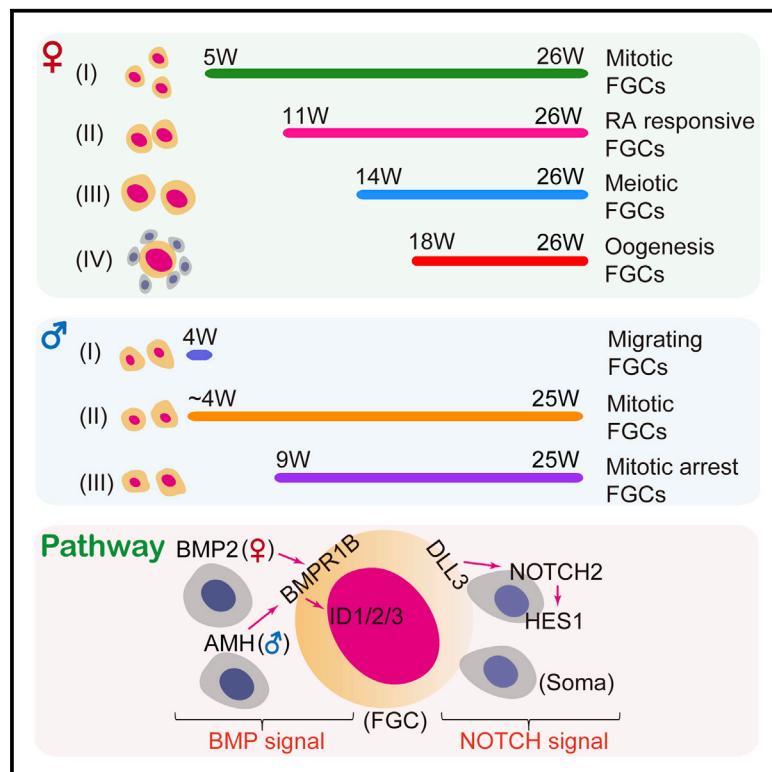


# Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions

## Graphical Abstract



## Authors

Li Li, Ji Dong, Liying Yan, ..., Lu Wen, Fuchou Tang, Jie Qiao

## Correspondence

tangfuchou@pku.edu.cn (F.T.), jie.qiao@263.net (J.Q.)

## In Brief

Li et al. interrogate the transcriptomes of over 2,000 fetal germ cells (FGCs) and their gonadal niche cells from male and female human embryos using single-cell RNA-seq analysis. They provide insights into the developmental trajectories and heterogeneity in FGCs over a wide range of developmental stages.

## Highlights

- RNA sequencing of 2,167 single FGCs identifies multiple stages in individual embryos
- Transcriptome landscapes decode TF networks underlying distinct developmental stages
- Reciprocal BMP and Notch signaling occurs between FGCs and their niche cells
- Female embryos contain four stages of FGCs, even in later developmental stages

# Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions

Li Li,<sup>1,2,6</sup> Ji Dong,<sup>1,2,6</sup> Liying Yan,<sup>1,3,6</sup> Jun Yong,<sup>1,3</sup> Xixi Liu,<sup>1,3</sup> Yuqiong Hu,<sup>1,2,5</sup> Xiaoying Fan,<sup>1,2</sup> Xinglong Wu,<sup>1,2</sup> Hongshan Guo,<sup>1,2</sup> Xiaoye Wang,<sup>1,3</sup> Xiaohui Zhu,<sup>1,3</sup> Rong Li,<sup>1,3</sup> Jie Yan,<sup>1,3</sup> Yuan Wei,<sup>1,3</sup> Yangyu Zhao,<sup>1,3</sup> Wei Wang,<sup>1,3</sup> Yixin Ren,<sup>1,3</sup> Peng Yuan,<sup>1,3</sup> Zhiqiang Yan,<sup>1,3</sup> Boqiang Hu,<sup>1,2</sup> Fan Guo,<sup>1,2</sup> Lu Wen,<sup>1,2</sup> Fuchou Tang,<sup>1,2,5,7,\*</sup> and Jie Qiao<sup>1,3,4,5,\*</sup>

<sup>1</sup>Beijing Advanced Innovation Center for Genomics (ICG), College of Life Sciences, Department of Obstetrics and Gynecology, Third Hospital, Peking University, Beijing 100871, China

<sup>2</sup>Biomedical Institute for Pioneering Investigation via Convergence and Center for Reproductive Medicine, Ministry of Education Key Laboratory of Cell Proliferation and Differentiation, Beijing 100871, China

<sup>3</sup>Key Laboratory of Assisted Reproduction, Ministry of Education, Beijing 100191, China

<sup>4</sup>Beijing Key Laboratory of Reproductive Endocrinology and Assisted Reproductive Technology, Beijing 100191, China

<sup>5</sup>Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China

<sup>6</sup>These authors contributed equally

<sup>7</sup>Lead Contact

\*Correspondence: tangfuchou@pku.edu.cn (F.T.), jie.qiao@263.net (J.Q.)

<http://dx.doi.org/10.1016/j.stem.2017.03.007>

## SUMMARY

Human fetal germ cells (FGCs) are precursors to sperm and eggs and are crucial for maintenance of the species. However, the developmental trajectories and heterogeneity of human FGCs remain largely unknown. Here we performed single-cell RNA-seq analysis of over 2,000 FGCs and their gonadal niche cells in female and male human embryos spanning several developmental stages. We found that female FGCs undergo four distinct sequential phases characterized by mitosis, retinoic acid signaling, meiotic prophase, and oogenesis. Male FGCs develop through stages of migration, mitosis, and cell-cycle arrest. Individual embryos of both sexes simultaneously contain several subpopulations, highlighting the asynchronous and heterogeneous nature of FGC development. Moreover, we observed reciprocal signaling interactions between FGCs and their gonadal niche cells, including activation of the bone morphogenic protein (BMP) and Notch signaling pathways. Our work provides key insights into the crucial features of human FGCs during their highly ordered mitotic, meiotic, and gametogenic processes *in vivo*.

## INTRODUCTION

Fetal germ cells (FGCs) are the embryonic precursors of mature gametes—the sperm and oocytes in mammals (Saitou and Yamaji, 2012). Therefore, the proper development of FGCs is crucial for faithfully transmitting genetic information to the next generation through the germline (Hirasawa et al., 2007; Saitou and Miyauchi, 2016; Seisenberger et al., 2013; Smallwood and

Kelsey, 2012). Recently, we and other groups have shown that DNA methylation is comprehensively removed from the genome in human FGCs, and within 10–11 weeks, the residual methylation in the germline genome becomes minimal, with only 5%–7% remaining (median level) (Gkountela et al., 2015; Guo et al., 2015; Tang et al., 2015; von Meyenn and Reik, 2015). Interestingly, during this global epigenomic reprogramming, the transcriptome of FGCs remains relatively stable from 4–11 weeks post-fertilization, with only a minor proportion of genes showing significant expression changes. We also showed that, in a 17-week female embryo, the majority of FGCs enter meiosis and show strong gene expression heterogeneity among individual FGCs (Guo et al., 2015). However, the transcriptome of individual FGCs in these human embryos has not been systematically analyzed, and how many subpopulations and developmental phases exist for human germ cell development remains largely unknown.

Human FGCs have a unique gene expression network that is very different from that of mice. For example, mouse early FGCs express the pluripotency master gene Sox2, whereas human FGCs do not express it but, instead, express SOX15 and SOX17 (Guo et al., 2015; Irie et al., 2015; Perrett et al., 2008). Another difference is that human female FGCs express the X chromosome inactivation master regulator *XIST* even after the inactivated X chromosome is reactivated. In fact, human male FGCs also express *XIST* at a significant level, which never occurs in mouse FGCs (Gkountela et al., 2015). Thus, the transcriptome of human FGCs needs to be analyzed at single-cell resolution for a wide range of developmental stages. This analysis will offer a solid reference dataset and sound basis for the *in vitro* differentiation of human pluripotent stem cells into FGC-like cells and provide new clues for further improving their differentiation efficiency *in vitro* (Chen and Clark, 2015; Irie et al., 2015; Nakaki et al., 2013; Sasaki et al., 2015; Zhou et al., 2016). Usually, the germ cells in human embryos before 11 weeks post-fertilization are called primordial germ cells (PGCs). The germ cells in female embryos after 11 weeks post-fertilization are called oogonia,

whereas those in male embryos are called gonocytes or prespermatogonia (De Felici and Barrios, 2013; Del-Mazo et al., 2013). Here, to make it easier to follow, we call all of these germ cells in human embryos FGCs. The expression pattern of our single-cell RNA sequencing (RNA-seq) data is accessible at <http://www.singlecell.pku.edu.cn/GermCell>.

## RESULTS

### The Gene Expression Landscapes of Human FGCs and Their Niche Cells in the Gonads

With ethics approval, we obtained 29 human embryos ranging from 4–26 weeks post-fertilization (17 female and 12 male embryos; see Figure 1A for details) to systematically investigate human FGC development and the regulatory relationships between FGCs and their neighboring niche cells (Methods S1). We used magnetic-activated cell sorting (MACS) and fluorescence-activated cell sorting (FACS) to enrich for c-KIT<sup>+</sup> FGCs and isolate larger c-KIT<sup>-</sup> cells because some FGCs are c-KIT-negative and are usually larger than gonadal somatic cells at the later developmental stages (Gkountela et al., 2013; Robinson et al., 2001). In addition, to elucidate the interaction between FGCs and their microenvironment cells, we also obtained c-KIT<sup>-</sup> cells of various sizes. For the 4-week and 5-week human embryos, we isolated c-KIT<sup>+</sup> FGCs from the aorta-gonad-mesonephros (AGM) region by MACS. We performed single-cell RNA-seq of 2,579 individual cells. After filtration, we retained 2,167 individual cells, including 1,490 FGCs and 677 somatic cells, for the subsequent analysis (Figure 1A). On average, we obtained approximately 1.9 million reads for each individual cell. The average number of genes expressed (transcripts per million [TPM] > 0) in each FGC cell and each gonadal somatic cell were 7,510 and 5,071, respectively. The average copy number of transcripts detected in each individual cell was 570,745 for FGC cells and 379,324 for gonadal somatic cells (Figure S1A; Table S1). We conducted a down-sampling analysis to verify that the sequencing depth was enough for the following analyses. Clearly, even when down-sampling of 30% of the reads, we still, on average, detected 90% of the expressed genes compared with using all of the reads, indicating that our sequencing depth was enough for a subsequent reliable analysis (Figure S1A).

We performed a t-distributed stochastic neighbor embedding (t-SNE) analysis and clearly identified 17 clusters (Figure 1B). Based on the expression patterns of the germline-specific genes, clusters 1–7 are classified as FGCs (Saitou and Miyauchi, 2016). Among them, clusters 1–4 were female FGCs, whereas clusters 5–7 were male FGCs.

First, we analyzed the gene expression dynamics of the female FGCs (clusters 1–4). According to the expression pattern of pluripotency and early FGC-specific markers such as *POU5F1* (also known as *OCT4*) and *NANOG*, cluster 1 cells were mitotic FGCs (Figure 1C). We then determined that clusters 2–4 were later-stage FGCs with much higher expression levels of *DAZL* and *DDX4* (also known as *VASA*) than cluster 1 (Figure 1C; Castrillon et al., 2000; Kee et al., 2009). Notably, cluster 1 female FGCs at the mitotic stage were isolated from embryos of a wide range of developmental stages (from 5 weeks to 26 weeks). They were relatively scattered on the t-SNE plot, suggesting that, even in the late gestation stage, fractions of the female FGCs remain in

the mitotic stage, expressing early FGC marker genes, and may retain the potential to replenish the germ cell population. In addition, in accordance with our previous work, clear cell-to-cell heterogeneity was observed in this cluster. The cells in cluster 2 were retinoid acid (RA) signaling-responsive FGCs and specifically expressed the key target of the RA signaling pathway, *STRA8* (Figure 1C; Soh et al., 2015). The cells in cluster 3 were meiotic prophase FGCs and clearly expressed synaptonemal complex proteins, such as *SYCP1* and *SYCP3*, and meiotic recombination proteins, such as *SPO11* and *PRDM9*, which play a central role in the transcriptional activation of genes during meiotic prophase (Figures 1C and 2; Figure S2; Soh et al., 2015). Finally, the cells in cluster 4 were FGCs at the oogenesis stage and specifically expressed the zona pellucida sperm-binding protein *ZP3* and the oocyte-secreted protein *OOSP2* (Figures 1C and 2; Figure S2; Sánchez and Smitz, 2012).

In general, starting at 11 weeks, the gonads from an individual female embryo contain mitotic phase and RA signaling responsive phase FGCs. From 14 weeks, meiotic prophase FGCs appeared in the embryo ovaries. From 18 weeks, the gonads from an individual female embryo contain FGCs of all four stages we identified (mitotic, RA signaling-responsive phase, meiotic prophase, and oogenesis phase), which is in agreement with previous studies and our unsupervised hierarchical clustering analysis (Figure 1B; Anderson et al., 2007; Heeren et al., 2016; Rajpert De Meyts et al., 2004).

Next, we analyzed the transcriptional characteristics of the male FGCs (clusters 5–7). The cells in clusters 5 and 6 were early male FGCs because they specifically expressed *POU5F1* and *NANOG* (Figure 1C). The cells in cluster 7 were late FGCs because they specifically expressed *DAZL* and *DDX4* (Figure 1C). 4-week male FGCs were mostly in cluster 5, whereas both clusters 6 and 7 included FGCs from 9–25 weeks. Based on the morphological information and their expression profiles, we classified the cells in cluster 5 as migration phase FGCs, the cells in cluster 6 as gonadal mitotic phase FGCs, and the cells in cluster 7 as mitotic arrest phase FGCs. Male and female FGCs at the mitotic stage (clusters 1 and 6) clearly share similar expression patterns and are mixed together in the t-SNE plot as one cluster because, at this stage, no sexual differentiation appeared for FGCs yet (Figure 1B). In contrast to female FGCs of similar developmental stages, late male FGCs in cluster 7 underwent mitotic arrest instead of entering meiosis, which hints at the distinct regulatory mechanism of male and female FGCs for precise cell fate determination. Collectively, these findings reveal the unique gene expression programs of female and male FGCs during human embryonic development.

