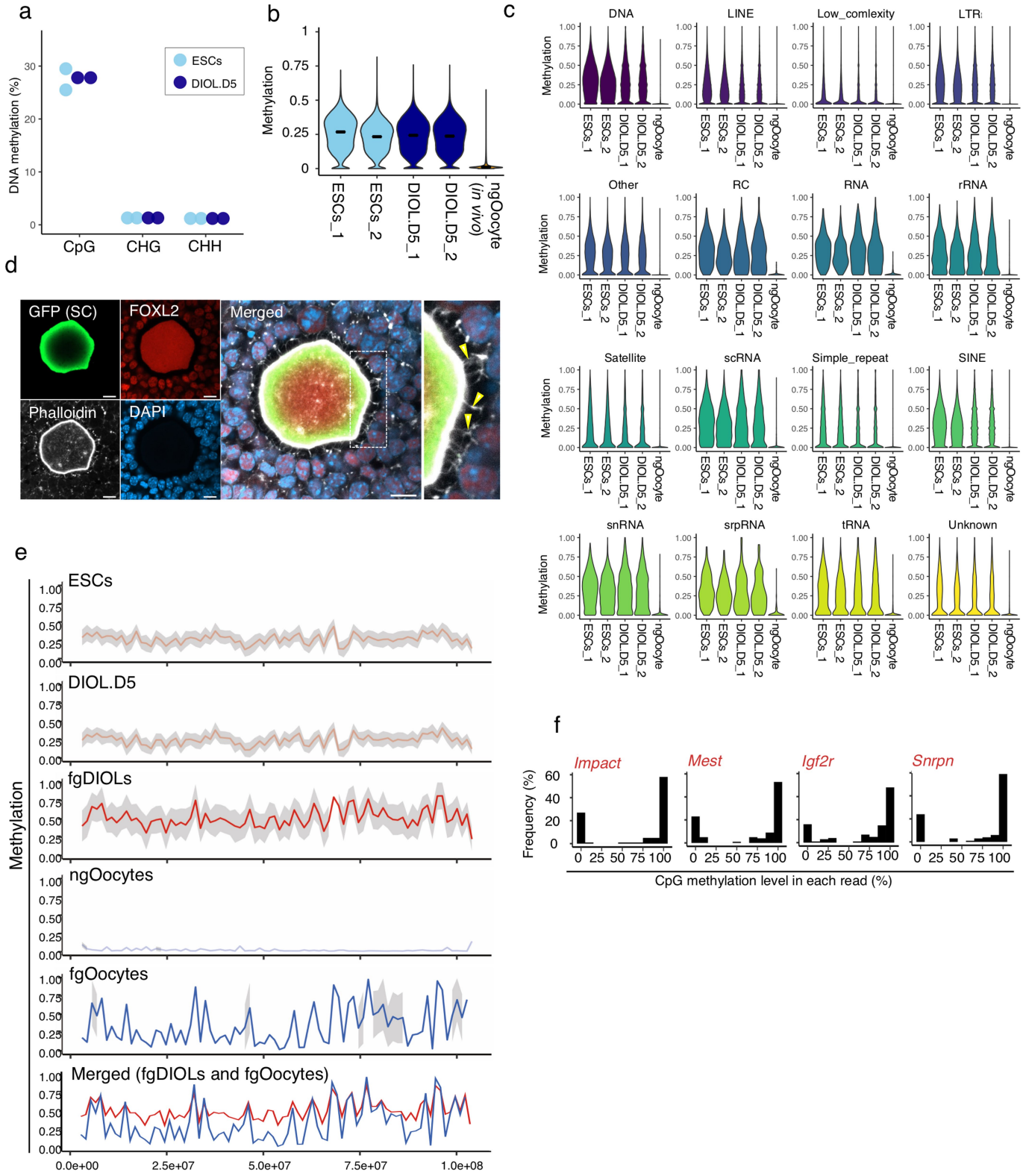
Extended Data Fig. 6 | Identification of a minimum set of genes for DIOL induction. a, Subtraction assay for eight transgenes. A total of 49 BVSCNCh-ES cell lines, which randomly lacked transgenes, were subjected to DIOL induction. The number of oocytes and distribution of oocyte size in each cell lines are shown. On the basis of the number of DIOLs induced from the ES cell lines, we found that *Dynll1*, *Sub1* and *Sohlh1* were dispensable for DIOL induction (see Mix8-Sub1_3, Mix8_8, Mix8-Dynll1_1, Mix8_3 and Mix8-Sohlh1_3). DIOLs were induced from ES cell lines containing 'NFTLS' (see Mix5_7 and Mix5_2) or 'NFTL' (see Mix5_3), whereas no DIOLs were induced from ES cell lines containing fewer than four transgenes among the lines tested. The values of the DIOL number and size were compiled from biologically triplicated experiments except Mix8-Sohlh1_3 (biologically duplicated experiments). Box plots as in Extended Data Fig. 1c. **b**, Representative images of NFTLS- and NFTL-induced DIOLs at 21 days of culture. Scale bars, 200 µm. $n = 3$, biologically independent experiments. **c**, Knock-in of *tdTomato* into the *Padi6* locus. For further subtraction assay, another reporter ES cells, BVSC'Ptd'-ES cells, were made by knocking-in *tdTomato* into the *Padi6* gene locus in BVSC ES cells. Primers for genotyping (arrows) and the expected size of the amplicons are shown. The images show the results of PCR using the primers and samples numbered. M, size marker. Cre-mediated *loxP* excision was made in the clone 3 shown in the left image. For gel source data, see Supplementary Fig. 1. **d**, The subtraction assay using BVSCPtd-ES cells. Seventeen ES cell lines, which randomly lacked transgenes among NFTLS, were subjected to DIOL induction. The images show expression of Stella-ECFP and PADI6-tdTomato (Ptd) in DIOL.D2 and PADI6-tdTomato in DIOL.D21 plus somatic cells. The set of transgenes is shown below each clone. Stella-ECFP expression was observed all clones containing *Figla*. Clones, which did not show PADI6-tdTomato expression in DIOLs, were not subjected to reaggregation with somatic cells. Note that no DIOLs were induced from BVSCPtd ES cells that lacked any of the NFTL genes. Scale bars, 100 µm (DIOL.D2) and 200 µm (DIOL+S.D21). $n = 3$, biologically independent experiments. **e**, Variation of DIOL induction among the transgenic BVSCNCh clones. All three of the BVSCNCh-ES cells harbouring the 8 factors showed a robust induction of DIOLs, whereas 2 of 6 NFTLS clones and 1 of 2 NFTL clones showed the DIOL induction.



