ZGLP1 is a determinant for the oogenic fate in mice

So I. Nagaoka,[1,2] Fumio Nakaki,[2,#] Hidetaka Miyauchi,[2,∫] Yoshiaki Nosaka,[1,2] Hiroshi Ohta,[1,2] Yukihiro Yabuta,[1,2] Kazuki Kurimoto,[2,∬] Katsuhiko Hayashi,[2,∭] Tomonori Nakamura,[1,2] Takuya Yamamoto,[1,3,4,5] and Mitinori Saitou.[1,2,3]

[1]Institute for the Advanced Study of Human Biology (ASHBi), Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan.

[2]Department of Anatomy and Cell Biology, Graduate School of Medicine, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan.

[3]Center for iPS Cell Research and Application (CiRA), Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan.

[4]AMED-CREST, AMED, 1-7-1 Otemachi, Chiyoda-ku, Tokyo, 100-0004, Japan.

[5]Medical-risk Avoidance based on iPS Cells Team, RIKEN Center for Advanced

Intelligence Project (AIP), Kyoto, 606-8507, Japan.

Present Addresses

[#]Multicellular Systems Biology Group, European Molecular Biology Laboratory (EMBL) Barcelona, C/ Dr. Aiguader, 88, PRBB Building, 08003 Barcelona, Spain.

[∫]Department of Surgery, Division of Hepato-Biliary-Pancreatic Surgery and Transplantation, Graduate School of Medicine, Kyoto University, 54 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan.

[∬]Department of Embryology, Nara Medical University, 840 Shijo-Cho, Kashihara, Nara 634-8521, Japan.

[∭]Department of Developmental Stem Cell Biology, Faculty of Medical Sciences, Kyushu University, Maidashi 3-1-1, Higashi-ku, Fukuoka 812-8582, Japan.

One Sentence Summary: *The BMP-ZGLP1 pathway confers the oogenic fate to sexually uncommitted germ cells in mice*

*Correspondence should be addressed to:

Mitinori Saitou, M.D., Ph.D.

E-mail:saitou@anat2.med.kyoto-u.ac.jp

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

ABSTRACT

Sex determination of germ cells is vital to creating the sexual dichotomy of germ-cell development, thereby ensuring sexual reproduction. However, the underlying mechanisms remain unclear. Here, we show that ZGLP1, a conserved transcriptional regulator with GATA-like zinc fingers, determines the oogenic fate in mice. ZGLP1 acts downstream of bone morphogenetic protein (BMP), but not retinoic acid (RA), and is essential for the oogenic program and meiotic entry. ZGLP1 overexpression induces differentiation of *in vitro* primordial germ cell-like cells (PGCLCs) into fetal oocytes by activating the oogenic programs repressed by Polycomb activities, whereas RA signaling contributes to the oogenic program maturation and PGC program repression. Our findings elucidate the mechanism for mammalian oogenic fate determination, providing a foundation for promoting in vitro gametogenesis and reproductive medicine.

In mammals the chromosomes specify the sex of somatic cells, which in turn activates the oogenic or spermatogenic program in developing germ cells. This manifests at around embryonic day (E) 12.5 in mouse germ cells that are colonized in embryonic ovaries or testes (*1*). A prevalent hypothesis posits that in females, retinoic acid (RA) induces the expression of STRA8, which acts as a key regulator for germ cells (oogonia) to adopt the female/oogenic fate and enter into the meiotic prophase, whereas in males, RA is degraded by CYP26B1 in embryonic Sertoli cells, and germ cells (pro-spermatogonia) ensheathed by such cells enter the male/spermatogenic pathway (*2-4*). However, in a recent *in vitro* analysis, we showed that RA and the expression of STRA8 are not sufficient to induce the oogenic fate in mouse primordial germ cell-like cells (mPGCLCs) derived from mouse embryonic stem cells (mESCs). Instead bone morphogenetic proteins (BMPs), which are expressed strongly in embryonic granulosa cells (*5, 6*), and RA synergistically confer the oogenic pathway and meiotic program on mPGCLCs (*7*). This finding not only defines a signaling principle that drives the oogenic pathway, but also creates an experimental framework for its systematic analysis. Here, we identify a key transcription factor (TF) downstream of BMP and establish an integrated paradigm

for the oogenic fate determination.

A system for identifying key TFs for the oogenic fate

In the *in vitro* system, mPGCLCs isolated by fluorescence-activated cell sorting (FACS) for the expression of *Blimp1-mVenus* and *Stella-ECFP* (BVSC) are cultured in the presence of forskolin and rolipram [which elevate intracellular cAMP levels and enhance PGC(LC) expansion] (*8*), and these cells are provided with BMP and RA from culture day 3 (c3) onward (Fig. 1A) (*7*). Under these conditions, mPGCLCs differentiate into fetal oocyte-like cells positive (+) for SYCP3, a key synaptonemal complex component, and proceed into meiotic prophase, reaching the pachytene stage at c9 (Fig. 1A, 1B) (*7*). BMP alone can also induce mPGCLCs into fetal oocyte-like cells, albeit with low efficiency (Fig. 1A, 1B) (*7*). Using RNA-sequence (RNA-seq) technology (*9*), we screened genes encoding TFs preferentially up-regulated in response to BMP (Fig. S1A, Table S1, S2). Considering genes known to function in meiotic initiation (*10, 11*), we selected eight genes (*Dazl*, *Gata2*, *Id1*, *Id3*, *Msx1*, *Msx2*, *Stra8*, *Zglp1*) for further analyses (Fig. S1A, S1B).

We established a system to overexpress the candidates in mPGCLCs in a doxycycline (Dox)-dependent fashion (Fig. 1A) (*12*, *13*). Immunofluorescence (IF) analyses revealed that the provision of Dox and RA to the eight-gene transfectants from c3 onward lead to the induction of $SYCP3^+$ fetal oocyte-like cells in a majority of SC^+ cells at c9 (Fig. 1A, 1B), suggesting that one or more of the eight factors act as BMP effectors and, together with RA, induce the oogenic fate. The provision of Dox alone also induced the $SYCP3^+$ cells (Fig. 1A, 1B). We therefore performed eight sets of experiments by providing Dox and RA to the seven-gene transfectants produced by withdrawing each of the eight genes in turn, which revealed that each of the seven-gene combinations, except that lacking *Zglp1*, resulted in the induction of $SYCP3^+$ cells (Fig. 1A, 1B). Accordingly, *Zglp1* overexpression, but not overexpression of the other genes, including *Gata2*, *Id1*, and *Msx1*, and the provision of RA resulted in a robust induction of $SYCP3^+$ cells (Fig. 1A, 1B, Fig. S1C, S1D). *Zglp1* overexpression without RA also generated $SYCP3^+$ cells as a significant fraction of SC^+ cells (Fig. 1A, 1B) (see below). We therefore explored the function of *Zglp1* in the oogenic fate determination.

ZGLP1 is essential for the oogenic fate

Zglp1 encodes a transcriptional regulator bearing a GATA-like zinc finger conserved across the metazoan phyla (Fig. S2) (*14, 15*). It has been reported that *Zglp1* is expressed in gonadal somatic cells, but not in germ cells, and is essential both for oogenesis and spermatogenesis (*14, 15*). However, our RNA-seq and IF analyses revealed that, during the embryonic period, *Zglp1*/ZGLP1 exhibits specific and transient expression in DDX4^{+} germ cells in females, but not in somatic cells or males, beginning at E12.0 and waning after E14.5, a key period for the sex determination of germ cells (Fig. 2A, 2B, Fig. S1B). ZGLP1 preceded STRA8 in expression by a period of more than one day (Fig. 2A, 2B). Postnatally, *Zglp1* showed specific expression in *Zbtb16*$^{+}$-undifferentiated and *Kit*$^{+}$-differentiating spermatogonia (Fig. S3A-C) (*16*).

We analyzed the *Zglp1* homozygous knockout$^{-/-}$ mice (Fig. S4A) (*14, 15*). Compared to wildtype ovaries, the *Zglp1*$^{-/-}$ ovaries at postnatal day (P) 8 and 6 weeks were highly atrophic with no ovarian follicles. Ovaries as early as E17.5 contained a drastically

reduced number of DDX4^{+} cells (Fig. 2C, 2D, Fig. S4B). At E15.5, although a significant number of DDX4^{+} cells remained in *Zglp1*$^{-/-}$ ovaries, a majority of them (> ~90%) were negative for SYCP3, even though they expressed STRA8 (Fig. 2D-F, Fig. S4C-E; see below for the level of *Stra8*/STRA8 in *Zglp1*$^{-/-}$ cells). The less prevalent SYCP3^{+} cells did not display features suggestive of meiotic entry, such as telomere clustering (*17*) or the expression of DMC1 (*18, 19*) or γH2AX (*20*) (Fig. 2E, 2F, Fig. S4F, S4G). Note that the *Zglp1*$^{-/-}$ female germ cells bore more severe phenotypes than the *Stra8*$^{-/-}$ cells, some of which survived to form mature ovarian follicles despite their failure in meiosis (*21*).

The *Zglp1*$^{-/-}$ male germ cells developed normally until at least P7 (Fig. S5A, S5B). Subsequently, they exhibited severe impairment in the first wave of spermatogenesis: unlike in females, they entered into meiotic prophase, but mostly failed to progress beyond the zygotene stage, resulting in a large fraction of abnormal pachytene cells both at P15 and P20 (Fig. S5A-F). The *Zglp1*$^{-/-}$ adults exhibited severe defects in spermatogenesis and were infertile (Fig. S5G, S5H) (*14*, *15*). Nonetheless, upon aging, a small fraction of the *Zglp1*$^{-/-}$ cells generated round/elongated spermatids, which, with

intracytoplasmic sperm injection (ICSI), contributed to apparently healthy offspring (Fig. S5I-O). Thus, in females, *Zglp1* is essential for activating the oogenic program, including meiotic prophase, whereas in males, it is dispensable for germ-cell sex determination, but is required in the spermatogonia for an efficient completion of meiotic prophase.

ZGLP1 is sufficient to create the foundation of the oogenic fate

Next, we explored the regulation and function of *Zglp1*. We examined the expression of *Zglp1* and other relevant genes by mPGCLC-derived cells under four conditions: control culture, or a culture with BMP2, RA, or both BMP2 and RA. *Zglp1* was undetectable in the control, whereas it was up-regulated as early as 24 hrs after BMP2 stimulation and elevated further thereafter (Fig. S6A). The manner of up-regulation of *Zglp1* in response to BMP2 was similar to that of *Id1*, a well-known BMP target (*22*) (Fig. S6A). There was no up-regulation of *Zglp1* in response to RA (Fig. S6A). Combined with the data that *Zglp1* up-regulation by BMP2 was impaired by LDN193189, a selective inhibitor of ALK2/3 receptors (Fig. S6B), these findings confirm that *Zglp1* is a downstream effector of BMP signaling. Although *Stra8* was up-regulated as early as 24 hrs after RA

stimulation and elevated further thereafter, it was also up-regulated, albeit at lower levels, in the control at late stages such as c7/c9 (Fig. S6C). Such up-regulation could be an effect of significant genome-wide DNA demethylation, including the demethylation of the *Stra8* promoter, during the mPGCLC culture (*8, 23*) (Fig. S6D). Note that BMP2 enhanced *Stra8* and *Rec8* expression at late stages such as c7/c9 (Fig. S6C), and these enhancements could have occurred, at least in part, via the action of ZGLP1 (see below). In contrast, RA-responsive genes (*24*) with no apparent roles in the oogenic pathway (*Cyp26a1*, *Rarb*, *Hoxa1*) were undetectable in the control or the culture with BMP2 (Fig. S6C), indicating that there is little, if any, RA activity in the culture.

We cloned mESC lines expressing high (A2), middle (C2) and low (H2) levels of *Zglp1*, respectively, upon Dox administration (Fig. S7A, S7B). ZGLP1 overexpression alone in mPGCLCs up-regulated key genes for the oogenic fate/meiotic entry (*Ddx4*, *Dazl*, *Stra8*, *Rec8*, *Sycp3*, *Spo11*, *Dmc1*) and induced $SYCP3^+$ cells in a dose-dependent manner: notably, the A2 clone with the highest *Zglp1* expression activated oogenic/meiotic genes robustly and induced as much as ~90% of mPGCLCs into $SYCP3^+$ cells (Fig. S7C-E).

RA enhanced such activations, particularly in the H2 and C2 clones with low and middle ZGLP1-overexpression levels, respectively (Fig. S7C, S7E). These findings suggest that ZGLP1 is sufficient to induce the oogenic fate, whereas RA augments the ZGLP1-activated oogenic program. We validated that while LDN193186 impaired the BMP-dependent up-regulation of oogenic/meiotic genes, it had no impact on that by ZGLP1 (Fig. S7F), corroborating that ZGLP1 is a key BMP effector for the oogenic fate.

We examined the consequences of ZGLP1 overexpression in a culture lacking fetal bovine serum and including WIN18,446, an inhibitor of aldehyde dehydrogenases and thereby of RA biosynthesis from retinol (*25*), a condition under which the presence of RA can be negated more formally. Even under this condition, ZGLP1 attained a robust induction of the $SYCP3^{+}$ cells (Fig. S8A-C). We explored the effect of BMS493, an inverse agonist of RA receptors (RARs) that enhances the interaction of RARs with their co-repressors (*26*). BMS493 therefore stabilizes a repressive chromatin state near RAR-bound genes, compromising their activation irrespective of the presence of RA (*26*). BMS493 inhibited the induction of $SYCP3^{+}$ cells by ZGLP1 to a substantial extent;

however, even in the presence of 1/10 μM of BMS493, which fully prevents the activity of 100 nM of RA (Fig. S8D), ZGLP1 induced ~20% of BV/SC^{+} cells into SYCP3^{+} cells, along with the up-regulation of relevant genes (Fig. S8E-G). Collectively, these findings indicate that ZGLP1 is sufficient to create the foundation of the oogenic fate in the absence of RA and this function involves the activation of oogenic/meiotic genes both through RAR-dependent (BMS-sensitive) and -independent (BMS-insensitive) mechanisms, and that RA enhances this pathway through RAR-mediated transcriptional regulation.

ZGLP1 reconstitutes key transcriptome programs for the oogenic fate

To examine the extent by which ZGLP1 overexpression recapitulates the oogenic program, we determined the transcriptome dynamics associated with ZGLP1 overexpression in mPGCLCs (the A2 clone) under various conditions examined thus far (Table S1, S3). Principal component analysis (PCA) revealed that mPGCLCs with ZGLP1 alone, as well as those under a serum-free condition with WIN18,446, progressed along the oogenic pathway, gaining, at c9, properties similar to those of E14.5/E15.5 fetal

oocytes (Fig. 3A, Fig. S9A). A spread analysis revealed that up to 3% of mPGCLCs with ZGLP1 alone proceed to the pachytene stage by c9 (Fig. 3B, 3C). mPGCLCs with ZGLP1 and BMS493 also adopted the oogenic program, gaining, at c9, more advanced properties than E13.5 fetal oocytes (Fig. S9B). The addition of RA facilitated the ZGLP1-activated oogenic program, with mPGCLCs with ZGLP1 and RA attaining a transcriptome similar to E14.5/E15.5 fetal oocytes at c7 (Fig. 3A) and up to 14% of them proceeding to the pachytene stage by c9 (Fig. 3C).

We defined the early PGC (315 genes: *Nanog*, *Sox2*, *Prdm1*, *Prdm14*, *Tfap2c*, etc.), late germ-cell (250 genes: *Dazl*, *Ddx4*, *Piwil2*, *Mael*, *Mov10l1*, etc.), and fetal oocyte genes (468 genes: *Stra8*, *Rec8*, *Sycp3*, *Dmc1*, *Sycp1*, etc.) that characterize the transition of transcription profiles for the oogenic program (*7*). mPGCLCs with ZGLP1 (without RA, serum-free with WIN18,446) activated late germ-cell/fetal oocyte genes to an extent comparable to fetal oocytes *in vivo* (Fig. S9C), and mPGCLCs with ZGLP1 and BMS493 activated such genes to a significant extent, with up-regulation of genes including *Stra8* and *Rec8* as early as 24 hrs after Dox administration (Fig. S9D). On the other hand,

mPGCLCs with ZGLP1 repressed early PGC genes moderately, and mPGCLCs with ZGLP1 and BMS493 failed, to a greater extent, to repress them (Fig. S9C). In contrast, mPGCLCs with ZGLP1 and RA activated late germ-cell/fetal oocyte genes and repressed the early PGC genes in a manner parallel to female germ cells *in vivo* (Fig. S9C), highlighting a role of the RA signaling in augmenting the oogenic program activation as well as in repressing the early PGC program.

Compared to the control, mPGCLCs with ZGLP1 up-regulated 489 genes, which were enriched in gene ontology (GO) terms key to the oogenic fate, such as “oogenesis”, “meiotic cell cycle”, and “piRNA metabolic process” (Fig. 3D). mPGCLCs with BMP2, with ZGLP1 (serum-free with WIN18,446), and with ZGLP1 and BMS493 up-regulated a similar number of genes (457, 505, and 413 genes, respectively) with key GO enrichment (Fig. 3E), and the ZGLP1-overexpressing cells showed no up-regulation of typical BMP-downstream genes, including *Id1/3* and *Msx1* (Fig. S9E, S9F). On the other hand, mPGCLCs with ZGLP1 and RA up-regulated a larger number of genes (1,178 genes), which encompassed nearly all the genes up-regulated in mPGCLCs with ZGLP1

(Fig. 3D). Nonetheless, the genes specifically up-regulated above the threshold [$\log_2$ (fold change) >2] by ZGLP1 and RA (713 genes, with enrichment for "cilium assembly", "protein ubiquitination/phosphorylation", and "gametogenesis") were up-regulated, albeit to a lesser extent, by ZGLP1 alone (Fig. S9G).

mPGCLCs with ZGLP1 down-regulated 294 genes, which were enriched in GO terms such as "regulation of transcription" and "glycolytic process" (Fig. 3F). Consistently, mPGCLCs with BMP2, with ZGLP1 (serum-free with WIN18,446), and with ZGLP1 and BMS493 down-regulated similarly sized sets of genes (107, 187, and 185 genes, respectively) (Fig. 3G). On the other hand, mPGCLCs with ZGLP1 and RA down-regulated 1,063 genes, which encompassed nearly all the genes down-regulated in mPGCLCs with ZGLP1 (Fig. 3F). The genes specifically down-regulated above the threshold [$\log_2$ (fold change) >2] by ZGLP1 and RA (795 genes, with enrichment for "metabolic process", "stem cell maintenance", and "G1/S transition of mitotic cell cycle") were down-regulated only to a small extent by ZGLP1 alone (Fig. S9H). Collectively, these findings demonstrate that the BMP signaling and ZGLP1 play a central role in

establishing the key programs for the oogenic fate, with the RA signaling assisting its maturation, including the repression of the early PGC program.

STRA8 activates meiotic genes in a ZGLP1-dependent context

To compare the role of *Zglp1* to that of *Stra8*, we generated *Zglp1* and *Stra8* homozygous knockout$^{-/-}$ mESCs (Fig. S10A-D). At least two *Zglp1*$^{-/-}$ and two *Stra8*$^{-/-}$ independent clones gave essentially identical phenotypes (see below), and we present representative data hereafter. The *Zglp1*$^{-/-}$ mESCs differentiated into apparently normal mPGCLCs, which propagated in culture in the presence of forskolin and rolipram; however, upon stimulation with BMP and RA, the *Zglp1*$^{-/-}$ mPGCLCs were retarded in down-regulating BVSC expression, failed to establish pre-meiotic DNA replication, and failed to differentiate into SYCP3^{+} fetal oocyte-like cells (Fig. S10E-H).

The transcriptome analyses revealed that in response to BMP2, *Zglp1*$^{-/-}$ mPGCLCs failed to up-regulate a vast majority of genes up-regulated in wild-type mPGCLCs, except such genes as *Id1* and *Id2* (Fig. 4A, 4B). Accordingly, *Zglp1*$^{-/-}$ mPGCLCs with BMP and RA

progressed very poorly along the female pathway, only gaining properties similar to those of ~E12.5 germ cells even at c9 (Fig. 4C). *Stra8*$^{-/-}$ mPGCLCs with BMP and RA also failed to establish pre-meiotic DNA replication and to fully differentiate into fetal oocyte-like cells (Fig. S10I, S10J) (*7*), yet they progressed further than *Zglp1*$^{-/-}$ mPGCLCs, acquiring the properties of germ cells between E13.5 and E14.5/E15.5 (Fig. S11A) (*7*). Although *Zglp1* expression was unaffected in *Stra8*$^{-/-}$ cells, *Stra8*/STRA8 was reduced to less than half in *Zglp1*$^{-/-}$ cells (Fig. 4D, 4E). Furthermore, although addition of RA at higher concentrations (1 and 10 μM) reinforced *Stra8* up-regulation in *Zglp1*$^{-/-}$ cells, it had no impact on the up-regulation of other oogenic/meiotic genes in these cells (Fig. 4F). Conversely, to explore whether *Zglp1* overexpression overcomes the defects in *Stra8*$^{-/-}$ cells, we disrupted *Stra8* in the A2 clone (Fig. S11B, S11C). The mPGCLCs from the *Stra8*$^{-/-}$ A2 clone with Dox and RA exhibited phenotypes similar to the *Stra8*$^{-/-}$ mPGCLCs with BMP2 and RA (Fig. S11D, S11E). Thus, although *Zglp1* is essential for launching the overall oogenic program, *Stra8* is an indispensable downstream effector of *Zglp1* for the meiotic program.

The number of genes that failed to be up-regulated in *Zglp1*$^{-/-}$ mPGCLC-derived cells with BMP and RA (530 genes) was much greater than that in *Stra8*$^{-/-}$ cells (224 genes) (Fig. 4G). Although such genes in *Stra8*$^{-/-}$ cells, a majority of which (199/224 genes, ~89%) also failed to be up-regulated in *Zglp1*$^{-/-}$ cells, were mainly associated with the processes for the meiotic prophase, such as “synapsis” and “meiotic cell cycle”, the genes that failed to be up-regulated exclusively in *Zglp1*$^{-/-}$ cells included those with wider implications for fetal oocyte development, such as “oogenesis”, “RNA secondary structure unwinding” and “transcription and chromatin modification”, in addition to those associated with the meiotic prophase (Fig. 4G). Similarly, the number of genes that failed to be down-regulated in *Zglp1*$^{-/-}$ cells with BMP and RA (281 genes) was greater than that in *Stra8*$^{-/-}$ cells (152 genes) (Fig. 4H). We conclude that *Zglp1* broadly regulates the oogenic program, whereas *Stra8*, which functions at least in part downstream of *Zglp1*, mainly regulates the meiotic program.