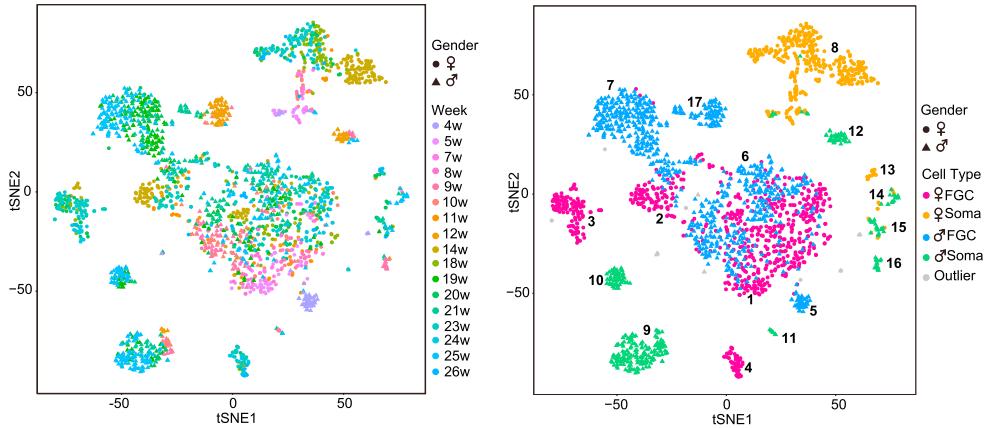
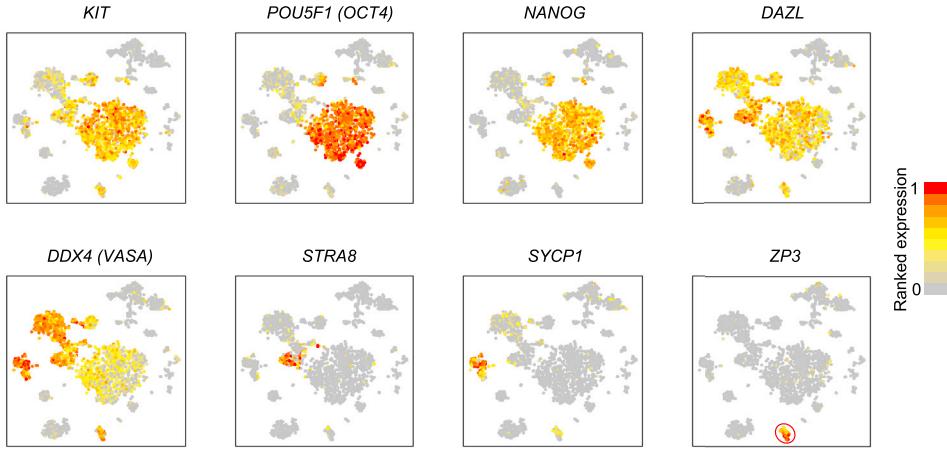
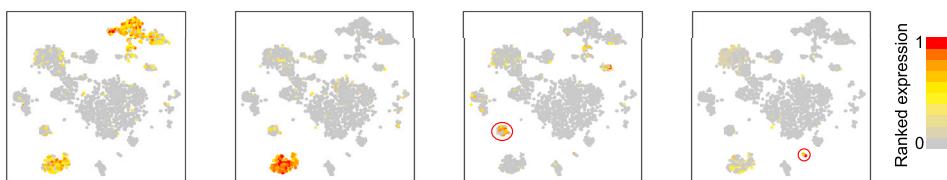
Furthermore, we also analyzed the niche cells of FGCs—the somatic cells in the gonads (clusters 8–16). In general, cluster 13 and most of cluster 8 were gonadal somatic cells from female embryos, whereas clusters 9–15 were gonadal somatic cells from male embryos. The majority of the cells in cluster 8 were granulosa cells and expressed the granulosa cell-specific master genes *WT1* and *FOXL2* (Figure 1D; Figure S1C). A small fraction of the cells in cluster 8 were from 4-week male and 5-week female embryos, and they expressed *PECAM1* (also known as *CD31*) and *CDH5* (also known as *VE-cadherin*, *CD144*); thus, they are probably endothelial cells in the gonads (Figure S1C).

**A**

Week	5w	7w	8w	10w	11w	12w	14w	18w		20w		23w		24w		26w	Total	
Embryo	twin-e1	twin-e2	e1	e1	e1	e1	e1	twin-e1	twin-e2	twin-e1	twin-e2	twin-e1	twin-e2	twin-e1	twin-e2	e1		
No. of ♀ FGC	34	15	34	41	50	30	48	75	31	34	47	59	64	52	46	52	20	732
No. of ♀ Soma	13	17	12	3	18	11	2	94	24	44	29	18	19	27	10	10	16	367

Week	4w	9w	10w	10w	12w	12w	19w	19w	20w	20w	21w	21w	21w	25w	25w	Total
Embryo	e1	e1	e1	e2	e1	e1	twin-e1	twin-e2	e1	e1	twin-e1	twin-e2	e3	e1		
No. of ♂ FGC	37	49	43	3	55	72	72	78	65	76	9	199				758
No. of ♂ Soma	4	25	2	16	30	20	18	10	20	1	17	147				310

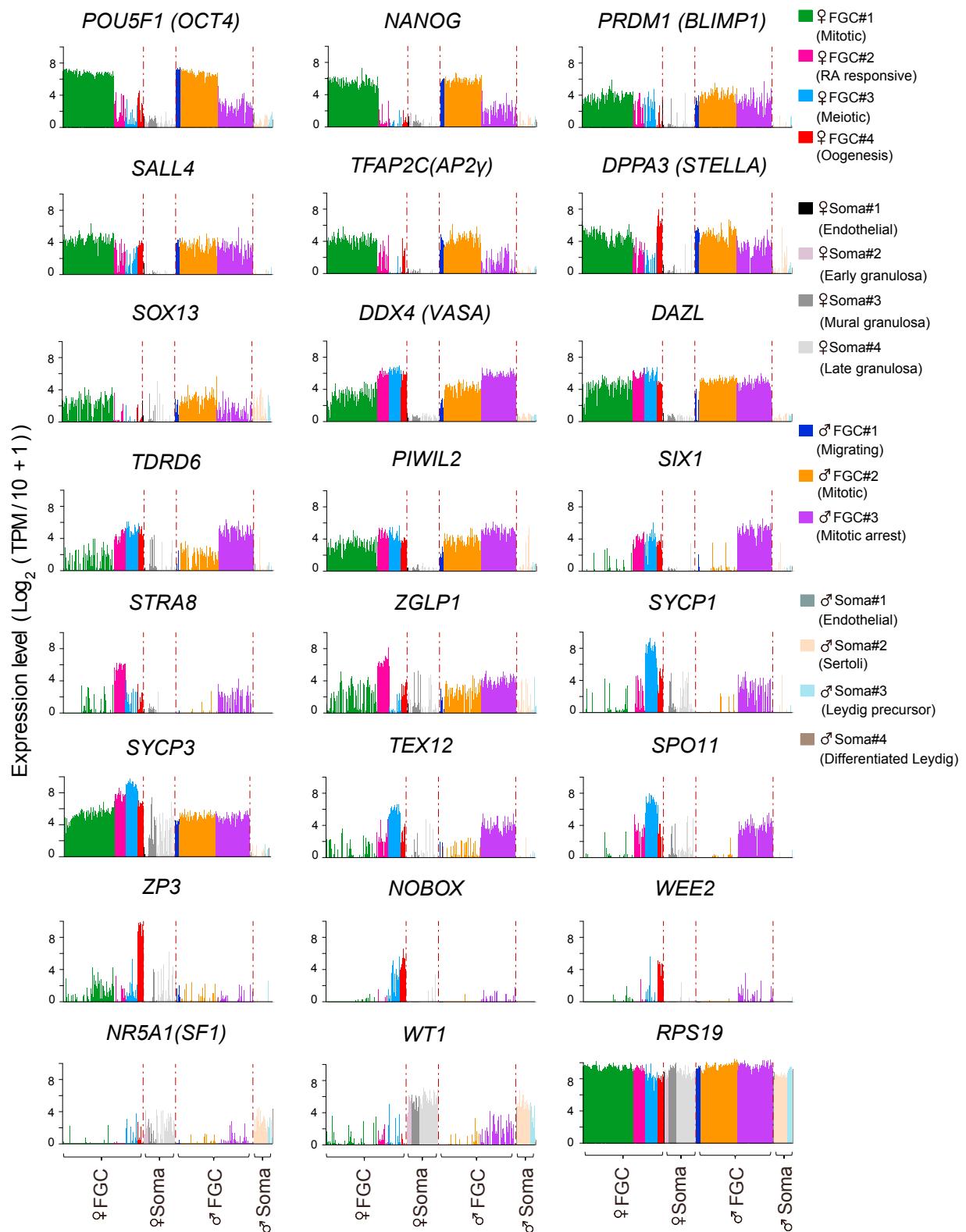
**B****C****D****Figure 1. Global Patterns of Single-Cell Expression Profiles and Identification of Cell Types**

(A) Overview of the FGCs and gonadal somatic cells analyzed in this study.

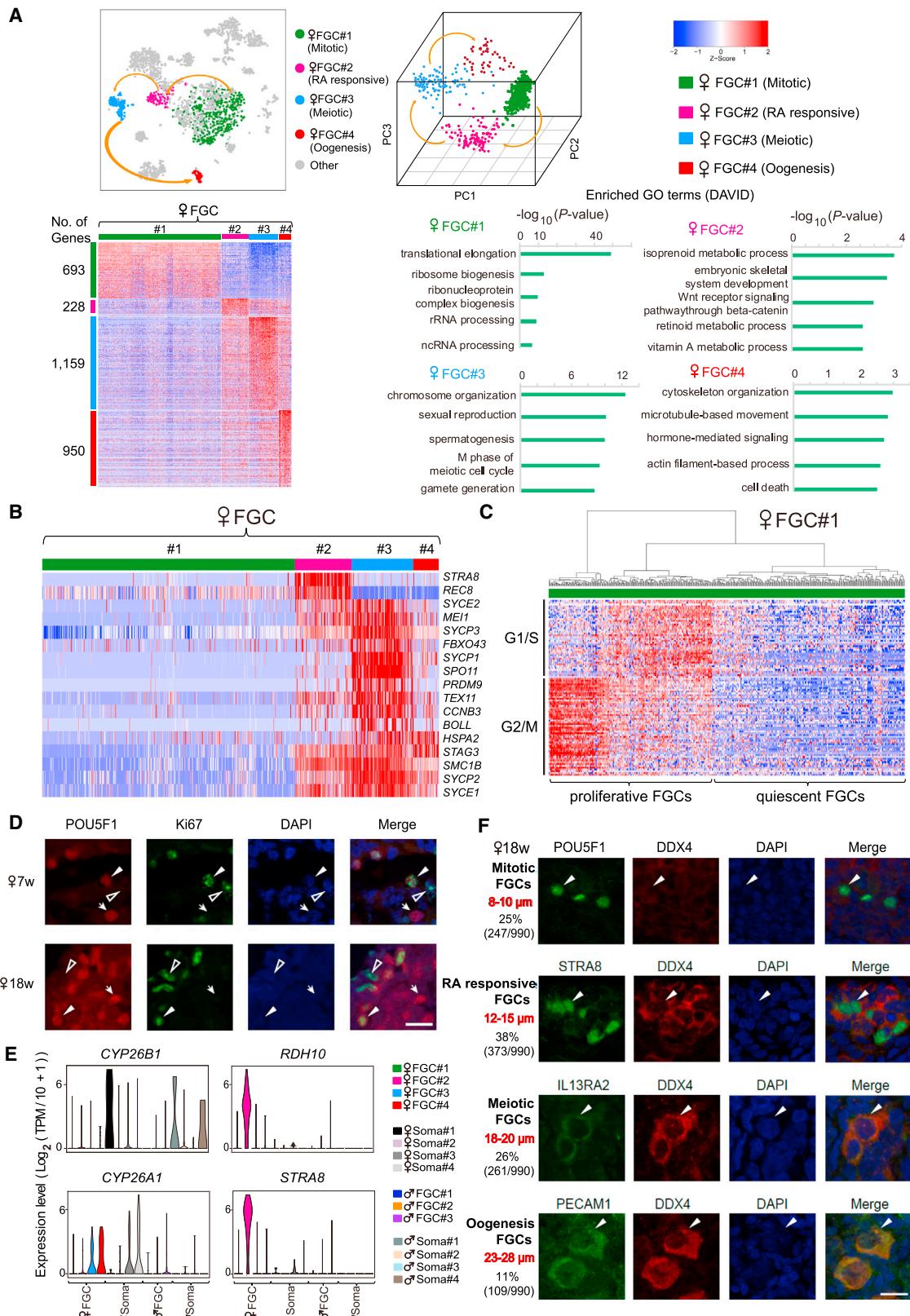
(B) t-SNE plots of FGCs and somatic cells. Left: cells colored by weeks post-fertilization. Right: cells colored by the identified cell types; the labeled numbers indicate the individual clusters: clusters 1–4 for female FGCs, clusters 5–7 for male FGCs, clusters 8–16 for neighboring somatic cells, and cluster 17 for possibly contaminated cells. See Table S1 for details.

(C) Expression patterns of FGC marker genes exhibited on t-SNE plots; a gradient of gray, yellow, and red indicates low to high expression.

(D) Expression patterns of gonadal somatic cell marker genes exhibited on t-SNE plots.

**Figure 2. Expression Patterns of Cell Type-Specific Genes**

Shown are histograms for the relative expression levels ( $\log_2[\text{TPM}/10+1]$ ) of selected marker genes in all analyzed cell types, including four types of female FGCs, four types of female gonadal somatic cells, three types of male FGCs, and four types of male gonadal somatic cells.

**Figure 3. Dynamic Gene Expression Patterns of Human Female FGCs**

(A) t-SNE plots, 3D PCA plots (top left), and heatmap of differentially expressed genes in four phases of female FGCs (bottom left). The enriched GO terms (biological processes) are shown at the right. The color key from blue to red indicates low to high gene expression levels, respectively. See Table S3 for details.

(legend continued on next page)

The cells in cluster 9 were Sertoli cells and specifically expressed *WT1* and *AMH*, two master genes for Sertoli cell development (de Santa Barbara et al., 2000). The cells in clusters 10 and 11 expressed *ARX* and *CYP17A1*, which are early and late Leydig cell markers, respectively, suggesting that cluster 10 cells are Leydig precursor cells, whereas cluster 11 cells are differentiated Leydig cells (Figure 1D).

In addition, cluster 12 were erythrocytes, clusters 13 and 14 were macrophages, clusters 15 and 16 were early T cells, and cluster 17 were FGCs that were probably contaminated by fragments of hemocytes. We excluded these cells in the subsequent analyses (Figure S1B).