Extended Data Fig. 7 | DIOL induction from somatic cells. a, A schematic protocol of DIOL induction from MEFs. BVSCNCh PPT8-ES cells (Mix8_6, see Extended Data Fig. 6a) were injected into blastocysts, followed by transplantation into pseudopregnant females. At 11 days after transplantation, MEFs were collected from female chimaera embryos, which had Stella-ECFP-positive PGCs or oocytes as shown in the images. BVSCNCh MEFs were selected by puromycin and cultured with Shield1 and gonadal somatic cells. **b**, No DIOL formation from MEFs. Images show MEFs and reaggregates at the indicated days of culture. Experiments were performed five times. Scale bars, 100 µm (MEFs) and 200 µm (reaggregates). **c**, A schematic protocol of DIOL induction from iPS cells. BVSC iPS cells were derived from the tail of a 10-week-old female BVSC mouse by overexpression of *Pou5f1*, *Sox2*, *Klf4* and *Myc*. Expression vectors for DD-tagged NOBOX, FIGLA, TBPL2 and STAT3 were transfected into BVSC iPS cells. BVSC iPS cells containing all the expression vectors were cultured with Shield1 and gonadal somatic cells. **d**, DIOL induction from BVSC iPS cells. Images show reaggregates on day 21 of culture. The results were reproducible in experiments repeated three times. Scale bars, 200 µm.