A mechanistic basis of ZGLP1 function

We determined the genome-wide binding profiles of ZGLP1 in c7/c8 mPGCLC-derived

cells overexpressing V5-tagged ZGLP1 from c3 onwards without or with RA by chromatin immunoprecipitation sequencing (ChIP-seq) (Fig. S12A-C, Table S4, S5), and analyzed them with reference to the chromatin modification states of cultured mPGCLCs at c7 (*8, 27*) (Figure S12D-F, Table S6). Under the condition without or with RA, we defined 16,576 and 14,662 ZGLP1-binding peaks in common in c7/c8 cells, respectively, and assigned 16,290 and 14,409 peaks to the transcription start sites (TSSs), respectively (Fig. S12G). With both conditions, the vast majority of the peaks (>90%) were located within 15 kb of the TSSs, and approximately half of them were associated with promoters, binding to a majority of the promoters/genes expressed at significant levels in c7/c8 cells [~74% and ~70% of genes expressed at $\log_2(RPM+1) > 4$, respectively] (Fig. S12G-I, Table S5). Accordingly, under the two conditions, a majority (≥79%) of the ZGLP1-bound genes overlapped (Fig. S12J, Table S5), suggesting that ZGLP1 functions at gene proximities independently from RA.

To identify ZGLP1-DNA interactions critical to the oogenic fate, we analyzed ZGLP1 binding to genes that were up-regulated in response to ZGLP1 overexpression. Upon

ZGLP1 overexpression at c3 without RA, 41, 181, 671, and 829 genes were up-regulated from c3 at c4, c5, c7, and c9, respectively (Fig. 5A, 5B, S12I). ZGLP1 bound to ~51% (21/41), ~54% (98/181), ~64% (426/671), and ~63% (520/829) of the c4, c5, c7, and c9 genes, respectively (Fig. 5A, 5B, Fig. S12I, Table S5). Notably, a dominant fraction of the c4 genes bound by ZGLP1 originally bore a repressive state; ~52% (11/21) and ~14% (3/21) showed bivalent modifications and H3K27me3 enrichment, respectively, and only ~29% (6/21) exhibited H3K4me3 enrichment (Fig. 5A, 5B, Table S6). This trend was also the case for the c5 genes bound by ZGLP1, with ~37% (36/98) and ~18% (18/98) being bivalent and H3K27me3-enriched, respectively (Fig. 5A, 5B, Table S6). In contrast, similar to the overall ZGLP1-binding profile in c7/c8 cells, most (>67%) of the c7/c9 genes bound by ZGLP1 were active genes with H3K4me3 enrichment (Fig. 5A, 5B, Table S6). Similarly, a dominant fraction of the c4 and c5 genes up-regulated by ZGLP1 and RA and bound by ZGLP1 originally bore a repressive state (Fig. S12K, S12L, Table S6).

The H3K27me3-enriched genes bound by ZGLP1 exhibited higher up-regulation at c4 and at c5 than those not bound by ZGLP1, whereas the H3K4me3-enriched genes were

up-regulated similarly irrespective of the ZGLP1 binding (Fig. 5C). The bivalent and H3K27me3-enriched genes bound by ZGLP1 and up-regulated by c5 were enriched with those for “multicellular organism development”, “transcription”, “meiotic cell cycle”, and “sex differentiation”, and included *Ddx25*, *Zbtb18*, *Zhx1*, *Meioc*, *Sycp1*, *Tdrd1*, *Dmrtc2*, *Dmrtb1*, *Taf9b*, and *Hdac11* (Fig. 5D, 5E). The trend for the preferential up-regulation of the repressed genes bound by ZGLP1 at c4 and c5 was also observed and even appeared to be augmented in mPGCLCs with ZGLP1 and RA (Fig. S12M, S12N), with bivalent genes also showing higher up-regulation at c4. Collectively, these findings support the idea that ZGLP1 activates key genes for the oogenic program, including meiotic prophase, which are in poised/repressed states with bivalent/H3K27me3 modifications in sexually uncommitted germ cells.

Discussion

Based on both *in vivo* and *in vitro* studies in mice, we have demonstrated that BMP signaling and its downstream effector, ZGLP1, specify the oogenic program, through the regulation of genes involved in “RNA processing”, “transcription and chromatin

modification", "retrotransposon regulation", "meiotic cell cycle", and "oocyte development", whereas RA signaling contributed to the overall maturation of the oogenic program as well as to the repression of the PGC program (Fig. 3, 6, Fig. S9). The finding that RA plays a role in repression of the PGC program, which involves a network of TFs for pluripotency, is consistent with a well-known function of RA in stem-cell differentiation (*28*). Thus, our findings provide an integrated paradigm for the signaling pathway and transcriptional mechanism for the oogenic program in mice (Fig. 6). A transcriptome analysis of human fetal germ cells and gonadal somatic cells revealed the potential involvement of the BMP-ZGLP1 pathway in human oogenic fate determination (*29*), suggesting that the BMP-ZGLP1 pathway may play a conserved function in the oogenic fate determination in a wide range of species. Clarification of the biochemical mechanism of action of ZGLP1 will be a critical future challenge.

We have shown that *Zglp1* plays a critical function in allowing the spermatogonia, not to enter, but to efficiently accomplish, the meiotic prophase, and only a small fraction of the *Zglp1*$^{-/-}$ cells generate round/elongated spermatids (*14*, *15*) (Fig. S5). The distinct

phenotypes between females and males represent a sexual dichotomy of the mechanism for the onset/progression of the oogenic and spermatogenic processes, including meiosis, which also warrants further investigation. Notably, the *Stra8*$^{-/-}$ male germ cells show more severe phenotypes: depending on the genetic background, they either fail to enter the meiotic prophase itself or fail to complete the meiotic prophase, with no formation of secondary spermatocytes or round/elongated spermatids (*30*, *31*). Thus, *Stra8* exhibits indispensable functions in ensuring the meiotic prophase both in females and males. A further understanding of the mechanism for the sex determination of germ cells will serve as a key not only for clarifying the etiology of relevant disease states, including infertility, but also for promoting in vitro gametogenesis in diverse mammalian species.

REFERENCES AND NOTES

1. C. Spiller, P. Koopman, J. Bowles, Sex Determination in the Mammalian Germline. *Annu Rev Genet* **51**, 265-285 (2017).

2. J. Bowles *et al.*, Retinoid signaling determines germ cell fate in mice. *Science* **312**, 596-600 (2006).

3. A. E. Baltus *et al.*, In germ cells of mouse embryonic ovaries, the decision to enter meiosis precedes premeiotic DNA replication. *Nat Genet* **38**, 1430-1434 (2006).

4. J. Koubova *et al.*, Retinoic acid regulates sex-specific timing of meiotic initiation in mice. *Proc Natl Acad Sci U S A* **103**, 2474-2479 (2006).

5. H. H. Yao *et al.*, Follistatin operates downstream of Wnt4 in mammalian ovary organogenesis. *Developmental dynamics : an official publication of the American Association of Anatomists* **230**, 210-215 (2004).

6. S. A. Jameson *et al.*, Temporal transcriptional profiling of somatic and germ cells reveals biased lineage priming of sexual fate in the fetal mouse gonad. *PLoS Genet* **8**, e1002575 (2012).

7. H. Miyauchi *et al.*, Bone morphogenetic protein and retinoic acid synergistically

specify female germ-cell fate in mice. *EMBO J* **36**, 3100-3119 (2017).

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

9. T. Nakamura *et al.*, SC3-seq: a method for highly parallel and quantitative measurement of single-cell gene expression. *Nucleic Acids Res* **43**, e60 (2015).

10. Y. Lin, M. E. Gill, J. Koubova, D. C. Page, Germ cell-intrinsic and -extrinsic factors govern meiotic initiation in mouse embryos. *Science* **322**, 1685-1687 (2008).

11. R. Le Bouffant *et al.*, Msx1 and Msx2 promote meiosis initiation. *Development* **138**, 5393-5402 (2011).

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

13. Y. Kojima *et al.*, Evolutionarily Distinctive Transcriptional and Signaling Programs Drive Human Germ Cell Lineage Specification from Pluripotent Stem Cells. *Cell Stem Cell* **21**, 517-532 e515 (2017).

14. S. Li, M. M. Lu, D. Zhou, S. R. Hammes, E. E. Morrisey, GLP-1: a novel zinc

finger protein required in somatic cells of the gonad for germ cell development. *Dev Biol* **301**, 106-116 (2007).

15. T. J. Strauss, D. H. Castrillon, S. R. Hammes, GATA-like protein-1 (GLP-1) is required for normal germ cell development during embryonic oogenesis. *Reproduction* **141**, 173-181 (2011).

16. B. P. Hermann *et al.*, The Mammalian Spermatogenesis Single-Cell Transcriptome, from Spermatogonial Stem Cells to Spermatids. *Cell reports* **25**, 1650-1667 e1658 (2018).

17. Y. Chikashige *et al.*, Telomere-led premeiotic chromosome movement in fission yeast. *Science* **264**, 270-273 (1994).

18. D. L. Pittman *et al.*, Meiotic prophase arrest with failure of chromosome synapsis in mice deficient for Dmc1, a germline-specific RecA homolog. *Mol Cell* **1**, 697-705 (1998).

19. K. Yoshida *et al.*, The mouse RecA-like gene Dmc1 is required for homologous chromosome synapsis during meiosis. *Mol Cell* **1**, 707-718 (1998).

20. S. K. Mahadevaiah *et al.*, Recombinational DNA double-strand breaks in mice

precede synapsis. *Nat Genet* **27**, 271-276 (2001).

21. G. A. Dokshin, A. E. Baltus, J. J. Eppig, D. C. Page, Oocyte differentiation is genetically dissociable from meiosis in mice. *Nat Genet* **45**, 877-883 (2013).

22. T. Lopez-Rovira, E. Chalaux, J. Massague, J. L. Rosa, F. Ventura, Direct binding of Smad1 and Smad4 to two distinct motifs mediates bone morphogenetic protein-specific transcriptional activation of Id1 gene. *J Biol Chem* **277**, 3176-3185 (2002).

23. K. Shirane *et al.*, Global Landscape and Regulatory Principles of DNA Methylation Reprogramming for Germ Cell Specification by Mouse Pluripotent Stem Cells. *Dev Cell* **39**, 87-103 (2016).

24. J. E. Balmer, R. Blomhoff, Gene expression regulation by retinoic acid. *J Lipid Res* **43**, 1773-1808 (2002).

25. J. K. Amory *et al.*, Suppression of spermatogenesis by bisdichloroacetyldiamines is mediated by inhibition of testicular retinoic acid biosynthesis. *J Androl* **32**, 111-119 (2011).

26. A. le Maire *et al.*, A unique secondary-structure switch controls constitutive gene repression by retinoic acid receptor. *Nat Struct Mol Biol* **17**, 801-807 (2010).

27. K. Kurimoto *et al.*, Quantitative Dynamics of Chromatin Remodeling during Germ Cell Specification from Mouse Embryonic Stem Cells. *Cell Stem Cell* **16**, 517-532 (2015).

28. D. R. Soprano, B. W. Teets, K. J. Soprano, Role of retinoic acid in the differentiation of embryonal carcinoma and embryonic stem cells. *Vitam Horm* **75**, 69-95 (2007).

29. L. Li *et al.*, Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions. *Cell Stem Cell* **20**, 858-873 e854 (2017).

30. E. L. Anderson *et al.*, Stra8 and its inducer, retinoic acid, regulate meiotic initiation in both spermatogenesis and oogenesis in mice. *Proc Natl Acad Sci U S A* **105**, 14976-14980 (2008).

31. M. Mark *et al.*, STRA8-deficient spermatocytes initiate, but fail to complete, meiosis and undergo premature chromosome condensation. *J Cell Sci* **121**, 3233-3242 (2008).

ACKNOWLEDGMENTS

We thank the members of our laboratory for their helpful input on this study, and Y. Nagai, Y. Sakaguchi, N. Konishi, and M. Kawasaki of the Saitou Laboratory, and T. Sato and M. Kabata of the Yamamoto Laboratory for their technical assistance. **Author contributions:** S.I.N. performed all the experiments and analyzed the data. F.N., H.M., Y.N., H.O., and K.H. assisted in the induction, expansion, and female differentiation by cytokines or transgene overexpression of PGCLCs. K.K. contributed to the ChIP-seq, T.N. and T.Y. contributed to the RNA-seq, and Y.Y. contributed to the analyses of RNA-seq/ChIP-seq data. S.I.N. and M.S. designed the experiments and wrote the manuscript. **Funding:** This work was supported by a Japan Society for the Promotion of Science (JSPS) Research Fellowship for Young Scientists to S.I.N., by a Grant-in-Aid for Specially Promoted Research from JSPS (17H06098), a JST-ERATO Grant (JPMJER1104), a Grant from HFSP (RGP0057/2018), and Grants from the Pythias Fund and Open Philanthropy Project to M.S. **Competing interests:** The authors declare no

competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary material. The accession number for the RNA-seq and ChIP-seq data generated in this study is GSE124262. All the cell lines are available upon request, with a MTA with Kyoto University.

SUPPLEMENTARY MATERIALS

Materials and Methods

Figs. S1 to S12

Tables S1 to S8

References (32-63)

Figure Legends

Fig. 1. Identification of a TF inducing the oogenic fate.

(A) (Left) (top) Scheme for the oogenic fate induction by BMP2 or doxycycline (Dox)-induced TF overexpression with/without RA in mPGCLCs. m220, m220 feeder cells; SCF, stem cell factor; FR10, forskolin and rolipram, 10 μM each. (bottom) Vectors for the Dox-inducible system. D4Z4, an insulator; EF1a, the elongation factor 1α promoter; CDS, a coding DNA sequence from a candidate gene; IRES, an internal ribosome entry site; Neo, neomycin-resistance gene; Pac, puromycin-resistance gene; pB-TR, *piggyBac* terminal repeat; rtTA, reverse tetracycline transactivator; tetO/mCMV, tetracycline operator fused to the minimal cytomegalovirus promoter.

(Right) Percentages of $SYCP3^+$ cell induction among GFP^+ (BV/SC^+) mPGCLC-derived cells at c9 under the conditions indicated. For each condition, mPGCLC-derived cells were imaged so that a minimum of 20 BV/SC^+ cells were included per image, and at least 10 such images were acquired from 2-4 biological replicates.

(B) Representative images for the expressions of GFP (BV/SC) (cyan), DDX4/mCherry (magenta), and SYCP3 (yellow) in mPGCLC-derived cells at c9 under the conditions

indicated. Cells were counterstained with DAPI (gray). Bars: 10 μm.

Fig. 2. Expression and requirement of ZGLP1 in oogenic fate determination.

(A) Expression of ZGLP1 (green) and STRA8 (yellow) in $DDX4^+$ (magenta) germ cells in fetal ovaries and testes from E11.5 to E15.5. Cells were counterstained with DAPI (gray). Yellow arrows indicate $ZGLP1^+$ cells in E12.0 fetal ovaries. Bars: 10 μm.

(B) Percentages of $ZGLP1^+$ or $STRA8^+$ cells among $DDX4^+$ cells per ovarian section (see **Materials and Methods**). The numbers in parentheses represent the number of embryos analyzed.

(C) Ovarian sections of *Zglp1*$^{+/+}$ and *Zglp1*$^{-/-}$ mice at E17.5 immunostained for DDX4 (magenta). Bars: 100 μm.

(D) Numbers of $DDX4^+$ cells per ovary from *Zglp1*$^{+/+}$ and *Zglp1*$^{-/-}$ mice at E13.5, E15.5, and E17.5 (see **Materials and Methods**). Two to three embryos were analyzed for each genotype at each time point. Black bars represent mean values. NS: not significant.

(E) Expression of SYCP3 (yellow) in $DDX4^+$ (magenta) cells in ovaries of *Zglp1*$^{+/+}$ and *Zglp1*$^{-/-}$ mice at E15.5. Bars: 10 μm.

(F) Percentages of DDX4^{+} cells in ovaries of *Zglp1*$^{+/+}$ and *Zglp1*$^{-/-}$ mice at E15.5 showing distinct SYCP3 staining as categorized in Fig. S4C. Two embryos for each genotype and at least 250 DDX4^{+} cells were analyzed per embryo. Error bars: standard deviations (SDs).

Fig. 3. Induction of the oogenic program by ZGLP1 and its augmentation by RA.

(A) PCA of the transcriptomes of germ cells *in vivo* (E9.5-E11.5 PGCs, E12.5-E15.5 female/male germ cells) and mPGCLCs from *Zglp1* over-expressing (O.E.) clones cultured with Dox or with Dox and RA. Each plot represents one transcriptome. Transcriptomes from two replicates (for E9.5 PGCs and E14.5/E15.5 oocytes, three replicates) are plotted. The color code is as indicated.

(B) Representative images showing the substages of meiotic prophase I in ZGLP1 O.E. mPGCLC-derived cells and E15.5 oocytes. Germ cells were spread and immunostained for SYCP1 (red), SYCP3 (green), and γH2AX (gray). An arrow indicates an unsynapsed X chromosome (this clone is XO) (Fig. S7B). Bars: ZGLP1 O.E. Cells, 5 μm; oocytes,10 μm.

(C) Meiotic progression of SYCP3^{+} mPGCLC-derived cells (at c7, c8, and c9) and E15.5 oocytes (see **Materials and Methods**). The numbers of SYCP3^{+} cells analyzed are shown in parentheses above each bar.

(D, F) The overlap of up/down-regulated genes between mPGCLCs over-expressing ZGLP1 without or with RA. The numbers of the genes up/down-regulated [$\log_2$ (RPM+1) >4, $\log_2$ (fold change) >2 from the control, either at c4, c5, c7, or c9] are shown. Key GO enrichments are shown.

(E, G) GO enrichment among genes up/down-regulated in mPGCLC-derived cells and in E15.5 oocytes. For mPGCLC-derived cells, the numbers of the genes up/down-regulated [$\log_2$ (RPM+1) >4, $\log_2$ (fold change) >2, either at c4, c5, c7, or c9] from the control culture (clone H18 for BMP2-treated and the *Zglp1* O.E clone A2 for other treatments) are shown. For E15.5 oocytes, the numbers of genes up/down-regulated from E9.5 PGCs are shown.

Fig. 4. Regulatory coverages of ZGLP1 and STRA8 for the oogenic program.

(A) The overlap of genes up-regulated by BMP2 between the wildtype and *Zglp1*$^{-/-}$

mPGCLC-derived cells. The numbers of the genes up-regulated [$\log_2$ (RPM+1) >4, $\log_2$ (fold change) >2 from the control, either at c4, c5, c7, or c9] in the respective clones are shown. Key GO enrichments are shown.

(B) Expression kinetics of the 457 up-regulated genes (A) in the wildtype and *Zglp1*$^{-/-}$ mPGCLC-derived cells at c4, c5, c7, and c9. Fold changes from the control culture are shown for the respective clones. Boxplots represent values for the 25th, 50th, and 75th percentiles.

(C) PCA of the transcriptomes of germ cells *in vivo* (E9.5-E11.5 PGCs, E12.5-E15.5 female/male germ cells) and wildtype and *Zglp1*$^{-/-}$ mPGCLC-derived cells (G10) cultured with BMP2 and RA from c3. Each plot represents one transcriptome. Transcriptomes from two replicates (for E9.5 PGCs and E14.5/E15.5 oocytes, three replicates) are plotted. The color code is as indicated.

(D) Expression dynamics of *Zglp1* and *Stra8* in the wildtype, *Zglp1*$^{-/-}$ and *Stra8*$^{-/-}$ mPGCLC-derived cells cultured with BMP2 and RA from c3 onward (relative to the wildtype mPGCLC-derived cells at c9).

(E) (Left) Western blot analysis of the STRA8 expression at c7 in the wildtype and *Zglp1*$^{-}$

$^{/-}$ mPGCLC-derived cells cultured with BMP2 and RA from c3 onward. (Right) The STRA8 levels (normalized by β-ACTIN) from two biological replicates. The mean value for the wildtype mPGCLC-derived cells is set as 1.0. Error bars: SDs.

(F) Expression of key oogenic/meiotic genes at c7 in the wildtype and *Zglp1*$^{-/-}$ mPGCLC-derived cells in the cultures as indicated. The expression in SC^{+} cells sorted by FACS was analyzed by qPCR. The ΔC_t from the average C_t values of two housekeeping genes, *Arbp* and *Ppia*, is shown for each gene. The mean values from two independent experiments, each with two technical replicates, are shown. Error bars: SDs. N.D.: not detected. The color code is as indicated.

(G, H) The overlap of differentially expressed genes (DEGs) [$\log_2$ (RPM+1) >4, $\log_2$ (fold change) >2, either at c4, c5, c7, or c9] between the wildtype and *Zglp1*$^{-/-}$ PGCLC-derived cells (common DEGs in clone E5 and G10; 530 up-regulated and 281 down-regulated) and DEGs between the wildtype and *Stra8*$^{-/-}$ mPGCLC-derived cells (common DEGs in clone B6 and F6; 224 up-regulated and 152 down-regulated). Key GO enrichments are shown.

Fig. 5. Activation of Polycomb-repressed genes by ZGLP1.

(A) The percentages of up-regulated genes [$\log_2$ (fold change) > 2 from c3, at c4, c5, c7, and c9] in mPGCLC-derived cells cultured with Dox from c3 onward, classified based on ZGLP1 binding and the enrichment of two histone modifications (H3K4me3 and H3K27me3) (see Table S6). The numbers of genes for each category are shown in parentheses. The color coding is as indicated.

(B) A scatter plot showing the IP levels for H3K4me3 and H3K27me3 around the transcription start sites (TSSs). The dotted lines represent the threshold IP levels for each histone modification (see **Materials and Methods**). Genes bound by ZGLP1 and up-regulated from c3 to c4 or from c3 to c5 by Dox treatment (101 genes in total) are colored as indicated. Representative ZGLP1 up-regulated genes are annotated.

(C) Expression dynamics of ZGLP1-bound/unbound genes. The expression levels [$\log_2$ (RPM+1)] for each gene are normalized by its peak expression levels between c3 and c9, and then the median values are plotted. **P*-value < 0.05. ***P*-value < 0.005.