### Dynamic Gene Expression Patterns of Human Female FGCs

As we mentioned, we have identified four clusters of human female FGCs: FGCs in the mitotic phase, RA signaling-responsive phase, meiotic prophase, and oogenesis phase (Figures 1B and 3A). To verify their diversity and dissect the differences in gene expression patterns, we identified 693, 228, 1,159, and 950 differentially expressed genes among these four clusters, respectively, and conducted a gene ontology (GO) analysis of these genes (Figure 3A; Table S3). Oogenesis phase FGC-specific genes were enriched in the categories of cytoskeleton organization, hormone-mediated signaling, and microtubule-based movement, among which were master transcription factors such as *NOBOX* and *FIGLA* (Figure 2; Figure S2). This pattern indicates that both the microtubule-based and microfilament-based cytoskeleton are reprogrammed when FGCs start oocyte formation and folliculogenesis (Figure 3A; Sánchez and Smitz, 2012).

As previously reported, human mitotic phase FGCs highly expressed pluripotency markers such as *POU5F1*, *LIN28A*, *SALL4*, and *DPPA5* (Figure 2; Figure S2; Houmard et al., 2009). They also expressed high levels of early germ cell-specific genes such as *PRDM1* (also called *BLIMP1*), *PRDM14*, *TFAP2C* (also known as *AP2γ*), *WNT3*, *NANOS3*, *DPPA3* (also known as *STELLA*), and *SOX17* (Figure 2; Figures S2 and S3C; Bowles et al., 2003; Julaton and Reijo Pera, 2011). Late FGC marker genes, such as *DDX4*, *DAZL*, *TDRD6*, and *PIWIL2*, were expressed in human RA-responsive phase, meiotic prophase, and oogenesis phase FGCs (Figure 2; Geijsen et al., 2004; Juliano et al., 2011; Kuramochi-Miyagawa et al., 2004; Saitou et al., 2002). In addition, meiotic prophase FGCs highly expressed gametogenesis-specific genes—*TEX12*, *TEX14*, *BOLL*, *SPO11*, *RAD51AP2*, and *PRDM9* (Figures 2 and 3B; Figure S2; Houmard et al., 2009; Soh et al., 2015). Oogenesis phase FGCs expressed *ZP3*, *NOBOX*, *OOSP2*, *GDF9*, *JAG1*, and *WEE2*, the master genes that are functionally crucial for oocyte development (Figure 2; Figure S2; Pepling, 2006; Tomizawa et al., 2012). More interestingly, we found that *SOX13* were expressed in human mitotic

phase FGCs, whereas *ZGLP1*, *ANHX*, *ASB9*, and *THRA/BTR* were specifically expressed in human RA-responsive FGCs (Figure 2; Figure S2). Although *Six1* is expressed in the somatic cells of the nascent genital ridges starting at 9.5 days post coitum (dpc) in mice, we found that *SIX1* was specifically expressed in human late FGCs (clusters 2–4) instead (Figure 2).

We then focused on the mitotic phase female FGCs (cluster 1). At this stage, female FGCs span a wide range of developmental stages from 4–26 weeks. We found developmental stage-specific gene expression patterns within 5-, 10-, 14-, 20-, and 26-week female mitotic FGCs by GO analysis (Figure S1D). In addition, we analyzed the cell-cycle status of FGCs using the previously reported G1/S and G2/M phase-specific genes (Tirshoh et al., 2016). For mitotic phase FGCs, approximately 40% were actively proliferating, with the remaining 60% at a relatively quiescent state (Figure 3C; Table S2). We co-stained *POU5F1* and *Ki67* in 7- and 18-week ovaries and found that 39% (69 of 176) and 42% (59 of 142) of *POU5F1*<sup>+</sup> cells were *Ki67*<sup>+</sup> cells in the 7- and 18-week samples, respectively (Figure 3D). The percentage of actively dividing FGCs was consistent with that estimated from our transcriptome analyses.

Next, we investigated the expression pattern of RA-responsive phase FGC-related genes to further decipher the mechanisms of the sex-specific timing of meiosis. Meiotic entry occurs at the mid-embryonic stage in female FGCs, whereas it occurs much later after birth in male germ cells (Koubova et al., 2006). A series of genetic ablation and pharmacological approach studies has demonstrated that, in the mouse embryonic ovary, RA is the driver of meiosis initiation via triggering *STRA8* and the expression of other downstream target genes. However, in the mouse embryonic testis, *CYP26B1*-expressing Sertoli cells form a barrier that degrades RA and inhibits the initiation of meiosis (Lin et al., 2008). Intriguingly, we found that the RA synthesis gene *RDH10* was pronouncedly upregulated in female gonadal cells rather than male gonadal cells, suggesting that RA synthesis may have an important role in female FGCs entering meiosis (Figure 3E; Duester, 2008; Sandell et al., 2007). Another remarkable finding is that *CYP26A1* expression was upregulated in RA signaling-responsive, meiotic prophase FGCs in female embryos, suggesting that *CYP26A1* may degrade RA quickly, which might be a result of negative feedback of meiosis initiation (Figure 3E). Collectively, we hypothesize that a balance of RA synthesis and degradation and its delicate regulation play a critical role in the sex-specific timing of meiosis entry of human FGCs.

Furthermore, we used principal-component analysis (PCA)-based method to find FGCs at the transition states between the two consecutive developmental stages. We conducted differential gene expression analysis and GO analysis of the transition state FGCs with their former- and later-stage cells, respectively (Figure S7A; Table S5).

(B) Heatmap of selected meiosis-related genes in four phases of female FGCs.

(C) Heatmap of cell-cycle-specific genes in female mitotic phase FGCs (FGC#1).

(D) Immunofluorescence of 7- and 18-week ovaries co-stained for *POU5F1* and *Ki67*. Filled triangles indicate *POU5F1*<sup>+</sup>*Ki67*<sup>+</sup> FGCs, arrows indicate *POU5F1*<sup>+</sup>*Ki67*<sup>−</sup> FGCs, and empty triangles indicate *POU5F1*<sup>−</sup>*Ki67*<sup>+</sup> ovary somatic cells. Scale bar, 20 μm.

(E) Violin plots showing relative expression levels ( $\log_2$  [TPM/10+1]) of selected genes involved in the RA synthesis and metabolism pathways.

(F) Immunofluorescence staining of the four types of female FGCs and their distribution in an 18-week ovary. The cell size and percentage of each type of FGCs are shown. Scale bar, 20 μm.

Intriguingly, the transition stage FGCs between meiotic prophase and oogenesis phase were enriched for GO terms such as the negative regulation of programmed cell death and apoptosis genes, whereas oogenesis phase FGCs were enriched for programmed cell death and apoptosis genes (Table S5). By immunostaining, we found that a fraction of the FGCs at the primordial follicle stage were positive for full-length CASP3 but not cleaved CASP3, indicating that they were probably primed for the apoptosis process (Figure S1H). This finding indicates that a proportion of human primordial follicles are probably ready to initiate apoptosis, which may contribute to germ cell attrition (Hussein, 2005; Jin et al., 2005).

Cell surface markers that are specific to human FGCs entering meiosis are still lacking. Fortunately, we found that *IL13RA2* (also known as *CD213A2*) and *PECAM1* were specifically expressed in meiotic prophase FGCs and oogenesis phase FGCs, respectively (Figure S2). We verified that *IL13RA2* and *PECAM1* were specifically expressed in *DDX4<sup>+</sup>* cells by immunostaining in an 18-week ovary (Figure 3F). Undoubtedly, isolation of the meiotic or oogenesis stage germ cells in female embryos will help us to understand the mechanisms of meiosis and oogenesis.

By immunofluorescence staining of the ovary cryosections from an 18-week human embryo, we validated the four types of female FGCs that we identified in transcriptome sequencing and further analyzed their size, location, and distribution in the ovary. We verified the four types of female FGCs in the sample: *POU5F1<sup>+</sup> DDX4<sup>-</sup>* mitotic FGCs, *STRA8<sup>+</sup>DDX4<sup>+</sup>* RA signaling-responsive FGCs, *IL13RA2<sup>+</sup>DDX4<sup>+</sup>* meiotic prophase FGCs, and *PECAM1<sup>+</sup>DDX4<sup>+</sup>* oogenesis phase FGCs (Figure 3F). Clearly, the *PECAM1<sup>+</sup>DDX4<sup>+</sup>* cells were primordial oocytes because they were enclosed by an ordered layer of flattened cells that were supposed to be granulosa cells. Intriguingly, with development proceeding, the FGCs gradually grew larger, from a diameter of about 8–10 µm at mitotic phase to approximately 23–28 µm at oogenesis phase (Figure 3F). Additionally, we found that mitotic and RA signaling-responsive FGCs were preferentially located at the peripheral cortex of the ovary, whereas meiotic prophase and oogenesis phase FGCs were located less peripherally than those earlier-stage FGCs (Figure S3D). In general, the percentages of the four types of female FGCs in the sample were about 25% (247 of 990, mitotic phase), 38% (373 of 990, RA signaling-responsive phase), 26% (261 of 990, meiotic prophase), and 11% (109 of 990, oogenesis phase), respectively. Moreover, the immunofluorescence assays clearly showed that *POU5F1* transferred from the nucleus to cytoplasm with the development process and finally disappeared at the primary follicle stage (Figure S1F). Nearly two-thirds of FGCs (156 of 227) that expressed *POU5F1* in the nucleus were *DDX4<sup>-</sup>* cells, with the remaining one-third (71 of 227) positive for *DDX4*.

In addition, we analyzed the FGCs in a 7-week ovary by co-staining of *POU5F1* and *DDX4*. The FGCs were relatively evenly distributed throughout the whole ovary. In general, *POU5F1<sup>+</sup> DDX4<sup>+</sup>*, *POU5F1<sup>+</sup>DDX4<sup>-</sup>*, and *POU5F1<sup>-</sup>DDX4<sup>+</sup>* FGCs accounted for 79% (160 of 203), 11% (23 of 203), and 10% (20 of 203), respectively (Figure S1G). Overall, the immunofluorescence assays validate the results of our transcriptome analyses

and offer us new insights into the divergence and heterogeneity of human fetal germ cells, even in the same ovary.

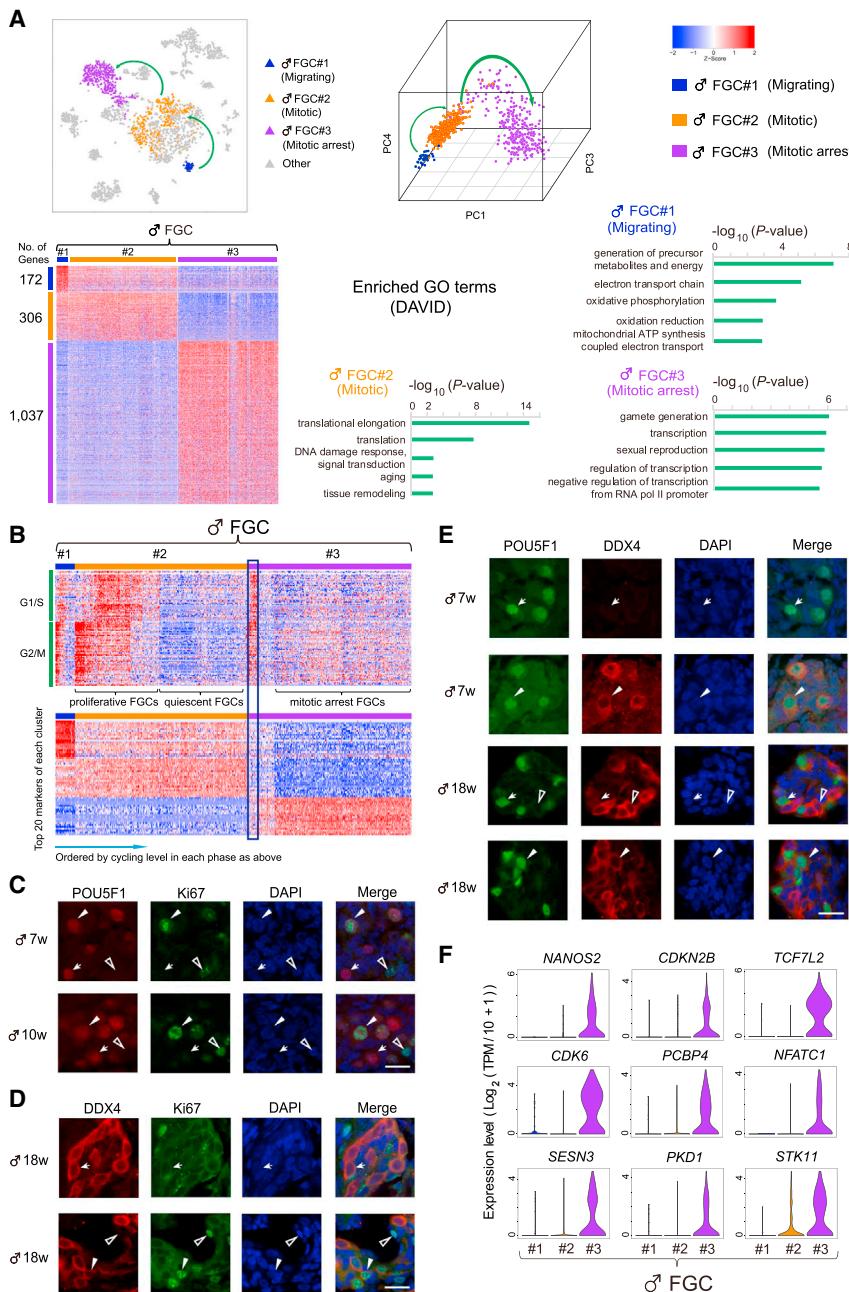
### Transcriptional Characteristics of Human Male FGCs

Based on the t-SNE analysis, human male FGCs could clearly be grouped into three clusters: migrating FGCs, gonadal mitotic FGCs, and FGCs entering mitotic arrest; these clusters were confirmed by unsupervised hierarchical clustering analysis (Figures 1B and 4A). According to the differentially expressed gene profiles, 172, 306, and 1,037 genes were specifically upregulated in migrating, mitotic, and mitotic arrest FGCs, respectively. Migrating FGC-specific genes were enriched in GO terms such as generation of precursor metabolites and energy, oxidative phosphorylation, and mitochondrial ATP synthesis-coupled electron transport, suggesting that migrating phase FGCs probably rely heavily on aerobic respiration and are different from other phases of FGCs in energy metabolism (Figure 4A; Table S3). Moreover, we used a PCA-based method to find FGCs at the transition states between the mitotic stage and mitotic arrest stage and analyzed their specific gene expression patterns (Figure S7A; Table S5).