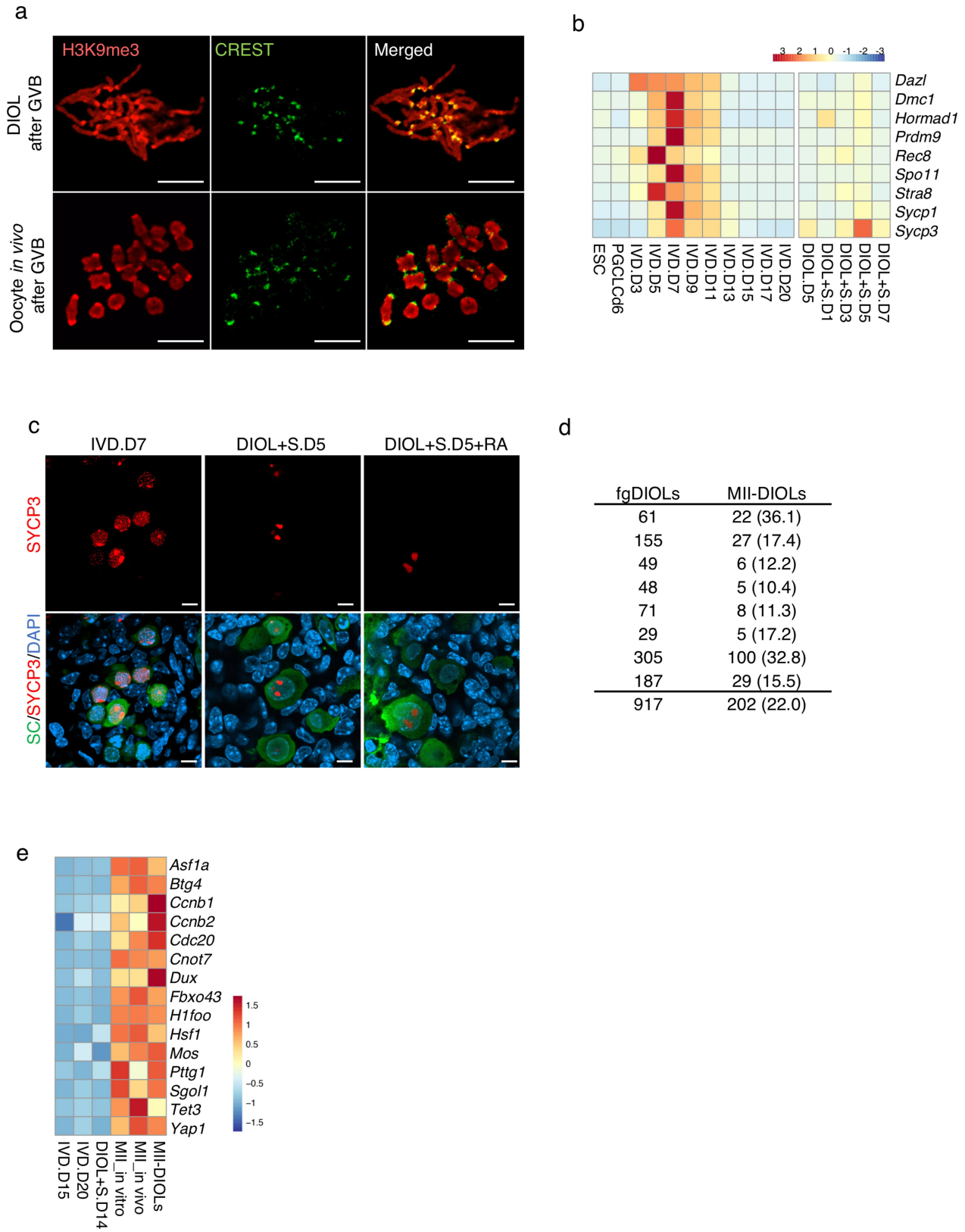


Extended Data Fig. 8 | Dispensability of PGC specification for DIOL formation. a, Heat map of genes essential for PGC specification. The expression profile is based on at least biologically duplicated samples. **b**, Deletion of the *Prdm1* gene by Cas9. gRNAs for the deletion of exons of the *Prdm1* gene and primers for detection of the deletions are shown. The numbers above the primer indicate locations in the genome. The right image shows PCR results using the primers. The numbers and M indicate the ES cell lines analysed and the size marker, respectively. Red dots indicate *Prdm1*-knockout ES lines. For gel source data, see Supplementary Fig. 1. **c**, DIOL induction with somatic cells from *Prdm1*-knockout ES cells. Note that Stella-ECFP- and NCh-positive oocytes were induced in the absence of *Prdm1*. Scale bars, 200 µm. $n = 6$, biologically independent experiments.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Dispensability of epigenetic reprogramming for DIOL formation and de novo methylation in full-grown DIOLs. a, Level of DNA methylation in the genome of DIOLs. Shown are the mean percentages with s.d. of methylated CpG, CHG and CHH (in which H correspond to A, T or C) in the genomes of ES cells and DIOL.D5 in the biologically duplicated samples. **b**, A violin plot showing the CpG methylation levels for each 10-kb window. **c**, Violin plots showing the distribution of CpG methylation in repetitive elements classified by repeat masker (http://www.repeatmasker.org). DNA, DNA repeat elements; LINE, long interspersed nuclear elements; RC, rolling circle; RNA, RNA repeats including RNA, tRNA, rRNA, small nuclear RNA (snRNA), small conditional RNA (scRNA) and signal recognition particle (srpRNA); SINE, short interspersed nuclear elements; simple repeats, microsatellites. **d**, Formation of transzonal projections between the DIOL and surrounding granulosa cells. Images show a representative DIOL at 9 days of IVG culture stained with anti-GFP antibody for Stella-ECFP (green), FOXL2 (red), phalloidin (white) and DAPI (blue). The box in the merged image is shown at the right image. Note that transzonal projections stained