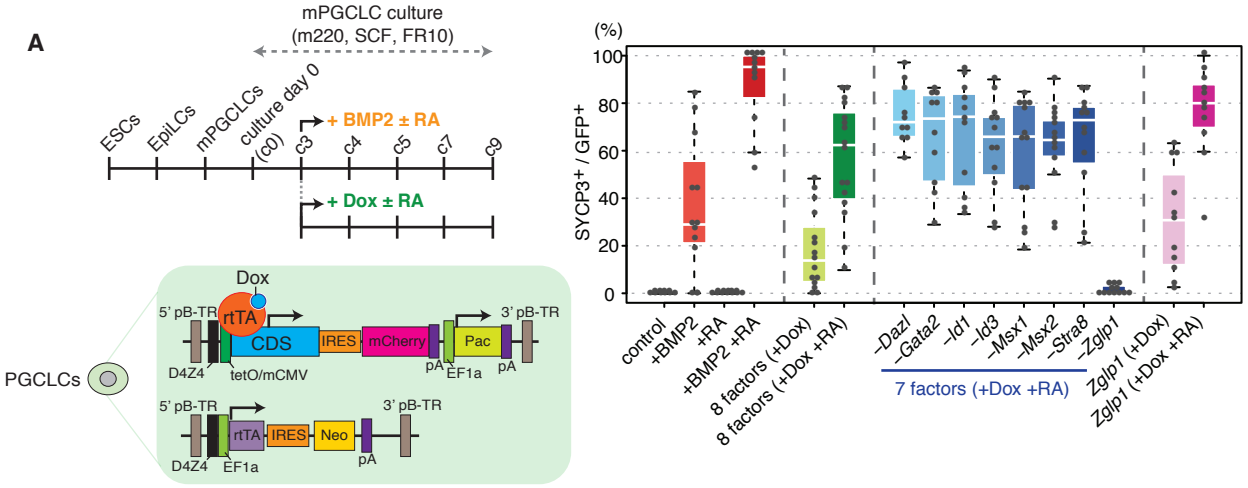
(D) GO term enrichment among bivalent and H3K27me3-enriched genes bound by ZGLP1 and up-regulated from c3 to c4 or from c3 to c5.

(E) ChIP-seq tracks for ZGLP1, H3K4me3, and H3K27me3 around *Zhx1*, *Dmrtc2*, and *Meioc*. The ChIP-seq data for the two histone modifications are from sexually uncommitted mPGCLCs during culture (*8*).

Fig. 6. A model for the oogenic fate determination in mice.

BMP2, most likely secreted from embryonic granulosa cells (*5*, *6*), up-regulates ZGLP1 in oogonia, which activates key oogenic programs, including those for "chromatin modification", "RNA processing", "retrotransposon repression", "folliculogenesis", and "meiotic cell cycle", whereas the RA signaling contributes to overall maturation of the oogenic program as well as to "repression of the PGC program". STRA8 mainly regulates the meiotic program in a ZGLP1-dependent context.

Figure 1, Nagaoka et al.



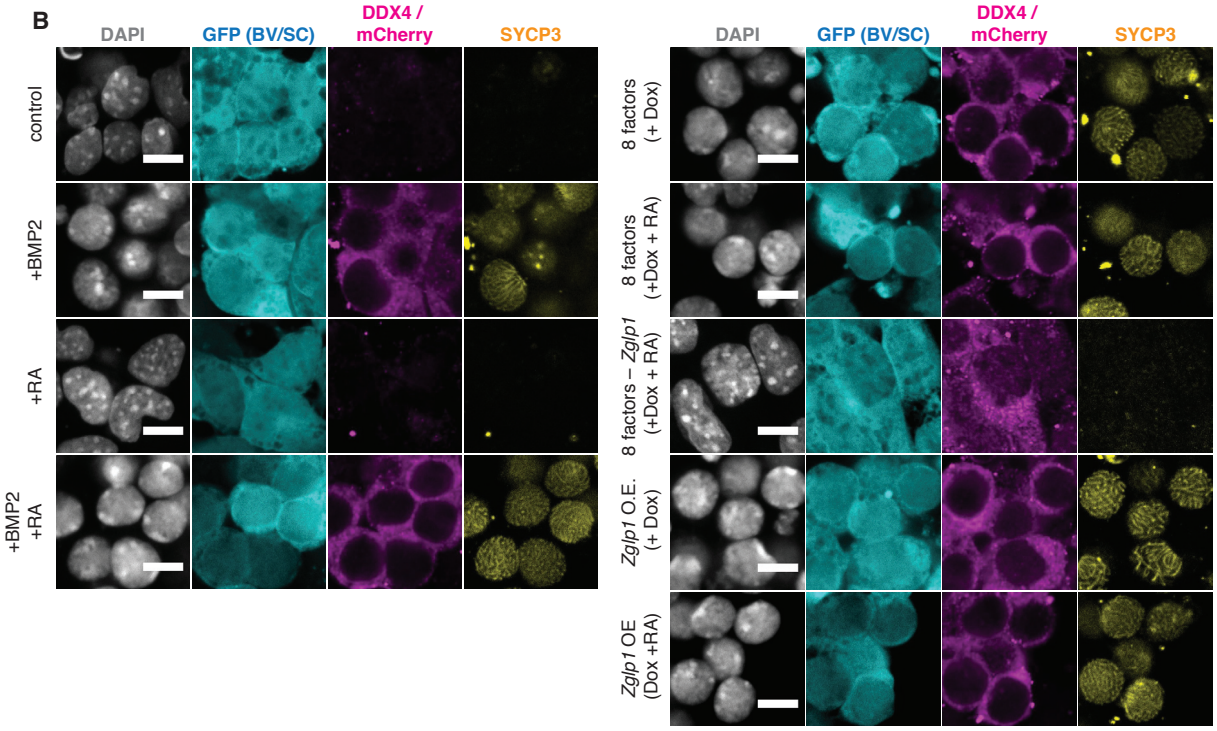


Figure 2, Nagaoka et al.

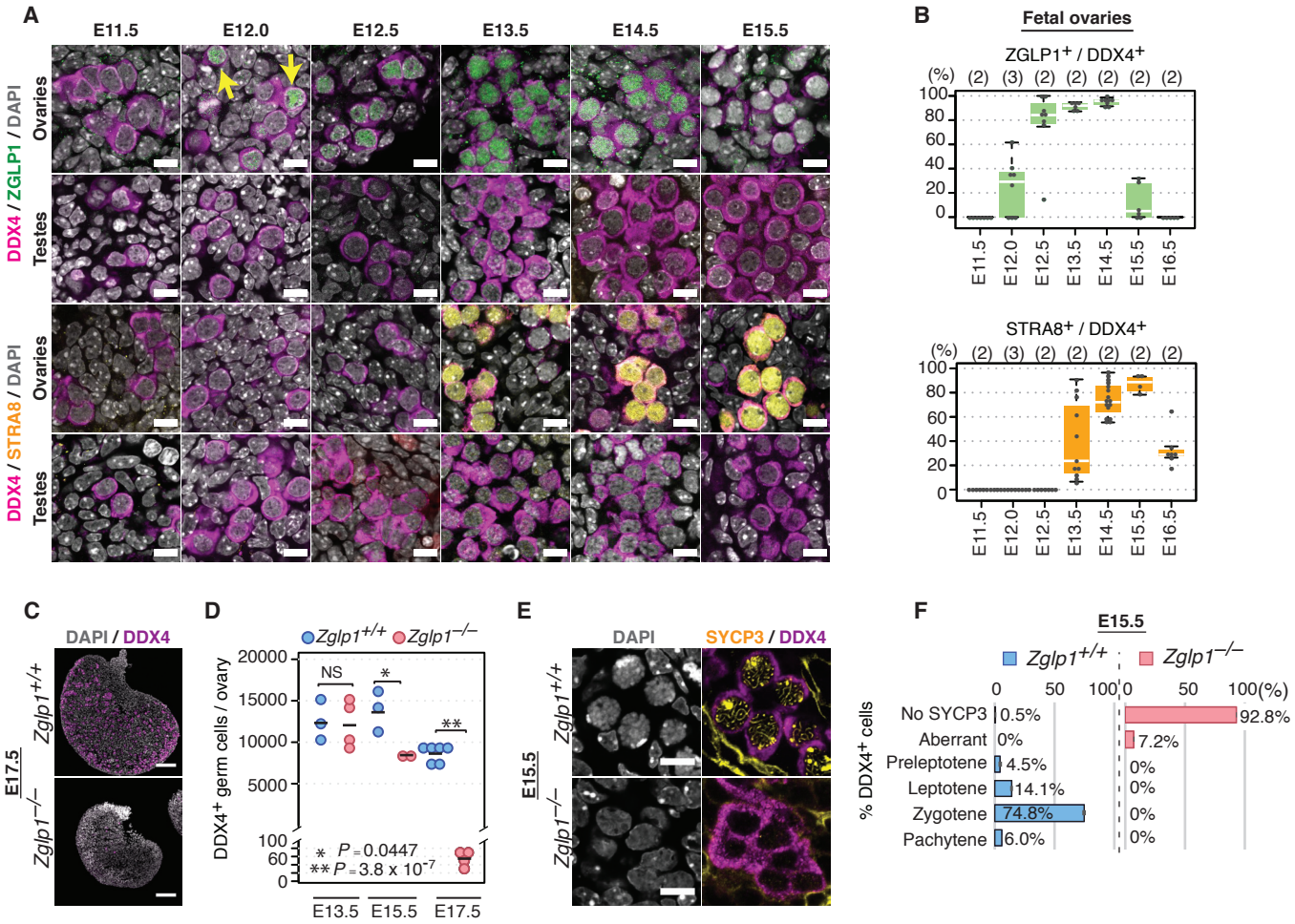


Figure 3, Nagaoka et al.

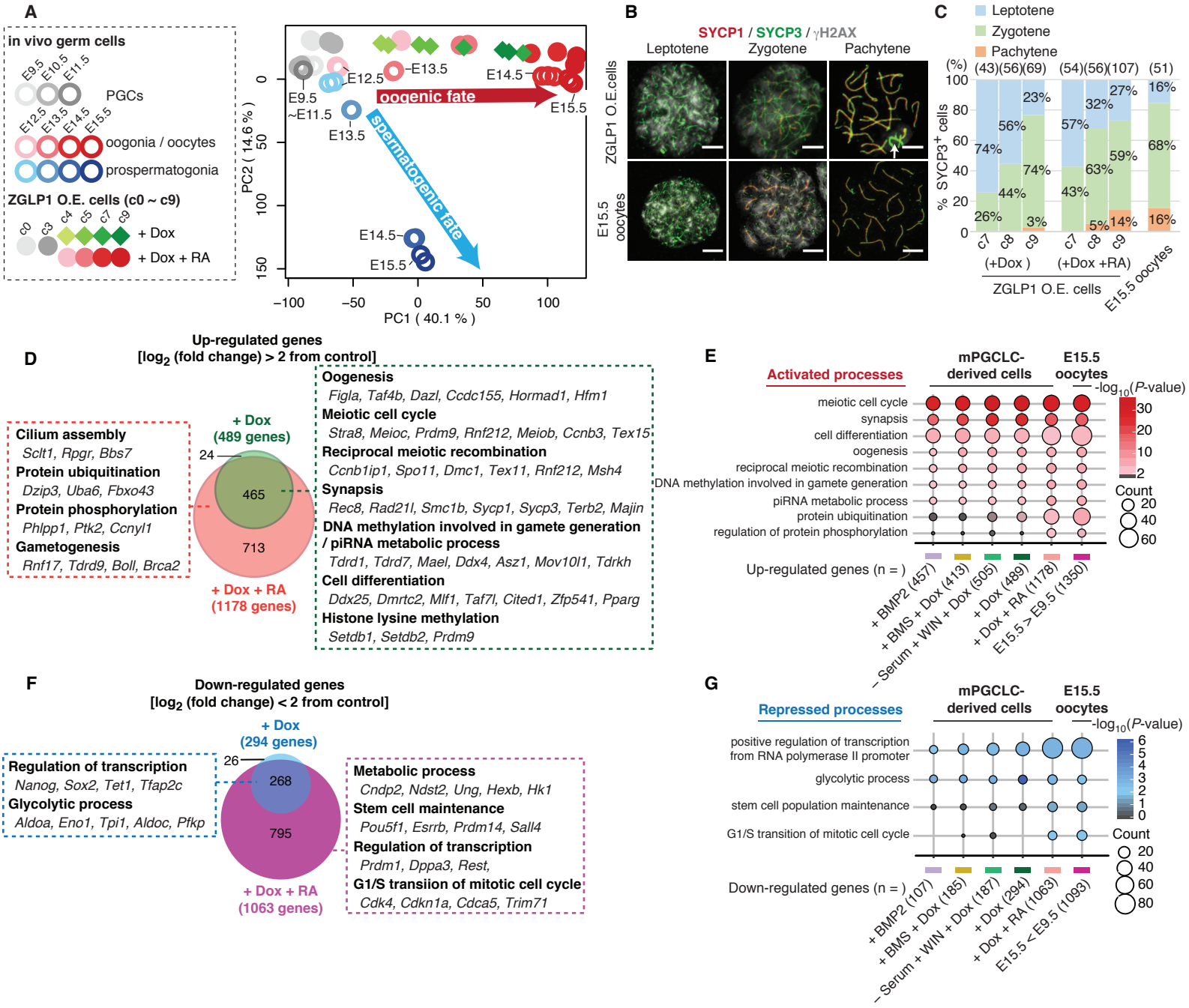


Figure 4, Nagaoka et al.

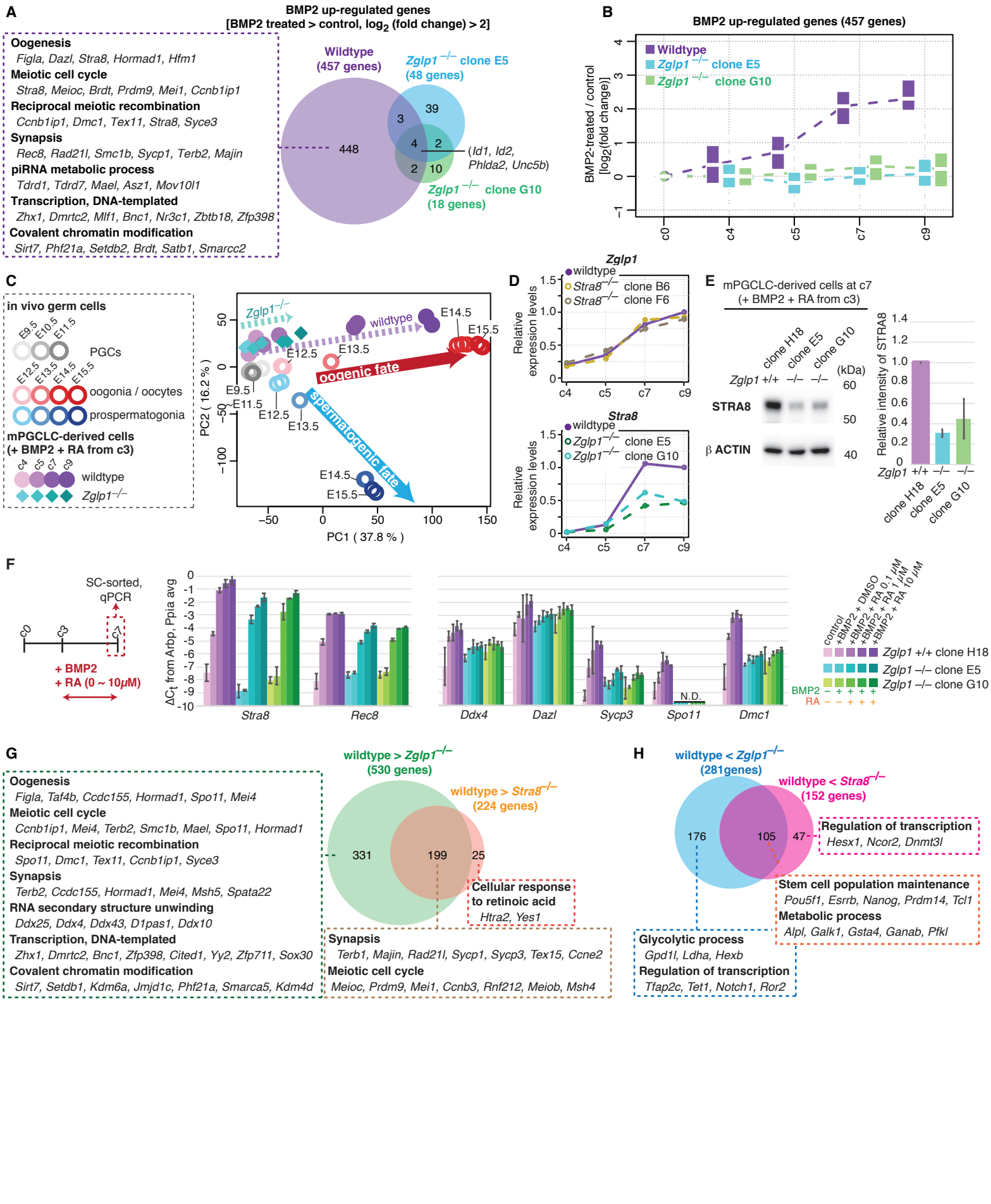


Figure 5, Nagaoka et al.

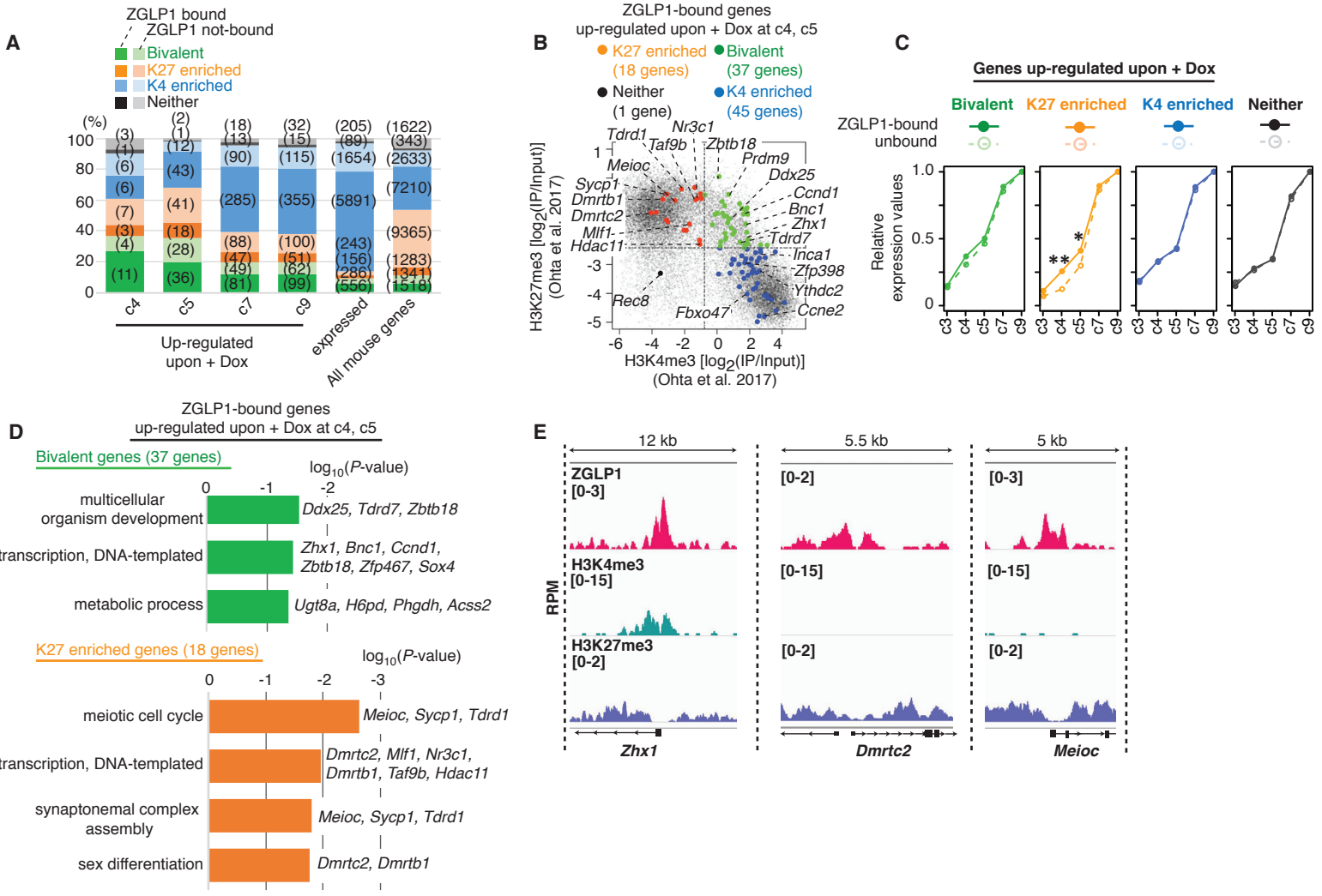
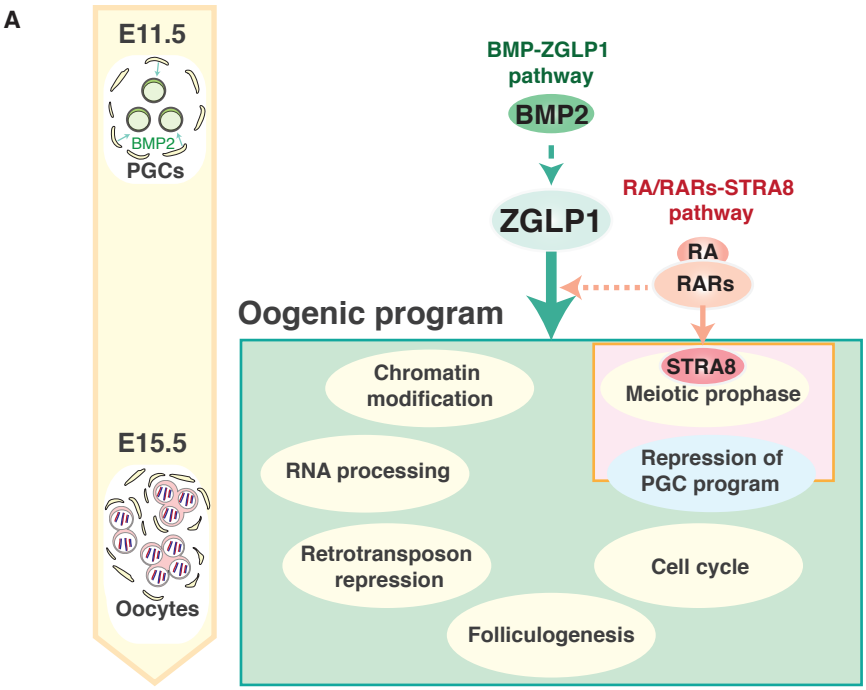


Figure 6, Nagaoka et al.



Supplementary Materials

Materials and Methods

Animals

All animal experiments were performed under the ethical guidelines of Kyoto University.