For the mitotic phase male FGCs, approximately 50% were actively proliferating, which was similar to the female FGCs at the mitotic stage (Figure 4B). In addition, essentially all mitotic arrest phase FGCs were in the quiescent state, as expected. Interestingly, a small portion of mitotic arrest FGCs expressed cell-cycle-related genes. These cells shared both the gene expression pattern of mitotic stage and mitotic arrest stage FGCs, indicating that they were at the transition stage of the two types of FGCs (Figure 4B). By immunostaining, we found that about 31% (122 of 388) and 34% (86 of 250) of *POU5F1<sup>+</sup>* cells were *Ki67<sup>+</sup>* in the 7- and 10-week samples, respectively (Figure 4C). To unveil genes that might play a role in mitotic arrest, we analyzed the expression of genes in GO terms related to cell-cycle arrest. We found that 17 cell-cycle-arrest-related genes were specifically upregulated in male mitotic arrest phase FGCs, including *NANOS2*, *CDKN2B*, and *CDK6* (Figure 4F; Figure S3A; Lukas et al., 1995; Suzuki and Saga, 2008).

Then we analyzed the distribution of FGCs in the testes of a 7- and an 18-week human embryo by co-staining of *POU5F1* and *DDX4*. The 7-week testis FGCs were scattered in every seminiferous cord (Figure S3D), which was clearly different from the more homogeneous distribution of the 7-week ovary FGCs. Additionally, in the 7-week testis, we observed that *POU5F1<sup>+</sup>DDX4<sup>+</sup>*, *POU5F1<sup>+</sup>DDX4<sup>+</sup>*, and *POU5F1<sup>-</sup>DDX4<sup>-</sup>* FGCs accounted for 14% (56 of 414), 15% (61 of 414), and 72% (297 of 414) of germ cells, respectively (Figure 4E). Clearly, the ratio of the three types of FGCs was drastically different from that of 7-week female FGCs.

Interestingly, in 18-week testis, *POU5F1<sup>+</sup>* FGCs were distributed more homogenously in the seminiferous cords, whereas *DDX4<sup>+</sup>* FGCs were mostly distributed at the peripheral zones (Figure S3D). It was clearly different from the peripheral-to-center distribution of the early- to late-stage female FGCs in the 18-week ovary. Moreover, in the 18-week embryo testis, *POU5F1<sup>+</sup>DDX4<sup>-</sup>* and *POU5F1<sup>-</sup>DDX4<sup>+</sup>* FGCs accounted for 35% (161 of 459) and 63% (288 of 459) of germ cells, respectively, whereas *POU5F1<sup>+</sup>DDX4<sup>+</sup>* FGCs only accounted for the remaining 2% (10 of 459) (Figure 4E). According to the



**Figure 4. Dynamic Gene Expression Patterns of Human Male FGCs**

(A) t-SNE, 3D PCA plots (top left), and heatmap of differentially expressed genes in three phases of male FGCs (bottom left). The enriched GO terms (biological processes) are shown at the right. The color key from blue to red indicates low to high gene expression levels, respectively. See Table S3 for details.

(B) Heatmap of cell-cycle-specific genes in three phases of male FGCs ordered by their expression levels in each phase (top). At the bottom is a heatmap showing the top 20 genes of each phase in the same order. Several cycling cells in FGC#3 enclosed by a black rectangle express both FGC#2 and #3 marker genes, indicating that they are at a transition state.

(C) Immunofluorescence of 7- and 10-week testes. Filled triangles indicate POU5F1<sup>+</sup>Ki67<sup>+</sup> FGCs, arrows indicate POU5F1<sup>+</sup>Ki67<sup>-</sup> FGCs, and empty triangles indicate POU5F1<sup>-</sup>Ki67<sup>+</sup> testis somatic cells. Scale bar, 20  $\mu$ m.

(D) Immunofluorescence staining of an 18-week testis. Arrows indicate DDX4<sup>+</sup>Ki67<sup>-</sup> FGCs, filled triangles indicate DDX4<sup>+</sup>Ki67<sup>+</sup> FGCs, and empty triangles indicate DDX4<sup>-</sup>Ki67<sup>+</sup> testis somatic cells. Scale bar, 20  $\mu$ m.

(E) Immunofluorescence staining of 7- and 18-week testes. Arrows indicate POU5F1<sup>+</sup>DDX4<sup>-</sup> FGCs, filled triangles indicate POU5F1<sup>+</sup>DDX4<sup>+</sup> FGCs, and empty triangles indicate POU5F1<sup>-</sup>DDX4<sup>+</sup> testis somatic cells. Scale bar, 20  $\mu$ m.

(F) Violin plots showing the relative expression levels ( $\log_2 [TPM/10+1]$ ) of representative cell-cycle-arrest-related genes.

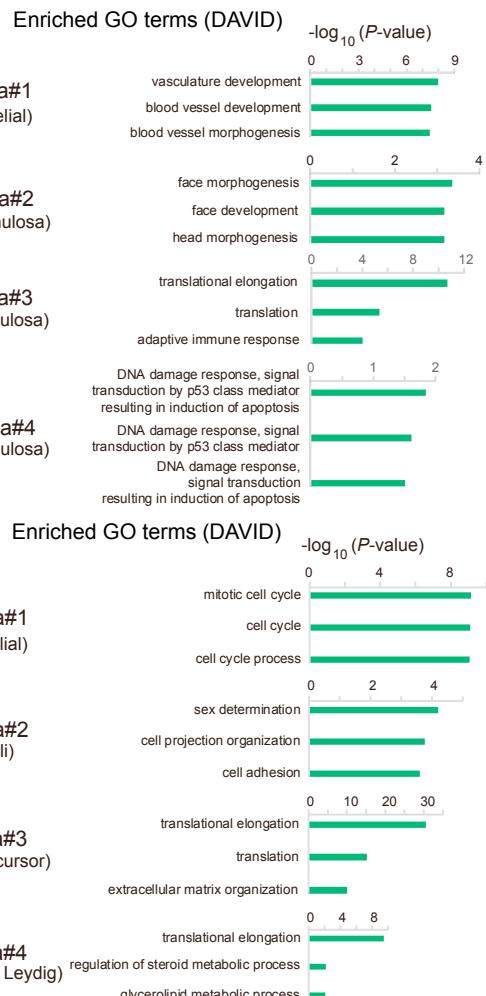
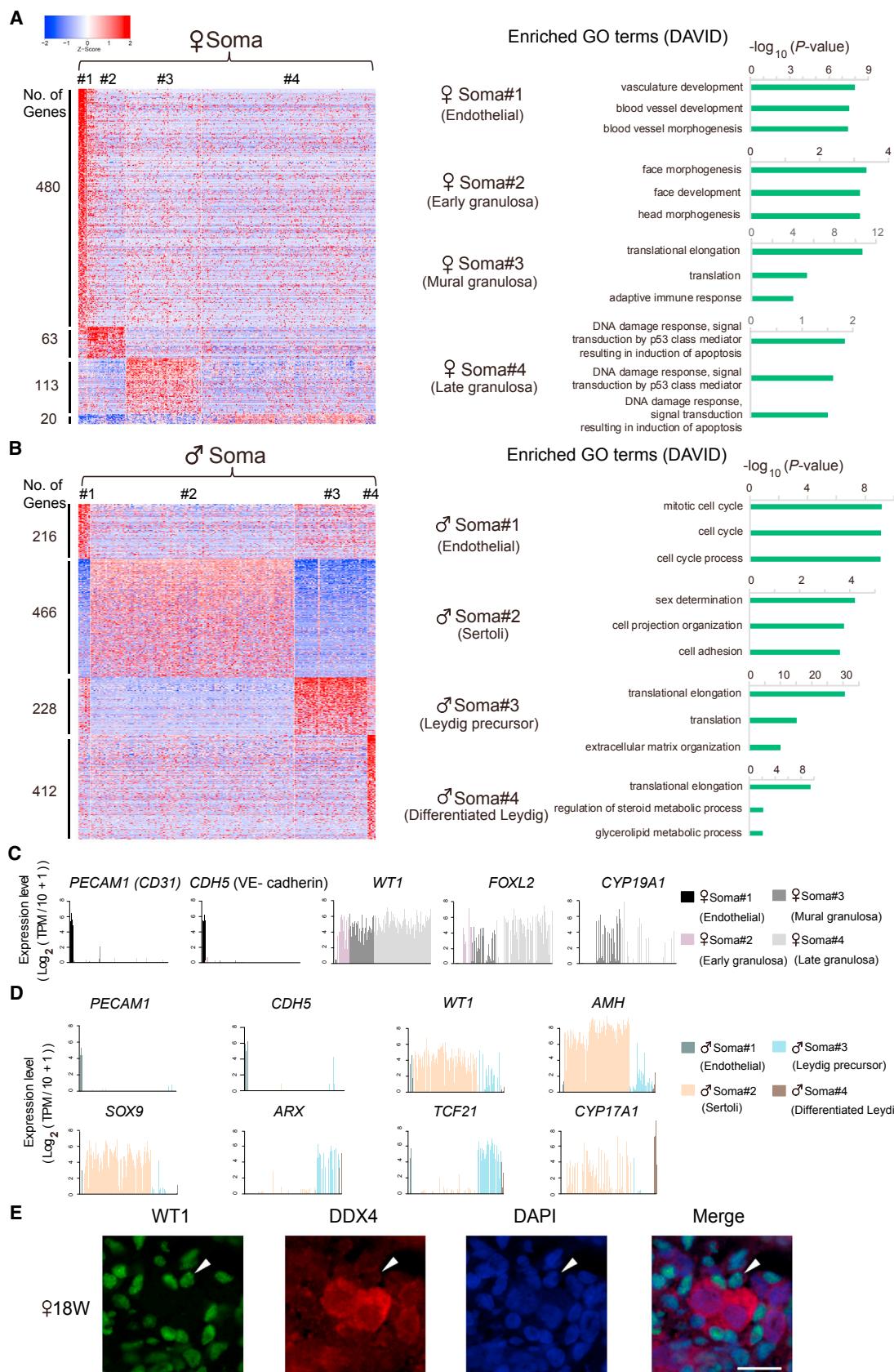
transcriptome data and the low rate of POU5F1<sup>+</sup>DDX4<sup>+</sup> FGCs, we supposed that POU5F1<sup>+</sup>DDX4<sup>-</sup> and POU5F1<sup>-</sup>DDX4<sup>+</sup> FGCs were mitotic stage and mitotic arrest stage FGCs, respectively. Co-staining of DDX4 and Ki67 in the 18-week embryo testis showed that 3% (6 of 224) of DDX4<sup>+</sup> FGCs were Ki67<sup>+</sup> (Figure 4D). We proposed that these DDX4<sup>+</sup> Ki67<sup>+</sup> cells were FGCs at the transition stage between the mitosis stage and mitosis arrest stage. Besides, the low percentage of POU5F1<sup>+</sup>DDX4<sup>+</sup> FGCs also implies that POU5F1 expression declines rapidly in the 18-week testis when development proceeds. The diversity between the male and female FGCs in the 7- and 18-week embryos concurs with the divergence of the anatomical structures and developmental processes between the testis and ovary.

We analyzed the pattern of X chromosome reactivation in FGCs and found that the result was consistent with our previous study (Guo et al., 2015). The total expression levels of all genes on the X chromosome in female FGCs were ~1.6-fold compared with those in male FGCs (Figure S4A). To analyze the sex-specific expressed genes in FGCs, we performed a differential gene expression analysis of mitotic phase FGCs in female and male embryos and found that 71 genes were differentially expressed between them; 30

of these genes were either on the X chromosome (for female FGCs) or on the Y chromosome (for male FGCs) (Figures S3B, S4B, and S4C).

### Transcription Profiles of FGC Niche Cells in the Gonads

First, we performed an unsupervised hierarchical clustering analysis of the somatic cells in female gonads and identified four types of cells (Figure 5A). Type 1 cells were gonadal endothelial cells that specifically expressed *PECAM1* and *CDH5*. Type 2–4 cells specifically expressed granulosa cell markers such as *WT1* and *FOXL2*. Type 3 cells also expressed *CYP19A1*, the mural granulosa cell marker (Figure 5C, and 5E; Hamel et al., 2008). Type 2, 3, and 4 cells were from 7- to



10-week, 10- to 20-week, and 20- to 26-week embryos, respectively, suggesting that they may be granulosa cells at different developmental stages. We conducted differential gene expression analysis and GO analysis of these cells. In general, the results of the GO analysis indicate that granulosa cells, at different time points, have different biological properties, such as unique translation or DNA repair features (Figure 5A; Table S4).

Next, based on the t-SNE analysis and the unsupervised hierarchical clustering analysis, the somatic cells in male gonads could be grouped into four clusters: endothelial cells (cluster 8), Sertoli cells (cluster 9), Leydig precursor cells (cluster 10), and differentiated Leydig cells (cluster 11) (Figures 1B, 1D, 5B, and 5D; Karl and Capel, 1998; Sharpe et al., 2003). According to the GO analysis, differentiated Leydig cells were enriched in the regulation of steroid metabolic process and glycerolipid metabolic process terms ( $p = 8.0 \times 10^{-3}$ ), indicating that differentiated Leydig cells have probably matured and can produce androgens (Figure 5B). The gonadal somatic cell master transcription factor *NR5A1* (also known as *SF1*) was expressed in granulosa cells, Sertoli cells, and Leydig cells (Figure 2). As expected, *XIST* was expressed in female gonadal somatic cells but not in male gonadal somatic cells (Plath et al., 2002). Interestingly, we found that *SOX9* was expressed not only in Sertoli cells but also in mural granulosa cells and late granulosa cells (Vidal et al., 2001). The male sex determination master gene *SRY* was expressed in Sertoli cells as expected (Sekido and Lovell-Badge, 2008). The transcription factor *Dmrt1* plays a significant role in sex determination and maintenance, mitosis/meiosis transition, and spermatogonial stem cell maintenance and replenishment in mice (Lindeman et al., 2015; Matson et al., 2010; Minkina et al., 2014; Zhang et al., 2016). We observed the divergent expression pattern of *DMRT1* in human male and female fetal gonads. In human testis, *DMRT1* was expressed highly throughout all of the developmental stages in male FGCs. However, in female FGCs, *DMRT1* was expressed highly in early-stage FGCs and decreased sharply when FGCs entered meiosis. The expression pattern is consistent with its role in mitosis/meiosis transitions in mice. In human gonadal somatic cells, *DMRT1* was expressed highly in Sertoli cells but not in granulosa cells. This agrees well with its role in sex determination and maintenance in mice (Figure S2).