with phalloidin bridge between the DIOL and surrounding granulosa cells (arrowheads). Formation of transzonal projections was observed in all 12 of the DIOL-granulosa cell complexes tested. Scale bar, 10 μm. $n = 10$, biologically independent DIOL–granulosa cell complexes. **e**, DNA methylation patterns across chromosome 15 estimated by a locally estimated scatterplot smoothing regression fitting. Red and blue lines indicate the mean DNA methylation levels of ES cells or DIOLs and in vivo oocytes, respectively. The shaded areas indicate the 95% confidence intervals. **f**, The frequency of reads with a different level of CpG methylation in imprinting loci. The *y* axis shows the frequency of reads at the maternal imprinting loci indicated. The *x* axis shows the percentage of CpG methylation in each read. Note that the frequency of completely methylated reads was more than 50 in all loci, suggesting that a portion of the MII-DIOLs completed the maternal imprinting. The methylome profile is based on biologically duplicated samples, except in the case of ngOocytes and fgOocytes, for which single samples were used.



fgDIOLs	MII-DIOLs
61	22 (36.1)
155	27 (17.4)
49	6 (12.2)
48	5 (10.4)
71	8 (11.3)
29	5 (17.2)
305	100 (32.8)
187	29 (15.5)
917	202 (22.0)

Extended Data Fig. 10 | Lack of meiosis in DIOLs. a, Chromosome structure in DIOLs. Shown are representative images of the immunofluorescence analysis of centrosomes and H3K9me3 in a DIOL and an oocyte in vivo soon after germinal vesicle breakdown. Bivalent structures of the chromosomes were observed in the oocyte in vivo ($n = 6$) but not in the DIOL ($n = 20$). Scale bar, 10 µm. **b**, Heat map of genes essential for meiosis. The expression profile is based on biologically duplicated samples. **c**, SYCP3 expression in DIOLs. A synaptonemal structure was seen in IVD.D7 but not at DIOL+S.D5 irrespective of the presence of retinoic acid (RA) (1 µM). Scale bar, 10 µm. A similar result was observed in three independent reconstituted ovaries. Scale bar, 10 µm. **d**, Percentage of DIOLs with a polar body. Shown are the number of fgDIOLs and MII-DIOLs in each experiment. **e**, Heat map of gene expression representing oocyte maturation.