C57BL/6 and ICR mice were purchased from SLC (Shizuoka, Japan). The BVSC [Tg(Blimp1-mVenus: BV), CDB0460T, and Tg(Stella-ECFP: SC), CDB0465T] (http://www.cdb.riken.jp/arg/TG%20mutant%20mice% 20list.html) and Tg(Stella-EGFP: SG) transgenic mice were established previously as reported {Payer, 2006 #317;Seki, 2007 #2348;Ohinata, 2008 #881}. *Zglp1* mutant mice were established previously and re-derived from cryo-preserved sperm and embryos {Li, 2007 #3122}. For re-derivation of the mutant carrier line, cryo-preserved sperm were purchased from the Mutant Mouse Resource and Research Center (MMRRC) (MMRRC strain ID: 36683; Strain Name: B6;129S-*Zglp1*^{tmEem} /Mmnc), and intra-cytoplasmic sperm injection was performed with oocytes obtained from 6–8 week-old C57BL/6 females. The resultant embryos were transferred to pseudo-pregnant C57BL/6 females. Additionally, cryo-preserved embryos were purchased from MMRRC, and embryo transfer was performed to obtain additional carrier mice. Mutant carriers were mated with C57BL/6 mice, and the colonies were maintained by mating heterozygous males with wild-type C57BL/6 females or mating between heterozygous animals. Genotyping for *Zglp1* alleles was performed using the primers in **Table S7**.

Noon of the day when a copulation plug was identified was designated as embryonic day (E) 0.5. All mice were housed in a specific pathogen-free animal facility under a 14 hr light/10 hr dark cycle.

ESC culture, EpiLC induction, mPGCLC induction, mPGCLC mitotic expansion, and cytokine-mediated oocyte differentiation

H18 BVSC ESCs [129 x C57BL/6 (XX)] and BDF1-2-1 BVSC ESCs [C57BL/6 x DBA/2 (XY)] were established as previously reported {Hayashi, 2012 #2336;Ohta, 2017 #3061}.

The ESC culture, EpiLC induction, and mPGCLC induction were performed as described previously {Hayashi, 2011 #1978;Hayashi, 2012 #2336;Ohta, 2017 #3061} with a few modifications. H18 BVSC and BDF1-2-1 BVSC ESCs were cultured in N2B27 medium (prepared based on previous reports {Ying, 2008 #1286;Hayashi, 2013 #2941}) with two chemical inhibitors—0.4 μM PD0325901 (Stemgent, 04-0006) and 3 μM CHIR99021 (BioVision, 1677-5)—and 1,000 U/ml LIF (Merck Millipore, ESG1107) on a dish coated with 0.01% poly-L-ornithine (Sigma-Aldrich, P3655) and Laminin (300 ng/ml for H18 and 100 ng/ml for BDF1-2-1, respectively; BD Biosciences, 354232). EpiLCs were induced by plating ESCs on a dish coated with 16.7 μg/ml human plasma fibronectin (Merck Millipore, FC010) in N2B27 medium containing 20 ng/ml activin A (Peprotech,

120-14), 12 ng/ml bFGF (Thermo Fisher Scientific, 13256-029) and 1% KSR (Thermo Fisher Scientific, 10828-028). At 42–46 hrs of EpiLC culture, mPGCLCs were induced under a floating condition by plating 4000 EpiLCs in the wells of a low-cell-binding U-bottom 96-well Lipidure-coat plate (Thermo Fisher Scientific, 81100525) in GMEM (Thermo Fisher Scientific, 11710-035) with 500 ng/ml BMP4 (R&D Systems, 314-BP), 100 ng/ml SCF (R&D Systems, 455-MC), 1000 U/ml LIF, 50 ng/ml EGF (R&D Systems, 2028-EG), 15% KSR, 0.1 mM nonessential amino acids (NEAA) (Thermo Fisher Scientific, 11140), 1 mM sodium pyruvate (Thermo Fisher Scientific, 11360), 2 mM L-glutamine (Thermo Fisher Scientific, 25030), 0.1 mM 2-mercaptoethanol (2-ME) (Thermo Fisher Scientific, 21985-023), and 100 U/ml penicillin and 0.1 mg/ml streptomycin (Thermo Fisher Scientific, 15140).

mPGCLC mitotic expansion culture was performed as previously described {Ohta, 2017 #3061}. Briefly, four days after mPGCLC induction, BV^{+} mPGCLCs were sorted by a flow cytometer (BD, FACSAria III) onto m220 feeder cells, which express a membrane-bound form of SCF {Dolci, 1991 #685;Majumdar, 1994 #2947}, in a 0.1% gelatin (Sigma-Aldrich, G1890)-coated well. The expansion culture was maintained in GMEM containing 100 ng/ml SCF, 10 µM forskolin (Sigma-Aldrich, F3917), 10 µM rolipram (Abcam, ab120029), 2.5% FBS (Thermo Fisher Scientific, 10437-028), 10% KSR, 0.1 mM NEAA, 1 mM sodium pyruvate, 2 mM L-glutamine, 0.1 mM 2-ME, 100 U/ml penicillin and 0.1 mg/ml streptomycin.

The induction of oogenesis was performed as previously reported {Miyauchi, 2017 #3092;Miyauchi, 2018 #3125} by a combined treatment of 300 ng/mL BMP2 (R&D Systems, 355-BM) and 100 nM all-trans retinoic acid (RA) (Enzo Life Sciences, BML-GR100, or Sigma-Aldrich, R2625). For the treatment, RA was dissolved in DMSO at a concentration of 50 mM and further diluted in the culture medium just prior to use. During culture, half the culture medium was exchanged at c3 with the medium containing 600 ng/ml BMP2 and 200 nM RA (final concentrations: 300 ng/ml BMP2 and 100 nM RA). After the initial medium change, half the medium was exchanged every two days with the medium containing 300 ng/ml BMP2 and 100 nM RA until the end of the culture period. For the inhibition of BMP and RA signaling, LDN193189 (Stemgent, 04-0074), WIN18,446 (Cayman Chemical, 14018) and BMS493 (Sigma-Aldrich, B6688) were dissolved in DMSO at the concentrations of 10 mM, 50 mM, and 100 mM, respectively. The inhibitors were further diluted in the mPGCLC culture medium prior to use.

Generation of ESC clones possessing doxycycline (Dox)-inducible candidate factors and Dox-mediated oocyte differentiation in mPGCLCs

The coding DNA sequences (CDS) for the candidate genes were PCR-cloned from cDNAs from SG-sorted germ cells from E12.5 or E13.5 female mice using high fidelity KOD Plus Neo DNA polymerase (TOYOBO, KOD-401) with the primer sets in **Table**

S7. The Kozak sequence GCCACC was added immediately 5' of the first ATG during the PCR cloning. The PCR products were TA-ligated into pGEM-T Easy vector (Promega, A1360), sequence-verified, and subcloned into KpnI/XhoI (for *Gata2*) or BamHI/XhoI (for other candidate genes) sites of pENTR1A vector (Thermo Fisher Scientific, A10462). Then, Kozak-CDS sequences were recombined into a *piggyBac*-based Dox-inducible expression vector {Woltjen, 2009 #2364}, which was insulated with D4Z4 sequences as reported previously {Kojima, 2017 #3088}, using LR Clonase II (Thermo Fisher Scientific, 11791). The protein expression from the expression vectors was confirmed in HEK293T cells by Western blot analysis. See the **Western blots** section for details.

The candidate gene vectors and the rtTA vectors were transfected into BVSC H18 ESCs along with the *piggyBac* transposase expression vector {Nakaki, 2013 #2456} by Lipofectamine2000 (Thermo Fisher Scientific, 11668-019). Transfectants were selected with 500 µg/ml G418 (Sigma-Aldrich, A1720) and 4 µg/ml puromycin (Thermo Fisher Scientific, A11138-03). For evaluation of the candidate factors, H18 BVSC ESCs were transfected with the expression vectors of all eight candidates and the rtTA expression vector. After drug selection, all the surviving ESCs were pooled and stored as a bulk stock. The resulting heterogeneous pools of ESCs were induced into PGCLCs, cultured under the expansion condition and treated with 1.5 µg/mL Dox (Clonetech, 631311) with or without 100 nM RA from c3. The differentiation capability towards oocytes was evaluated by immunofluorescence (IF) at c9. See **Immunofluorescence analyses of cultured mPGCLCs** for details of the culture and staining conditions. mPGCLCs in a culture dish were imaged so that a minimum of 20 GFP^+ cells were included per image and percentages for each culture condition/transgene combination were calculated (Fig. 1A). In Fig. 1A, each dot represents the percentage of $SYCP3^+$ cells from a single image. After the evaluation of the "8 factor" transfectants, another round of transfection was conducted: eight different combinations of the candidate genes in which a single candidate gene was subtracted from the cocktail ("7 factor" combinations), and ESC bulk stocks for each combination were established. The oocyte differentiation was evaluated by IF for each combination. Subsequently, the bulk transfectants with the *Zglp1* expression vector and the rtTA expression vector were established, and mPGCLCs derived from such bulk transfectants exhibited an ability to differentiate towards oocytes with Dox alone or with Dox and RA. Finally, ESC clones harboring a single expression vector (*Zglp1*, *Gata2*, *Id1*, or *Msx1)* and the rtTA expression vector were established for detailed analyses. For establishing independent ESC clones, colonies were picked on the 6^{th} or 7^{th} day during drug selection and expanded individually. The presence of transgenes in the chosen ESC clones were evaluated by PCR using the primer sets in **Table S7**. The transgene expression level of each clone was evaluated by mCherry fluorescence upon 1.5 µg/ml Dox administration under the ESC culture condition, and clones with high transgene expression were selected.

Karyotyping of *Zglp1* O.E. ESCs

Metaphase chromosome spreads of ESCs were prepared by treating the cells with 0.03 µg/mL demecolcine (Wako, 045-18761) for 3 hrs at 37 °C, followed by hypotonic treatment with a solution containing 75 mM potassium chloride (Wako, 163-03545) and 5 mM HEPES (Thermo Fisher Scientific, 15630-106) for 25 min at room temperature. Cells were fixed in Carnoy's solution [3:1 mixture of methanol (Nacalai Tesque, 21915-93) and acetic acid (Nacalai Tesque, 00211-95)] and dropped onto slides (Matsunami, S2124) that were pre-soaked with 100% ethanol (Nacalai Tesque, 14713-53) and placed on a dampened Kim Towel (Nippon Paper Crecia, 61000). Slides were air-dried overnight at room temperature, mounted in VECTASHIELD mounting medium with 4',6-diamidino-2-phenylindole (DAPI) (Vector Laboratories, H-1200), and observed under a confocal laser microscope (Olympus, FV1000).

Establishment of *Zglp1* KO ESCs

For generating *Zglp1* KO ESCs, exon 3 of *Zglp1* was targeted with the aim of introducing out-of-frame mutations in front of the evolutionarily conserved GATA-type zinc finger domain of *Zglp1* {Li, 2007 #3122} (Fig. S2). Guide sequence oligos targeting the exon 3 of *Zglp1* were designed using the CRISPR Design Tool (http://tools.genome-engineering.org) {Ran, 2013 #3005}, and the guide oligos were cloned into pX335 plasmid containing Cas9 nickase (Cas9n) (Addgene plasmid #42335) {Ran, 2013 #3004} fused in frame with GSG-p2A-mCherry as previously reported {Kojima, 2017 #3088;Miyauchi, 2017 #3092}. The activities of the guide RNAs (gRNA)/Cas9n pairs were evaluated by a single-strand annealing (SSA) assay, and a gRNA-Cas9n plasmid pair with higher efficiency (pair #1) was selected and electroporated (500 ng each) into one million BVSC H18 ESCs by an electroporator (NEPA GENE, Super Electroporator NEPA21 type II). Three days after the electroporation, ESCs with high mCherry expression (top 0.1%) were sorted using a FACSAria III, and single cells were seeded onto MEFs in a 96-well plate (0.1% gelatin-coated). Eight days after the seeding, expanded colonies were split into two 96-well plates: one plate for genotyping and the other for making frozen-stocks.

For genotyping, sequences surrounding the exon 3 of *Zglp1* were PCR-amplified using KOD Plus Neo DNA polymerase, and the PCR products were analyzed on a microfluidic capillary electrophoresis system (Perkin Elmer, LabChip GX). Clones with potential deletions (i.e., clones with PCR products smaller than the 498 bp product of the wild-type parental clone) were selected for sequencing. For sequencing, the PCR products were cloned into pGEM-T Easy vector, and colonies were picked for direct PCR with KOD FX Neo polymerase (TOYOBO, KFX-201) and the M13-RV and M13-M4 primers. The PCR products were sequenced by Eurofins Genomics (Tokyo) with the SP6 primer, and seven ESC clones harboring out-of-frame deletions were selected for subsequent analyses. The

selected clones were evaluated for their PGCLC induction efficiency, and two clones were selected to investigate phenotypes upon cytokine-mediated oocyte differentiation (Fig. S10A) (E5: IF, RNA-seq; G10: IF, RNA-seq, cell-cycle). Primers/oligos used for the establishment/analyses of *Zglp1* KO ESCs are listed in **Table S7**.

Establishment of *Stra8* KO ESCs

For generating *Stra8* KO ESCs, gRNA/Cas9n expressing vectors targeting the exon 6 of *Stra8* were used as previously reported (Miyauchi, 2017). The gRNA/Cas9n expressing vectors were electroporated into one million H18 BVSC ESCs, and single cells were seeded onto MEFS and expanded. For genotyping, sequences surrounding the exon 6 of *Stra8* were analyzed as previously reported (Miyauchi, 2017), and two clones were selected for further analyses (clone B6 and F6) (Fig. S10B). Primers/oligos used for the establishment/analyses of *Stra8* KO ESCs are listed in **Table S7**.

Fluorescence-activated cell sorting (FACS) and cell cycle analysis

At day 4 of mPGCLC induction, floating aggregates of EpiLCs were incubated with TrypLE Express (Thermo Fisher Scientific, 12604-021) for 8 min at 37 °C, with periodical tap-mixing. The reaction was diluted/quenched with a 5 times volume of DMEM/F12 (Thermo Fisher Scientific, 11320-082) containing 0.1% bovine serum albumin (BSA) fraction V (Thermo Fisher Scientific, 15260-037). The cell suspension was centrifuged at 220×g (Eppendorf, 5702) for 3 min, re-suspended in DMEM/F12 with 0.1% BSA, and passed through a cell strainer (Falcon, 352235). BV^+ cells were sorted with a FACSAria III (BD Biosciences), and collected in mPGCLC expansion culture medium.

The cell-cycle analysis was performed using a Click-iT EdU Flow Cytometry Assay Kit (Thermo Fisher Scientific, C10424) according to the manufacturer's instructions. Cultured mPGCLC-derived cells were treated with 10 µM EdU for 30 min and stained with 0.5 µg/mL 7-AAD (BD Biosciences, 559925).

Immunofluorescence analyses of ovarian germ cells

For immunofluorescence analyses of ovarian germ cells, ovaries were fixed with 2% paraformaldehyde (PFA) (Sigma-Aldrich, 158127 or TAAB Laboratories, P001)/PBS on ice for 1 hr (for ZGLP1 staining), or with 4% PFA/PBS on ice for 1 hr (for other stainings), followed by three washes in PBS containing 0.2% Tween-20 (PBST), and submerged in serial concentrations (10% and 30%) of sucrose (Nacalai Tesque, 30404-45)/PBS overnight at 4 °C. The samples were embedded in the OCT compound (Sakura Finetek, 4583), snap-frozen in liquid nitrogen, and cryo-sectioned at a thickness of 10 µm at -20 °C on a cryostat (Leica, CM1850). The sections were placed on a coated glass slide (MAS-GP type A coated slide glass; Matsunami, S9901) and dried completely.

For immunostaining, the slides were incubated with PBS containing 10% normal donkey serum (NDS) (Jackson Laboratories, 017-000-121) or 10% normal goat serum (NGS) (Vector Laboratories, S-1000), 3% BSA (Sigma-Aldrich, A3059), and 0.2% Triton X-100 (Sigma-Aldrich, T9284) for 1 hr at room temperature, followed by incubation with the primary antibodies diluted in 5% NDS or NGS overnight at room temperature. The slides were washed three times with PBS, then incubated with the secondary antibodies diluted in 5% NDS or NGS with DAPI at 1 µg/mL for 1 hr at room temperature. Following three washes in PBS, the samples were mounted in VECTASHIELD (Vector Laboratories, H-1000) and observed under confocal laser microscopes (Olympus, FV1000 or Zeiss, LSM780). All images were analyzed using Fiji/Image J software {Schindelin, 2012 #3126}. All antibodies used in this study are listed in **Table S8**.

For calculating the percentages of ZGLP1^{+} and STRA8^{+} germ cells, sections containing at least 20 DDX4^{+} cells were analyzed (Fig. 2B). In Fig. 2B, each dot represents the percentage from a single section.

Numbers of DDX4^{+} cells per ovary from *Zglp1*$^{+/+}$ and *Zglp1*$^{-/-}$ embryos at E13.5, E15.5, and E17.5 (Fig. 2D) were estimated as follows: For each ovary, every tenth section of the entire ovary was immunostained for DDX4 and the numbers of DDX4^{+} cells were counted per section. The sum of DDX4^{+} cells from all the sections covering the entire ovary was multiplied by 10 to yield the estimated numbers of DDX4^{+} cells per ovary.

Immunofluorescence analyses of cultured mPGCLC-derived cells

For immunofluorescence analyses of cultured mPGCLC-derived cells, mPGCLCs were cultured on m220 feeder cells seeded on a 0.1% gelatin-coated plate used specifically for imaging (Eppendorf, 0030741005 or Thermo Fisher Scientific, 165305). The cultured mPGCLC-derived cells were fixed at c7 or c9 with 4% PFA/PBS at room temperature for 30 min, followed by three washes in PBS. The fixed cells were incubated in PBS containing 10% NDS, 3% BSA, and 0.2% Triton X-100 for 1 hr, then incubated with the primary antibodies at room temperature for 3 hrs. After three washes in PBS, they were incubated with the secondary antibodies and DAPI at room temperature for 1 hr, washed three times in PBS, and then mounted in VECTASHIELD. Immunostained samples were observed under a Zeiss LSM780 confocal microscope.

Preparation of meiotic cell spread from gonadal germ cells and cultured PGCLC-derived cells

Meiotic cell spread was performed mostly as described in a previous study {Peters, 1997 #3127}. For oocyte spread, ovaries from E15.5 ICR females were treated with a hypotonic extraction solution [30 mM Tris-HCl (Sigma-Aldrich, T5941), 50 mM sucrose (Nacalai Tesque, 30404-45), 17 mM trisodium citrate (Nacalai Tesque, 31404-15), 5 mM ethylenediaminetetraacetic acid (EDTA) (Dojindo, 342-01875), 2.5 mM dithiothreitol

(DTT) (Nacalai Tesque, 14128-46), 0.5 mM phenylmethylsulfonylfluoride (PMSF) (Nacalai Tesque, 27327-94), pH: 8.2-8.4] at 37 °C for 15 min. Ovaries were then dissociated using needles in 100 mM sucrose solution to release oocytes, and the cell suspension was dispersed onto slide glasses (Matsunami, S9901) covered with 1% PFA (TAAB Laboratories, P001)/0.2% Triton X-100 (pH: 9.2-9.4). The slides were incubated at room temperature overnight in a humidified chamber. Finally, the slides were air dried and washed with 0.5% DRIWEL (Fujifilm, 009517) for 2 min at room temperature.

For spermatocyte spread from P20 males, the tunica albuginea was removed from testes, and the tubules were gently separated and treated with hypotonic solution for 45 min at room temperature. Spermatocytes were released in 100 mM sucrose solution, and the cell suspension was fixed and processed in the same way as the oocyte preparation.

For cell spread from mPGCLC culture, cultured mPGCLC-derived cells were treated with TrypLE Express at 37°C for 5 min, washed/quenched with DMEM/F12(1:1) containing 0.1% BSA fraction V, and harvested from culture wells. Harvested cells were centrifuged at 220×g for 3 min to form cell pellets. Cell pellets were dislodged by gentle tapping and then washed in PBS (Thermo Fisher Scientific, 20012-027). Following centrifugation and pellet dislodging, cells were treated with the hypotonic solution at 37°C for 1 hr, with a tap mix at 30 min. Subsequently, the cells were fixed and processed in the same way as the oocyte and spermatocyte preparation. We noticed that the cell preparations from cultured PGCLC-derived cells spread less efficiently than the oocyte and spermatocyte preparation, making them appear “smaller” than the oocyte and spermatocyte preparation.

The spread slides were incubated in PBS containing 10% NDS, 3% BSA, or 0.2% Triton X-100 for 1 hr, then incubated with the primary antibodies at room temperature overnight. After three washes in PBS, they were incubated with the secondary antibodies and DAPI at room temperature for 1 hr, washed three times in PBS, and then mounted in VECTASHIELD. Immunostained cells were observed under a Zeiss LSM780 confocal microscope.

Staging of meiotic prophase in ovarian sections and meiocyte spread

Substages of meiotic prophase I in ovarian sections (Fig. 2F) were classified based on DAPI and SYCP3 staining as follows (see Fig. S4C). Preleptotene: strong dot-like accumulations of SYCP3 staining in the nucleus with weak DAPI accumulations at the nuclear periphery. Leptotene: numerous short threads of SYCP3 staining with some accumulation of DAPI at the nuclear periphery deriving from the “telomere bouquet” formation {Chikashige, 1994 #3139;Fernandez-Capetillo, 2003 #3141;Tankimanova, 2004 #3140}. Zygotene: numerous long thin threads of SYCP3 staining with several prominent DAPI accumulations at the nuclear periphery. Pachytene: thick SYCP3 staining with prominent DAPI accumulations at the nuclear periphery. Aberrant cell

configuration was characterized by numerous DAPI-dense chromocenters with faint dot-like accumulations of SYCP3 staining.

Substages of meiotic prophase I from cell spread (Fig. 3B, S5E) were classified based on SYCP1 and SYCP3 staining as follows. Leptotene: short segments of SYCP3 staining, with no SYCP1 staining. Zygotene: thin long lines of SYCP3, with small segments of SYCP1 staining co-localizing with portions of SYCP3 staining. Pachytene: thick and short lines of SYCP3 staining, with SYCP1 staining co-localizing with most SYCP3 staining.