We then analyzed the expression patterns of cell-cycle-related genes in FGCs and their niche cells. We found that FGC gonadal somatic cells generally expressed fewer cell-cycle-related genes than FGCs (Figure S5A). In addition, by calculating the percentage of Ki67<sup>+</sup> cells in POU5F<sup>-</sup> cells, we found that approximately one-tenth of gonadal somatic cells were actively proliferating (Figures 3D and 4C).

Moreover, we analyzed the expression characteristics of ribosome-associated genes, RNA-binding proteins in FGCs, and

their niche cells. We focused on the genes expressed divergently between mouse and human (Figures S3C and S5B–S5D).

### Transcription Factor Regulatory Network in Human FGC Development

Transcription factors usually play key regulatory roles during a wide variety of biological processes. The algorithm for the reconstruction of accurate cellular networks (ARACNe) is a powerful algorithm that can identify master players or “hub” genes in a gene regulatory network based on a large set of gene expression data, especially transcription factors. Thus, we used the ARACNe method to analyze all 1,568 known transcription factors from Animal TFDB 2.0. We found that *ZNF208*, *YBX1*, and *ZNF791* might specifically play a critical role in female mitotic phase FGCs, whereas *HES6*, *MAEL*, *ZGLP1*, *ZNF362*, *ZBTB11*, *HOXA5*, *HOXB6*, *HMGB3*, and *PBX1* are the top candidate genes for regulating the gene expression network in RA-responsive phase FGCs (Figures 6A–6D). In addition, *LHX8*, *NR4A2*, *ZNF382*, *MGA*, *RLF*, *ZIC4*, *PAXBP1*, *HSF2*, *DMRTA2*, and *L3MBTL1* potentially regulate the development of meiotic prophase FGCs, and *FIGLA*, *STAT1*, *TBX3*, *AFF1*, *JARID2*, *NFKB2*, and *NR3C2* most likely initiate the unique transcription network in oogenesis phase FGCs (Figures 6C–6F). Moreover, *HMGN3* and *CARHSP1* might be in charge of the gene expression network in male migrating FGCs. In contrast, *SIX2*, *CLOCK*, *TEF*, *ZBTB43*, *KDM5A*, *NKRF*, *MAEL*, *MEF2D*, *SOX15*, and *KLF8* might play an important role in male mitotic FGCs, and *EBF3*, *SATB2*, *ZNF267*, *FEZF1*, *SOX12*, *DMRTB1*, *ZSCAN5A*, and *MAEL* may have a regulatory role in male mitotic arrest FGCs (Figures S6A–S6D). In addition, we found that *SOX30*, *GATA3*, and *FOXD1* were expressed in female and male late-stage FGCs—female RA-responsive FGCs, meiotic prophase FGCs, oogenesis FGCs, and male mitotic arrest FGCs (Figure S6E). Interestingly, some of the pluripotency master transcription factors that were highly expressed in female mitotic phase FGCs declined drastically at the RA-responsive phase but recovered at the oogenesis phase, including *NANOG*, *TFCP2L1*, and *KLF4* (Figure S6F). This finding indicates that these pluripotency master genes are not only important for early-stage FGCs but also play crucial roles during later oogenesis.

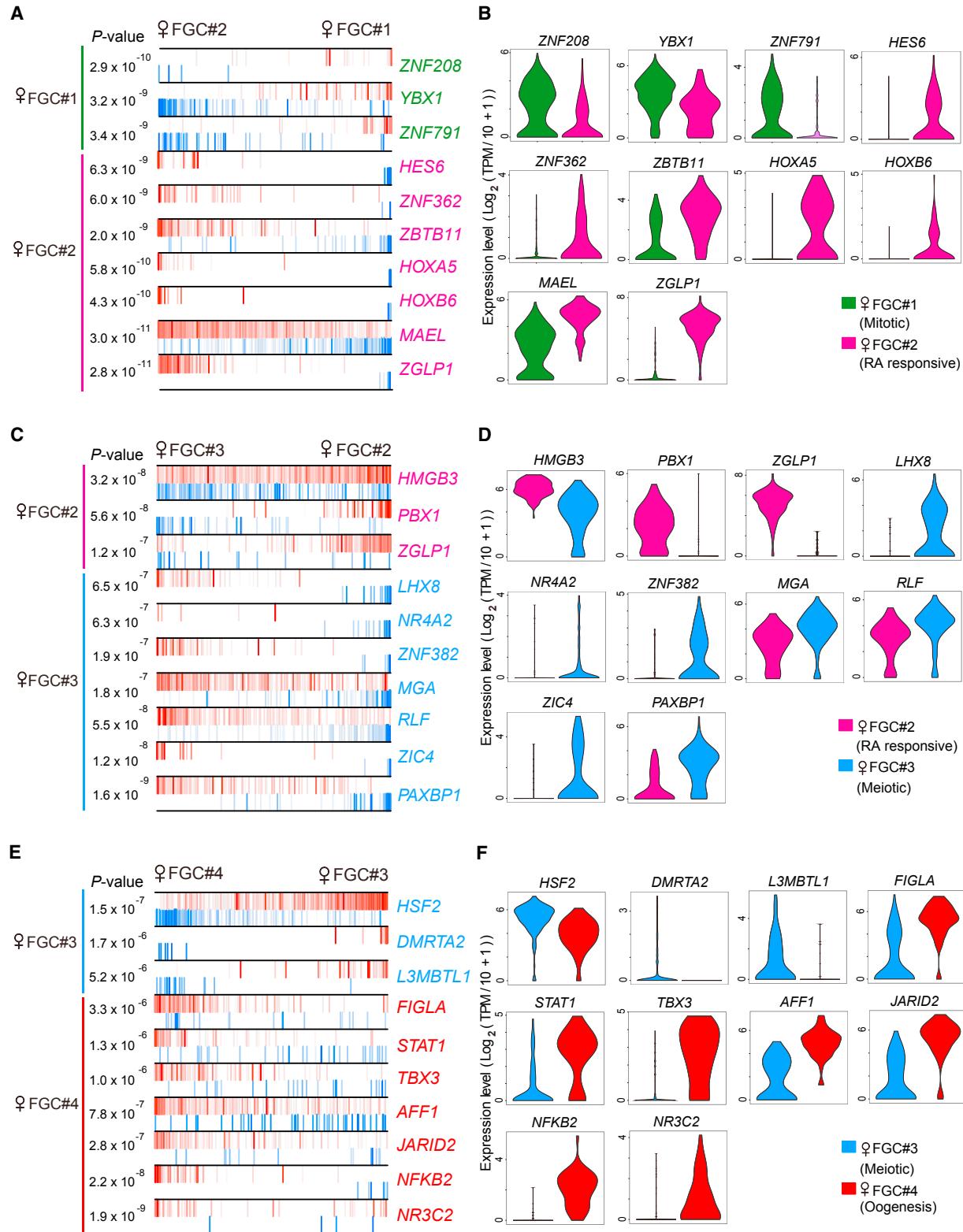
### Signaling Pathway Features in Human FGC and Gonadal Niche Cell Development

Our sequencing data showed that *KITL* (also known as *SCF*), the ligand of the KIT signaling pathway, was specifically expressed in gonadal somatic cells, whereas the receptor *KIT* (also known as *c-KIT* or *CD117*) was highly expressed in FGCs, which is consistent with previous findings (Figure S3C; Driancourt et al., 2000; Mauduit et al., 1999).

Gene set enrichment analysis (GSEA) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis showed that the

### Figure 5. Dynamic Gene Expression Patterns of Human Gonadal Somatic Cells

- (A) Heatmap of differentially expressed genes in four types of female gonadal somatic cells. The color key from blue to red indicates low to high expression levels, respectively. Shown at the right are the enriched GO terms (biological processes).
- (B) Heatmap of differentially expressed genes in four types of male gonadal somatic cells. Shown at the right are the enriched GO terms (biological processes).
- (C) Histograms showing the relative expression levels ( $\log_2 [\text{TPM}/10+1]$ ) of representative female gonadal somatic cell marker genes.
- (D) Histograms showing the relative expression levels ( $\log_2 [\text{TPM}/10+1]$ ) of representative male gonadal somatic cell marker genes.
- (E) Immunofluorescence of ovary cryosections from an 18-week human embryo. Triangles indicate gonadal somatic cells specifically expressing WT1. Scale bar, 20  $\mu\text{m}$ .

**Figure 6. Inferred Master Regulators for Four Phases of Female FGCs**

(A–F) The top 10 candidate master regulators for each type of FGCs identified by an algorithm for master regulator analysis algorithm (MARINa) (A, C, and E). Violin plots show the relative expression levels ( $\log_2 [\text{TPM}/10 + 1]$ ) of each master regulator (B, D, and F). MARINa plots show activated targets colored red and repressed targets colored blue for each potential master regulator (vertical lines on the x axis). On the x axis, genes were rank-sorted by their differential expression in the two developmental phases. The p values on the left indicate the significance of enrichment, calculated by permutating two developmental phases.

transforming growth factor  $\beta$  (TGF- $\beta$ ) signaling pathway, including the bone morphogenic protein (BMP) and NODAL signaling pathway, was involved in female RA-responsive, meiotic prophase, and oogenesis phase FGCs as well as in male mitotic arrest phase FGCs (Figure 7A; Shi and Massagué, 2003). For the BMP signaling pathway, in the fetal ovary, the ligand *BMP2* was highly expressed in mural granulosa cells and late granulosa cells, and the targets *ID2* and *ID3* were specifically upregulated in RA-responsive, meiotic prophase, and oogenesis FGCs. This is compatible with the possibility that the specificity of ligand expression leads to BMP signaling pathway activation in female late-stage FGCs, although the receptor *BMPR1B* was expressed in both granulosa cells and FGCs. This pattern indicates that BMP signaling is gradually activated when FGCs enter meiosis and that it probably plays a key role during this mitosis-meiosis transition (Figure 7A). In contrast, in fetal testis, the ligand *AMH* was highly expressed in Sertoli cells, whereas the receptor *BMPR1B* and targets *ID1*, *ID2*, and *ID3* were strongly upregulated when FGCs entered mitotic arrest, indicating that BMP signaling is probably important for male FGCs to enter mitotic arrest (Figure 7A). It seems that the BMP signaling pathway is activated in female meiotic FGCs by *BMP2* secreted from the neighboring granulosa cells, whereas it is activated in male mitotic arrest FGCs by *AMH* secreted from Sertoli cells.

Moreover, we performed a PCA-based analysis and obtained the pseudotime of both male and female FGCs and somatic cells. We plotted the trends of the BMP signaling pathway along the pseudotime to show the potential somatic cell-germline cell interaction. In the fetal ovary, *BMP2* expression gradually increased in granulosa cells along the pseudotime. Correspondingly, *BMPR1B*, *ID1*, *ID2*, and *ID3* gradually increased in FGCs in general. In the fetal testis, *AMH* expression gradually increased in Sertoli cells along the pseudotime. The expressions of *BMPR1B*, *ID1*, *ID2*, and *ID3* strongly increased at the transition stage between mitosis and mitotic arrest FGCs. In late mitotic arrest FGCs, *ID1* and *ID2* expression was relatively stable, whereas *ID3* decreased to some extent. The results are consistent with our GSEA and give new insights into the dynamic interaction of the signal pathway between FGCs and their niche cells (Figures S7B and S7C).

In the case of the NODAL signaling pathway, intriguingly, the ligand NODAL was expressed in both female and male mitotic FGCs, whereas the receptor *ACVR1C* and the target gene *PITX2* were specifically expressed in female meiotic and male mitotic arrest FGCs (Figure 7A; James et al., 2005). We performed immunofluorescence staining of *ACVR1C* in the ovary of an 18-week embryo and found that *ACVR1C* was specifically expressed in *DDX4*<sup>+</sup> cells (Figure 7C). Thus, we propose that late-stage FGCs respond to a NODAL signal secreted from early-stage FGCs, which maintains or even enhances the differences between these two subpopulations in the same embryo.

More interestingly, we analyzed gene expression features related to the NOTCH signaling pathway and found that the NOTCH signaling pathway might be activated in gonadal somatic cells by ligand expressed by FGCs (Bray, 2006). In the fetal ovary and testis, ligand *DLL3* was highly expressed in all phases of FGCs we analyzed and *JAG1* was specifically expressed in oogenesis phase FGCs. Correspondingly, the receptor *NOTCH2* and the target gene *HES1* were highly expressed in nearly all fe-

male and male somatic cells (Figure 7B). Hence, these results are compatible with the hypothesis that the NOTCH signaling pathway is activated in human gonadal somatic cells by receiving ligands expressed by neighboring germ cells.

