

Revealing cellular and molecular transitions in neonatal germ cell differentiation using single cell RNA sequencing

Jinyue Liao^{1,2}, Shuk Han Ng^{1,2}, Alfred Chun Luk^{1,2}, Hoi Ching Suen^{1,2}, Yan Qian^{1,2}, Annie Wing Tung Lee^{1,2}, Jiajie Tu^{1,2}, Jacqueline Chak Lam Fung^{1,2}, Nelson Leung Sang Tang³, Bo Feng^{1,4}, Wai Yee Chan^{1,2,4,5,6}, Pierre Fouchet^{7,8,9}, Robin M. Hobbs¹⁰ and Tin Lap Lee^{1,2,4,5,6,*}

ABSTRACT

Neonatal germ cell development provides the foundation of spermatogenesis. However, a systematic understanding of this process is still limited. To resolve cellular and molecular heterogeneity in this process, we profiled single cell transcriptomes of undifferentiated germ cells from neonatal mouse testes and employed unbiased clustering and pseudotime ordering analysis to assign cells to distinct cell states in the developmental continuum. We defined the unique transcriptional programs underlying migratory capacity, resting cellular states and apoptosis regulation in transitional gonocytes. We also identified a subpopulation of primitive spermatogonia marked by CD87 (plasminogen activator, urokinase receptor), which exhibited a higher level of self-renewal gene expression and migration potential. We further revealed a differentiation-primed state within the undifferentiated compartment, in which elevated *Oct4* expression correlates with lower expression of self-renewal pathway factors, higher *Rarg* expression, and enhanced retinoic acid responsiveness. Lastly, a knockdown experiment revealed the role of *Oct4* in the regulation of gene expression related to the MAPK pathway and cell adhesion, which may contribute to stem cell differentiation. Our study thus provides novel insights into cellular and molecular regulation during early germ cell development.

KEY WORDS: Single cell, Transcriptomes, Spermatogonia, Gonocytes, Differentiation

INTRODUCTION

In mammals, male fertility is sustained by male germline stem cells (GSCs) known as spermatogonial stem cells (SSCs) that either


self-renew or differentiate to progenitors to initiate spermatogenesis (Oatley and Brinster, 2008). The current model suggests that the initial pool of SSCs is derived from the transition of postnatal gonocytes arising from primordial germ cells (PGCs) during embryogenesis, which enter a quiescent state after embryonic day 13.5 (Kluin and Derooij, 1981). Shortly after birth, gonocytes resume mitotic activity and begin migrating to the basal lamina of testicular cords. Once they reside at the basement membrane, they become spermatogonia, including self-renewing SSCs (Culty, 2009; Manku and Culty, 2015; Saitou and Yamaji, 2012). Violating this cell fate decision process can lead to serious clinical consequences (Kamisawa et al., 2012; Rajpert-de Meyts and Høie-Hansen, 2007). SSC self-renewal and differentiation must also be balanced to maintain normal spermatogenesis and avoid tumorigenesis (Ishii et al., 2012).

Although much attention has been dedicated to the characterization of adult SSCs, the molecular relationship between the different germ cell populations at the neonatal stage has not been sufficiently addressed. The main challenge is that the germ cells in neonatal testes are not developmentally synchronized, resulting in the co-existence of germ cells at different developmental stages (Nagano et al., 2000). Moreover, there is significant heterogeneity regarding marker expression among the germ cell populations, where different markers are restricted to distinct or overlapping subsets of cells (Hermann et al., 2015; Nakagawa et al., 2007; Suzuki et al., 2009; Yoshida et al., 2006). Recently, another layer of heterogeneity in marker expression also received substantial attention: SSC markers can display a gradient of expression level and define different cellular states. For example, adult SSCs are present in populations of cells expressing a high level of ID4, whereas progenitor populations express relatively lower levels of ID4 (Helsel et al., 2017; Komai et al., 2014). To understand this further, it will be important to clarify whether expression gradients can be applied to other markers. OCT4 (POU5F1) is of particular relevance as it can exert a dose-dependent effect in mouse embryonic stem cells, whereby different expression levels of OCT4 define differentiation, de-differentiation or self-renewal (Zeineddine et al., 2014).

Single cell RNA sequencing (RNA-Seq) technology is increasingly used to deconstruct cellular subpopulations and transcriptional heterogeneity. Here, we profiled single cell transcriptomes of germ cells from mouse testes on postnatal day (PND) 5.5. We provide a snapshot of the cellular composition of undifferentiated germ cells, reveal global patterns of gene expression heterogeneity, and present potential novel transcriptional regulators during differentiation. Further analysis reveals a subpopulation within the most primitive spermatogonia labeled by CD87 (also known as uPAR and PLAU), as well as a differentiation state marked by higher *Oct4* expression and the potential role of OCT4 in regulation of the MAPK and cell adhesion pathway.

¹Developmental and Regenerative Biology Program, School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China. ²The Chinese University of Hong Kong – Shandong University (CUHK-SDU) Joint Laboratory on Reproductive Genetics, Shatin, Hong Kong SAR, China. ³Department of Chemical Pathology, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China. ⁴Key Laboratory for Regenerative Medicine, Ministry of Education, School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China. ⁵Joint CUHK-UoS (University of Southampton) Joint Laboratories for Stem Cells and Regenerative Medicine, School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China. ⁶CUHK-BGI Innovation Institute of Trans-omics Hong Kong, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China. ⁷CEA DRF IBFJ IRCM, Laboratoire des Cellules Souches Germinales, 92265 Fontenay-aux-Roses, France. ⁸Université Paris Diderot, Sorbonne Paris Cité, INSERM, UMR 967, 92265 Fontenay-aux-Roses, France. ⁹Université Paris Sud, INSERM, UMR 967, 92265 Fontenay-aux-Roses, France. ¹⁰Development and Stem Cells Program, Monash Biomedicine Discovery Institute and Department of Anatomy and Developmental Biology, Monash University, Melbourne, Victoria 3800, Australia.

*Author for correspondence (leel@cuhk.edu.hk)

 T.L.L., 0000-0002-6654-0988

RESULTS**Single cell transcriptional profiling of Oct4-GFP⁺/KIT⁻ cells from mouse testes at PND5.5**

To resolve cellular heterogeneity in early germ cell development, we focused on PND5.5 when germ cells consist of gonocytes and primitive spermatogonia, including SSCs and differentiating spermatogonia. We found that a portion of Oct4-GFP⁺ cells resided within the center of testicular cords or were positioned adjacent to the basement membrane with protrusions towards the periphery, which confirmed the existence of transitioning gonocytes yet to finish the migratory phase at PND5.5 (Fig. 1). Oct4-GFP⁺ cells

exhibit significant overlap with PLZF (ZBTB16), a marker for the undifferentiated state (Fig. 1A,D), suggesting that Oct4-GFP⁺ cells indeed label most, if not all, germ cells at this stage. Co-expression of Oct4-GFP with the self-renewal factor GFRA1 was also observed in a subset of Oct4-GFP⁺ cells (Fig. 1B,D). In addition, many Oct4-GFP⁺ cells begin to express the differentiation marker KIT, which suggests that PND5.5 testis contains differentiating spermatogonia (Fig. 1C). Altogether, Oct4-GFP⁺-labeled germ cells at this single time point should represent a mixture of germ cells at different cellular states and single cell transcriptome profiling of these cells would allow reconstruction