RNA extraction, cDNA synthesis and qPCR analysis

Cells were lysed and total RNAs were purified using an RNeasy Micro Kit (Qiagen, 74004) or NucleoSpin RNA XS (Macherey-Nagel, 740902.50) according to the manufacturer's instructions. RNA was analyzed/quantified by a UV-Vis spectrophotometer (Thermo Fisher Scientific, NanoDrop 2000), or quantified by a fluorometer (Thermo Fisher Scientific, Qubit 2.0). Synthesis of cDNAs from poly(A)$^+$ RNA was performed from 100 ng RNA using a SuperScript III First-Strand synthesis kit (Thermo Fisher Scientific, 18080-51) with oligo-dT primer according to the manufacturer's instructions or reverse transcribed/amplified from 1 ng RNA as previously described {Kurimoto, 2006 #213;Nakamura, 2015 #2714}. qPCR on cDNAs was performed using the Power SYBR Green Master Mix (Applied Biosystems, 4367659) on a real-time qPCR system (Biorad, CFX384). The gene expression levels are presented as ΔC_t (in a $\log_2$ scale) normalized with the mean C_t values of two housekeeping genes, *Arbp* and *Ppia*. The primer sequences for the genes examined in this study are listed in **Table S7**.

Western blots

Snap-frozen cell pellets were re-suspended in cell re-suspension buffer [20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 10 mM KCl] supplemented with 1x Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific, 87786). The re-suspended cell pellets were sonicated using a Bioruptor sonicator on the "High" setting for 10 seconds for 3 cycles. Trition X-100 was added at the concentration of 1% and the sonicates were kept on ice for 30 min. The sonicated solution was then centrifuged at 16,500G for 15 min at 4 °C. The supernatant was then treated with 1×Laemmli's Sampling Buffer [62.5 mM Tris-HCl (pH 6.8), 140 mM 2-ME (Nacalai Tesque, 21438-82), 2% SDS (Nacalai Tesque, 02873-75), 0.025% bromophenol blue (Nacalai Tesque, 05808-32), 10% glycerol (Wako, 075-00616)] and heated at 98 °C for 5 min. Lysates from 5 x 10^4 cells were separated on 4-12% NuPAGE Bis-Tris Gels (Thermo Fisher Scientific, NP0322BOX), blotted onto an Immobilon-P PVDF transfer membrane (Merck Millipore, IPVH00010), and incubated with primary antibodies.

Chemiluminescence reaction was performed using Chemi-Lumi One Super (Nacalai Tesque, 02230), and luminescent signals were detected by a LAS4000IR (FUJIFILM) or FUSION Solo S (Vilber Lourmat).

Histology

For histological analyses, testes and ovaries were fixed with Bouin's solution for 16 hrs, washed with 70% ethanol three times, dehydrated with serial concentrations of ethanol, and embedded in paraffin. Tissues were sectioned at a thickness of 7 μm. After de-paraffinization with Xylene (Nacalai Tesque, 36612-93) three times and dehydration with a graded series of ethanol, the sections were stained with hematoxylin and eosin.

Sample preparation and library preparation for RNA-seq analyses

For transcriptome analysis, total RNAs of ESCs, EpiLCs, day 4 (c0) mPGCLCs (BV-sorted), and c3, c4, c5, c7, and c9 cultured mPGCLC-derived cells (SC-sorted) were extracted and purified using an RNeasy Micro Kit or NucleoSpin RNA XS. 1 ng of total RNAs from each sample was reverse transcribed and amplified as previously reported {Kurimoto, 2006 #213;Nakamura, 2015 #2714}. The cDNA library preparation for Nextseq500 (Illumina) was performed as previously reported {Ishikura, 2016 #2985}.

Mapping of sequenced reads from RNA-seq and conversion to gene-expression levels

RNA-sequence reads were processed mainly as described previously {Nakamura, 2015 #2714}. In brief, all RNA-seq reads were processed with cutadapt v1.9.1 {Martin, 2011 #2665} to remove library adaptor and poly-A tail sequences and low quality bases with the options "-e 0.1", "-q 20", "-n 2", "-O 1", "-m 30", "-a CTCGAGGGCGCGCCGGATCC", "-g CTCGAGGGCGCGCCGGATCC", "-a AAAAAAAAAAAAAAAAAAAA", and "-a TTTTTTTTTTTTTTTTTTTT". Reads of 30 bases or longer were used for mapping onto the mouse genome (mm10) and ERCC spike-in RNA sequences using TopHat v2.1.0/Bowtie v1.1.2 {Trapnell, 2009 #3130;Langmead, 2009 #2199} with the "--bowtie1", "--library-type fr-secondstrand" and "--no-coverage-search" options. Mapped reads were sorted by genome coordinates using SAMtools v1.3 {Li, 2009 #2634}, and read counts per gene were counted using HTSeq-count (HTSeq v0.9.1 {Anders, 2015 #2699}) with the default settings and converted to RPM. All transcript references were extended up to 10 kb at the transcription termination site for maximum read recovery as described previously {Nakamura, 2015 #2714}. The read mapping statistics for the RNA-seq are summarized in Table S1.

Data analysis of the RNA-seq

The RNA-seq data were analyzed by the R software package (version 3.5.3.) and EXCEL (Microsoft). The RPM values were $\log_2$-converted, and $\log_2$(RPM+1) values were defined as the gene expression levels. Genes showing $\log_2$(RPM+1) values > 2 [greater than ~ 3 copies per cell {Nakamura, 2015 #2714}] in at least one sample were used for

analysis. Principal component analysis (PCA) was performed using the prcomp function. Differentially expressed genes (DEGs) were defined as those bearing $\log_2$(RPM+1) > 4 and a difference in the $\log_2$(RPM+1) values greater than 2 [i.e., fold change > 4] between samples of interest. Gene ontology (GO) analysis was performed using DAVID {Huang da, 2009 #2627}.

Generation of V5-Zglp1-expressing ES clones

For ChIP experiments, the D4Z4-insulated *Zglp1* expression cassette was modified to incorporate the V5 tag sequence upstream of *Zglp1* CDS. First, a sequence containing the BamHI-Kozak-V5-[S(GGGGS)$_{x3}$]-SpeI was PCR-cloned from the V5-*Tfap2c* expression vector {Nakaki, 2013 #2456}, and *Zglp1* CDS was PCR-cloned while incorporating the SpeI and NotI sites at the 5' and 3' ends from the *Zglp1* expression vector. The two PCR products were TA-ligated into the pGEM T Easy vector, and the sequences were verified. Subsequently, the V5-(G4S)linker fragment and the SpeI-*Zglp1*CDS-NotI fragment were subcloned into the pENTR1A vector using BamHI, SpeI, and NotI. The resultant V5-(G4S)linker-*Zglp1*CDS was recombined into the D4Z4-insulated destination vector by LR Clonase II.

The modified expression vector was transfected along with the rtTA expression vector and the transposase expression vector into BVSC BDF1-2-1 ESCs by Lipofectamine2000. The BVSC BDF1-2-1 (XY) ESC line was chosen because PGCLCs derived from this ESC line possess high proliferative capability (~25-fold increase) upon mitotic expansion culture {Ohta, 2017 #3061}, and the expansion culture yields experimental materials sufficient for ChIP experiments (4 x 10^6 cells/ChIP). Lines with high transgene expression were selected based on mCherry expression during ESC culture and evaluated for PGCLC induction. A clone (V5-Zglp1 O.E. clone A12) possessing a robust PGCLC induction efficiency (20% BV+ cells / aggregate) was selected.

Chromatin immunoprecipitation (ChIP)

During the expansion culture, 1.5 µg/mL Dox with/without 100 nM RA was added at c3, and SC^+ cells with high mCherry expression (top 50%) were sorted by FACS at c7 and c8. For ESC ChIP, ESCs were treated with 1.5 µg/mL Dox and 100 nM RA for 24 hrs during 2i+LIF culture and harvested. The harvested cells were centrifuged at 220×g for 3 min, suspended in 1 mL of PBS and fixed by the addition of 27 µL of 36.5% formalin (Sigma-Aldrich, F8775) for 12 min at room temperature with constant mixing on a tube rotator. Fixation was terminated by the addition of 110 µL of 1.5 M glycine (Wako, 077-00735). The fixed cells were washed twice in PBS (Thermo Fisher Scientific, 20012-027) containing 0.1% BSA fraction V and were lysed in 400 µL of SDS-lysis buffer [50 mM Tris-HCl (pH 8.0) (Nacalai Tesque, 35435-11), 1% SDS, 10 mM EDTA, and 1 mM PMSF] on ice for 10 min. The lysate was split amongst 4 tubes (Diagenode, 315-81401) at 100 µL/tube, and chromatins were sonicated using Picoruptor (Diagenode, 316-81311)

at 4 °C for 19 cycles, yielding DNA fragments of 100–300 bp. The sonicated products from 4 tubes were combined and centrifuged at 16,000×g for 5 min. The supernatants were diluted in 1 mL ChIP dilution buffer [16.7 mM Tris-HCl (pH 8.0), 0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA (Nacalai Tesque, 14347-21) and 167 mM NaCl (Nacalai Tesque, 31334-51)] and divided into 0.1 mL aliquots as input DNA and 1.3 mL aliquots as ChIP samples.

For ChIP samples, 15 µg of rabbit polyclonal anti-V5 antibody (Merck Millipore, AB3792) was reacted with 40 µL of M280 Dynabeads Protein G (Thermo Fisher Scientific, 10003D) at room temperature for 40 min. ChIP samples were then mixed with the antibody-Dynabeads complexes and incubated at 4 °C for 22 hrs with gentle mixing on a tube rotator. The chromatin-antibody-Dynabeads complexes were recovered using a DynaMag-2 magnet (Thermo Fisher Scientific, 123.21D) and washed with 200 µL of a low salt buffer [20 mM Tris-HCl (pH 8.0), 0.1% SDS, 1% Triton X-100, 2 mM EDTA, and 150 mM NaCl], a high salt buffer [20 mM Tris-HCl (pH 8.0), 0.1% SDS, 1% Triton X-100, 2 mM EDTA, and 500 mM NaCl], RIPA buffer [10 mM Tris-HCl (pH 8.0), 1% sodium deoxycholate (Nacalai Tesque, 10712-96), 1% NP40 (Nacalai Tesque, 23640-94), 250 mM LiCl (Wako, 244-00461), and 1 mM EDTA], and TE buffer [10 mM Tris-HCl (pH 8.0), and 1 mM EDTA]. After transferring to a new tube, the chromatin was incubated in 50 µL of elution buffer [0.1 M $NaHCO_3$ (Nacalai Tesque, 31213-15), 1% SDS, 10 mM DTT Nacalai Tesque, 14128-91], and 60 ng/µL tRNA (Roche, 10109517001)] for 15 min at room temperature. The eluted chromatins were separated from the Dynabeads using DynaMag-2, then reverse-crosslinked by the addition of 8 µL of 2.5 M NaCl and incubated at 65 °C for 16 hrs. Then 2 µL of 0.5 M EDTA, 4 µL of 1 M Tris-HCl (pH 6.5), and 0.2 µL of 20 mg/mL proteinase K (Sigma-Aldrich, P2308) were added to the reaction mix, and the reaction mix was further incubated at 45 °C for 1 hr. The ChIP'd DNAs were purified with Qiaquick PCR purification columns (Qiagen, 28104) using buffer EB (Qiagen, 19086) containing 4 ng/µL of tRNA. Input chromatins were reverse-crosslinked, treated with proteinase K, and purified in the same way as the ChIP'd samples.

ChIP qPCR analysis

To assess the quality of ChIP experiments, qPCR on ChIP'd DNAs and input DNAs was performed using the Power SYBR Green Master Mix on a BioRad CFX384 qPCR system with the primer pairs listed in **Table S7**. The open reading frame 2 (ORF2) of the long interspersed nuclear element 1 (LINE1) was used as a negative control as previously reported {Kurimoto, 2015 #2715;Mitani, 2017 #3101}. ChIP'd samples with a signal-to-noise ratio more than 32-fold that of the negative control were used for the ChIP-seq experiments.

Library preparation for ChIP-seq

Library preparations for the next-generation sequencing were performed as previously described {Mitani, 2017 #3101}, except that ChIP'd DNAs were not sonicated further by Covaris. The libraries were sequenced on NextSeq (Illumina) to generate 86 bp single-end reads using a custom read primer {Mitani, 2017 #3101}. A total of 8 samples [ESC Dox+RA (ChIP'd), ESC Dox+RA (input), c7 Dox only (ChIP'd), c8 Dox only (ChIP'd), c8 Dox only (input), c7 Dox+RA (ChIP'd), c8 Dox+RA (ChIP'd), and c8 Dox+RA (input)] were sequenced.

ChIP-seq mapping, peak identification, and peak selection

All reads were processed with the trim_galore program (Trim Galore! version 0.4.1 and cutadapt 1.9.1) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) using the "-a ATCACCGACTGCCACGCCTTGGCCGTACAGCA GCCT" and "-q 20" options to remove the Barcode Internal+12 mer adaptor sequences {Mitani, 2017 #3101} and low quality reads, respectively. The processed reads of 20 bases and longer were mapped on the mouse genome mm10 by Bowtie v1.1.2 {Langmead, 2009 #2199} with the "-n 3" and "–m 1" options. The mapped reads in Sequence Alignment/Map (SAM) format were sorted by chromosome by Picard-tools v2.1.0 (http://piard.sourceforge.net), then processed with SAMtools v1.3 ("-s" option {Li, 2009 #2634}) and bedtools v2-2.25.0 {Quinlan, 2010 #3134} (command line: bamToBed) to remove duplicate reads and converted to bed format, respectively.

Peak calling for V5-ZGLP1 was performed using MACS v2.1.1.20160309 with the default settings {Zhang, 2008 #2635}. For MACS peak calling, samples were processed against the respective input read data for each cell type and treatment. Following the peak calling for each sample, we consolidated peaks detected in close proximity (within 500 bp) across different samples as a single peak as follows. 1) We first listed peak summits from all samples, and two neighboring peak summits located within 100 bp were joined as a single summit (i.e., a new summit was created halfway between the two original summits. 2) Newly called summits within a distance of 200 bp were subsequently processed in the same way. 3) The process was repeated for the distances of 300, 400, and 500 bp. Following the peak summit-bundling process, 32,927 genomic loci (i.e., 32,927 peak summits) were considered as potential ZGLP1-binding sites.

All the sequenced reads and their genomic coordinates were listed in a MySQL database table (version 4.6.35). The database was then queried for the coordinates of genomic loci, and hit counts were identified as read counts when the center of a read, which is extended for 250 bp from the first base of the sequenced read (i.e., the center is located at the 125th base from the first base), was overlapped with a genomic coordinate of interest by a single base. For ChIP samples, reads within 500 bp of the peak summits were counted and then normalized as reads per million (RPM). For input samples, reads within 5,000 bp of the peak summits were counted and then normalized in RPM.

The following steps were taken for the selection of ZGLP1-binding sites. 1) We first averaged the read tag density (in RPM) from three input samples (i.e., ESC, c8 Dox only, and c8 Dox+RA) at each potential ZGLP1-binding site and selected 32,725 sites with (0.1 < averaged input RPM < 0.7) for subsequent analysis. 2) RPM values across ChIP'd samples were normalized so that the mean of each sample matched the mean of the c8 Dox+RA sample. 3) The ZGLP1 IP level (IP/input) at each binding site was defined as the ratio between the normalized RPM value of a ChIP'd sample and the averaged RPM value from the three input samples. 4) The IP/input values in mPGCLC-derived cells for Dox only and Dox+RA samples were averaged between c7 and c8 and $\log_2$-converted [i.e., $\log_2((IP/input) + 0.1)$]. 5) Lastly, ZGLP1-binding sites in mPGCLC-derived cells were defined as sites with [averaged $\log_2((IP/Input) + 0.1) > 1$ and $\log_2((IP/Input) + 0.1)$ differences between c7 and c8 samples < 0.585].

The selection process yielded 16,576 ZGLP1-binding sites for the Dox only treatment and 14,662 binding sites for the Dox+RA treatment in mPGCLC-derived cells. Association of ZGLP1-binding sites with nearby genes was performed using Homer v 4.9.1 {Heinz, 2010 #3135} (command line: annotatePeaks.pl) with the default settings. For annotation of the binding sites, we focused our analysis on binding sites that were associated with genes included in our RNA-seq analysis (25,315 genes; 16,290 sites for Dox only treatment, 14,409 sites for Dox+RA treatment) and binding sites that were within 15 kb of the transcription start sites (TSSs) of the associated genes (14,869 sites for Dox only treatment, 13,379 sites for Dox+RA treatment). IGV 2.3.91 was used for the visualization of ChIP-seq tracks (http://www.broadinstitute.org/igv).

ChIP-seq for histone H3K4me3 and H3K27me3 modifications

ChIP-seq experiments of H3K4me3 and H3K27me3 for ESCs, EpiLCs, d2 mPGCLCs, d6 mPGCLCs (GSE60204) {Kurimoto, 2015 #2715}, and d4c7 mPGCLCs (GSE87645) {Ohta, 2017 #3061} were performed previously. Previously, the read data were mapped on the mouse mm10 genome using Bowtie v1.1.2, processed using Picard-tools v2.1.0 and SAMtools v1.3, and peak-called using MACS v2.1.0. The sequenced read data were listed in a MySQL database table. For the current study, read counts around TSSs (within 500 bp for H3K4me3 and within 1000 bp for H3K27me3) and around ZGLP1-binding sites (within 500 bp for H3K4me3 and within 1000 bp for H3K27me3) were calculated from the MySQL database as described for the V5-ZGLP1 ChIP-seq for ChIP'd samples. Read counts within 5000 bp around TSSs and ZGLP1-binding sites were calculated for the input samples. We first excluded 419 sites that possess no sequenced reads from input samples for both H3K4me3 and H3K27me3 ChIP. Subsequently, at each TSS and ZGLP1-binding site, the read density (in RPM) for input reads from two studies {Kurimoto, 2015 #2715;Ohta, 2017 #3061} was averaged to provide a combined input read, and the read density for each ChIP'd sample was normalized by the mean of ESC

rep1. At each TSS and ZGLP1-binding site, IP levels (IP/input) were calculated as the ratio between the read density of the ChIP'd sample and the read density of the averaged input and converted into a log scale [$\log_2$((IP/input) + 0.1)]. To classify a chromatin state around TSSs for sexually uncommitted PGCLCs (c7 PGCLCs), threshold IP levels for each histone mark were defined as follows. RNA-seq data for c7 PGCLCs were downloaded from GSE87644 {Ohta, 2017 #3061}, and genes with expression levels [$\log_2$(RPM+1)] between 1.5 and 3.5 (990 genes) were retrieved. Threshold IP levels were defined as the mean IP levels around the TSSs of the 990 genes for each histone mark.

Supplementary Figure Legends

Fig. S1. Screening of key effectors of the BMP signaling for the oogenic fate.

(A) Scheme for the selection of key effectors of the BMP signaling for the oogenic fate induction. At day 4 of mPGCLC induction, BV^+ cells were sorted by FACS and cultured on m220 feeder cells. At c3 of the mitotic expansion culture, mPGCLCs were treated with BMP2 or cultured without BMP2 (control). $BVSC^+$ cells were sorted at c4 and processed for RNA-seq. The following criteria were used to select candidate genes: 1) genes encoding transcription factors (TFs), 2) genes up-regulated [$\log_2$ (RPM+1) >4, $\log_2$ (fold change) >2] in BMP2-treated mPGCLC-derived cells compared to the control, and 3) genes expressed more highly in female germ cells than in male germ cells at E12.5 [$\log_2$ (RPM+1) >4, $\log_2$ (fold change) >2]. The criteria resulted in the selection of *Id1* and *Zglp1*. Additionally, *Msx1*, *Msx2*, *Gata2*, *Stra8*, which fulfill the criteria 1) and 3), and *Id3*, which is a paralogue of *Id1* and has been known as a downstream effector of BMP signaling, {Ho, 2011 #2424} were included. Finally, we included an RNA regulator, *Dazl*, which has been shown to function as a licensing factor for germ-cell sex specification {Lin, 2008 #1389}.

(B) Expression dynamics of the eight candidate genes in PGCs (black, sex undetermined) and female (red) and male (blue) germ cells between E9.5 and E15.5.

(C) Expression of GFP(BV/SC), DDX4/mCherry and SYCP3 in mPGCLC-derived cells over-expressing *Gata2*, *Id1*, and *Msx1*. Secondary antibodies for DDX4 and SYCP3 were conjugated with Alexa Fluor 568 and 647, respectively. Thus, the fluorescent signals for DDX4 (cytoplasmic) include the signals from mCherry protein (nuclear and cytoplasmic). Bars: 10 μm.

(D) The numbers of $SYCP3^+$ cells in mPGCLC-derived cells over-expressing *Gata2*, *Id1*, and *Msx1*. Three independent over-expressing clones for each gene were analyzed at c9 upon Dox and RA treatment from c3.

Fig. S2. Evolutionary conservation of the ZGLP1 protein across metazoan phyla.

(A) Scheme for the mouse ZGLP1 protein.

(B) Alignment of the amino acid sequences around the two zinc fingers of ZGLP1 proteins across metazoan phyla. The alignment was performed by Clustal Omega {Sievers, 2011 #3137}. Light red shading indicates fully conserved amino acid residues, while light orange shading indicates residues with conservation between groups of amino acids with similar properties.

Fig. S3. The expression of *Zglp1* in post-natal spermatogenic cells.

(A) The expression of β-galactosidase protein (green) and DDX4 (magenta) in a testicular section from post-natal day (P) 15 $Zglp1^{+/-}$ heterozygous mice. The *Zglp1* null allele was generated by replacing all five exons of *Zglp1* by the bacterial β-galactosidase cDNA {Li, 2007 #3122}. Bar: 10 μm.

(B) The expression of *Zbtb16*, *Zglp1*, *Stra8*, and *Sycp3* in unsorted spermatogenic cells

from adult male mice. Single-cell transcriptomes derived from adult male mice {Hermann, 2018 #3160} were re-analyzed using The Loupe Cell Browser v3.0.1 (10x Genomics).
(C) The expression of *Gfra1*, *Zglp1*, *Stra8*, and *Kit* in *Id4*-EGFP^{+} spermatogenic cells from P6 male mice. Single-cell transcriptomes {Hermann, 2018 #3160} were re-analyzed using The Loupe Cell Browser v3.0.1 (10x Genomics).