Corresponding author(s): Katsuhiko Hayashi
Nobuhiko Hamazaki

Last updated by author(s): Oct 14, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist .

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used except for Illumina basecalling and demultiplexing software.
Data analysis	Freely available DNA sequencing data analysis software was used to analyze data, as described in Methods: STAR_2.6.0c, subread_v1.6.0, edgeR_3.28.0, FactoMineR_2.0, pheatmap_1.0.12, DAVID_Bioinformatics_Resources_6.8, clusterProfiler_3.14.3, MEME_4.12.0, WGCNA_1.61, imageJ1.52, FACS_DIva_6.1.3, Flowjo_10.6.1 Custom code used in this article can be accessed on https://github.com/nhamazaki/2020_DIOL.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data shown in this study can be downloaded in raw and processed forms from the NCBI Gene Expression Omnibus (GSE143218 and GSE143219) and DDBJ Sequence Read Archive (DRS001541 and DRS001547). The raw data is associated with Figures 1a, 1e-f, 2h, 3a-b, 3d-e, 3f, 4b and Extended Data Figures 1b-g, 2a, 3a-e, 5h-k, 8a, 9a-c, 9e-f, 10b, 10e. There is no restriction on data availability.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical method was used to pre-determine sample size. Sample sizes were determined to obtain reproducibility and reliable distribution among biologically independent samples, based on extensive laboratory experience and literatures in this field (Li et al. Nature, 564:136–140, 2018; Ota et al. EMBO J 36:1888-1907, 2017). The number of independent experiments is described in the legend of each Figure.
Data exclusions	No data were excluded from the study.
Replication	We made biological duplicates or more for all samples to be tested in the experiments in this study. All attempts at replication were successful.
Randomization	All samples are randomly collected or sorted from their population. In addition, experiment and sample collections were performed with controls, e.g. wild-type and knocked-out cells.
Blinding	Sequencing and sample collection were performed by more than two different researchers. Sample names and group allocations were blinded during data collection. Sample names were blinded until the gene expression matrix or DNA methylation matrix were generated. The investigators were not blinded during data analysis, due to feasibility of data analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

All antibodies are commercially available.
1st Antibody
Anti-RFP pAb, Rabbit, MBL, PM005, 1:300 dilution
Anti-Foxl2, Goat, Thermo Fisher Scientific, NB1001277, 1:200 dilution
Anti-GFP monoclonal, Rat, Nacalai, 04404-84, 1:200 dilution
Anti-GFP Chicken polyclonal to GFP, Chick, Abcam, ab13970, 1:200 dilution
Anti-GDF9 monoclonal, Mouse, Santa Cruz, sc-514933, 1:200 dilution
Anti-Centromere protein IgG protein A purified, Human, Antibodies Incorporated, 15-235-0001, 1:200 dilution
Anti-α-Tubulin antibody, Mouse, Sigma, T9026-100ul, 1:200 dilution
Anti-Histone H3 (tri methyl K9) antibody H3K9me3, Rabbit, abcam, ab8898, 1:500 dilution
Anti-FOXO3a monoclonal Rabbit, Cell signaling, #2497, 1:200 dilution
Anti-Ddx4 (vasa), Mouse, Abcam, ab27591, 1:200 dilution
Anti-5-methylcytosine (5-mC) monoclonal [33D3], Mouse, Abcam, ab10805, 1:250 dilution
Anti-5-Hydroxymethylcytosine (5-hmC) (pAb), Rabbit, Active Motif, 39770, 1:250 dilution
Anti-SYCP3 mAb, Mouse, Abcam, Ab97672, 1:750 dilution