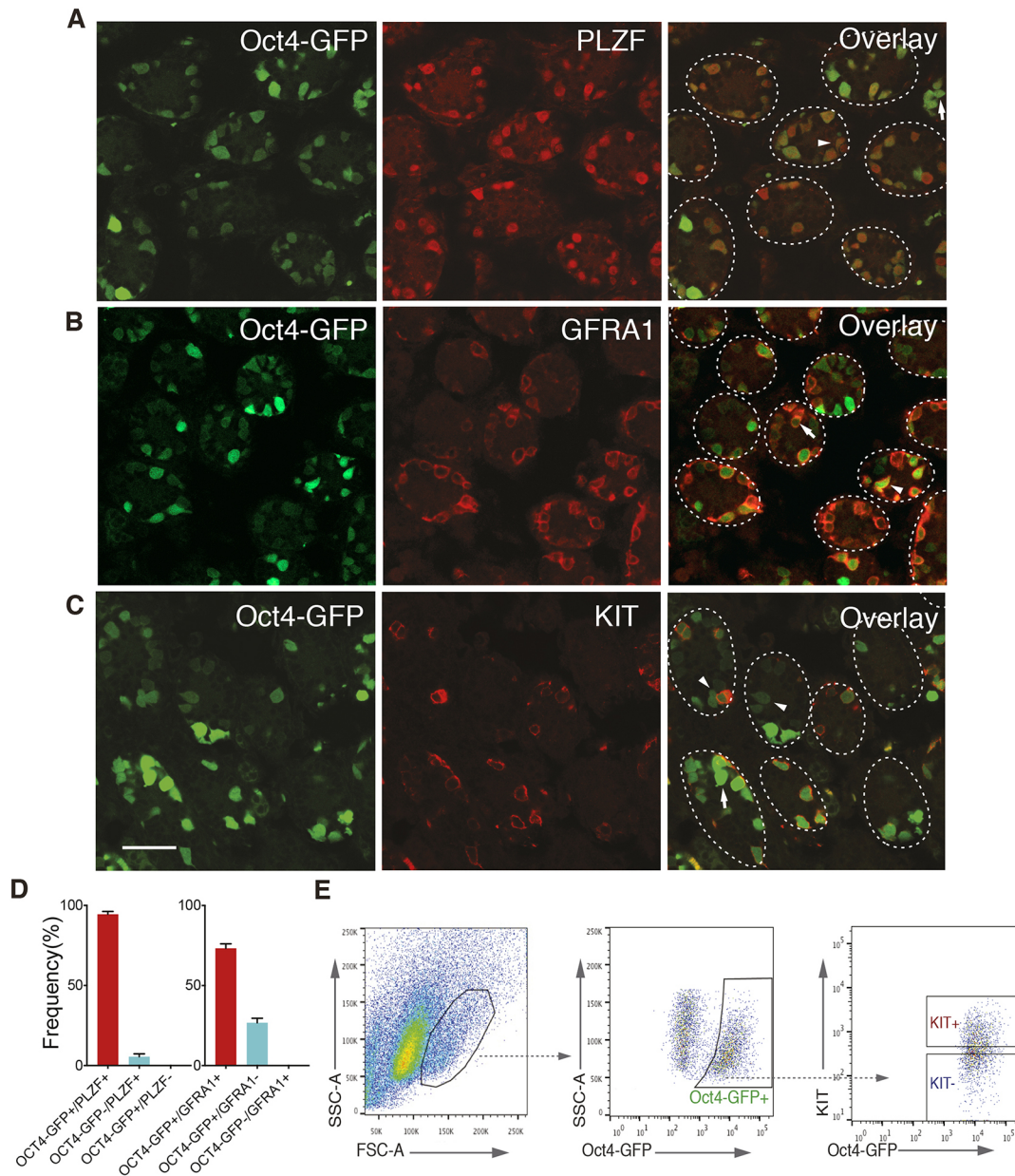


Fig. 1. Characterization and isolation of Oct4-GFP⁺ germ cells from mouse testes at PND5.5. (A-C) Testicular sections from PND5.5 Oct4-GFP transgenic mice showing immunofluorescent staining of PLZF (A), GFRA1 (B) and KIT (C). Oct4-GFP expression was observed in cells that were co-stained with the undifferentiated spermatogonial markers GFRA1 and PLZF (A,B). Oct4-GFP-expressing cells can be classified as KIT⁻ or KIT⁺ (C). Transitional gonocytes reside in the center of the testicular cord (arrows) or near the basement membrane with cytosolic projection towards the periphery (arrowheads). Representative of at least three independent experiments. (D) Quantification of the overlap of Oct4-GFP with PLZF or GFRA1 in testis section immunostaining as shown in A and B. Results are expressed as mean ± s.e.m. of four testis sections. (E) Sorting strategy for isolation of undifferentiated (Oct4-GFP⁺/KIT⁻) germ cells at PND5.5. Scale bar: 50 μm.

of dynamic gonocyte-to-spermatogonia transition (GST) and subsequent differentiation *in vivo* (Fig. S1A).

To maximize the representation of the most undifferentiated germ cells, we isolated Oct4-GFP⁺/KIT⁻ cells from PND5.5 testis by fluorescence-activated cell sorting (FACS) (Fig. S1B-F). We used isotype control to confirm specific sorting (Fig. S1B). qRT-PCR analysis of *Kit* mRNA expression also showed a more than 10-fold increase in the KIT⁺ sorting fraction, which confirms the specific enrichment of Oct4-GFP⁺/KIT⁻ cells (Fig. S1F). The sorted Oct4-GFP⁺/KIT⁻ cells were subjected to Fluidigm C1 single cell RNA-Seq (Fig. 1E, Fig. S2, Tables S1 and S2). The germ cell markers *Dazl* and *Ddx4* were abundantly expressed in all cells, indicating no contamination from somatic cells (Fig. S3A). A significant portion of germ cell-related genes, such as *Bcl6b*, *Egr2*, *Sohlh1* and *Csf3r*, showed substantial variability, which revealed the significant gene expression heterogeneity in early male germ cell populations (Fig. S3A,B).

Single cell RNA-Seq analysis recapitulates germ cell subpopulations and cellular dynamics

We next sought to interrogate the heterogeneous cellular composition. Principal component analysis (PCA) was used to identify the top 500 most variable genes that accounted for the majority of the variance in expression (Table S3) (Pollen et al., 2014). Unbiased clustering with this reduced gene list revealed three clusters of cells (Fig. 2A): one main cluster containing 56 cells, and two smaller clusters (I and IV: nine cells and six cells, respectively). It is notable that Cluster IV cells lack expression of several self-renewal genes (e.g. *Gfra1*, *Etv5* and *Ccnd2*) (Fig. 2A), which were named as differentiation-primed spermatogonia. The aforementioned characterization (Fig. 1) implies that the majority of cells captured at PND5.5 should be undifferentiated spermatogonia and may be linked to the main cluster. This cluster can be further divided into two subclusters (Cluster II and Cluster III) based on the dendrogram (Fig. 2A). To reconstruct the transitional dynamics and ontogeny relationship of the cell subpopulations, we used the single cell analysis tool Monocle to order cells along a hypothetical timeline (pseudotime) of development (Trapnell et al., 2014). The resulting trajectory path began with Cluster I, passing through undifferentiated spermatogonia (Cluster II), then Cluster III, and ending at Cluster IV (Fig. 2B, upper panel). This indicated that Cluster II might sit at an earlier developmental stage than Cluster III in the pseudotime. Consistently, our PCA results implied that the first component represents the developmental transition from less differentiated cells (Cluster II) to the differentiation-primed spermatogonia (Cluster IV) on the *x*-axis (Fig. 2C). The PC scores of the top 500 genes revealed SSC self-renewal genes (*Etv5* and *Ccnd2*) were expressed at higher levels in Cluster II cells, whereas differentiation regulators (*Sohlh1* and *Trib3*) were expressed at higher levels in Cluster III cells (Fig. 2D). The majority of cells with high *Id4* expression fell into Cluster II {15 out of 23 cells, log₂[transcripts per million (TPM)]>10}, consistent with the observation that cells with a high level of *Id4* represent true SSCs (Helsel et al., 2017). *ErbB3* and *Pax7* were also preferentially expressed in Cluster II (Fig. 2E). Together with the Monocle ordering result, these data indicate that Cluster II cells are likely to be the more stem cell-like spermatogonia (hereafter referred to as ‘primitive stem cells’) and Cluster III cells are more progenitor-like spermatogonia (hereafter referred to as ‘progenitors’). Cluster I corresponds to gonocytes, based on their earliest pseudotime stage and the fact that the existence of a gonocyte population should be reflected in single cell transcriptomes of the captured cells (Fig. 2B,

upper panel). It is notable that *Cd81*, which mediates migratory gonocyte movement (Tres and Kierszenbaum, 2005), showed higher expression in Cluster I (mean TPM: Cluster I=62, cluster II=18, Cluster III=22, Cluster IV=10) (Fig. 2E). Monocle analysis also revealed three major cell states, which corresponded to GST (State 1), SSC self-renewal (State 2) and differentiation priming (State 3) (Fig. 2B, lower panel).

To validate the single cell RNA-Seq result, we sorted the second batch of PND5.5 Oct4-GFP⁺/KIT⁻ germ cells and measured the expression of a panel of genes that were preferentially expressed in each subpopulation using single cell RT-qPCR (Fig. S4A-C, Table S4). In line with the single cell RNA-Seq results, four clusters were recovered from hierarchical clustering (Fig. S4D) and the PCA plot showed a similar structure with the transition from stem cell subpopulation (green) to progenitor (red) and then to differentiation-primed cells (orange) (Fig. S4E). Similarly, the gene expression data generally recapitulated patterns we identified in the RNA-Seq experiment (Fig. S4F).

Unique transcriptional signature in the gonocyte population

To investigate the transcriptional signature of the gonocyte population, we first focused on the cells in GST state (State 1) and compared gene expression between the first half (nine cells from Cluster I) and the second half (nine cells from Cluster II) in the trajectory (Fig. 3A). Top-ranked differentially expressed genes identified were postulated to regulate cell migration. For example, this analysis revealed *Cd81* as one of the top genes (Table S5). Cluster I also expressed higher *Crk*, which encodes the adapter protein of PDGFR and might be involved in PDGF-mediated gonocyte survival and migration (Fig. 3A) (Basciani et al., 2008). Moreover, integrin-binding associated genes, such as *Nf2*, *Itgb1*, *Utrn*, *Mfge8* and *P4hb*, are also upregulated in the gonocytes (Fig. 3A). In contrast, the gonocyte population expressed a lower level of *Tnfrsf1* (also known as *Bacurd2*) (Fig. 2D, Table S5), which influences the directional cellular migration through the regulation of RHOA protein turnover (Sailland et al., 2014). Interestingly, top-ranked genes enriched in migratory gonocytes included *Trpm7*, indicating its possible role in gonocyte migration through calcium-mediated cell movement (Chen et al., 2010).

Gene set enrichment analysis (GSEA) further revealed that gene sets associated with cell cycle and protein synthesis (aminoacyl-tRNA biosynthesis) were under-represented in the gonocyte population (Fig. 3B). Interestingly, gonocytes were shown to express lower levels of basal transcriptional factors (TFIIB) such as *Taf4b*, which encodes a chromatin-associated factor and has been demonstrated to regulate early germ stem cell specification (Falender et al., 2005). In line with the presence of 12.6% (9/71) cells (Cluster I) within the total cells captured in the single cell RNA-seq experiment, we observed a similar fraction that was negative for Ki-67 (*Mki67*) expression (Fig. 3C). Immunostaining also revealed that the centrally located cells contained more Ki-67-negative cells than peripherally located cells (Fig. 3D,E). These results implied a resting phase of these gonocytes prior to their transition to spermatogonia.