Fig. S4. *Zglp1* is essential for the oogenic fate determination in pre-natal ovaries.
(A) Expression of ZGLP1 (green) and DDX4 (magenta) in ovaries from *Zglp1*$^{+/+}$ and *Zglp1*$^{-/-}$ females at E13.5. Bar: 10 μm.
(B) Ovarian sections from *Zglp1*$^{+/+}$ and *Zglp1*$^{-/-}$ females at P8 and 6 weeks stained by hematoxylin and eosin. Ovaries are encircled with black dotted lines. Black arrows: primordial follicles. Black arrowheads: primary follicles. A red arrow: antral follicles. Bars: 100 μm (left), 10 μm (middle), 200 μm (right).
(C) Representative images of germ cells immunostained for DDX4 (magenta) and SYCP3 (yellow) in ovaries from *Zglp1*$^{+/+}$ and *Zglp1*$^{-/-}$ females at E15.5. Substages of meiotic prophase I are classified based on DAPI and SYCP3 staining as described in **Materials and Methods**. White arrows: telomere clustering. Bars: 5 μm.
(D) Expression of STRA8 (yellow) in DDX4^{+} (magenta) cells in ovaries from *Zglp1*$^{+/+}$ and *Zglp1*$^{-/-}$ females at E15.5. Bars: 10 μm.
(E) Percentages of STRA8^{+} cells among DDX4^{+} cells from two embryos for each genotype at E15.5. The numbers of DDX4^{+} cells analyzed for each embryo are denoted in parentheses.
(F) Expression of DAZL (magenta), DMC1 (green) and γH2AX (cyan) in ovaries from *Zglp1*$^{+/+}$ and *Zglp1*$^{-/-}$ females at E15.5. Bars: 10 μm.
(G) Percentages of DMC1^{+} and/or γH2AX^{+} cells among DAZL^{+} cells from two embryos for each genotype at E15.5. The numbers of DAZL^{+} cells analyzed for each embryo are denoted in parentheses.

Fig. S5. *Zglp1* is critical for meiotic progression in post-natal testes.
(A) Testicular sections from *Zglp1*$^{+/+}$ and *Zglp1*$^{-/-}$ males at P7, P15 and P20 stained by hematoxylin and eosin. Bars: 10 μm.
(B) Analysis of the seminiferous tubules from *Zglp1*$^{+/+}$ and *Zglp1*$^{-/-}$ males at P7, P15 and P20. The tubules are categorized based on the most advanced cell types in the tubule or the presence of dead/abnormal cells.
(C) Expression of DAZL (magenta), DMC1 (green) and γH2AX (cyan) in testes from *Zglp1*$^{+/+}$ and *Zglp1*$^{-/-}$ males at P15. Bars: 10 μm.
(D) Expression of DAZL (magenta) and SYCP3 (yellow) in testes from *Zglp1*$^{+/+}$ and *Zglp1*$^{-/-}$ males at P15. Bars: 10 μm.
(E) Spermatocyte spreads from *Zglp1*$^{+/+}$ and *Zglp1*$^{-/-}$ males at P20 immunostained with SYCP1 (red), SYCP3 (green), and γH2AX (cyan), and counterstained with DAPI (gray).

Substages of meiotic prophase I are defined as in **Materials and Methods**. White arrowheads: XY bodies. White arrows: abnormal chromosomes with residual γH2AX staining. Bars: 10 μm.
(F) Quantification of the $SYCP3^+$ cells at substages of meiotic prophase I from (E).
(G) Testes from *Zglp1*$^{+/+}$ and *Zglp1*$^{-/-}$ males at 6 weeks of age. Bars: 2 mm.
(H, I) Testicular sections from *Zglp1*$^{+/+}$ and *Zglp1*$^{-/-}$ males at 6 weeks and 6 months of age stained by hematoxylin and eosin. Black arrows: elongated spermatids. Black arrowheads: round spermatids. Red arrows: dead or abnormal cells. Bars: 10 μm.

Fig. S6. Roles of BMP signaling for the activation of *Zglp1*.
(A) Expression of RA-responsive genes {Balmer, 2002 #3161}{Mahony, 2011 #3162} in response to titrated concentrations of RA during mPGCLC culture. mPGCLCs were provided with DMSO, 0.1, 1, 10, or 100 nM of RA from c3 to c4. At c4, SC^+ cells were sorted and the expression of RA-responsive genes were analyzed by qPCR. For Figs. S6A, S6B, S6C, and S6D, the ΔC_t from the average C_t values of two housekeeping genes, *Arbp* and *Ppia*, is shown on the $\log_2$ scale for each gene. The mean values from two independent experiments, each with two technical replicates, are shown. The color coding is as indicated. Error bars: SDs. N.D.: Not Detected.
(B) Expression of *Zglp1* and *Id1* in mPGCLC-derived cells during control culture, culture with BMP2, with RA, and with BMP2 and RA. SC^+ cells were sorted at c3.5, c4, c5, c7, c9 and the expression of *Zglp1* and *Id1* was analyzed by qPCR.
(C) Expression of *Zglp1* in mPGCLC-derived cells cultured with BMP2 in the presence of titrated concentrations of LDN193186. mPGCLCs were cultured with BMP2 and DMSO, or BMP2 and 1, 10, or 100 nM of LDN193186 from c3 to c7. At c7, SC^+ cells were sorted and the expression of *Zglp1* was analyzed by qPCR.
(D) Expression of RA-responsive genes in PGCLC-derived cells under a control culture condition, or in culture with BMP2, RA, or both BMP2 and RA. SC^+ cells were sorted at c3.5, c4, c5, c7, and c9, and the expression of RA-responsive genes was analyzed by qPCR.
(E) Kinetics of 5-methylcytosine (5mC) levels within the *Stra8* promoter CpGs (c0, c3, c7) {Ohta, 2017 #3061} and the expression of *Stra8* (c0, c3, c7, c9) during control culture. For the promoter 5mC levels, whole-genome bisulfite sequencing data from a published study {Ohta, 2017 #3061} were used. The RNA-seq data from a published study {Miyauchi, 2017 #3092} were used for the expression levels of *Stra8*.

Fig. S7. Characterization of *Zglp1* over-expressing (O.E.) clones.
(A) Expression of ZGLP1 and α-tubulin in the parental clone (clone H18) and three *Zglp1* O.E. clones in response to Dox treatment (1.5 μg/mL) for 24 hrs during ESC culture.
(B) Karyotype of the parental clone and the selected *Zglp1* O.E. clones. The *Zglp1* O.E. clones most likely lost one X chromosome during the cloning. We assume this has little, if any, influence on the female sex determination of mPGCLCs, since we showed

previously that both XY and XX germ cells entered into the female pathway in a similar manner {Miyauchi, 2017 #3092}, and because the meiotic prophase stage distribution and the transcriptome that the XO ESC-derived ZGLP1-overexpressing cells acquired at c7/c9 were comparable to those that the parental XX ESC (BVSC H18)-derived cells acquire at c9 (Fig. 3A-C) {Miyauchi, 2017 #3092}.

(C) Expression of the key genes for the oogenic fate/meiotic entry in mPGCLC-derived cells at c7 from the parental clone and three *Zglp1* O.E. clones (H2, C2, A2) under the indicated culture conditions. At c7, SC^+ cells were sorted and the gene expression was analyzed by qPCR. For Fig. S7C and S7F, the ΔC_t from the average C_t values of two housekeeping genes, *Arbp* and *Ppia*, is shown on the $\log_2$ scale. The mean values from two independent experiments, each with two technical replicates, are shown. The color coding is as indicated. Error bars: SDs. N.D.: Not Detected.

(D) Expression of Stella-ECFP (SC, cyan), DAZL/mCherry (magenta), STRA8 (red), and SYCP3/Blimp1-mVenus (BV) (yellow) in c9 mPGCLC-derived cells from *Zglp1* O.E. clone A2 cultured as indicated. Secondary antibodies for DAZL and SYCP3 were conjugated with Alexa Fluor 568 and 488, respectively. Thus, the fluorescent signals for DAZL (cytoplasmic) include the signals from mCherry protein (nuclear and cytoplasmic). The signals for SYCP3 are nuclear, while the signals for BV are mostly on the cellular membrane. The expression of SYCP3 can therefore be distinguished from that of BV. Bars: 10 μm.

(E) Percentages of $SYCP3^+$ cells among SC^+ cells at c9 from the parental clone and three *Zglp1* O.E. clones cultured as indicated. Data from three biological replicates, each with at least 100 SC^+ cells, are shown. The color coding is as indicated. Error bars: SDs. N.D.: Not Detected.

(F) Expression of the key genes for the oogenic fate/meiotic entry in c9 mPGCLC-derived cells from the parental clone and *Zglp1* O.E. clone A2 under the indicated culture conditions. mPGCLCs from the parental clone were cultured with BMP2 and RA in the presence of titrated concentrations of LDN193186. mPGCLCs from clone A2 were cultured with Dox and RA in the presence of titrated concentrations of LDN193186. At c9, SC^+ cells were sorted and the gene expression was analyzed by qPCR.

Fig. S8. Activation of the oogenic/meiotic program by ZGLP1 under the inhibition of RA/RAR activity.

(A) Expression of GFP (BV/SC) (cyan), DDX4/mCherry (magenta) and SYCP3 (yellow) in c9 mPGCLC-derived cells from *Zglp1* O.E. clone A2 cultured under a serum-free condition with WIN18,446 (1 μM), as indicated. The secondary antibody for DDX4 is conjugated with Alexa Fluor 568. Thus, the fluorescent signals for DDX4 (cytoplasmic) include the signals from mCherry protein (nuclear and cytoplasmic). Bars: 10 μm.

(B) Percentages of $SYCP3^+$ cells among SC^+ cells at c9 from *Zglp1* O.E. clone A2 cultured as indicated. Data from two biological replicates, each with at least 100 SC^+ cells, are shown. Error bars: SDs. N.D.: Not Detected.

(C) Expression of key genes for the oogenic fate/meiotic entry and RA-responsive genes in mPGCLC-derived cells at c4, c5, c7, and c9 from *Zglp1* O.E. clone A2 cultured under a serum-free condition with WIN18,446 (1 μM), as indicated. SC^+ cells were sorted and the gene expression was analyzed by qPCR. For Figs. S8C, S8D and S8G, the ΔC_t from the average C_t values of two housekeeping genes, *Arbp* and *Ppia*, is shown on the $\log_2$ scale. The mean values from two independent experiments, each with two technical replicates, are shown. The color coding is as indicated. Error bars: SDs. N.D.: Not Detected.
(D) Expression of *Stra8*, *Rec8*, and *Cyp26a1* in mPGCLCs (clone H18) cultured with 100 nM RA and titrated concentrations of BMS493 from c3 to c4. At c4, SC^+ cells were sorted and the gene expression was analyzed by qPCR.
(E) Expression of GFP (BV/SC) (cyan) and SYCP3 (yellow) in ZGLP1 O.E. mPGCLC-derived cells from the clone A2 cultured with BMS483, as indicated. Bars: 10 μm.
(F) Percentages of $SYCP3^+$ cells among $GFP(BV/SC)^+$ cells at c7 from *Zglp1* O.E. clone A2 cultured as indicated. Data from two biological replicates, each with at least 100 GFP^+ cells, are shown. Error bars: SDs.
(G) Expression of key genes for the oogenic fate/meiotic entry and RA-responsive genes in mPGCLC-derived cells at c4, c5, c7, and c9 from *Zglp1* O.E. clone A2 cultured with BMS493, as indicated. SC^+ cells were sorted and the gene expression was analyzed by qPCR.

Fig S9. Transcriptome properties of mPGCLC-derived cells overexpressing ZGLP1 under relevant conditions.
(A, B) Principal component analysis (PCA) of the transcriptomes of germ cells *in vivo* (E9.5-E11.5 PGCs, E12.5-E15.5 female and male germ cells) and mPGCLC-derived cells of *Zglp1* O.E. clone A2 cultured under a serum-free condition with WIN18,446 (1 μM) (A) or cultured with BMS483 (1 μM). Each plot represents one transcriptome. Transcriptomes from two replicates (except in the case of E9.5 PGCs, and E14.5 and E15.5 female germ cells, for which there were three replicates each) are plotted. The color code is as indicated.
(C) Heat map of the Pearson correlation coefficients regarding the expression levels of early PGC genes (315 genes), late germ cell genes (250 genes), and fetal oocyte genes (468 genes) between E15.5 oocytes and ZGLP1 O.E. mPGCLC-derived cells cultured as indicated.
(D) Scatter-plot comparison of the gene-expression levels between mPGCLCs at c4 from the A2 clone cultured with BMS493 and mPGCLCs at c4 from the A2 clone cultured with BMS493 and Dox. Differentially expressed genes [$\log_2$ (RPM+1) >4, $\log_2$ (fold change) >2] are colored (red: up-regulated with BMS493 and Dox, 57 genes; blue: down-regulated with BMS493 and Dox, 36 genes). The correlation coefficient is indicated above the scatter plot.
(E) Scatter-plot comparison of the gene-expression levels between E15.5 oocytes and

ZGLP1 O.E. mPGCLC-derived cells cultured with Dox and RA at c9. Differentially expressed genes [$\log_2$ (RPM+1) >4, $\log_2$ (fold change) >2] are colored (red: up-regulated in E15.5 oocytes, 71 genes; blue: down-regulated in E15.5 oocytes, 162 genes).
(F) GO term enrichment of genes that show higher (orange) or lower (blue) expression in E15.5 oocytes compared to ZGLP1 O.E. mPGCLC-derived cells at c9, as in (E).
(G, H) Expression kinetics of genes specifically up/down-regulated above the threshold level [$\log_2$ (RPM+1) >4, $\log_2$ (fold change) >2] in the ZGLP1 O.E. mPGCLC-derived cells cultured with RA at c9 compared to those without RA at c9 (up-regulated genes from Fig. 3D: 713 genes; down-regulated genes from Fig. 3F: 795 genes). Fold changes at c4, c5, c7 and c9 from c3 are shown in the $\log_2$ scale for the respective culture conditions. Boxplots represent values for the 25th, 50th, and 75th percentiles.

Fig. S10. Generation and characterization of *Zglp1* and *Stra8* knockout$^{-/-}$ mESCs.
(A) Generation of two independent *Zglp1*$^{-/-}$ ESC lines (clone E5, G10). Gray shaded boxes represent *Zglp1* exons 1 to 5. Regions encoding the conserved GATA-type zinc finger are shown in green boxes spanning between exons 4 and 5. Sequences for exon 3 for the parental clone (clone H18) and *Zglp1*$^{-/-}$ clones E5 and G10 are shown. Target sequences for the guide RNA (gRNA) pair are shown in red. PAM sequences for each gRNA are shown in blue. Deleted sequences in the knockout lines are depicted by dashes.
(B) Generation of two independent *Stra8*$^{-/-}$ ESC lines (clones B6, F6). Sequences around the target site in the exon 6 of the parental clone (clone H18) and *Stra8*$^{-/-}$ clones B6 and F6 are shown. Target sequences for the gRNA pair are shown in red. PAM sequences for each gRNA are shown in blue. Deleted sequences in the knockout lines are depicted by dashes.
(C) Western blot analysis of the expression of ZGLP1 and β-ACTIN in mPGCLC-derived cells (cultured with BMP2 and RA from c3) from the parental clone and *Zglp1*$^{-/-}$ clones E5 and G10.
(D) Western blot analysis of the expression of STRA8 and β-ACTIN in ESCs cultured with RA for 24 hrs from the parental clone and *Stra8*$^{-/-}$ clones B6 and F6.
(E) FACS analysis of BVSC expression for *Zglp1*$^{+/+}$ (clone H18) and *Zglp1*$^{-/-}$ mPGCLC-derived cells (clone G10) cultured with BMP2 and RA from c3 to c9.
(F) Cell-cycle analysis of mPGCLC-derived cells at c9 from *Zglp1*$^{+/+}$ (clone H18) and *Zglp1*$^{-/-}$ (clone G10) cultured under the control condition or with BMP2 and RA from c3 to c9. Representative FACS plots from two independent experiments are shown. The vertical axis represents EdU incorporation and the horizontal axis represents DNA content (7-AAD). Cells in the G1, S, and G2/M phases are boxed, along with the percentages of cells in the respective phases. The percentages of the cells in the G1, S, and G2/M phase in two independent experiments are shown in stacked bar charts.
(G) Expression of GFP (BV/SC) (cyan) and SYCP3 (yellow) in mPGCLC-derived cells at c9 with BMP2 and RA from the parental clone and *Zglp1*$^{-/-}$ clones E5 and G10. Bars: 10 μm.

(H) The numbers of SYCP3$^+$ cells among GFP$^+$ cells at c9 with BMP2 and RA from the parental clone and *Zglp1*$^{-/-}$ clones E5 and G10.
(I) Expression of GFP (BVSC) (cyan) and SYCP3 (yellow) in mPGCLC-derived cells at c9 with BMP2 and RA from the parental clone and *Stra8*$^{-/-}$ clones B6 and F6. Bars: 10 μm.
(J) The numbers of SYCP3$^+$ cells among GFP$^+$ cells at c9 with BMP2 and RA from the parental clone and *Stra8*$^{-/-}$ clones B6 and F6.

Fig. S11. Analysis of mPGCLC-derived cells from *Stra8*$^{-/-}$ and *Zglp1* O.E.; *Stra8*$^{-/-}$ clones.
(A) PCA of the transcriptomes of germ cells *in vivo* (E9.5-E11.5 PGCs, E12.5-E15.5 female and male germ cells) and mPGCLC-derived cells from the parental clone and *Stra8*$^{-/-}$ clone F6 cultured with BMP2 and RA from c3. Each plot represents one transcriptome. Transcriptomes from two replicates (except in the case of E9.5 PGCs, and E14.5 or E15.5 female germ cells, for which there were three replicates each) are plotted. The color code is as indicated.
(B) Generation of a *Zglp1* O.E.; *Stra8*$^{-/-}$ ESC clone. Sequences around the target site in the exon 6 of the parental clone (*Zglp1* O.E clone A2) and *Zglp1 O.E.; Stra8*$^{-/-}$ clone A11 are shown. Target sequences for the gRNA pair are shown in red. PAM sequences for each gRNA are shown in blue. Deleted sequences in the knockout lines are depicted by dashes.
(C) Western blot analysis of the expression of STRA8 and β-ACTIN in ESCs cultured with RA for 24 hrs from the parental clone (*Zglp1* O.E clone A2) and *Zglp1 O.E.; Stra8*$^{-/-}$ clone A11.
(D) Immunofluorescence analysis of the expression of GFP (BV/SC) (cyan) and SYCP3 (yellow) in mPGCLC-derived cells at c9 from the parental clone (*Zglp1* O.E clone A2) and *Zglp1 O.E.; Stra8*$^{-/-}$ clone A11 cultured with Dox and RA from c3. Bars: 10 μm.
(E) Scatter-plot comparison of gene expression levels between c9 *Zglp1* O.E; *Stra8*$^{-/-}$. mPGCLC-derived cells cultured with Dox and RA at c9 and *Stra8*$^{-/-}$ mPGCLC-derived cells with BMP2 and RA at c9. Differentially expressed genes [$\log_2$ (RPM+1) >4, $\log_2$ (fold change) >2] are colored (red; 42 genes, blue; 79 genes).

Fig. S12. Genome-wide binding profiles of ZGLP1.
(A) Expression of Stella-ECFP (SC) (cyan), DDX4 (magenta) and V5-ZGLP1 (green) in mPGCLC-derived cells (*V5-Zglp1* O.E. clone A12) at c5. Bar: 10 μm.
(B) Western blot analysis of V5-ZGLP1 and Lamin in mPGCLC-derived cells at c7 from the parental clone and *V5-Zglp1* O.E. clone A12. mPGCLCs were treated with Dox and RA from c3 to c7, and SC$^+$ cells were sorted by FACS at c7. V5-ZGLP1 was detected by anti-V5 antibody.
(C) The numbers of SYCP3$^+$ cells among SC$^+$ cells at c9 (cultured with Dox and RA from c3 to c9) in *V5-Zglp1* O.E. clone A12.

(D) Heat map representation of ZGLP1-binding levels and the corresponding H3K4me3 and H3K27me3 levels around ± 5kb of the top 50% ZGLP1-binding sites. The data for the enrichment levels of H3K4me3 and H3K27me3 in c7 PGCLCs (control culture) are from a published study {Ohta, 2017 #3061}. RPM: reads per million.
(E) Averaged intensity plots of ZGLP1, H3K4me3 and H3K27me3 near genes [± 3 kb from transcription start sites (TSS) and transcription termination sites (TTS)].
(F) Scatter-plot showing the IP levels for H3K4me3 and H3K27me3 around the TSSs of all mouse genes (25315 genes). Genes are classified and colored according to the associated H3K4me3 and H3K27me3 IP levels as: K27 enriched, K4 enriched, Bivalent, and neither K27 enriched nor K4 enriched (see **Materials and Methods**). Genes that are up-regulated by ZGLP1 are indicated.
(G) Cumulative distribution of the distances of the ZGLP1 peaks to the nearest TSSs in V5-ZGLP1 O.E. PGCLCs cultured with Dox or with Dox and RA. See **Materials and Methods** for peak assignment to the nearest gene.
(H) Genomic annotation of the peaks within 15 kb (+ Dox: 14869 peaks, + Dox + RA: 13379 peaks) from the nearest TSSs as defined in (G). See **Materials and Methods** for peak annotations.
(I) The numbers of ZGLP1-bound genes among up/down-regulated [$\log_2$ (RPM+1) >4, $\log_2$ (fold change) >2 from c3, either at c4, c5, c7, and c9], expressed [$\log_2$ (RPM+1) >4, either at c4, c5, c7, and c9] and all mouse genes.
(J) Venn diagram showing the overlap between ZGLP1-bound genes in PGCLCs cultured with Dox or with Dox and RA.
(K) Stacked bar charts showing the percentages of up-regulated genes [$\log_2$ (fold change) > 2 from c3, up-regulated at c4, c5, c7, and c9] in PGCLCs cultured with Dox and RA from c3. Genes are classified based on ZGLP1 binding and the enrichment of two histone modifications (H3K4me3 and H3K27me3). The numbers of genes for each category are shown in parentheses. The color coding is as indicated.
(L) A scatter plot showing the IP levels for H3K4me3 and H3K27me3 around the TSSs. The dotted lines represent the threshold IP levels for each histone modification (see **Materials and Methods**). Genes bound by ZGLP1 and up-regulated from c3 to c4 or from c3 to c5 by Dox and RA treatment are colored as indicated (333 genes). Representative ZGLP1 up-regulated genes are indicated.
(M) Expression dynamics of ZGLP1-bound/unbound genes. The expression levels [$\log_2$ (RPM+1)] for each gene are normalized by its peak expression levels between c3 to c9, and then the median values are plotted. **P*-value < 0.01. ***P*-value < 0.001.
(N) GO term enrichment among bivalent and H3K27me3-enriched genes bound by ZGLP1 and up-regulated from c3 to c4 or from c3 to c5 upon culture with Dox and RA.

Supplementary Tables

Table S1. Mapping statistics of RNA-seq data generated/used in this study.

Table S2. RNA-seq data for the identification of BMP target genes.

Table S3. RNA-seq data for the analysis of ZGLP1 function.

Table S4. Mapping statistics of ChIP-seq data generated/used in this study.