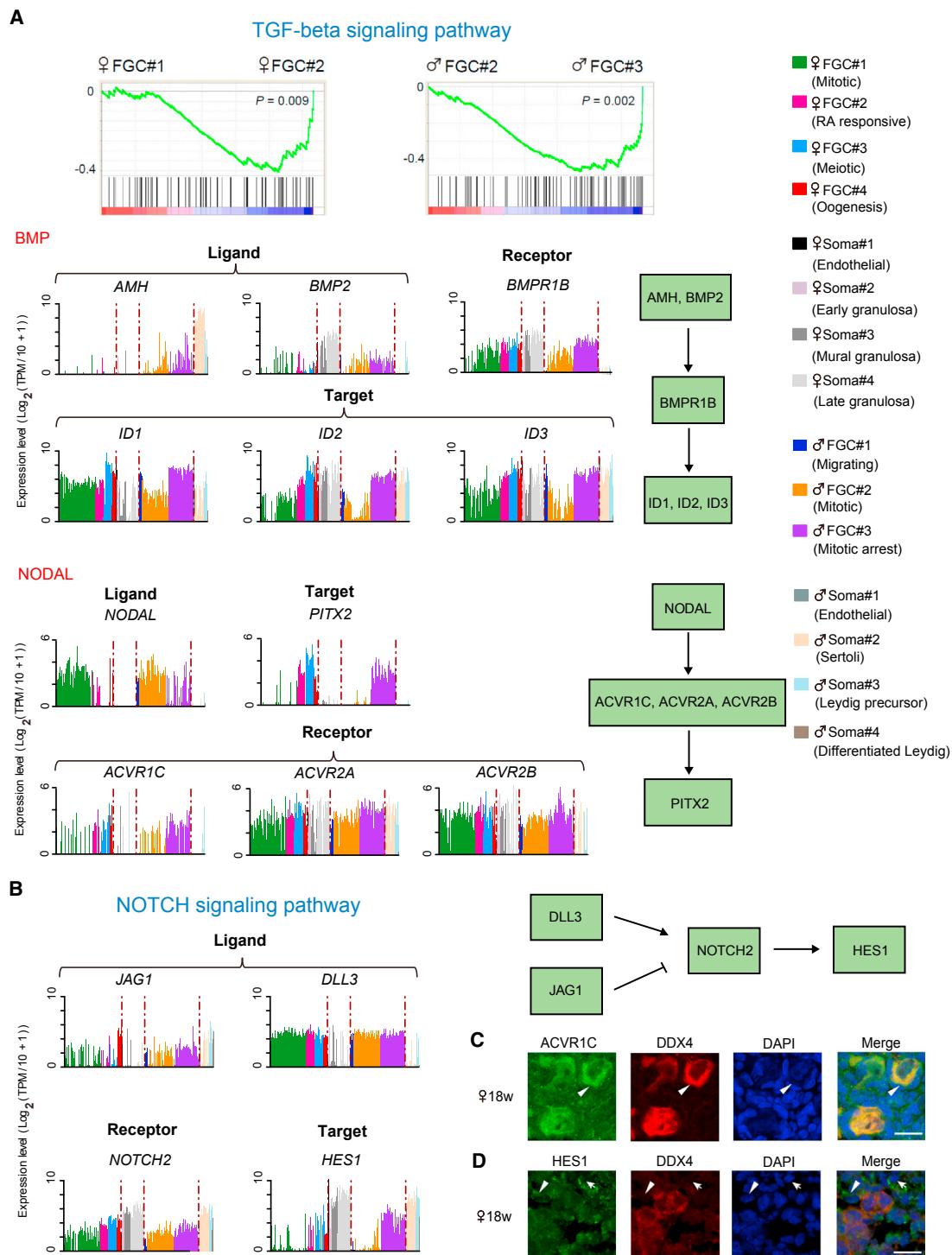
We found that *HES1* was highly expressed in some of the gonadal somatic cells, both those surrounded by FGCs and those not surrounded by FGCs, by immunostaining (Figure 7D). Considering that a direct cell-cell interaction is necessary for the Notch signaling pathway and that the number of FGCs is much lower than that of somatic cells in the gonads, we supposed that the gonadal somatic cells might receive NOTCH input from both surrounding FGCs and other adjacent somatic cells. Based on our single-cell RNA-seq data, one of the NOTCH ligands, *DLL3*, was expressed not only in FGCs but also in a fraction of gonadal somatic cells, providing a foundation for NOTCH signaling interaction between different gonadal somatic cells.

In addition, by GSEA, we found that human FGCs show diverse sex-specific and developmental stage-specific biological characteristics in basal metabolism, immune response, DNA replication, and DNA damage repair response (Figure S5B).

In summary, every subpopulation of FGCs, female or male, possessed specific transcription and metabolism states that were regulated by a coordinated and highly dynamic group of master transcription factors that are dedicatedly orchestrated by reciprocally interacting signaling pathways. At the same time, the FGCs also express and secrete a large set of ligands and program the transcription network and differentiation process of their niche cells in the gonads and vice versa. Thus, germline-nongermline and subpopulation-subpopulation interactions together fulfilled the exceptionally complex yet highly robust and accurately ordered development of the germ cells within the gonads. The important candidate factors we have found should greatly facilitate our further understanding of the mechanisms involved in FGC and gonadal somatic cell development *in vivo* and provide novel clues to FGC-like cell differentiation *in vitro* and potential therapy solutions for reproduction-related diseases (Kurimoto and Saitou, 2015; Saitou and Yamaji, 2012).

## DISCUSSION

We previously performed single-cell RNA-seq analysis for 319 human FGCs and their neighboring somatic cells for 15 embryos between 4 and 19 weeks post-fertilization (Guo et al., 2015). However, the numbers of embryos (two embryos) and single cells (100 cells) after 11 weeks were too sparse in this previous study, and the subpopulations of meiotic FGCs have not been properly identified. Here we performed single-cell RNA-seq analysis for 2,167 single FGCs and their niche cells from the gonads of 29 embryos between 4 and 26 weeks. We clearly identified four subpopulations for female FGCs: mitotic phase, RA signaling-responsive phase, meiotic prophase, and oogenesis phase. Correspondingly, we also identified three subpopulations for male FGCs: migrating phase, gonadal mitotic phase, and mitotic arrest phase. For their niche cells, we identified endothelial cells, early granulosa cells, mural granulosa cells, and late granulosa cells for female gonads and endothelial cells, Sertoli cells, Leydig precursor cells, and differentiated Leydig cells for male gonads. To our knowledge, this is the first time that human FGCs and their niche cells have been so comprehensively and

**Figure 7. Selected Signaling Pathways in the Development of Human FGCs and Their Niche Cells**

(A) TGF- $\beta$  signaling pathway. Shown is a GSEA enrichment plot of the KEGG TGF- $\beta$  signaling pathway. For the BMP and NODAL signaling pathways, the relative expression levels ( $\log_2$  [TPM/10+1]) of the specific ligands, receptors, and target genes are shown (left). The diagrams at the right show the relationship among these genes.

(B) For the NOTCH signaling pathway, the relative expression levels ( $\log_2$  [TPM/10+1]) of the specific ligands, receptors, and target genes are shown (left). The diagrams at the right show the relationship among these genes.

(C) Immunofluorescence staining of an 18-week ovary. Triangles indicate the ACVR1C $^{+}$ DDX4 $^{+}$  FGCs. Scale bar, 20  $\mu$ m.

(D) Immunofluorescence staining of an 18-week ovary. Triangles indicate HES1 $^{+}$  somatic cells surrounding FGCs, and arrows indicate HES1 $^{+}$  somatic cells not surrounding FGCs. Scale bar, 20  $\mu$ m.

extensively analyzed, covering 22 weeks of development for human germ cells.

Second, we found that several signaling pathways are coordinately and reciprocally expressed between FGCs and their gonadal niche cells. For example, male gonadal somatic cells specifically express *AMH*, and female gonadal somatic cells specifically express *BMP2*, the ligands for the BMP signaling pathway, whereas subpopulations of FGCs express its receptor, *BMPR1B*, to receive and respond to this signal. *ID1*, *ID2*, and *ID3*, the target genes of the BMP signaling pathway, are highly and specifically expressed in FGCs. Moreover, FGCs specifically express *DLL3*, the ligand for the NOTCH signaling pathway, whereas their gonadal niche cells express *NOTCH2* to respond to the signal. In fact, we found that one of the major target genes of NOTCH signaling, *HES1*, is highly and specifically expressed in gonadal somatic cells, verifying that the NOTCH signaling pathway is specifically active in gonadal somatic cells (Song et al., 2007).

Third, using ARACNe, we identified the top candidate master transcription factors that are potentially crucial for the stage-specific development of both female and male FGCs. These candidate transcription factors are most likely the key hub or node genes that establish a robust regulatory gene expression network in fetal germ cells. In the future, further analysis of these genes, their upstream signaling pathway, and their downstream target genes in FGCs *in vivo* will give us a deeper understanding of the unique gene regulation network of germline cells.

Our work gives new insights into the crucial features of human fetal germ cells during their highly ordered mitotic, meiotic, and gametogenesis processes *in vivo*. It also lays a solid foundation for assessing the cell identity and property of FGC-like cells that have been differentiated from pluripotent stem cells *in vitro* (Reik and Surani, 2015; Tomizawa and Sasaki, 2012). More importantly, the reciprocal relationship between the signaling pathways of FGCs and their niche cells will give important clues for further optimizing and improving the efficiency of germ cell differentiation *in vitro* (Wang and Zhou, 2016). Our work offers a comprehensive roadmap for germ cell development *in vivo* under physiological conditions and may help dissect germ cell-related diseases, such as infertility and teratomas, under pathological conditions (Stevens, 1967).

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- COLLECTION OF HUMAN FETAL SAMPLES
- METHOD DETAILS
  - Isolation of Human FGCs and Gonadal Somatic Cells by MACS and FACS
  - Single-cell RNA-Seq Library Preparation and Sequencing
  - Processing of Single-cell RNA-seq Data
  - Down-sampling Analysis
  - Nonlinear Dimensional Reduction (tSNE)
  - Identification of Erythrocytes

- Identification of Differentially Expressed Genes
- Cell-Cycle Analysis
- Inference of Master Regulators with ARACNe and MARINA
- GSEA Analysis
- PCA-based Developmental Pseudotime and Characterization of the Transition State of FGCs
- Reactivation of X Chromosome in the Female FGCs
- Immunofluorescence Staining of the Frozen Sections
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

## SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, five tables, and methods and can be found with this article online at <http://dx.doi.org/10.1016/j.stem.2017.03.007>.

## AUTHOR CONTRIBUTIONS

F.T. and J.Q. conceived the project. L.L., L.Y., J. Yong, Y.H., X.F., X. Wu, H.G., X.L., X. Wang, X.Z., R.L., J. Yan, Y.W., Y.Z., W.W., Y.R., P.Y., Z.Y., F.G., and L.W. performed the experiments. J.D. and B.H. conducted the bioinformatics analyses. F.T., L.L., J.D., and L.Y. wrote the manuscript with help from all authors.

## ACKNOWLEDGMENTS

L.Y., J.Q., and F.T. were supported by grants from the National Natural Science Foundation of China (31625018, 31230047, 31522034, 31571544, 81561138005, and 81521002) and the National Basic Research Program of China (2014CB943200). J.Q. and F.T. were also supported by a grant from the Beijing Municipal Science and Technology Commission (D151100002415000). L.Y. was supported by a grant from the National High Technology Research and Development Program (2015AAQ020407). The work was supported by the Beijing Advanced Innovation Center for Genomics at Peking University. We sincerely thank Shujing Wang, Xuefang Zhang, and Fei Wang from the National Center for Protein Sciences Beijing (Peking University) for assistance with FACS. We also thank Chunyan Shan from the Core Facilities at School of Life Sciences, Peking University for assistance with confocal microscopy. Part of the bioinformatics analysis was performed on the Computing Platform of the Center for Life Science.

Received: August 31, 2016

Revised: November 30, 2016

Accepted: March 15, 2017

Published: April 27, 2017; corrected online: May 18, 2017

## REFERENCES

- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.
- Anderson, R.A., Fulton, N., Cowan, G., Coutts, S., and Saunders, P.T. (2007). Conserved and divergent patterns of expression of DAZL, VASA and OCT4 in the germ cells of the human fetal ovary and testis. *BMC Dev. Biol.* 7, 136.
- Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. *Nat Genet.* 37, 382–390.
- Bowles, J., Teasdale, R.P., James, K., and Koopman, P. (2003). Dppa3 is a marker of pluripotency and has a human homologue that is expressed in germ cell tumours. *Cytogenet. Genome Res.* 101, 261–265.
- Bray, S.J. (2006). Notch signalling: a simple pathway becomes complex. *Nat. Rev. Mol. Cell Biol.* 7, 678–689.
- Castrillon, D.H., Quade, B.J., Wang, T.Y., Quigley, C., and Crum, C.P. (2000). The human VASA gene is specifically expressed in the germ cell lineage. *Proc. Natl. Acad. Sci. USA* 97, 9585–9590.