Lastly, gene ontology (GO) analysis identified a significant upregulation of genes associated with the death receptor pathway in Cluster I cells, such as *Fadd*, *Tnfrsf1b*, *Icam1*, *Madd*, *Stx4a* and *Skil*. We also found elevated expression of genes involved in intrinsic apoptosis signaling (*Itm2b*) (Fig. 3F, Table S5) (Fleischer et al., 2002). Genes associated with the DNA damage/repair pathway were shown to be repressed in gonocytes, including *Ercc1*, *Trp53* and *Bre* (Fig. 3F, Table S5). These results are in line with the notion that

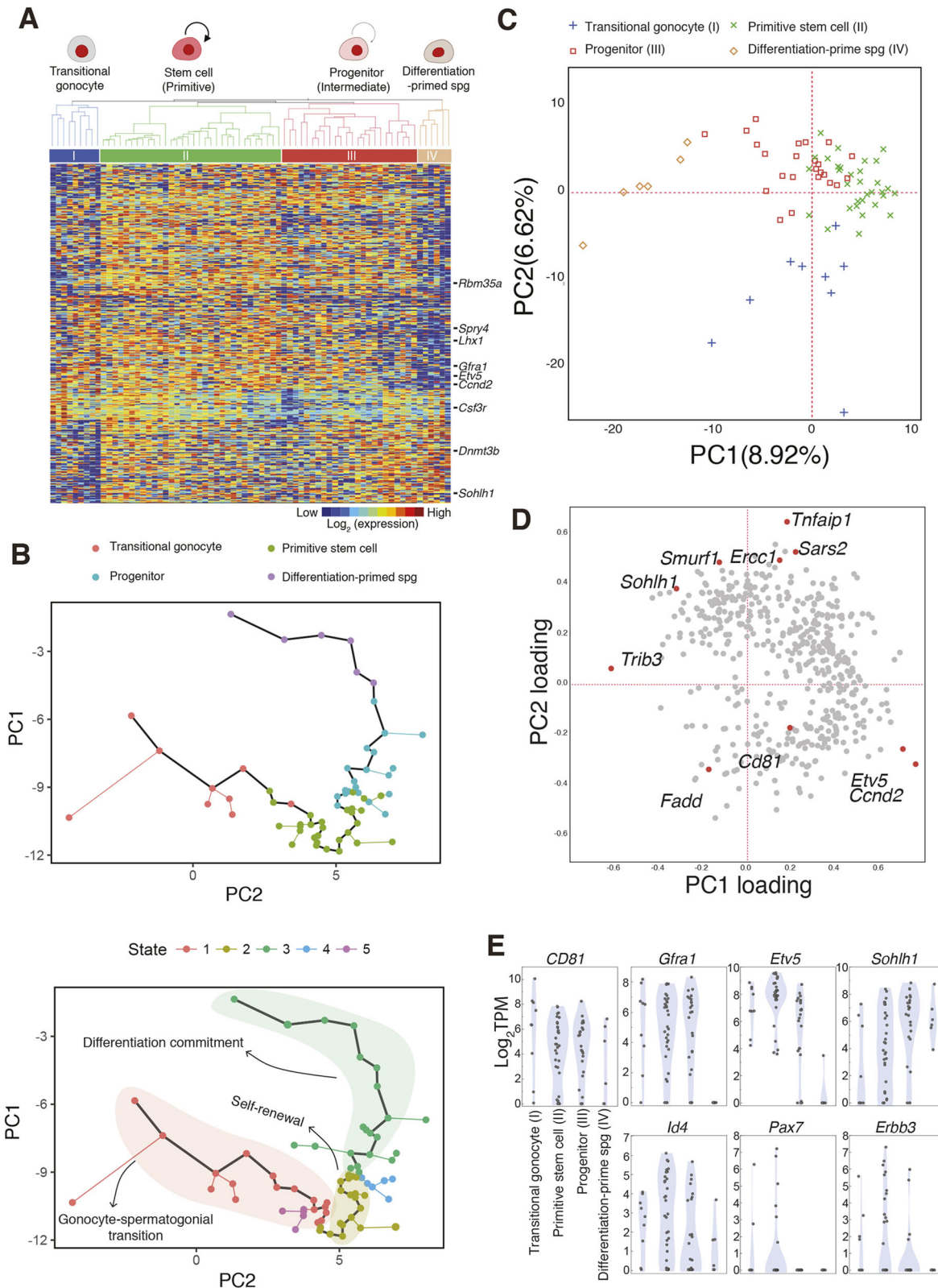


Fig. 2. Single cell transcriptome analysis reveals subpopulations and cellular transitions. (A) Heatmap showing the expression of the top 500 most variable genes. Germ cell subpopulation identities (Clusters I-IV) are resolved as described in the main text. Dendrogram resulting from hierarchical clustering with the top 500 most variable genes shows the relative distances between the cell clusters. (B) Reconstructing continuous trajectory and assigning pseudotime for germ cell development using Monocle. Colors are based on the group assignment from hierarchical clustering. Three main developmental states were revealed by Monocle (lower panel). (C) Principal component (PC) projection of individual cells based on the gene expression profiles of the top 500 most variable genes. Each point represents a single cell and is colored according to the clustering result as shown in A. (D) PC projections of the top most 500 variable genes, showing the contribution of each gene to the first two PCs shown in C. Examples of genes associated with germ cell development are marked in red. (E) Violin plots illustrating the distribution of representative genes of subpopulations (top) as well as adult A_s marker genes (bottom). spg, spermatogonia.

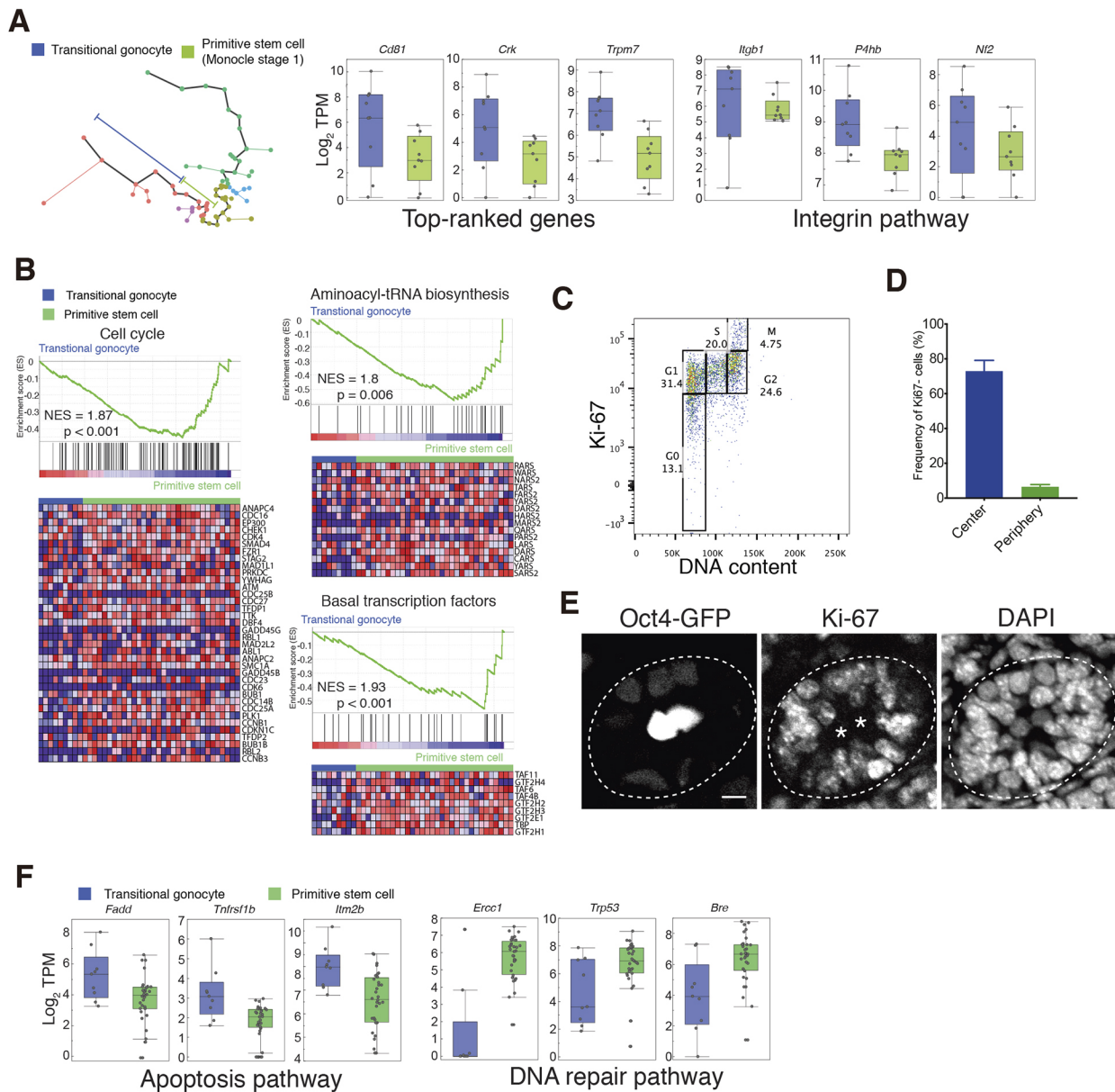


Fig. 3. Characterization of gene expression and signaling pathways in the gonocyte-spermatogonial transition. (A) Box plots for representative genes identified as differentially expressed ($FDR < 0.05$) between cells in the first half and second half of stage 1 as indicated in Fig. 2B. Median (centre line), first and third quartiles (bounds of box) and 1.5 interquartile range (whiskers) and individual data points (dots) are indicated. (B) Signaling pathways enriched in primitive stem cells compared with transitional gonocytes by GSEA/KEGG analysis. The values indicated on individual plots are the normalized enrichment score (NES) and P -value of the enrichment (permutation analysis). (C) Exemplary FACS plots of Ki-67 and Hoechst (DNA content) of Oct4-GFP⁺/KIT⁻ cells to assess resting and cycling cells. The data indicate the percentage of cells in each phase of the cell cycle. (D) The percentage of Ki-67-negative cells in the cells located at the tubular basement membrane and at the center of the seminiferous tubules ($n=3$). (E) Cross-sectional images of Ki-67 immunostaining of testis from Oct4-GFP mice. The asterisks denote Oct4-GFP⁺ cells that reside at the center of the tubules and lack Ki-67 expression. (F) Box plots for apoptosis-related genes identified as differentially expressed between transitional gonocytes (Cluster I) and stem cell spermatogonia (Cluster II). Median (centre line), first and third quartiles (bounds of box) and 1.5 interquartile range (whiskers) and individual data points (dots) are indicated. Scale bar: 10 μ m.

apoptosis plays an integral role in gonocyte cell fate determination (Wang et al., 1998).