Table S5. ChIP-seq data for V5-ZGLP1.

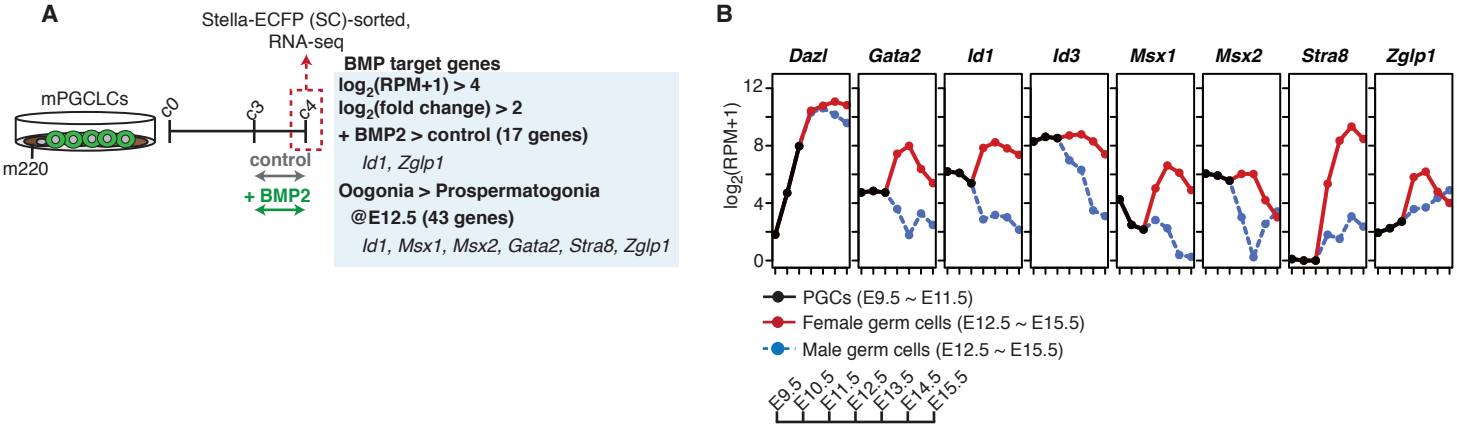
Table S6. Re-analysis of ChIP-seq data for histone modifications from previous studies.

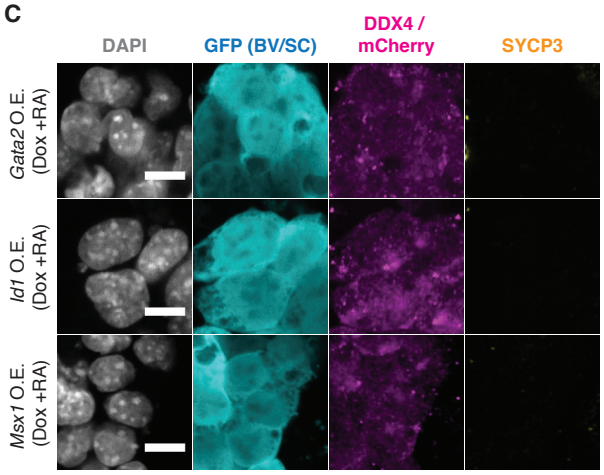
Table S7. Primers used in this study.

Table S8. Antibodies used in this study.

References

Supplementary Figure 1, Nagaoka et al.





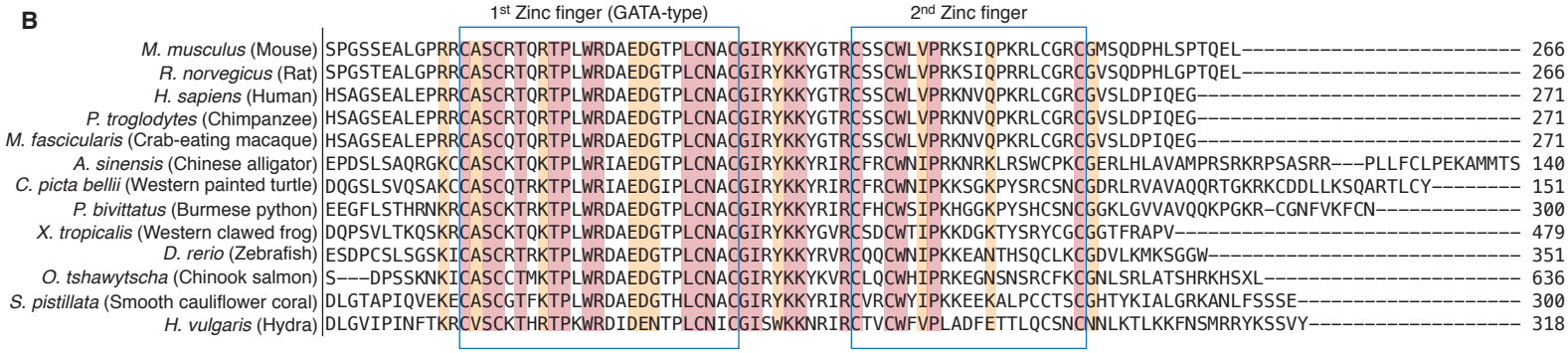
D

		SYCP3+ / GFP+
Gata2 OE (Dox +RA)	clone C5	0 / 139
	clone C8	0 / 94
	clone C11	0 / 97
Id1 OE (Dox +RA)	clone F2	0 / 91
	clone F4	0 / 64
	clone F8	0 / 65
Msx1 OE (Dox +RA)	clone A4	0 / 94
	clone A9	0 / 102
	clone A11	0 / 135

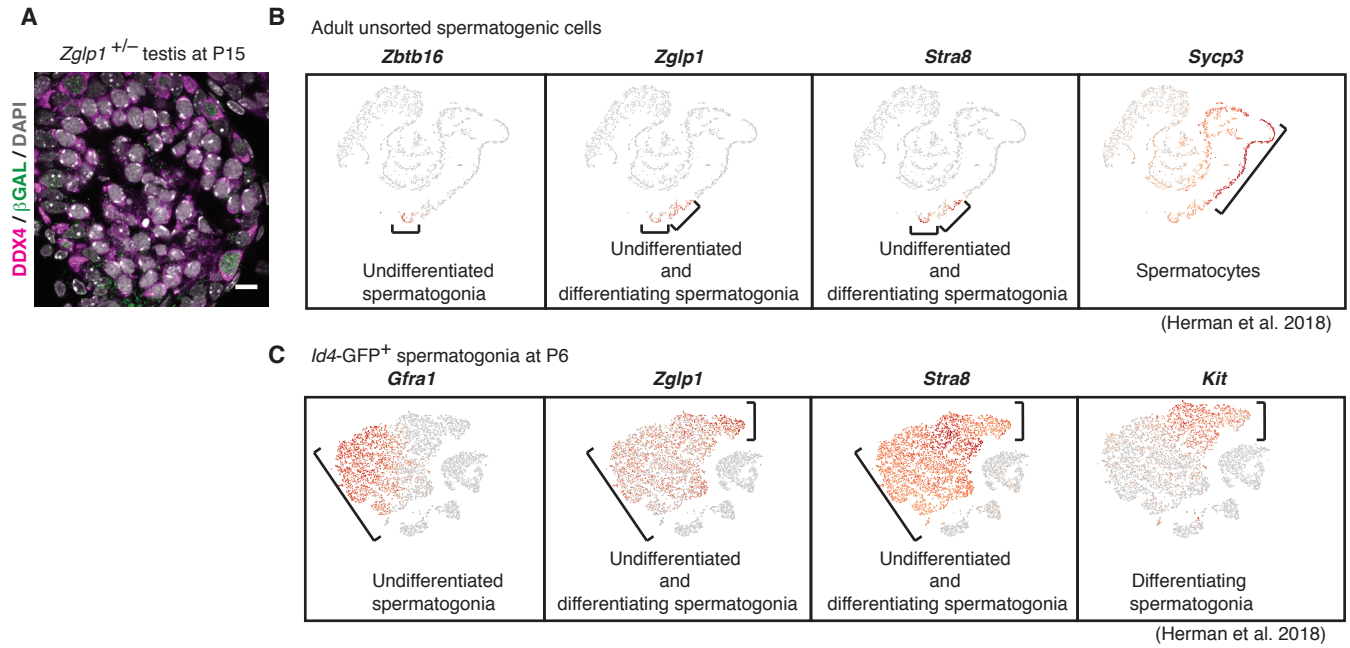
Supplementary Figure 2, Nagaoka et al.

A

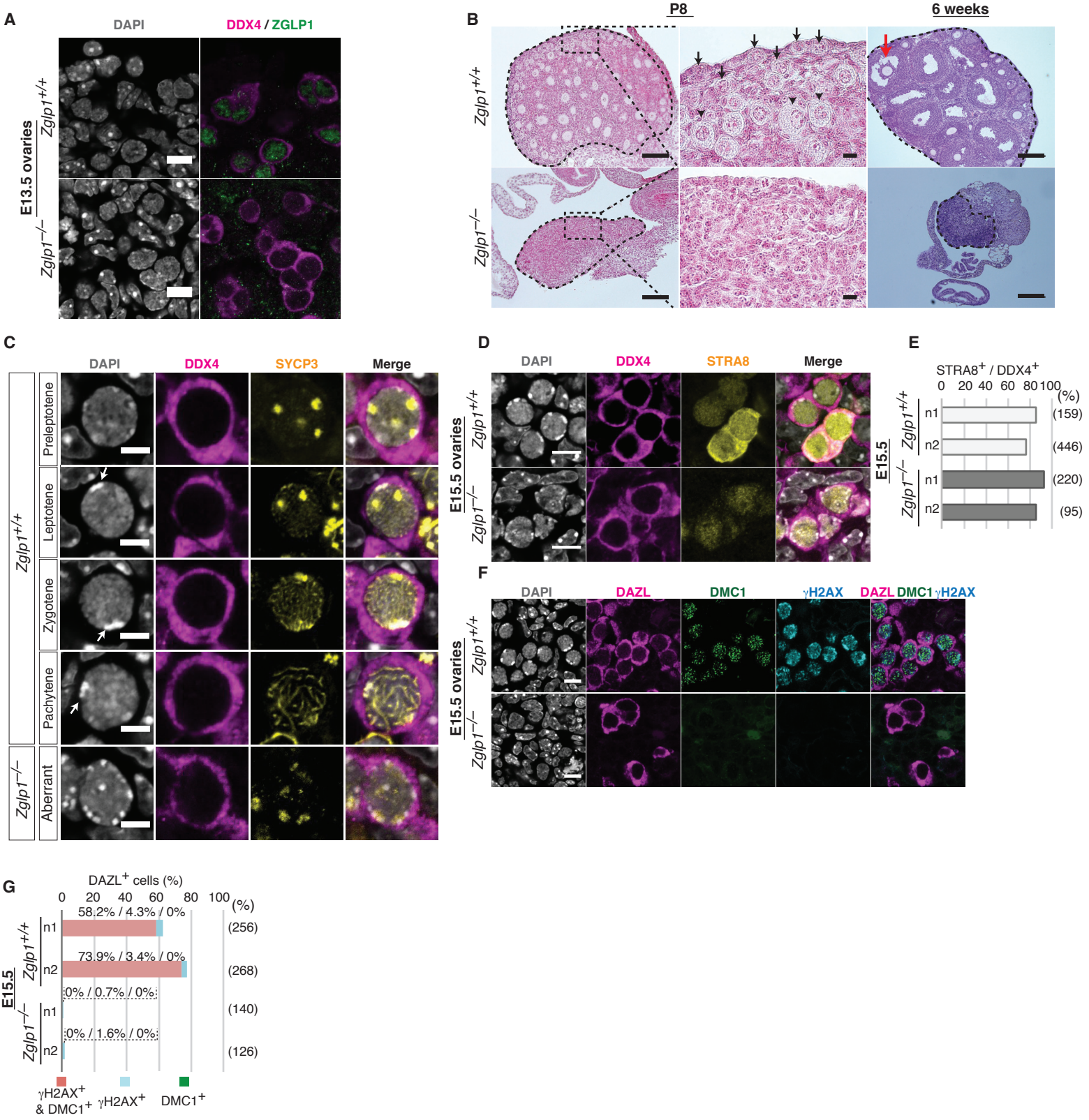
B



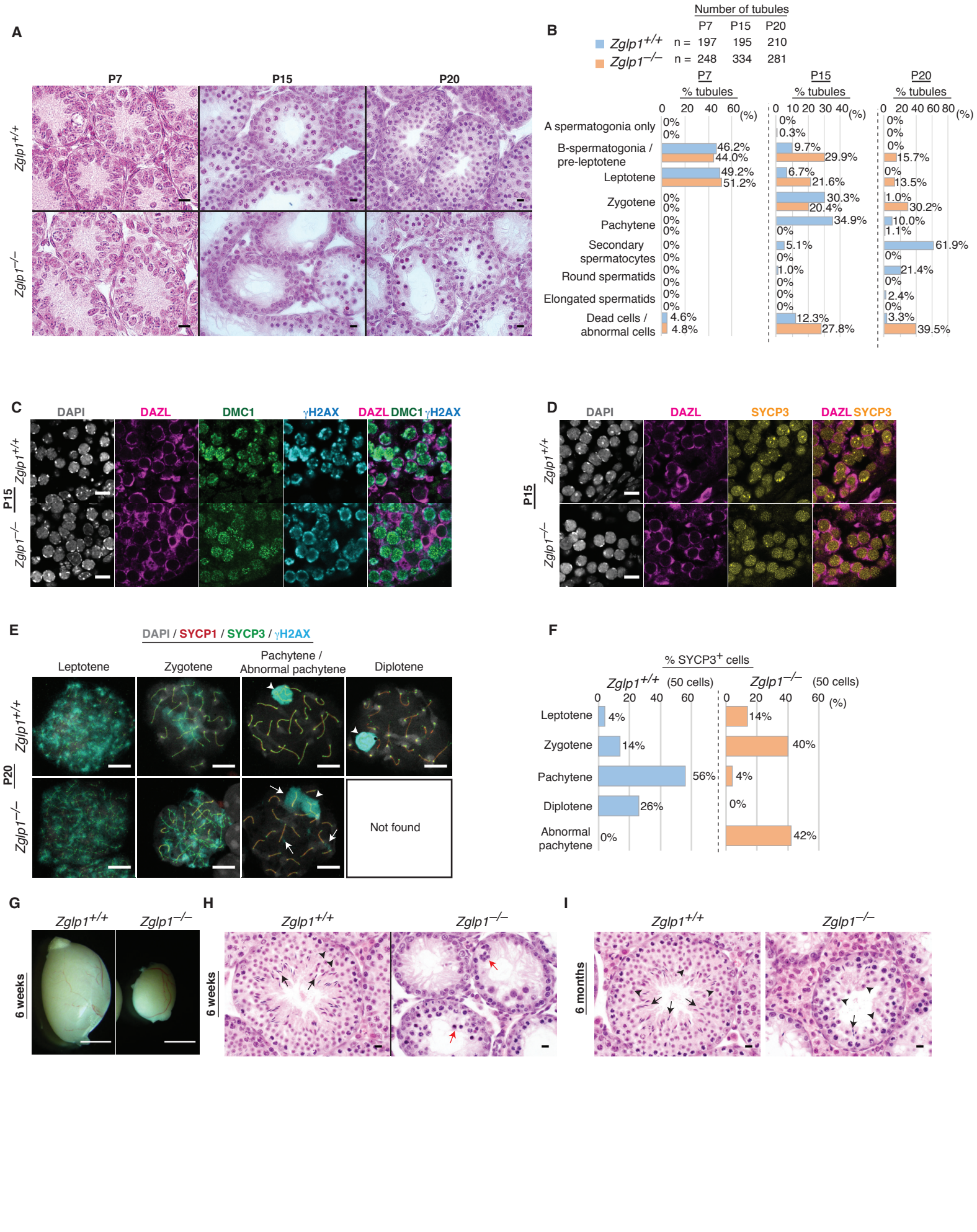
Supplementary Figure 3, Nagaoka et al.



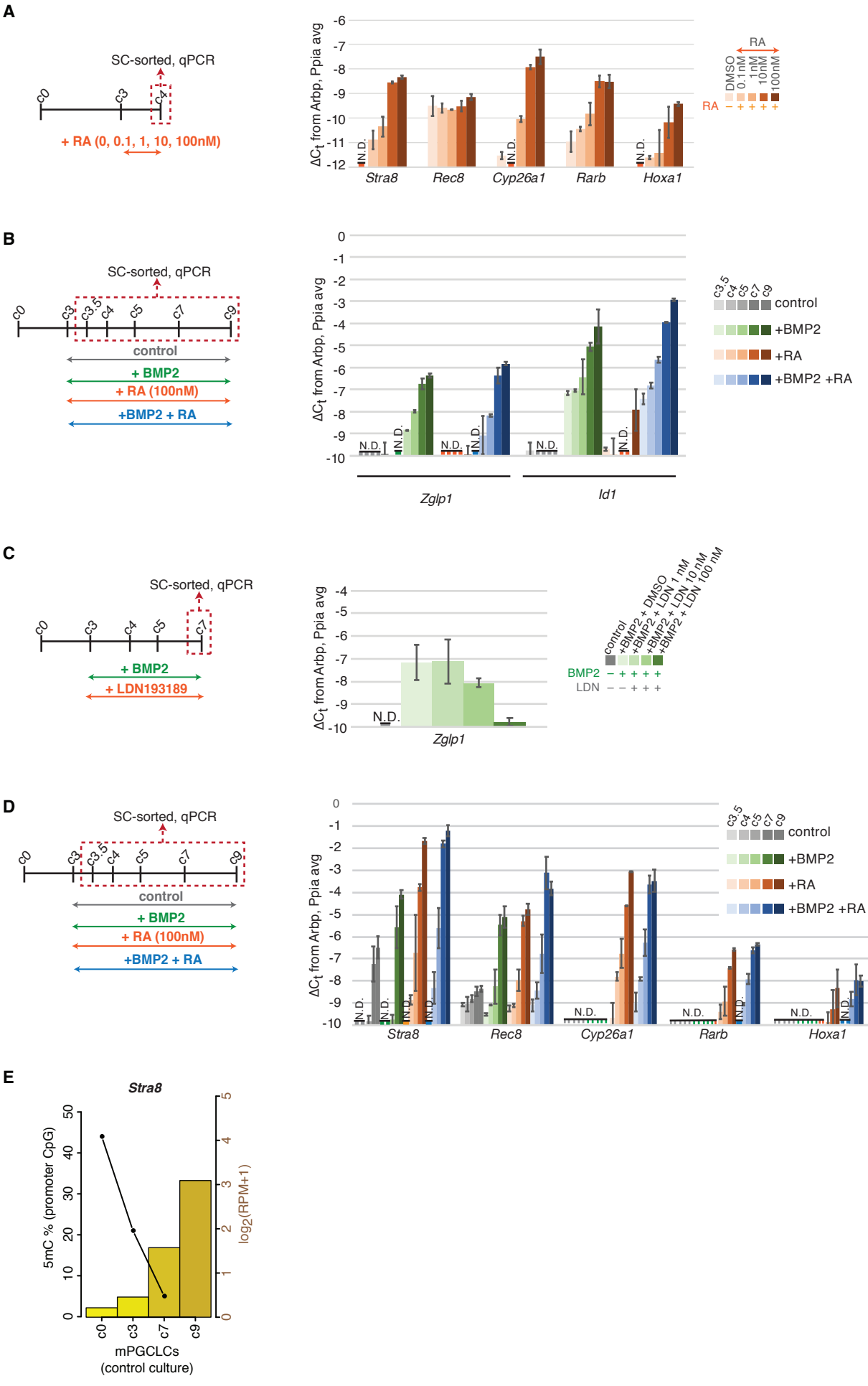
Supplementary Figure 4, Nagaoka et al.



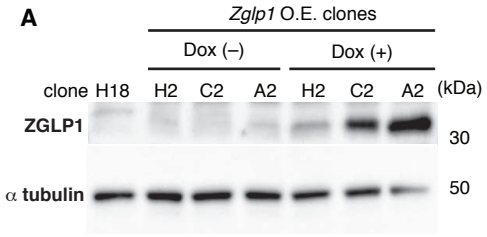
Supplementary Figure 5, Nagaoka et al.



Supplementary Figure 6, Nagaoka et al.

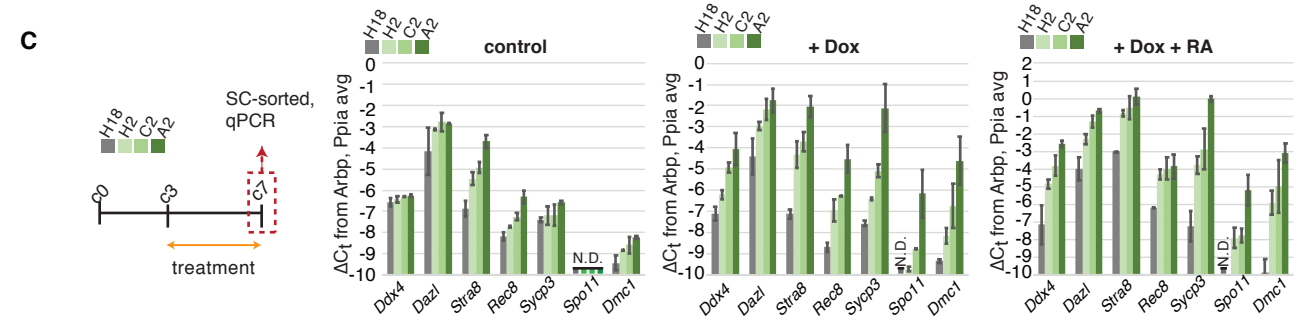


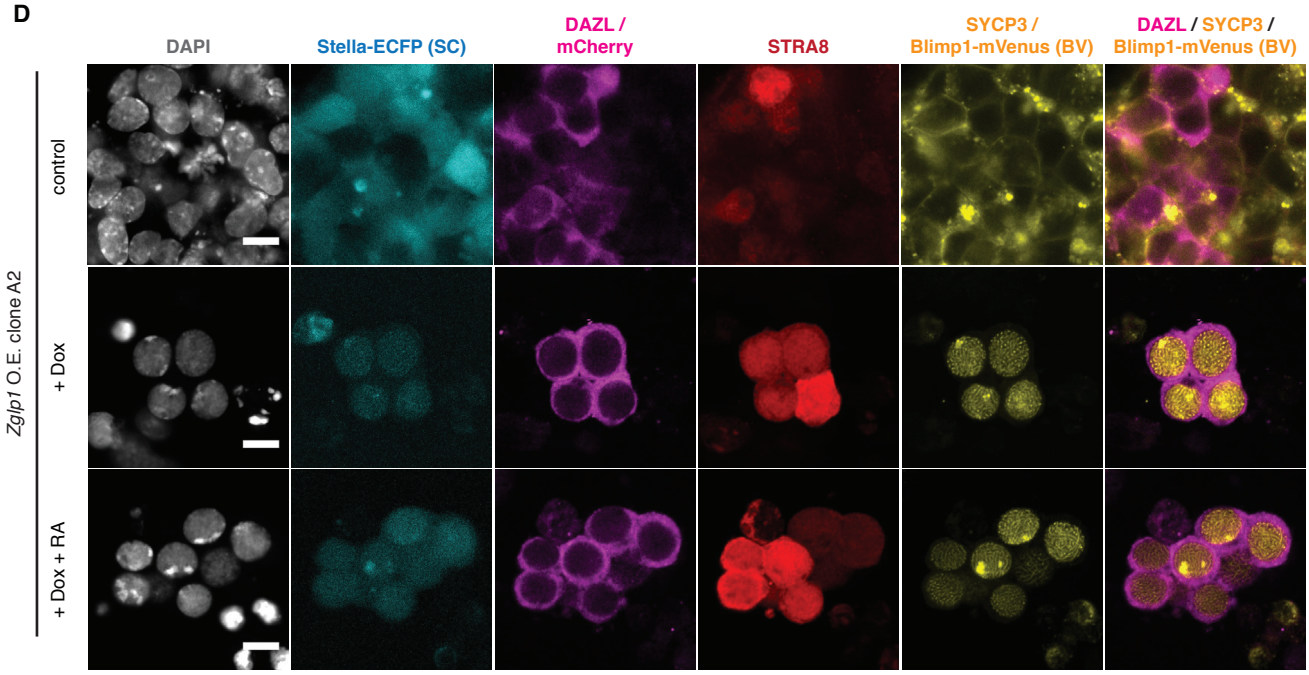
Supplementary Figure 7, Nagaoka et al.