2nd Antibody
A488, D-aRat, Donkey, Thermo Fisher Scientific, A21208, 1:500 dilution

A568, D-aRabbit, Donkey, Thermo Fisher Scientific, A10042, 1:500 dilution
A647, D-aMouse, Donkey, Thermo Fisher Scientific, A31571, 1:500 dilution
A488, G-aHuman, Goat, Thermo Fisher Scientific, A-11013, 1:500 dilution
A568, D-aMouse, Donkey, Thermo Fisher Scientific, A10037, 1:500 dilution
A488 D-aChicken, Donkey, Jackson, 703-545-155, 1:500 dilution
A647, D-aGoat, Donkey, Thermo Fisher Scientific, A21447, 1:500 dilution
A488, D-aMouse, Donkey, Thermo Fisher Scientific, A21202, 1:500 dilution
A488, D-aRabbit, Donkey, Thermo Fisher Scientific, A21206, 1:500 dilution
A633, G-aRabbit, Goat, Thermo Fisher Scientific, A21070, 1:500 dilution

MACS Antibody
Anti-SSEA-1(CD15) MicroBeads, human and mouse, Mouse, Miltenyi Biotec, 130-094-530, 1:20 dilution
CD31 MicroBeads, Rat, Miltenyi Biotec, 130-097-418, 1:20 dilution

Validation

Validation statements of all antibodies are available in the manufacturers websites (www.mblintl.com; www.nacalaiusa.com/products/view/101/anti-gfp-rat-igg2a-monoclonal-gf090r; www.abcam.com/; www.thermofisher.com/us/en/home/life-science/antibodies.html ; www.scbt.com/home; www.antibodiesinc.com/; www.sigmaaldrich.com/united-states.html; www.cellsignal.com/; www.activemotif.com/; www.jacksonimmuno.com/).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

ESCs H18 from Dr. Mitinori Saitou (Kyoto University). iPSCs are made by ourselves.

Authentication

Transgenes (BV and SC) and sex chromosome of the cell lines were tested by genomic PCR at their establishment from mice embryos. Proper Knock-in of fluorescent proteins were confirmed by genomic PCR and Sanger sequencing.

Mycoplasma contamination

The cell lines are not tested for Mycoplasma contamination

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

This study used Slc:ICR mice for collection of gonadal somatic cells, blastocysts or sperm, and F1(129svj xC57BL/6) mice for collection of MEF or derivation of iPS cells. Mice used for collection of sperm were males older than 12 weeks of age. Mice used for for collection of gonadal somatic cells or blastocysts were females older than 8 weeks of age. All mice are maintained in constant temperature (22-25°C) and humidity (50-70%) under a 12h light-12h dark cycle with light onset at 8am.

Wild animals

No wild animal is used in this study.

Field-collected samples

This study does not involve field-collected samples.

Ethics oversight

All animal experiments were approved by a relevant committee (Kyushu University, approved nos. #A28-109-4 and #26-74). All animal experiments followed all relevant guidelines and regulations.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

The cells were dissociated by trypsin digestion and gentle pipetting and filtered by using a 70 µm-pored nylon mesh to remove cell clumps.

Instrument

FACS Aria Fusion or FACS Aria SORP

Software	FACSDiva (version 6.1.3) and FlowJo (version 10.6.1) were used for data collection and analysis, respectively.
Cell population abundance	The purity of sorted cells were confirmed by their reporter fluorescence activity, e.g. Stella-CFP, Blimp1-Venus, and Npm2-mCherry, under inverted microscopy.
Gating strategy	FSC-A and SSC-A scatter plot were used to separate cell events from debris and/or dead cells. FSC-H and FSC-W were used to separate singlets and doublets.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.