Collectively, our analyses revealed transcriptional signatures that might be involved in cell fate decision and migration of gonocytes during GST.

Distinct transcriptional profiles of different cellular states in spermatogonia

We next investigated the underlying molecular features of the progression from the stem-like stage to the progenitor state of

spermatogonia. We identified 511 genes differentially expressed when comparing Cluster II with Cluster III (Table S6). GO analysis revealed significant enrichment for signatures associated with integrin complex, Rap1 signaling and lipid metabolism in primitive stem cell pools (Fig. S5A, Table S6), whereas TGF signaling genes (e.g. *Smad3* and *Creb1*) were over-represented in progenitor pool (Fig. S5B, Table S6). GSEA analysis further revealed that ‘JAK/STAT signaling’, ‘cell adhesion molecules (CAMs)’, ‘neuroactive ligand-receptor interaction’, and ‘RIG-I-like receptor signaling pathway’ were enriched in the stem cell cluster, whereas

the ribosome biogenesis pathway was upregulated in the progenitor subset (Fig. 4A, Fig. S6A).

The transition of SSCs into a differentiating state requires exposure to retinoic acid (RA), leading to the onset of KIT expression. The differentiation-primed population (Cluster IV) lacks both *Gfra1* and *Kit* expression (*Gfra1*⁻/*Kit*⁻), suggesting that these spermatogonia are exiting self-renewal but uncommitted to differentiation. In support of this, GSEA analysis showed a strong repression of glial cell line-derived neurotrophic factor (GDNF)-regulated genes in *Gfra1*⁻/*Kit*⁻ cells (Fig. 4B) (Oatley et al., 2006). Along with this, previously identified pathways involved in the maintenance of SSCs, such as the Ras signaling pathway, cell-cell communication and the gonadotropin-releasing hormone pathway, are also inhibited in *Gfra1*⁻/*Kit*⁻ cells (Fig. S7, Table S7) (Lee et al., 2009). Moreover, *Gfra1*⁻/*Kit*⁻ cells express an increased level of differentiation regulator *Rarg*, indicating that they begin to acquire responsiveness to RA (Fig. S4F, Table S7) (Ikami et al., 2015). To understand better the molecular signatures associated with the priming state, we searched for genes with increased expression in *Gfra1*⁻/*Kit*⁻ cells. We identified several transcriptional regulators, such as *Cfp1* (*Cxxc1*), *Ctcf*, *Ehmt2* (also known as *G9a*), *Tdrd5* and *Trib3* (Fig. 4C, Table S7).

Pseudotime and network analysis reveals molecular dynamics in spermatogonia differentiation

An advantage of sequencing single cells is that it allows reconstruction of the gene expression dynamics along the pseudotime. We applied Monocle analysis and identified 1223 genes with significant variation in expression levels along the differentiation trajectory ($P < 0.05$). These genes are clustered into four groups according to their expression trend (Fig. 4D). During the first stage in the trajectory, the expression of known self-renewal genes, such as *Ccnd2* and *Etv5*, as well as genes involved in MAPK/Erk1 and VEGF signaling began to decrease (Group 1 genes). Concurrently, a series of genes highly enriched for cell cycle and cell division was activated (Group 2 genes). Second, the cells passed through an intermediate phase and transiently expressed an elevated level of *Nanos3* and genes related to the DNA-repair pathway (*Bre* and *Trp53*) (Group 3 genes). Third, the priming phase featured complete repression of SSC self-renewal genes and elevation of differentiation genes, such as *Sohlh1* (Group 4 genes). Genes involved in the mTOR pathway (*Eif4ebp1*, *Pdcd4*, *Cycs* and *Prkab2*) were also upregulated, which is in line with enhanced mTOR pathway activity in adult spermatogonia differentiation (Fig. 4E) (Hobbs et al., 2010).

To identify novel regulatory networks associated with SSC self-renewal and differentiation, we applied the weighted gene co-expression network analysis (WGCNA) algorithm on transcriptomes of spermatogonia excluding gonocytes (Cluster I cells) (Langfelder and Horvath, 2008). In the search for modules that correlate with our differentiation model, we identified the pink module, which was enriched in primitive stem cells and decreased across pseudotime (Fig. S6B, left). GO terms of this module are associated with lipoprotein particle receptor activity, cell differentiation and development (Table S8). This unbiased approach highlighted key regulators as central to this network, including *Etv5*, *Ccnd2* and *Gfra1*, and GDNF signaling regulation (19 out of 107 genes are GDNF regulated), which confirmed the current knowledge of SSC self-renewal (He et al., 2009). The second module was mostly repressed in differentiation-primed cells and was enriched for genes involved in mitotic cell cycle and RNA binding (Fig. S6B, right, Table S8). Hub-gene-network analysis indicated *Tex13* and *Rbm35a* (*Esrp1*) are the most highly connected genes, which might exert potential functional roles in spermatogonia differentiation.

Identification of a unique subpopulation in undifferentiated spermatogonia marked by CD87 expression

We also identified putative markers or regulators that were specific to the earlier, potentially self-renewing spermatogonia (Cluster II) (Fig. S8A). We found CD87 (PLAUR), the uPA receptor (uPAR), particularly interesting as it has been implicated in hematopoietic cell migration and homing and has been used as a marker for a subset of hematopoietic stem/progenitor cells (HSPCs) (Tjwa et al., 2009). Immunostaining of CD87 protein confirmed that expression is restricted to a subset of Oct4-GFP⁺ cells (Fig. 5A). In addition, RNA-Seq analysis of FACS-sorted cells showed that CD87⁺ cells expressed a higher level of self-renewal genes (*Id4* and *Gfra1*) and a reduced level of early differentiating genes (*Crabp1*, *Akt1*) (Fig. 5B,C, Table S9). Importantly, in the CD87⁺ subpopulation, genes involved in cell migration were more abundantly expressed (*Mmp9*, *Hes1*, *Chd1*, *Lrp1* and *Procr*). We then explored whether CD87 participated in germ-cell migration using a matrix adhesion assay. A portion of sorted cells seeded on Matrigel formed cellular protrusions, which represents an essential step for cell migration (Ladoux and Nicolas, 2012). Interestingly, CD87⁺ cells displayed cellular protrusions more frequently than CD87⁻ cells, suggesting a greater migration potential (Fig. 5D, Fig. S8B). The pseudotime analysis revealed a potential interaction between the cell cycle genes (Group II genes, Fig. 4D,E) and differentiation. Therefore, we investigated whether CD87⁺ cells, as a proxy to primitive stem cells, differed from CD87⁻ cells with regard to cell cycle activity. FACS analysis showed a higher frequency of cells in the G2/M phase in CD87⁺ versus CD87⁻ cells, likely reflecting a faster proliferation rate (Fig. 5E). Indeed, we observed a significantly higher growth rate for germ cell cultures derived from CD87⁺ cells (Fig. 5F). Interestingly, all Oct4-GFP⁺ cells expressed CD87 at early stages and the proportion gradually decreased from PND4.5 to PND7.5, and only a very small fraction of cells expressed CD87 at PND13.5 (Fig. 5G). Collectively, these results indicate that CD87⁺ cells in neonatal testis mark a unique cell state with higher expression of self-renewal genes and greater migration potential.

Elevated Oct4 expression signifies the priming of spermatogonia differentiation and enhanced RA responsiveness

Oct4 has been shown to be crucial for SSC maintenance as *Oct4* is downregulated by RA-induced differentiation and knockdown of *Oct4* results in differentiation (Dann et al., 2008). Surprisingly, differentiation-primed cells expressed a notably higher level of *Oct4* and *Rarg* (Fig. 6A). We confirmed that Oct4-GFP fluorescence intensity positively correlates with endogenous levels of both *Oct4* mRNA and protein in single cells (Figs S4B, S9A). We then sorted the Oct4-GFP⁺/KIT⁻ cells into fractions with low, medium and high GFP expression levels and examined their gene expression profile by bulk RNA-Seq. Notably, the gene expression signature of Oct4-GFP-high cells strongly correlates with the differentiation-primed population, which is characterized by decreased expression of SSC self-renewal genes (e.g. *Gfra1*, *Id4* and *Etv5*) and increased expression of genes associated with differentiation [e.g. *Ngn3* (*Neurog3*), *Rarg* and *Sohlh1*] as well as the aforementioned candidates upregulated in differentiation-primed cells (e.g. *Ehmt2*, *Trib3* and *Tdrd5*) (Fig. 6B). We validated this observation by FACS and qRT-PCR analyses on cells from PND7.5 testes, which did not contain gonocytes, and found consistently higher expression of *Rarg*, *Sohlh1* and *Ngn3* in Oct4-GFP-high cells (Fig. 6C). We further associated Oct4-GFP intensity to GFRA1 protein level by FACS analysis and found that