- Chen, D., and Clark, A.T. (2015). Human germline differentiation charts a new course. *EMBO J.* 34, 975–977.
- Chung, N.C., and Storey, J.D. (2015). Statistical significance of variables driving systematic variation in high-dimensional data. *Bioinformatics* 31, 545–554.
- De Felici, M., and Barrios, F. (2013). Seeking the origin of female germline stem cells in the mammalian ovary. *Reproduction* 146, R125–R130.
- de Santa Barbara, P., Moniot, B., Poulat, F., and Berta, P. (2000). Expression and subcellular localization of SF-1, SOX9, WT1, and AMH proteins during early human testicular development. *Dev. Dyn.* 217, 293–298.
- Del-Mazo, J., Brieño-Enríquez, M.A., García-López, J., López-Fernández, L.A., and De-Felici, M. (2013). Endocrine disruptors, gene deregulation and male germ cell tumors. *Int. J. Dev. Biol.* 57, 225–239.
- Driancourt, M.-A., Reynaud, K., Cortvrindt, R., and Smitz, J. (2000). Roles of KIT and KIT LIGAND in ovarian function. *Rev. Reprod.* 5, 143–152.
- Duester, G. (2008). Retinoic acid synthesis and signaling during early organogenesis. *Cell* 134, 921–931.
- Geijsen, N., Horoschak, M., Kim, K., Gribnau, J., Eggan, K., and Daley, G.Q. (2004). Derivation of embryonic germ cells and male gametes from embryonic stem cells. *Nature* 427, 148–154.
- Gkountela, S., Li, Z., Vincent, J.J., Zhang, K.X., Chen, A., Pellegrini, M., and Clark, A.T. (2013). The ontogeny of cKIT+ human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and in vitro differentiation. *Nat. Cell Biol.* 15, 113–122.
- Gkountela, S., Zhang, K.X., Shafiq, T.A., Liao, W.W., Hargan-Calvopiña, J., Chen, P.Y., and Clark, A.T. (2015). DNA Demethylation Dynamics in the Human Prenatal Germline. *Cell* 161, 1425–1436.
- Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., Hu, Y., Wang, X., Wei, Y., et al. (2015). The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. *Cell* 161, 1437–1452.
- Hamel, M., Dufort, I., Robert, C., Gravel, C., Leveille, M.C., Leader, A., and Sirard, M. (2008). Identification of differentially expressed markers in human follicular cells associated with competent oocytes. *Hum. Reprod.* 23, 1118–1127.
- Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* 2, 666–673.
- Heeren, A.M., He, N., de Souza, A.F., Goercharn-Ramlal, A., van Iperen, L., Roost, M.S., Fernandes, M.M.G., van der Westerlaken, L.A., and de Sousa Lopes, S.M.C. (2016). On the development of extragonadal and gonadal human germ cells. *Biology Open*. <http://dx.doi.org/10.1242/bio.013847>.
- Hirasawa, R., Kato, Y., and Sasaki, H. (2007). Establishment of genomic imprints in the mammalian germline and their maintenance in preimplantation embryos. *Tanpakushitsu Kakusan Koso* 52, 2136–2141.
- Huang, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13.
- Huang, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57.
- Houmard, B., Small, C., Yang, L., Naluai-Cecchini, T., Cheng, E., Hassold, T., and Griswold, M. (2009). Global gene expression in the human fetal testis and ovary. *Biol. Reprod.* 81, 438–443.
- Hussein, M.R. (2005). Apoptosis in the ovary: molecular mechanisms. *Hum. Reprod. Update* 11, 162–177.
- Irie, N., Weinberger, L., Tang, W.W., Kobayashi, T., Viukov, S., Manor, Y.S., Dietmann, S., Hanna, J.H., and Surani, M.A. (2015). SOX17 is a critical specifier of human primordial germ cell fate. *Cell* 160, 253–268.
- Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.-B., Lönnberg, P., and Linnarsson, S. (2012). Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. *Nat. Protoc.* 7, 813–828.
- Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* 11, 163–166.
- James, D., Levine, A.J., Besser, D., and Hemmati-Brivanlou, A. (2005). TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* 132, 1273–1282.
- Jin, X., Han, C.S., Yu, F.Q., Wei, P., Hu, Z.Y., and Liu, Y.X. (2005). Anti-apoptotic action of stem cell factor on oocytes in primordial follicles and its signal transduction. *Mol. Reprod. Dev.* 70, 82–90.
- Julaton, V.T.A., and Reijo Pera, R.A. (2011). NANOS3 function in human germ cell development. *Hum. Mol. Genet.* 20, 2238–2250.
- Juliano, C., Wang, J., and Lin, H. (2011). Uniting germline and stem cells: the function of Piwi proteins and the piRNA pathway in diverse organisms. *Annu Rev Genet.* 45, 447–469.
- Kannanayakal, T.J., and Eberwine, J. (2005). mRNA methods used in dissecting gene expression of the brain. *Ageing Res. Rev.* 4, 513–528.
- Karl, J., and Capel, B. (1998). Sertoli cells of the mouse testis originate from the coelomic epithelium. *Dev. Biol.* 203, 323–333.
- Kee, K., Angeles, V.T., Flores, M., Nguyen, H.N., and Reijo Pera, R.A. (2009). Human DAZL, DAZ and BOULE genes modulate primordial germ-cell and haploid gamete formation. *Nature* 462, 222–225.
- Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161, 1187–1201.
- Koubova, J., Menke, D.B., Zhou, Q., Capel, B., Griswold, M.D., and Page, D.C. (2006). Retinoic acid regulates sex-specific timing of meiotic initiation in mice. *Proc. Natl. Acad. Sci. USA* 103, 2474–2479.
- Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., et al. (2004). Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* 131, 839–849.
- Kurimoto, K., and Saitou, M. (2015). Mechanism and Reconstitution In Vitro of Germ Cell Development in Mammals. *Cold Spring Harbor Symp. Quant. Biol.* 80, 147–154.
- Lachmann, A., Giorgi, F.M., Lopez, G., and Califano, A. (2016). ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics* 32, 2233–2235.
- Lefebvre, C., Rajbhandari, P., Alvarez, M.J., Bandaru, P., Lim, W.K., Sato, M., Wang, K., Sumazin, P., Kustagi, M., Bisikirska, B.C., et al. (2010). A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol. Syst. Biol.* 6, 377.
- Lin, Y., Gill, M.E., Koubova, J., and Page, D.C. (2008). Germ cell-intrinsic and -extrinsic factors govern meiotic initiation in mouse embryos. *Science* 322, 1685–1687.
- Lindeman, R.E., Gearhart, M.D., Minkina, A., Krentz, A.D., Bardwell, V.J., and Zarkower, D. (2015). Sexual cell-fate reprogramming in the ovary by DMRT1. *Curr. Biol.* 25, 764–771.
- Lukas, J., Aagaard, L., Strauss, M., and Bartek, J. (1995). Oncogenic aberrations of p16INK4/CDKN2 and cyclin D1 cooperate to deregulate G1 control. *Cancer Res.* 55, 4818–4823.
- Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161, 1202–1214.
- Margolin, A.A., Nemenman, I., Bass, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and Califano, A. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7 (Suppl 1), S7.
- Matson, C.K., Murphy, M.W., Griswold, M.D., Yoshida, S., Bardwell, V.J., and Zarkower, D. (2010). The mammalian doublesex homolog DMRT1 is a transcriptional gatekeeper that controls the mitosis versus meiosis decision in male germ cells. *Dev. Cell* 19, 612–624.
- Mauduit, C., Hamamah, S., and Benahmed, M. (1999). Stem cell factor/c-kit system in spermatogenesis. *Hum. Reprod. Update* 5, 535–545.
- Minkina, A., Matson, C.K., Lindeman, R.E., Ghyselinck, N.B., Bardwell, V.J., and Zarkower, D. (2014). DMRT1 protects male gonadal cells from retinoid-dependent sexual transdifferentiation. *Dev. Cell* 29, 511–520.

- Nakaki, F., Hayashi, K., Ohta, H., Kurimoto, K., Yabuta, Y., and Saitou, M. (2013). Induction of mouse germ-cell fate by transcription factors in vitro. *Nature* 501, 222–226.
- Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 27, 29–34.
- Pepling, M.E. (2006). From primordial germ cell to primordial follicle: mammalian female germ cell development. *Genesis* 44, 622–632.
- Perrett, R.M., Turnpenny, L., Eckert, J.J., O'Shea, M., Sonne, S.B., Cameron, I.T., Wilson, D.I., Rajpert-De Meyts, E., and Hanley, N.A. (2008). The early human germ cell lineage does not express SOX2 during *in vivo* development or upon *in vitro* culture. *Biol. Reprod.* 78, 852–858.
- Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* 10, 1096–1098.
- Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9, 171–181.
- Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A., and Panning, B. (2002). Xist RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.* 36, 233–278.
- Rajpert-De Meyts, E., Hanstein, R., Jørgensen, N., Graem, N., Vogt, P.H., and Skakkebaek, N.E. (2004). Developmental expression of POU5F1 (OCT-3/4) in normal and dysgenetic human gonads. *Hum. Reprod.* 19, 1338–1344.
- Reik, W., and Surani, M.A. (2015). Germline and pluripotent stem cells. *Cold Spring Harb. Perspect. Biol.* 7, a019422.
- Robinson, L.L., Gaskell, T.L., Saunders, P.T., and Anderson, R.A. (2001). Germ cell specific expression of c-kit in the human fetal gonad. *Mol. Hum. Reprod.* 7, 845–852.
- Saitou, M., and Miyauchi, H. (2016). Gametogenesis from Pluripotent Stem Cells. *Cell Stem Cell* 18, 721–735.
- Saitou, M., and Yamaji, M. (2012). Primordial germ cells in mice. *Cold Spring Harb. Perspect. Biol.* 4, a008375.
- Saitou, M., Barton, S.C., and Surani, M.A. (2002). A molecular programme for the specification of germ cell fate in mice. *Nature* 418, 293–300.
- Sánchez, F., and Smitz, J. (2012). Molecular control of oogenesis. *Biochim. Biophys. Acta* 1822, 1896–1912.
- Sandell, L.L., Sanderson, B.W., Moiseyev, G., Johnson, T., Mushegian, A., Young, K., Rey, J.-P., Ma, J.X., Staehling-Hampton, K., and Trainor, P.A. (2007). RDH10 is essential for synthesis of embryonic retinoic acid and is required for limb, craniofacial, and organ development. *Genes Dev.* 21, 1113–1124.
- Sasaki, K., Yokobayashi, S., Nakamura, T., Okamoto, I., Yabuta, Y., Kurimoto, K., Ohta, H., Moritoki, Y., Iwatani, C., Tsuchiya, H., et al. (2015). Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells. *Cell Stem Cell* 17, 178–194.
- Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33, 495–502.
- Seisenberger, S., Peat, J.R., and Reik, W. (2013). Conceptual links between DNA methylation reprogramming in the early embryo and primordial germ cells. *Curr. Opin. Cell Biol.* 25, 281–288.
- Sekido, R., and Lovell-Badge, R. (2008). Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature* 453, 930–934.
- Sharpe, R.M., McKinnell, C., Kivlin, C., and Fisher, J.S. (2003). Proliferation and functional maturation of Sertoli cells, and their relevance to disorders of testis function in adulthood. *Reproduction* 125, 769–784.
- Shi, Y., and Massagué, J. (2003). Mechanisms of TGF-β signaling from cell membrane to the nucleus. *Cell* 113, 685–700.
- Smallwood, S.A., and Kelsey, G. (2012). De novo DNA methylation: a germ cell perspective. *Trends Genet.* 28, 33–42.
- Soh, Y.Q., Junker, J.P., Gill, M.E., Mueller, J.L., van Oudenaarden, A., and Page, D.C. (2015). A Gene Regulatory Program for Meiotic Prophase in the Fetal Ovary. *PLoS Genet.* 11, e1005531.
- Song, X., Call, G.B., Kirilly, D., and Xie, T. (2007). Notch signaling controls germline stem cell niche formation in the *Drosophila* ovary. *Development* 134, 1071–1080.
- Spaethling, J.M., Sanchez-Alavez, M., Lee, J., Xia, F.C., Dueck, H., Wang, W., Fisher, S.A., Sul, J.-Y., Seale, P., Kim, J., et al. (2016). Single-cell transcriptomics and functional target validation of brown adipocytes show their complex roles in metabolic homeostasis. *FASEB J.* 30, 81–92.
- Stevens, L.C. (1967). Origin of testicular teratomas from primordial germ cells in mice. *J. Natl. Cancer Inst.* 38, 549–552.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550.
- Suzuki, A., and Saga, Y. (2008). Nanos2 suppresses meiosis and promotes male germ cell differentiation. *Genes Dev.* 22, 430–435.
- Tang, W.W., Dietmann, S., Irie, N., Leitch, H.G., Floros, V.I., Bradshaw, C.R., Hackett, J.A., Chinnery, P.F., and Surani, M.A. (2015). A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. *Cell* 161, 1453–1467.
- Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352, 189–196.
- Tomizawa, S., and Sasaki, H. (2012). Genomic imprinting and its relevance to congenital disease, infertility, molar pregnancy and induced pluripotent stem cell. *J. Hum. Genet.* 57, 84–91.
- Tomizawa, S., Nowacka-Woszuk, J., and Kelsey, G. (2012). DNA methylation establishment during oocyte growth: mechanisms and significance. *Int. J. Dev. Biol.* 56, 867–875.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111.
- Vidal, V.P., Chaboissier, M.-C., de Rooij, D.G., and Schedl, A. (2001). Sox9 induces testis development in XX transgenic mice. *Nat. Genet.* 28, 216–217.
- von Meyenn, F., and Reik, W. (2015). Forget the parents: Epigenetic reprogramming in human germ cells. *Cell* 161, 1248–1251.
- Wang, J., and Zhou, Q. (2016). Derivation and application of pluripotent stem cells for regenerative medicine. *Sci. China Life Sci.* 59, 576–583.
- Yang, Q., Kathiresan, S., Lin, J.-P., Tofler, G.H., and O'Donnell, C.J. (2007). Genome-wide association and linkage analyses of hemostatic factors and hematological phenotypes in the Framingham Heart Study. *BMC Med. Genet.* 8 (Suppl 1), S12.
- Zhang, H.M., Chen, H., Liu, W., Liu, H., Gong, J., Wang, H., and Guo, A.Y. (2012). AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Res.* 40, D144–D149.
- Zhang, H.M., Liu, T., Liu, C.J., Song, S., Zhang, X., Liu, W., Jia, H., Xue, Y., and Guo, A.Y. (2015). AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Res.* 43, D76–D81.
- Zhang, T., Oatley, J., Bardwell, V.J., and Zarkower, D. (2016). DMRT1 is required for mouse spermatogonial stem cell maintenance and replenishment. *PLoS Genet.* 12, e1006293.
- Zhou, Q., Wang, M., Yuan, Y., Wang, X., Fu, R., Wan, H., Xie, M., Liu, M., Guo, X., Zheng, Y., et al. (2016). Complete meiosis from embryonic stem cell-derived germ cells *in vitro*. *Cell Stem Cell* 18, 330–340.

**STAR★METHODS****KEY RESOURCES TABLE**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
PE Mouse Anti-Human CD117	BD PharMingen	#555714, clone YB5.B8, RRID: AB_396058
mouse anti-POU5F1	Santa Cruz	#sc-5279, RRID: AB_628051
rabbit anti-DDX4	abcam	#ab13840, RRID: AB_443012
mouse anti-DDX4	abcam	#ab27591, RRID: AB_11139638
rabbit anti-STR8	EMD Millipore	#ABN1656
anti-IL13RA2	Miltenyi Biotec	#130-104-670
mouse anti-human PECAM1	BD Biosciences	#558068
rabbit anti-Ki67	abcam	#ab15580, RRID: AB_443209
rabbit anti-WT1	abcam	#ab89901, RRID: AB_2043201
rabbit anti-HES1	abcam	#ab71559, RRID: AB_1209570
mouse anti-CASPASE-3	abcam	#ab2171, RRID: AB_302870
<b>Critical Commercial Assays</b>		
CD117 MicroBeads	Miltenyi Biotec	#130-091-332
Kapa Hyper Prep Kit	Kapa Biosystems	KK8505
<b>Deposited Data</b>		
RNA-seq data	This paper	GEO: GSE86146
<b>Software and Algorithms</b>		
FACSDiva software	BD Bioscience	N/A
LAS AF software	Leica	N/A
TopHat (v2.0.12)	Trapnell et al., 2009	N/A
HTSeq package	Anders et al., 2015	N/A
Seurat	Satija et al., 2015	<a href="http://satijalab.org/seurat/old-get-started/">http://satijalab.org/seurat/old-get-started/</a>
GO (DAVID)	Huang et al., 2009a, 2009b	<a href="https://david.ncifcrf.gov/home.jsp">https://david.ncifcrf.gov/home.jsp</a>
ARACNe-AP software	Lachmann et al., 2016	N/A
ssmarina package	Lefebvre et al., 2010	<a href="http://dx.doi.org/10.6084/m9.figshare.785718">http://dx.doi.org/10.6084/m9.figshare.785718</a>
GSEA	Subramanian et al., 2005	<a href="http://www.broadinstitute.org/gsea/index.jsp">http://www.broadinstitute.org/gsea/index.jsp</a>
KEGG pathways	Ogata et al., 1999	N/A
PCA, clustering, scatterplot3d	N/A	R software
Animal TFDB 2.0	Zhang et al., 2012; Zhang et al., 2015	<a href="http://bioinfo.life.hust.edu.cn/AnimalTFDB/">http://bioinfo.life.hust.edu.cn/AnimalTFDB/</a>
Cell-cycle gene list	Macosko et al., 2015; Tirosh et al., 2016	N/A

**CONTACT FOR REAGENT AND RESOURCE SHARING**

Further information and requests for reagents may be directed to, and will be fulfilled by, the Lead Contact, F.T. ([tangfuchou@pku.edu.cn](mailto:tangfuchou@pku.edu.cn)).

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**

The donors in this study were pregnant women who underwent medical termination of pregnancy (due to conditions such as cervical insufficiency, inevitable abortion, infection, eclampsia, etc.). All of the patients signed informed consents and voluntarily donated the fetal tissues for this study, as detailed in the informed consent forms (formal Chinese version) and summary of them (informal English translation) in **Methods S1**. The experiments performed in this study were approved by the Reproductive Study Ethics Committee of Peking University Third Hospital (2012SZ013).