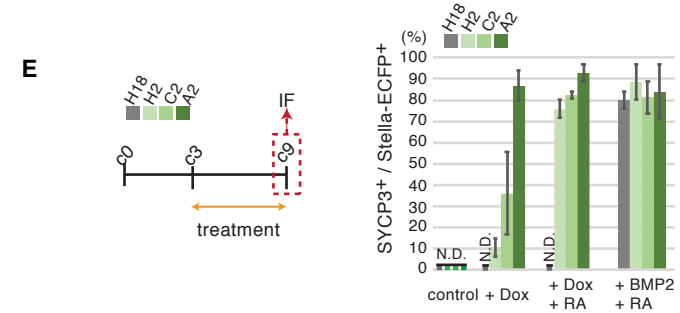


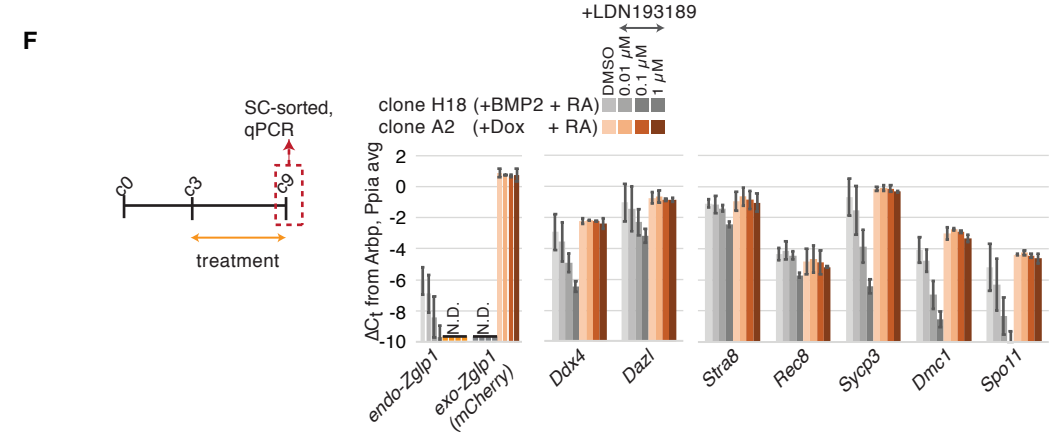
B

		Number of chromosomes (ESCs)		
		39	40	41
	clone H18	1	24	0
Zglp1 O.E.	clone H2	23	1	1
	clone C2	24	0	0
	clone A2	27	0	0

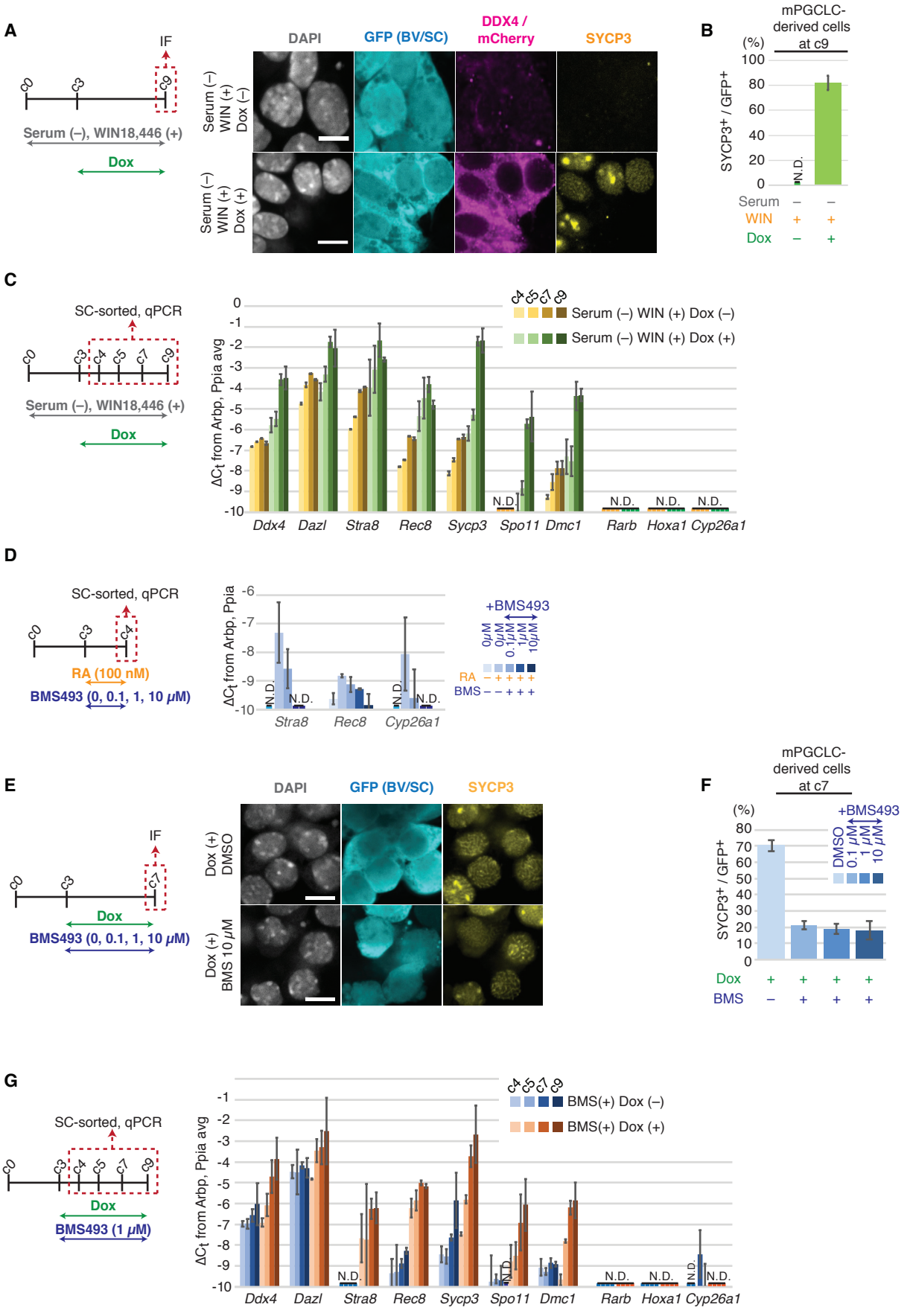




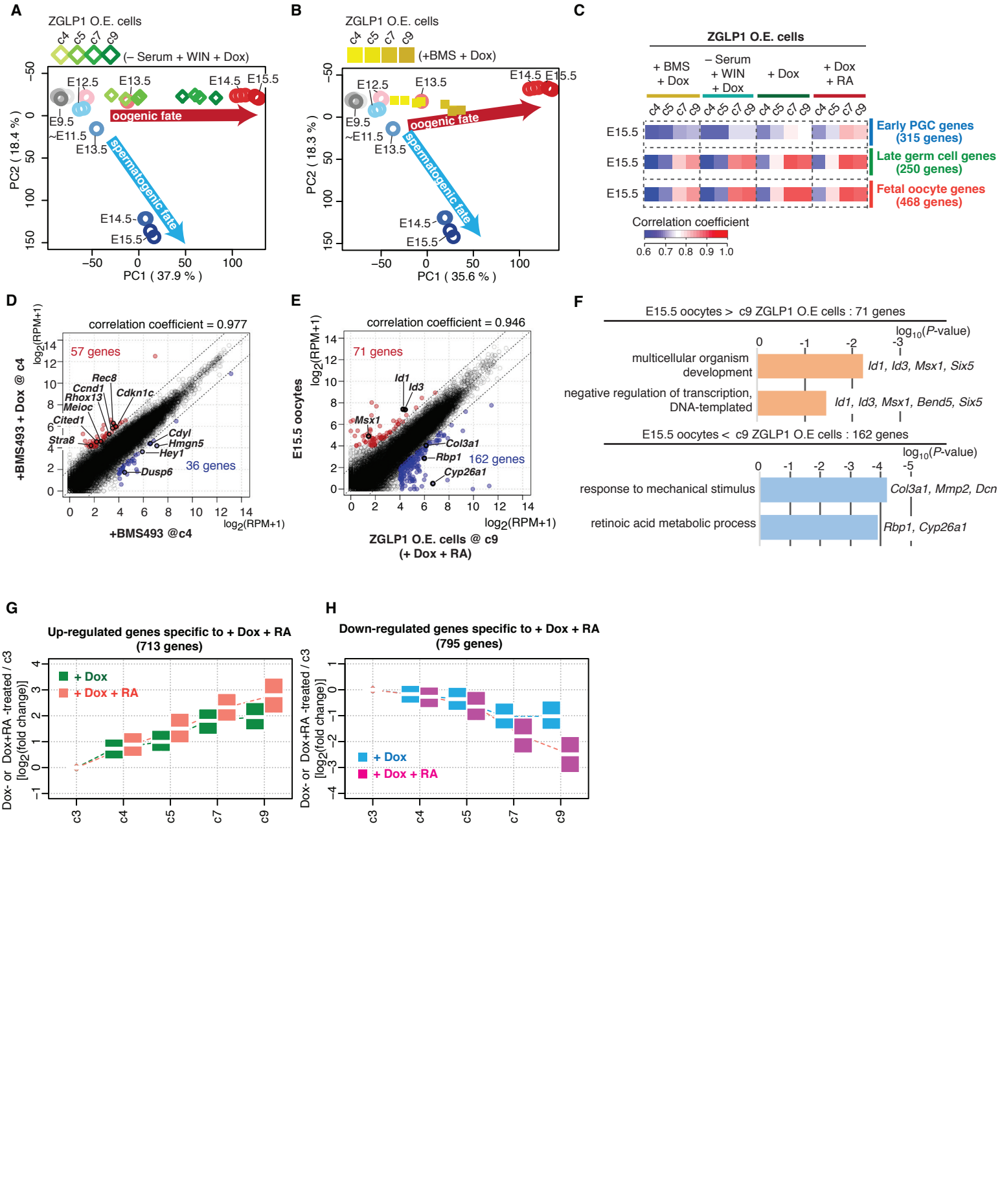




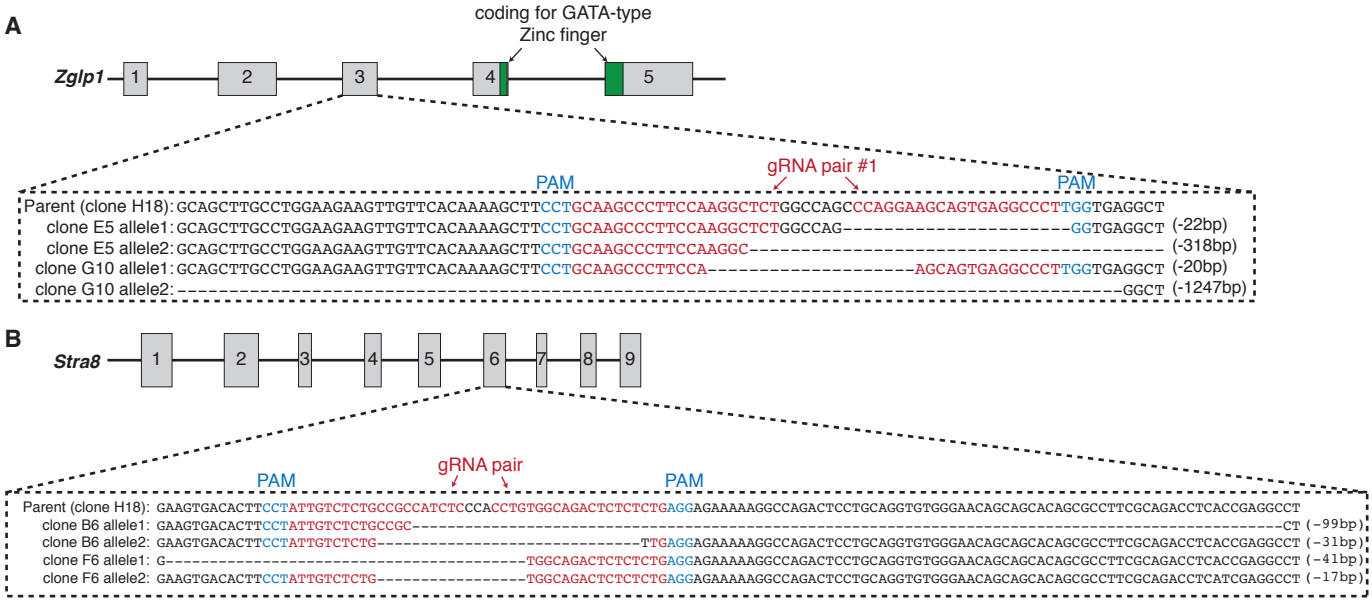
Supplementary Figure 8, Nagaoka et al.

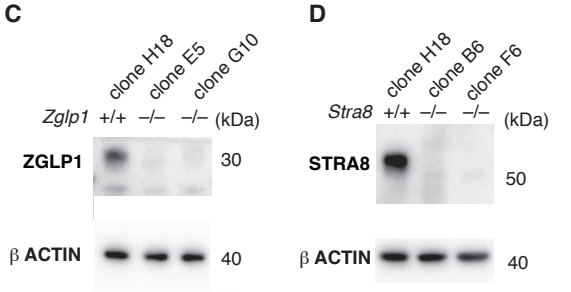


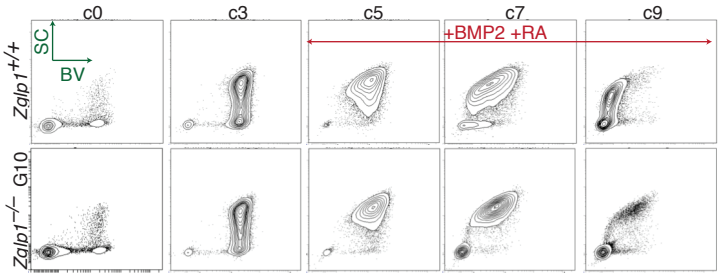
Supplementary Figure 9, Nagaoka et al.

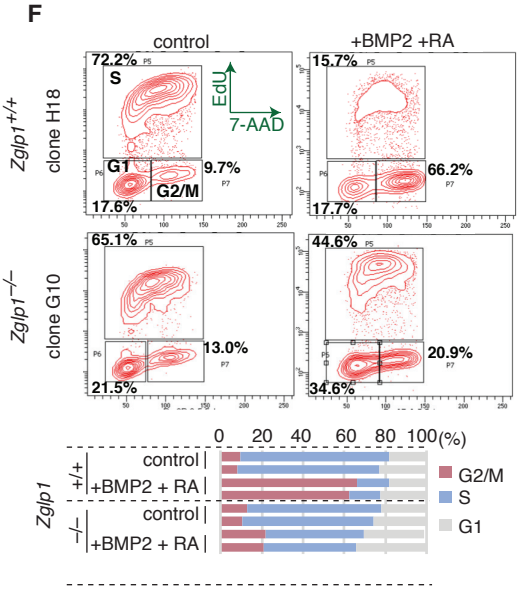


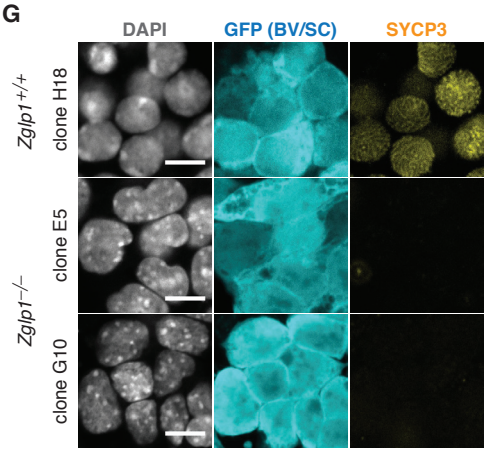
Supplementary Figure 10, Nagaoka et al.





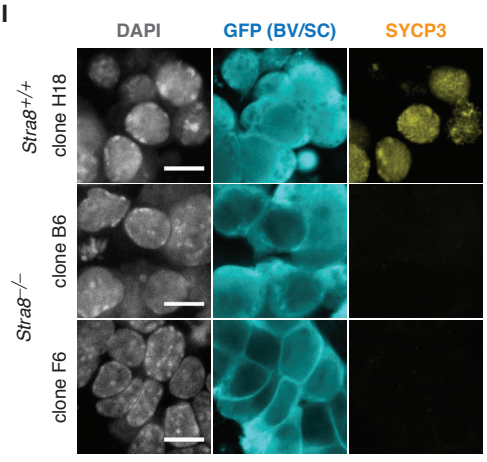






H

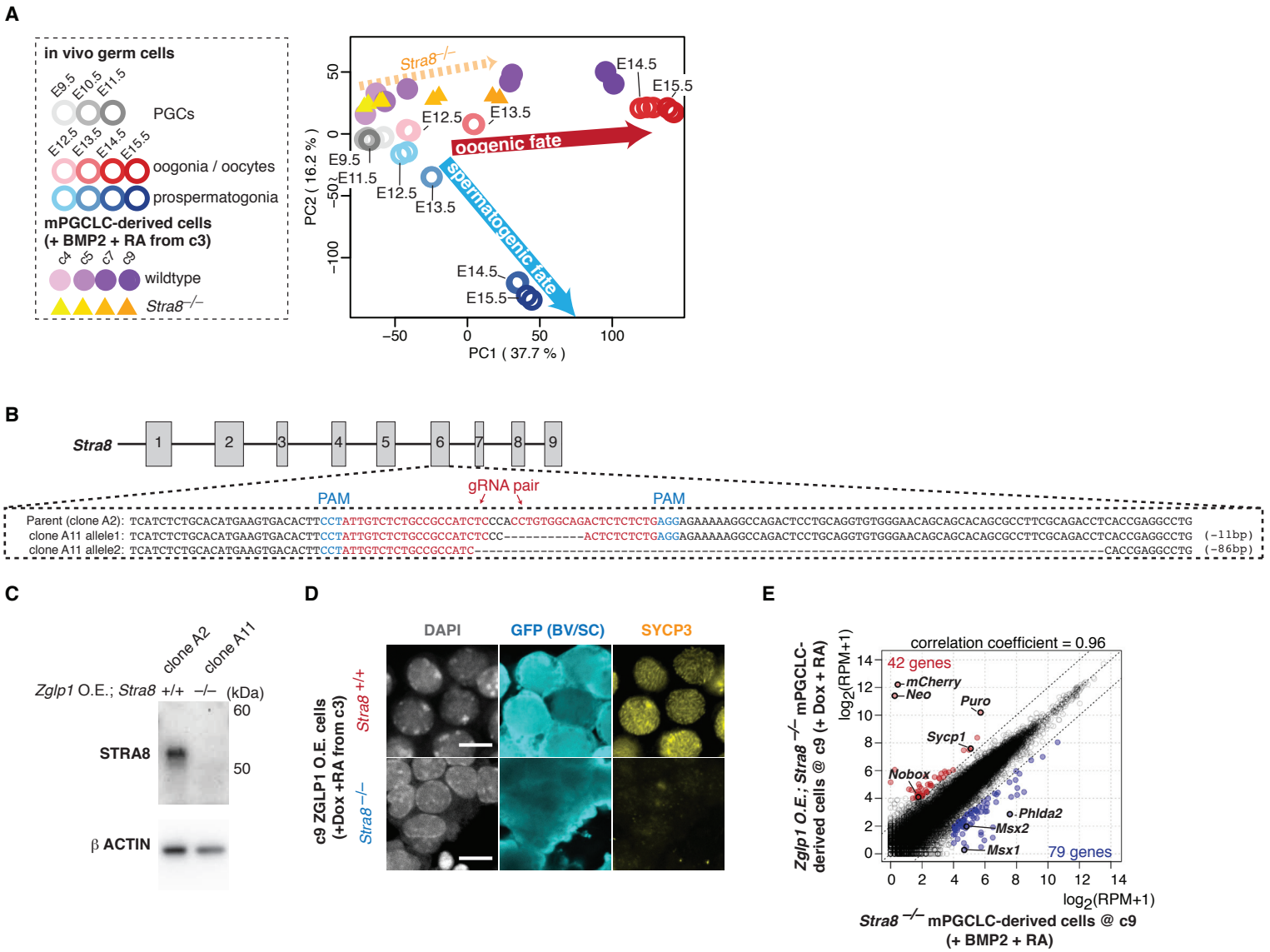
	SYCP3+	GFP+ (BV/SC)
Zglp1$^{+/+}$	205	214
Zglp1$^{-/-}$ E5	0	223
Zglp1$^{-/-}$ G10	0	172



J

	SYCP3+	GFP+ (BV/SC)
Stra8$^{+/+}$	114	135
Stra8$^{-/-}$ B6	0	267
Stra8$^{-/-}$ F6	0	221

Supplementary Figure 11, Nagaoka et al.



Supplementary Figure 12, Nagaoka et al.