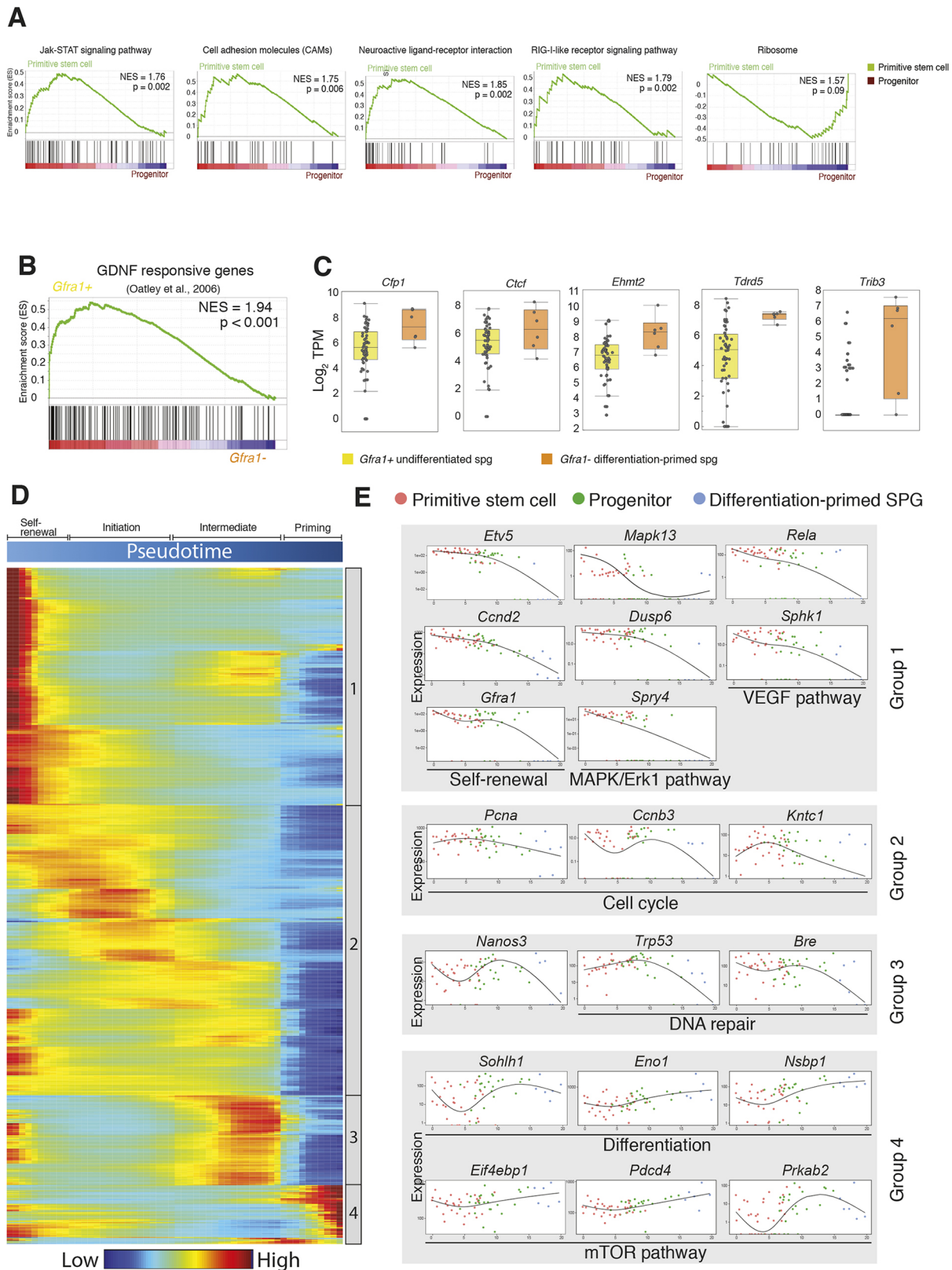


Fig. 4. Molecular cascades regulating spermatogonia self-renewal and differentiation. (A) Signaling pathways enriched in stem cell spermatogonia compared with progenitors by GSEA/KEGG analysis. The values indicated on GSEA enrichment plots are the normalized enrichment score (NES) and *P*-value of the enrichment. (B) Enrichment plots from GSEA analysis using GDNF-responsive gene sets. (C) Box plots for representative genes identified as differentially expressed (FDR<0.05) between *Gfra1*⁺ (Cluster II and III) and *Gfra1*⁻ (Cluster IV) spermatogonia. Median (centre line), first and third quartiles (bounds of box) and 1.5 interquartile range (whiskers) and individual data points (dots) are indicated. (D) Hierarchical clustering of genes using expression kinetic curves calculated by Monocle. The bar on the right shows four distinct kinetic trends recovered. (E) Pseudotime profiles of representative genes showing different kinetics of expression. Each dot represents a cell, and the solid line represents the LOESS regression of the expression level.

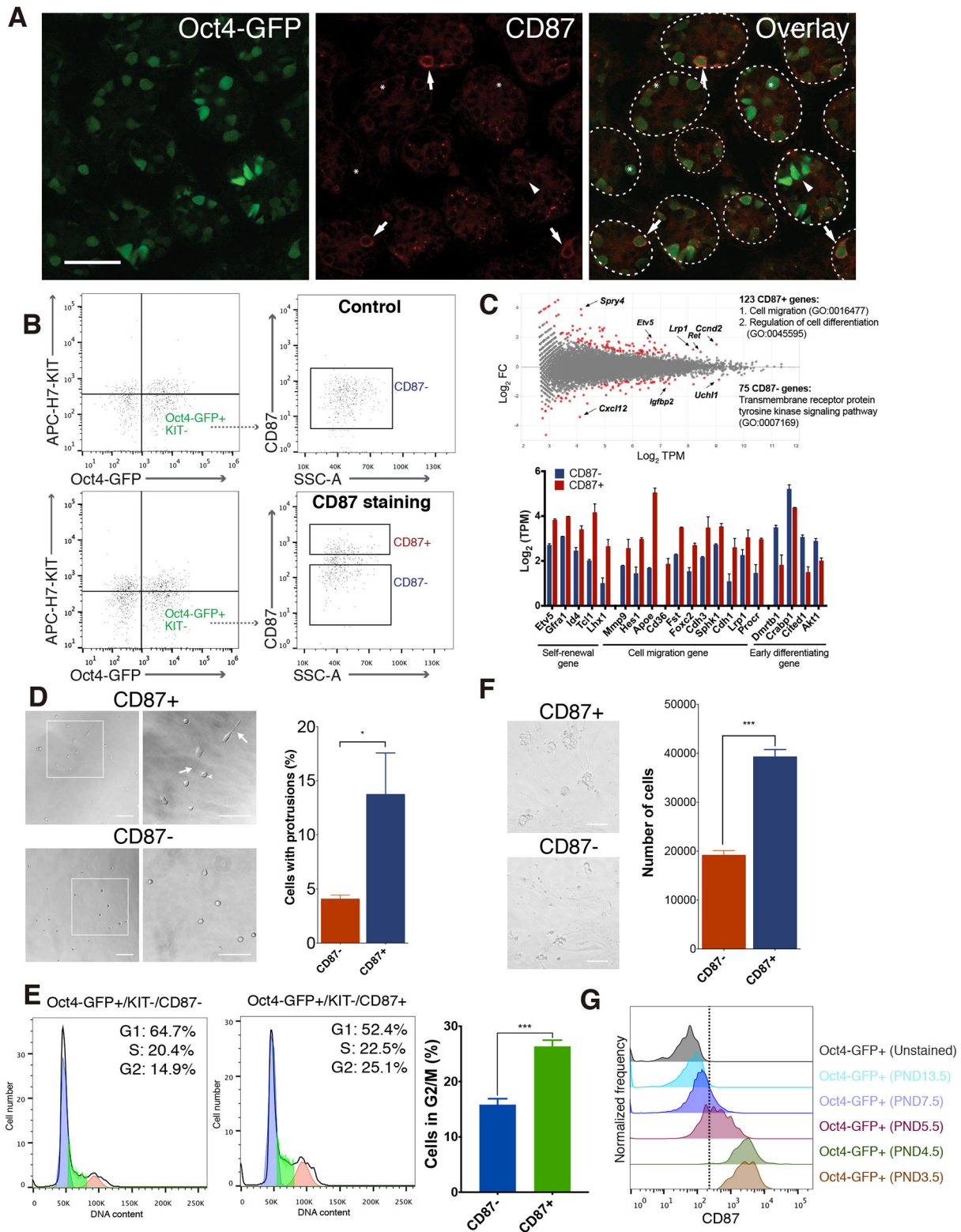


Fig. 5. See next page for legend.

Oct4-GFP-high cells contained significantly fewer GFRA1⁺ cells (Fig. 6D, Fig. S9B,C).