## COLLECTION OF HUMAN FETAL SAMPLES

All human embryos from 4 to 26W of gestation used in this study were obtained with informed consent from the donors. We collected male and female fetal gonads with intact morphology and reasonable cell viability from embryos spanning 4W to 26W. The stages of human embryos in this study were calculated from the estimated fertilization time, rather than last menstruation bleeding time. The clinicians who obtained the samples made this determination.

The total numbers of both male and female embryos were 10~20, respectively. The number of replicates for each developmental stage was no more than three. In total, we collected 12 male embryos and 17 female embryos in this study. For female embryos, 5W, 18W, 20W, 23W and 24W samples had two biological replicates, and 7W, 8W, 10W, 11W, 12W 14W and 26W had one biological replicate. For male embryos, 21W samples had three biological replicates, 10W, 19W samples had two biological replicates, and 4W, 9W, 12W, 20W and 25W had one biological replicate.

## METHOD DETAILS

### **Isolation of Human FGCs and Gonadal Somatic Cells by MACS and FACS**

For 7W to 12W human embryos, the fetal gonads were dissected in L15 medium (plus 10% FBS). After digestion of the gonads by 500 µL Accutase Cell Detachment Solution (Millipore #SCR005) for 5 min at 37°C, the tissue was pipetted up and down for 2 min, and the cell suspension was filtered through 50-µm Pre-Separation Filters (Miltenyi Biotec #130-041-407). Then, 500 µL of L15 medium (plus 10% FBS) was added to stop digestion, the cell suspension was centrifuged at 300 × g for 6 min at 4°C, and the cells were resuspended in 300 µL of L15 medium (plus 10% FBS). 100 µL of FcR Blocking Reagent, 100 µL of CD117 MicroBeads (Miltenyi Biotec #130-091-332) and 5 µL (1:100 dilution) of PE Mouse Anti-Human CD117 antibody (BD PharMingen #555714, clone YB5.B8) were added to the 300 µL of cell suspension and were mixed well. After incubation for 30 min in a 4°C refrigerator, the cell suspension was centrifuged at 300 × g for 6 min, the supernatant was aspirated completely, and the cells were resuspended in 500 µL of L15 medium (plus 10% FBS). Through magnetic separation (Miltenyi Biotec) and BD FACSAria (Special Order Research Product) sorting, we collected both the CD117-positive fraction (FGCs) and the CD117-negative fraction (gonadal somatic cells and FGCs).

For 14W to 26W human embryos, the gonads were dissected in L15 medium (plus 10% FBS), and the tissues were digested in Accutase for 5–15 min at 37°C. In addition to isolating the CD117-positive FGCs as described above, we also collected the larger CD117-negative cell fraction by FACS.

For 4W and 5W human embryos, the AGM region was dissected in L15 medium (plus 10% FBS), and the tissue was digested in 500 µL of Accutase for 2 min at 37°C. To minimize cell loss, we performed only magnetic enrichment (Miltenyi Biotec) by CD117 and did not perform FACS.

### **Single-cell RNA-Seq Library Preparation and Sequencing**

A modified Smart-seq2 protocol was applied for single-cell RNA-seq (Kannanayakal and Eberwine, 2005; Picelli et al., 2013, 2014; Spaethling et al., 2016). Briefly, after MACS and FACS purification, a single FGC or gonadal somatic cell was placed into the lysis buffer by mouth pipette. The reverse transcription reaction was performed with 25 nt oligo(dT) primer anchored with an 8 nt cell-specific barcode (Table S2) and 8 nt unique molecular identifiers (UMIs) (Hashimshony et al., 2012; Islam et al., 2012, 2014; Klein et al., 2015). After the first-strand synthesis, the second-strand cDNAs were synthesized, and the cDNAs were amplified by 17 cycles of PCR. The amplified cDNAs of the single cells were then pooled together for the following steps. Biotinylated pre-indexed primers were used to further amplify the PCR product by an additional 4 cycles of PCR to introduce biotin tags to the 3' ends of the amplified cDNAs. Approximately 300 ng cDNA was sheared to approximately 300 bp by Covaris S2, and the 3' terminal of the cDNA was captured by Dynabeads® MyOne Streptavidin C1 beads (Thermo Fisher). The RNA-seq library was constructed using a Kapa Hyper Prep Kit (Kapa Biosystems) and subjected to 150 bp paired-end sequencing on an Illumina HiSeq 4000 platform (sequenced by Novogene).

### **Processing of Single-cell RNA-seq Data**

Raw reads were first separated by specific cell barcode information attached in read 2 of the pair-ended reads. UMI information was aligned to the corresponding read 1, which was then trimmed to remove the template switch oligo (TSO) sequence and polyA tail sequence. Subsequently, we also removed reads with adaptor contaminants and low-quality bases (N > 10%). Next, the clean reads were aligned to the hg19 human transcriptome (UCSC) using TopHat (version 2.0.12) (Trapnell et al., 2009). Uniquely mapped reads were then counted using htseq-count in the HTSeq package and grouped by the cell-specific barcodes (Anders et al., 2014). Based on the UMI information, duplicated transcripts with the same UMI sequences of each gene were removed. Finally, within a given individual cell, distinct UMIs of each gene were counted as the transcript copy number of that gene.

For all 2,579 sequenced single cells, we quantified the numbers of genes and transcripts in each cell. Cells with either fewer than 2,000 genes or fewer than 100,000 transcripts detected were filtered out. In total, 2,210 cells passed the filter standards. The expression levels were normalized by log<sub>2</sub> (TPM/10+1), where TPM (transcripts-per-million) was calculated as the number of UMIs of each

gene divided by all UMIs of a given cell, then multiplying by 1,000,000. We divided the TPM values by 10 because the UMI number of most of our single cell samples did not reach the order of 1,000,000 transcripts. Thus, we could avoid counting each transcript several times.

### Down-sampling Analysis

To check whether the sequencing depth is enough for the subsequent analyses, we selected an embryo (male, 20 week, 88 single cells sequenced) to down-sample the raw data to 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% of its original data. Then we analyzed the percentages of the numbers of detected genes in these down-sampling data compared to the numbers of detected genes in the original data, which was exhibited in boxplot as shown in Figure S1A.

### Nonlinear Dimensional Reduction (tSNE)

The Seurat method was applied to analyze our single-cell data (2,210 cells) on log2 (TPM/10+1) expression values (Satija et al., 2015). Specifically, only the genes with an expression level greater than 1 that were expressed in at least 10 single cells were considered, whereas single cells with less than 2,000 genes expressed were excluded, leaving 2,187 cells and 18,919 genes for the downstream analysis. Three hundred sixty highly variable genes with average expression more than 2 and a dispersion greater than 2 were used as inputs for initial principal component analysis (PCA). We then performed a jackstraw analysis to select principal components (PCs) that hold the most differences to separate the cells using the function jackstraw with 1,000 replicates (Chung and Storey, 2015). PCs 1–16 were selected because their *P*-values were below 1e-5. tSNE was performed through the run\_tsne function (dims.use = 1:16, max\_iter = 2,000). To cluster cells, we used the DBclust\_dimension function with parameters G.use = 4.2. After this step, 20 single cells were unassigned to any groups and treated as outliers, leaving 2,167 single cells used for subsequent analyses.

To complement the tSNE clustering and obtain more accurate clusters, we also separately clustered the four types of cells—namely, female FGCs, female somatic cells, male FGCs and male somatic cells—with hierarchical clustering using the R software program. Briefly, PCA analysis was performed based on highly variable genes selected by the above methods. Subsequently, the top 20 or 30 genes within the first three or four principal components were chosen. We then performed hierarchical clustering on these identified genes through Euclidean or correlation distance metrics. For the subsequent analysis, we considered only the cells that were consistent with the two clustering methods. Somatic cells with a high expression of NANOG, POU5F1, DAZL or DDX4 were also removed because these cells might be contaminated by FGC fragments.

### Identification of Erythrocytes

Cells in cluster 12 showed specific expression of erythrocyte-specific genes, such as HBA1, HBA2, HBB, HBG1, HBG2 and HBM, implying that they were erythrocytes (Yang et al., 2007). In addition, the cells in clusters 13 and 14 express CD4, CD45, CD300A, CD14, CD68 and CD204 and were macrophages. The cells in clusters 15 and 16 expressed CD4, CD45, CD300A, and CD132 and were early T cells (Figure S1B; Tirosh et al., 2016). Cells in cluster 17 expressed not only germ cell markers, but also hemocyte markers, such as HBA1, HBA2, HBB, HBG1, HBG2 and HBM, and essentially all of them were from one 18W male embryo, which was probably of low quality (Figure S1B). We supposed that these cells were FGCs that were contaminated by fragments of hemocytes and so we excluded them from the subsequent analyses.

### Identification of Differentially Expressed Genes

Unique cluster-specific marker genes were identified by running the Seurat find\_all\_markers function (thresh.test = 1, test.use = “roc”) on log2 (TPM/10+1) expression values. For a certain gene, the roc test would return a value ranging from 0 to 1, representing the ‘classification power’, where 0 for ‘random’ and 1 for ‘perfect’, and genes with fold-change  $\geq 2$  and power  $\geq 0.4$  were selected. To identify differentially expressed genes between two given clusters, the Seurat function find.markers with parameters thresh.= 1, test.use = “roc” was used, and genes with fold-change  $\geq 2$  or  $\leq 0.5$  and power  $\geq 0.4$  were selected. Heatmaps and Bar-plots of select genes were plotted using R. Violin plots were plotted using Seurat. Gene ontology analysis was performed using DAVID (<https://david.ncifcrf.gov/home.jsp>) (Huang et al., 2009a, 2009b).

### Cell-Cycle Analysis

In our cell-cycle analysis, we used cell-cycle-related genes, including a previously defined core set of 43 G1/S and 54 G2/M genes (Table S2; Macosko et al., 2015; Tirosh et al., 2016).

### Inference of Master Regulators with ARACNe and MARINA

The ARACNe is widely used to accurately reconstruct gene regulatory networks (Basso et al., 2005; Margolin et al., 2006). To infer the transcription factor regulatory network of this study, we used all 1,568 human transcription factors of Animal TFDB 2.0 (<http://bioinfo.life.hust.edu.cn/AnimalTFDB/>; Zhang et al., 2012, 2015). We first performed regulatory network analysis for female and male embryos separately, each with corresponding expression data including FGCs and somatic cells, using ARACNe-AP software (Lachmann et al., 2016). Second, we performed master regulator analysis using the ssmarina package (available from <http://dx.doi.org/10.6084/m9.figshare.785718>), a modification of the MARINA algorithm (Lefebvre et al., 2010). Enrichment of the predicted targets was assessed by comparing the gene expression between two clusters. Regulators with FDR-corrected *P*-values below 0.01 were inferred as candidate master regulators between two given clusters.

### GSEA Analysis

We used Gene Set Enrichment Analysis (GSEA) to identify gene sets that show significant differences between two given clusters (<http://www.broadinstitute.org/gsea/index.jsp>; Subramanian et al., 2005). GSEA was performed in the KEGG pathways (Ogata et al., 1999). We chose the ‘Signal2Noise’ ranking metric, a threshold nominal  $P$ -value  $< 0.05$ .

### PCA-based Developmental Pseudotime and Characterization of the Transition State of FGCs

3D PCA plots for both female and male FGCs were generated in R scatterplot3d package using high variable genes identified by Seurat. The developmental phases for both female and male FGCs were well unfolded in the 3D plots (as shown in Figures 3A and 4A), thus we used the PCA-based routes as developmental pseudotime (female: PC3 for FGC#1, PC1 for FGC#2, PC2 for FGC#3 and #4; male: PC3 or FGC#1 and #2, PC1 or PC1 for FGC#3). And for Sertoli and granulosa cells, we also used PCA-based route as the developmental pseudotime. Finally, we plotted the trends of the BMP signaling pathway along the pseudotime to show the somatic-germline interaction in Figure S7A.

We performed PCA analysis for two consecutive developmental stages of the FGCs and found that the first PCA component was associated with the developmental transition. Thus, we regarded the positions along PC1 as the developmental route, which was then plotted using R package beeswarm. As shown in Figure S7A, for each analysis, we selected three groups: top 20 positive and top 20 negative cells along PC1 were treated as two original groups (namely 1 and 3 in Figure S7), cells from the intermediate position along PC1 were chosen as the transition state group (namely 2 in Figure S7). Differential gene expression and gene ontology analysis were performed between 1 and 2, as well as 2 and 3, and the full results were shown in Table S5.

### Reactivation of X Chromosome in the Female FGCs

For the FGC and somatic cells at each developmental stage, we calculated and plotted the percentage of total expression levels of all genes on X chromosome to those of all genes in the genome, except for the developmental stages in which the total cell numbers were less than 5.

### Immunofluorescence Staining of the Frozen Sections

The gonads were fixed in 4% cold fresh paraformaldehyde for 4 hr and embedded with Tissue-Tek® O.C.T. Compound (Sakura #4583). 16  $\mu\text{m}$  cryosections were prepared and immunofluorescence staining assays were performed as described previously (Guo et al., 2015). Images were obtained using a confocal microscope (Leica TCS SP8, Leica Microsystems, Germany) equipped with a HCX PL APO 63  $\times$  1.40 NA oil objective by LAS AF software (Leica).

### QUANTIFICATION AND STATISTICAL ANALYSIS

In Figure S4A, quantification data are presented as mean  $\pm$  SEM, error bar indicates  $\pm$  95% confidence intervals ( $\pm 1.96 \text{ SEM}$ ). To do tSNE clustering, a jackstraw analysis was performed to select principal components (PCs) whose  $P$ -values were below 1e-5 using the function jackstraw with 1,000 replicates in Seurat. The roc test was used to identify marker genes in Seurat, and genes with fold-change  $\geq 2$  or  $\leq 0.5$  and power  $\geq 0.4$  were selected. To infer the Master Regulators with ARACNe and MARINa, the FDR-corrected  $P$ -values below 0.01 was set as threshold. In GSEA analysis, we chose a threshold nominal  $P$ -value  $< 0.05$ .

### DATA AND SOFTWARE AVAILABILITY

The accession number for all sequencing data reported in this paper is GEO: GSE86146. The expression pattern of our single cell RNA-seq data is accessible at <http://www.singlecell.pku.edu.cn/GermCell>. The expression value on log2 (TPM/10+1) data, UMI counts data and supplemental discussion are available at <http://github.com/zorrodong/germcell>.