Having established that Oct4-GFP-high cells represent differentiation-primed cells in neonatal testis, we investigated whether this population differed in their potential for RA responsiveness. To test this, primary cultures derived from PND5.5

Oct4-GFP⁺/KIT⁻ cells were treated with all-trans RA or vehicle. Control cultures without RA exposure showed signs of spontaneous differentiation and Oct4-GFP-high cells contained a higher proportion of KIT⁺ cells (Fig. 6E). RA treatment did not alter the intensity of Oct4-GFP or the proportions of cell subsets with different Oct4-GFP intensities. This suggested that Oct4-GFP level was stable

Fig. 5. CD87 marks a unique subpopulation of undifferentiated germ cells in neonates. (A) Confocal images from mouse at PND5.5 immunostained for CD87. Arrows indicate cells with strong CD87 expression. The asterisks and arrowheads denote spermatogonia and transitional gonocytes lacking CD87 expression, respectively. Representative of three independent experiments. (B) Oct4-GFP⁺/KIT⁻/CD87⁺ and Oct4-GFP⁺/KIT⁻/CD87⁻ subpopulations were sorted by FACS. Representative flow cytometry data to illustrate the gating strategy for FACS purification of CD87⁺ and CD87⁻ germ cells. (C) Differential gene expression between the CD87⁺ and CD87⁻ subpopulations of PND5.5 Oct4-GFP⁺/KIT⁻ spermatogonia. Upper panel shows an MA plot of RNA-Seq data showing differentially expressed genes. Lower panel shows the expression level of differentially expressed genes in different categories shown for comparisons between CD87⁺ and CD87⁻ cells. For each gene, expression levels in CD87⁺ versus CD87⁻ are significantly different (FDR<0.05 from two biological replicates). (D) Brightfield images showing that CD87⁺ cells display more spreading and lamellipodia formation on Matrigel. Right-hand panels are magnifications of the boxed areas to the left. Arrows indicate the filopodia-like protrusive structures; arrowhead indicates the spreading of cultured cells. Quantification of cell protrusions comparing CD87⁺ and CD87⁻ cells is shown on the right (three independent experiments). **P*<0.05. (E) The representative result of cell cycle distribution analysis comparing CD87⁺ and CD87⁻ cells. The percentages of cell populations of different cell cycle phases are shown in the upper-right corner. Quantification from three independent experiments (right-hand graph) shows a significantly higher percentage of G2/M cells in the CD87⁺ population. ****P*<0.001. (F) The appearance of germ cell cultures derived from freshly isolated CD87⁺ and CD87⁻ cells 3 days after seeding. Quantification of three independent experiments shows a significantly higher proliferation rate of CD87⁺ cells. ****P*<0.001. (G) The CD87 expression is dynamic throughout spermatogenesis initiation. Scale bars: 50 μ m.

in this experimental setup and that cells of the same intensity in control or treatment groups could be directly compared. Indeed, there were significantly more Oct4-GFP-high cells that became KIT⁺ compared with Oct4-GFP-low and Oct4-GFP-mid cells (Fig. 6E).

Collectively, these data show that Oct4-GFP-high cells display molecular and functional properties associated with cells primed to differentiation.

Role of Oct4 in the regulation of spermatogonia differentiation potential

To enable a more systematic analysis of the gene expression program accompanied by the Oct4-GFP intensity, we conducted Fuzzy clustering using all the differentially expressed genes among the three groups (Fig. 7A) (Kumar and Futschik, 2007). Interestingly, over 90% of the differentially expressed genes either show a gradually decreasing (Clusters 5 and 6) or increasing trend (Clusters 1, 3 and 4) from Oct4-GFP-low to Oct4-GFP-high transition (Fig. 7A, Table S10). Notably, Clusters 5 and 6 contain self-renewal genes, such as *Gfra1*, *Fgfr1*, *Ret* and *Ccnd2*. Further GO analyses revealed over-representation of the MAPK cascade (e.g. *Dusp4*, *Dusp5*, *Nefl*, *Psmc3*) (Fig. 7B). These genes are co-enriched in several pathways related to SSC self-renewal, such as RET signaling and signaling to RAS and ERKs (Table S10). Surprisingly, stronger Oct4-GFP intensity also correlated with higher expression of mesenchymal markers (e.g. *Vim*, *Dab2* and *Emp3*) along with genes significantly enriched in extracellular matrix organization, cell migration and cellular response to growth factor stimulus (Figs 6B and 7A,B). These results indicate that SSCs may modify their cellular adhesion property when they are primed for differentiation. Taken together, these results reveal a gradient of *Oct4* expression coupled with self-renewal and differentiation transcriptional programs.

We hypothesized that higher *Oct4* expression is important for the SSC differentiating potential and intended to reduce *Oct4* expression closer to the ‘ground level’. Mild knockdown of *Oct4*

using siRNA showed that there is a significant reduction but not complete loss of *Oct4* mRNA and protein level (Fig. S9D). The expression of key stem cell markers (i.e. *Gfra1* and *Id4*) remained unchanged, while 590 differentially transcribed genes (FDR<0.05, fold change>1.5, Fig. 7C) were identified. Interestingly, the majority of differentially expressed genes were upregulated (*n*=500, 85%). The top genes included *Etv4*, a downstream factor of GDNF, and *Egr4*, which is important for male fertility (Fig. S10) (Lu et al., 2009; Tourtellotte et al., 1999). GO analysis clearly revealed significant enrichment for focal adhesion, which may be important for stem cell-niche interaction (e.g. *Rhou*, *Ctmb1* and *Itgb1*) (Table S11, Fig. S10). Of particular interest is *Itgb1* (β 1-integrin), which has been implicated as an essential adhesion receptor required for both gonocyte migration and SSC homing (Kanatsu-Shinohara et al., 2008; Tres and Kierszenbaum, 2005). The upregulated genes are also highly enriched in FGF signaling pathway components (e.g. *Fgfr1*, *Etv4*, *Mapk1*, *Ctmb1* and *Fst*) and MAPK protein kinases (e.g. *Mapk1*, *Mapk12*, *Wee1* and *Map3k2*), which are downstream regulators of GDNF and FGF signaling (Fig. 7D) (Hasegawa et al., 2013). Only 90 downregulated genes were identified as a consequence of *Oct4* knockdown (Fig. 7C). Among these, *Pten* is a direct target of OCT4 in cancer cells and in SSC maintenance (Huang et al., 2011; Tang et al., 2015). *Dll3* is a ligand of the Notch signaling pathway (Fig. S10). To characterize further the functional consequence of the altered gene expression, we analyzed the ability of cultured germ cells to respond to short-pulse RA-induced differentiation with decreased *Oct4* expression. Knockdown of *Oct4* led to a 30% reduction of RA-induced *Kit* mRNA expression (Fig. 7E). These results suggest *Oct4* may contribute to the differentiation-primed state of SSCs through transcriptional regulation.

DISCUSSION

Unraveling the heterogeneity of SSC by high-throughput single cell analysis has recently drawn attention. Studies on neonatal germ cells using single cell qRT-PCR analysis or *Id4*-GFP⁺ cells using single cell RNA-Seq has provided a better understanding of their heterogeneity. However, these studies focused on a panel of genes or a restricted population (*Id4*-GFP⁺) (Hermann et al., 2018, 2015; Song et al., 2016). Here, we extended these analyses to whole-transcriptome profiling on a broader undifferentiated germ cell population in PND5.5 testes.

To aid in the interpretation of these data, we first discuss challenges and limitations. Comprehensive characterization in an unbiased manner requires isolation of all subtypes during development. The Oct4-GFP⁺ population harbors GFRA1⁺ and PLZF⁺ cells, making it a good surrogate for undifferentiated germ cells (Fig. 1). However, we cannot rule out the possibility that rare or unknown subpopulations (i.e. Oct4-GFP⁻, or GFRA1⁻ or PLZF⁻) may not fit into the current classification scheme and yet contribute to the establishment of spermatogenesis. It is also possible that rare subpopulations within the Oct4-GFP⁺ compartment may not have been captured in our 71 samples during profiling. In this sense, our profiling might not be completely exhaustive and the existence of other subtypes remains a possibility. Nevertheless, our study presents an outlook on the heterogeneity of the main germ cell populations in neonatal testis.

Unique transcriptional program in gonocytes

Our study provides a global view of the gene expression signature in transitional gonocytes. First, we discovered several molecules that might be involved in gonocyte migration. Second, the

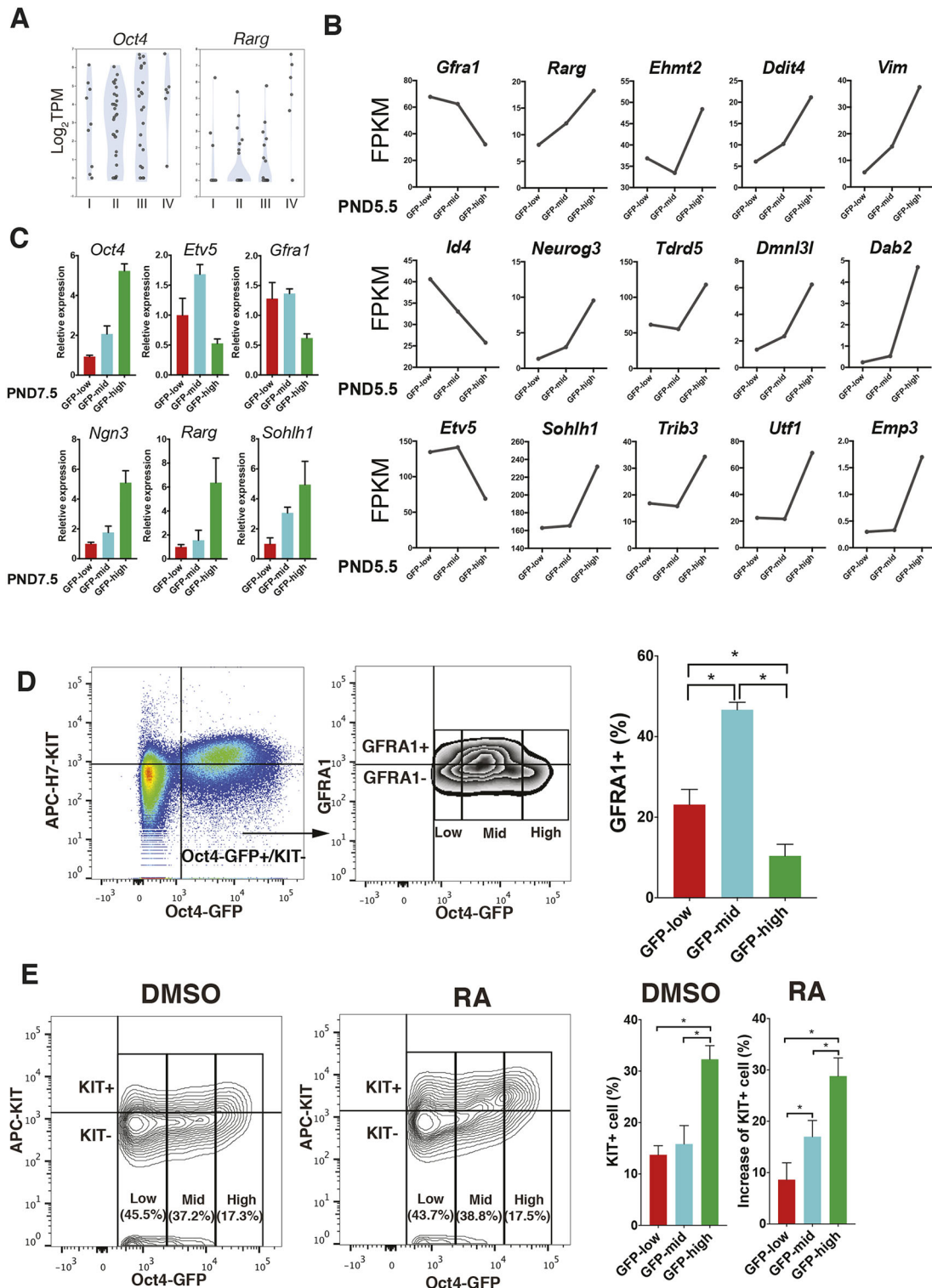


Fig. 6. Differentiation-primed cells display enhanced RA responsiveness. (A) Both *Oct4* and *Rarg* mRNA were expressed at higher levels in Cluster IV than other clusters. (B) Line plots depicting the dynamics of gene expression in relation to the Oct4-GFP gradient in PND5.5 testes. (C) Patterns of gene expression measured by RT-qPCR in subpopulations of Oct4-GFP⁺/KIT⁻ cells sorted from PND7.5 testes. (D) Indirect staining and FACS analysis of GFRA1 expression. FACS plot showing the GFRA1 expression in Oct4-GFP⁺/KIT⁻ cells (left). Quantification from three independent experiments of the distribution of GFRA1⁺ cells in different subpopulations based on GFP intensity (right). **P*<0.005. (E) Representative plots from three independent experiments showing KIT staining in germ cell culture derived from Oct4-GFP⁺/KIT⁻ cells with or without RA treatment showing a significantly enhanced RA response in Oct4-GFP-high cells. (F) Quantification of the FACS data shown in E. **P*<0.005.

gonocyte and spermatogonia population is distinguished by global transcription activity and change of cell cycle status. Third, our results showed that the expression of apoptosis-related genes is

actively regulated during GST. Particularly, transitional gonocytes seem to have a lower threshold for initiating apoptosis (elevated death receptor pathway) and increased susceptibility to DNA

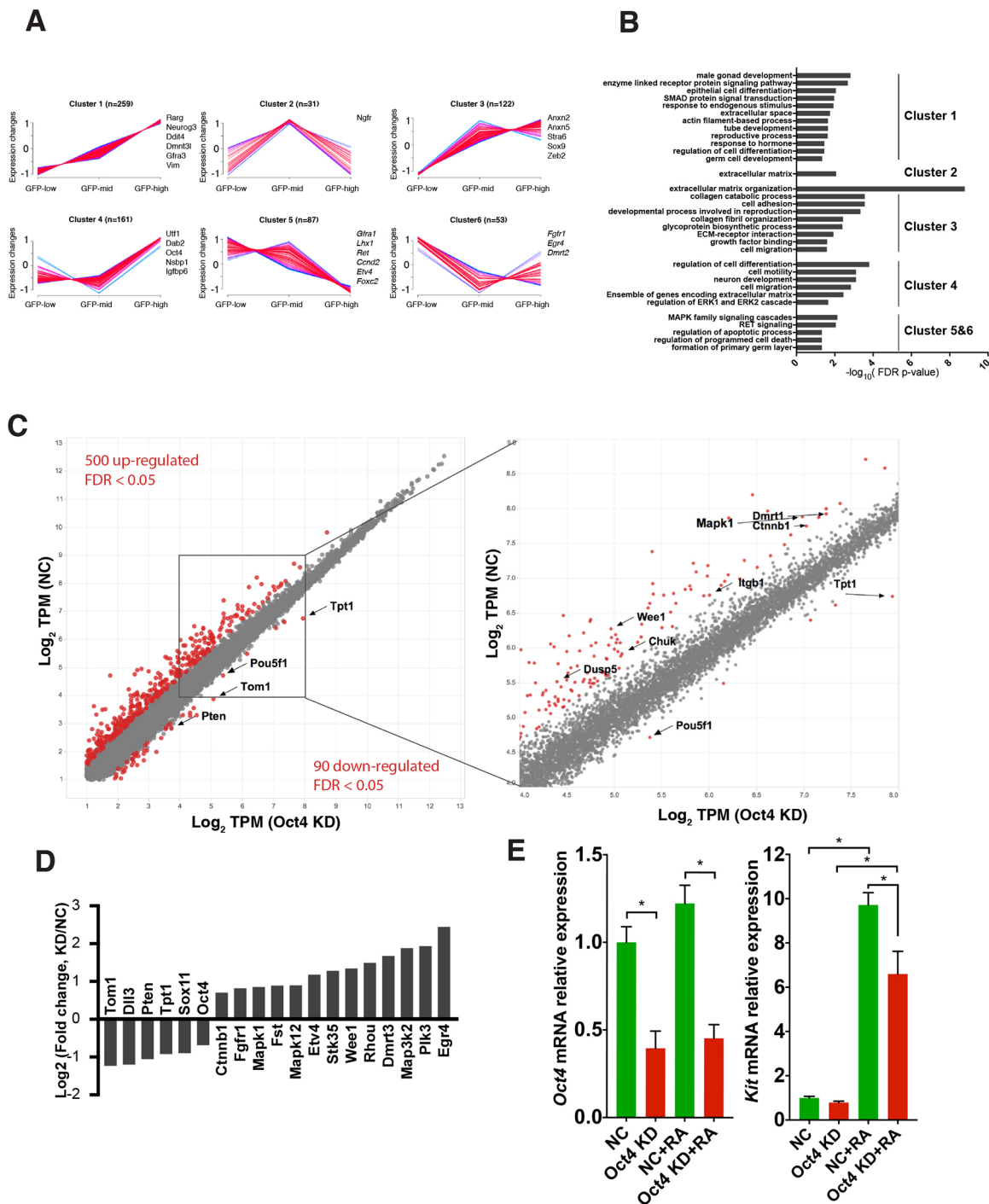


Fig. 7. The potential role of *Oct4* in SSC differentiation. (A) Differentially expressed genes were identified among Oct4-GFP-low, Oct4-GFP-mid and Oct4-GFP-high groups and z-score-normalized ratios of the gene expression were subjected to soft clustering with Mfuzz software. The six clusters represent different expression profiles. (B) GO-term enrichment for Clusters 1-6. (C) Scatter plot of RNA-Seq data comparing mRNA abundance in Oct4-GFP⁺/KIT⁻ germ cell cultures after *Oct4* knockdown (KD) and negative control (NC). (D) Expression change of selected genes in knockdown compared with negative control as determined by RNA-Seq. (E) qRT-PCR analysis of *Kit* mRNA expression in Oct4-GFP⁺/KIT⁻ germ cell cultures after knockdown of *Oct4*. Culture cells were treated with RA for 4 h and sorted by FACS to remove MEF feeder cells and RNA was extracted for qRT-PCR analysis. Data are mean±s.e.m. for *n*=3 different cultures (biological replicates). **P*<0.005.

damage-induced apoptosis (reduced DNA-repair pathway). This raises the possibility that transitional gonocytes possess poised apoptotic machinery, which renders them more vulnerable to apoptotic cell death. Further studies using our resources will help to shed light on the mechanistic nature of such transcriptional regulation in GST.

Transcriptional regulation during the SSC-to-progenitor transition and differentiation priming

In corroboration of a recent ‘hierarchical model’ of adult SSCs, we showed that neonatal primitive stem cell and progenitor populations are distinguished by expression gradients of SSC maintenance genes (*Id4*, *Gfra1*, *Etv5* and *Tspan8*) and differentiation-promoting

genes (*Sohlh1* and *Rarg*) (de Rooij, 2017). The ‘ultimate’ stem cells proposed in this model might reside in the primitive stem cell population we identified. CD87 is a glycosylphosphatidylinositol-anchored protein and acts as the receptor of urokinase-type plasminogen activator (uPA), which induces cell adhesion and migration (Ploug et al., 1991). We showed that CD87⁺ cells are most abundant within the first week after birth and have greater potential for migration *in vitro*. Interestingly, CD87 is also a ligand for several integrins including ITGB1 (β 1-integrin). Thus, it is tempting to speculate that CD87 may play an important role in the initial establishment of the SSC pool, potentially through the regulation of migration.

Our results also provided an opportunity to reflect on whether the regulations of neonatal and adult spermatogonia are distinct. The stem cell-progenitor-differentiation-primed spermatogonia transition is reminiscent of the adult differentiation trajectory reported recently (Green et al., 2018; Hermann et al., 2018). Interestingly, Cluster I cells (including stem cells) expressed a higher level of a set of genes that overlapped with the FGF target genes (e.g. *Spry4*, *Dusp6*, *Lhx1* and *Etv4*) as reported recently (Kitadate et al., 2019). Kitadate et al. demonstrated the role of FGFs in the cell fate decision of adult SSCs to either self-renew or differentiate. Echoing this, we found that syndecan genes (*Sdc1* and *Sdc2*, regulators of FGF signaling) are more abundantly expressed in Cluster I primitive stem cells (Fig. S6A), suggesting that Cluster I cells may be better furnished with the reception machinery for FGFs. Therefore, it is possible that the mechanism of mitogen (FGF) competition in adult SSCs is also shared by neonatal spermatogonia.

At the same time, this high-resolution data set provides a more comprehensive characterization of transcriptome-wide features and differences between stem cells and progenitors. Our analysis showed that genes related to lipid metabolism are highly enriched in the stem cell pool, which extends previous findings on the role of fatty acid metabolism in stem cell maintenance (Ito and Suda, 2014). Our results also indicated that ribosome biogenesis increased in progenitors compared with stem cell spermatogonia, implying an enhanced need for global protein synthesis during differentiation. Our results further demonstrated that the SSC-to-progenitor transition is intertwined with cell cycle states.

Although the action of RA signaling on the neonatal spermatogonia differentiation process is known, the transcriptional processes that regulate the first steps in differentiation are not well defined. Surprisingly, our study uncovered a differentiation-primed population with a higher level of *Rarg*. Higher *Rarg* expression may enhance the sensitivity of these cells to RA exposure, which closely resembles the recently described differentiation model in adult spermatogenesis (Ikami et al., 2015). These observations revealed the earliest event of SSC differentiation in which cells turn off the self-renewal program before committing to differentiation, which can only be resolved by single cell analysis.

The role of *Oct4* in the regulation of SSC differentiation

Differentiation-primed cells notably display a surprisingly higher level of *Oct4* (Fig. 6A, Fig. S4D,F). The association between *Oct4* and SSC differentiation has been overlooked but it has been recently demonstrated that Oct4-GFP⁻ and Oct4-GFP⁺ undifferentiated spermatogonia population in adult testis represent stem cells and progenitors, respectively (La et al., 2018). Our study also showed that higher Oct4-GFP was associated with a more differentiated state and that the gradient of *Oct4* induction was paralleled by the gradual loss of self-renewal genes and gain of differentiation-promoting genes. Functionally, we showed that Oct4-GFP-high cells displayed

stronger RA responsiveness, which shared the functional property of Id4-GFP-dim progenitors as reported previously (Lord et al., 2018).

A prominent unanswered question is whether and how OCT4 contributes to the enhanced RA response of SSCs. In our mild knockdown experiment, the expression of key self-renewal markers (i.e. *Gfra1* and *Id4*) does not appear to be affected, suggesting that there was little change in the relative abundance of cell subpopulations. It is also unlikely that OCT4 directly controls the level of *Rarg* and positively regulates the RA pathway because we did not identify a significant change in the expression of differentiation-related genes (i.e. *Rarg*, *Kit* and *Stra8*). Instead, OCT4 appears to modulate downstream signaling of the SSC self-renewal pathway, such as the MAPK pathway. Our results also raised the possibility that OCT4 might modulate the stem cell-niche interaction in SSCs. Oct4-GFP-high cells express a higher level of mesenchymal markers and enrichment of genes associated with cell migration. In addition, many OCT4-regulated genes were associated with focal adhesion (e.g. *Igfb1*). Thus, it is possible that OCT4 could modulate the exposure of SSCs to niche-secreted short-range signals and cell contact-dependent signals. Taken together, we speculate that different levels of OCT4 expression may act as a molecular rheostat to regulate self-renewal versus differentiation of SSCs in neonatal testis. Elevation of *Oct4* in the earliest stage of SSC differentiation leads to the repression of several downstream components of the GDNF/FGF pathway, as well as the exposure of SSCs to niche signals, facilitating the response to the differentiation cue. Our work thus provides insights into the important issue of the coupling between cell adhesion, stem cell maintenance and differentiation.

In conclusion, we envisage that knowledge of transcriptional dynamics and cell fate decisions during early germ cell establishment should provide hints to aid further dissection of the heterogeneity in adult human and mouse spermatogonia. It will also help the study of SSCs in the context of aging and disease.

MATERIALS AND METHODS

Animals

All animal procedures were performed according to protocols approved by the Animal Experimentation Ethics Committee (AEEC) of the Chinese University of Hong Kong and following the Animals (Control of Experiments) Ordinance (Cap. 340) licensed from the Hong Kong Government Department of Health. Oct4-GFP transgenic mice [B6; CBA-Tg(Pou5f1-EGFP)2Mnn/J, Stock no.: 004654] were acquired from The Jackson Laboratory and maintained in the CUHK Laboratory Animal Services Centre.

Immunofluorescence staining of testis tissue sections

Neonatal testes were harvested and fixed in 4% paraformaldehyde (PFA) solution for 2 h followed by dehydration in 30% sucrose. Then, the testes were embedded in OCT (Sakura) and snap-frozen. Cryosections of 10 μ m thickness were collected on positively charged slides (Thermo Fisher Scientific). Sections were blocked with PBS supplemented with 0.2% Triton X-100 and 10% normal donkey serum (Jackson ImmunoResearch) for 1 h at room temperature, followed by overnight incubation with antibodies against KIT, GFRA1, PLZF, APC-Ki-67 and CD87 at 4°C. After washing in PBS, sections were incubated with the corresponding secondary antibody (except for Ki-67 staining). Antibodies used are listed in Table S12. Finally, tissue slides were washed with PBS and mounted with mounting medium (Thermo Fisher Scientific). A minimum of three different animals were tested and negative controls lacking primary antibody were used. Fluorescent microscopy and brightfield images were acquired using a Leica SP2 confocal microscope. All images were captured using Leica Application Suite Software. For quantification of cells with overlapping expression of markers, different localization or Ki-67 expression, each staining was performed in triplicate and images were taken for at least three sections for

each testis. Cells were counted in ten randomly selected testis cords in each testis section.

Isolation and analysis of undifferentiated and differentiating spermatogonia by FACS

Undifferentiated germ cells from testes of Oct4-GFP transgenic mice at PND5.5 or PND7.5 were purified using the method described previously (Garcia and Hofmann, 2012). For the Fluidigm C1 single cell capture experiment, testes of five pups at PND5.5 from one litter were pooled before FACS sorting. In the Biomark qRT-PCR experiment, testes of four pups at PND5.5 from another litter were pooled before FACS sorting. Isolated seminiferous tubules were digested with 1 mg/ml type 2 collagenase (Worthington), 1 mg/ml hyaluronidase (Sigma-Aldrich) and 5 µg/ml DNase I (Sigma-Aldrich) at 37°C for 20 min with occasional shaking. The suspension was passed through a 40-µm strainer cap (BD Falcon) to yield a uniform single cell suspension. After incubation with antibodies against KIT or/and CD87 at 4°C for 30 min, cell fractions were collected with a FACS Aria II cell sorter (BD Biosciences). In order to obtain pure populations, sorting gates were set within the boundaries defined by to ‘fluorescence minus one’ (FMO) controls. Indirect labeling was used for GFRA1 and OCT4 analysis. Cells were stained for 30 min on ice with APC-H7-conjugated anti-KIT antibody and washed with PBS. Cells were then fixed in 2% PFA, permeabilized in Perm buffer (0.5% Tween 20, 2% fetal bovine serum in PBS). Washed cells were incubated with anti-GFRA1 or anti-OCT4 antibodies at 4°C for 30 min. After extensive washing, cells were incubated with Alexa Fluor 647 secondary antibody at 4°C for 30 min. Positive GFRA1 or OCT4 antibody labeling was determined by comparison with staining with Alexa Fluor 647 antibody only. Age- and strain-matched non-reporter mice, i.e. C57BL/6 mice, were used for cell purification as a negative control devoid of GFP expression for the cell-sorting experiment. The antibodies used in this study can be found in Table S12.

Quantitative real-time PCR

Quantitative real-time PCR (qRT-PCR) was performed using the Fast SYBR Green Master Mix (Applied Biosystems) on a ViiA 7 Real-Time PCR System (Applied Biosystems) according to the manufacturer’s instructions. qRT-PCR was carried out with at least three biological replicates per group, and each biological replicate was performed in duplicate or triplicate. Primers used are listed in Table S13.

Capturing of single cells, preparation of libraries and sequencing

Single cells were captured on a medium-sized (10–17 µm cell diameter) microfluidic RNA-Seq chip (Fluidigm) using the Fluidigm C1 Auto Prep system. Cells were loaded onto the chip at a concentration of 300–500 cells/µl and stained for viability (LIVE/DEAD cell viability assay; Molecular Probes, Life Technologies). The Fluidigm C1 IFC Chip after cell capture was inspected under the microscope and dead cells and doublets captured sites were eliminated. cDNAs were prepared using the SMARTer Ultra Low RNA Kit for Illumina (Clontech). Libraries were generated using the Nextera XT (Illumina) protocol and were sequenced as 50-bp single-end reads on Illumina HiSeq 2000.

Analysis of single cell RNA-Seq data

Raw reads were pre-processed with trim galore for trimming Illumina adaptor sequences and mapped to mouse mm9 genome. We estimated the expression level for all RefSeq genes using Salmon (Patro et al., 2017) and selected the 500 top-ranked PCA genes for clustering. Pseudotime ordering was performed with Monocle (v1.2.0) as previously described (Trapnell et al., 2014). Differential gene expression analysis was performed with DESeq2 method (Love et al., 2014). R package ‘WGCNA’ was used to construct the weighted gene co-expression network. Further details can be found in supplementary Materials and Methods.

Bulk RNA-Seq

Oct4-GFP⁺/KIT⁻ cells were FACS-sorted from PND5.5 testes based on different GFP intensity levels. Germ cell culture in the *Oct4* knockdown experiments were FACS-sorted to remove MEF feeder cells. Oct4-GFP⁺/KIT⁻/CD87⁺ and Oct4-GFP⁺/KIT⁻/CD87⁻ cells were also sorted by FACS.

Total RNA was isolated from sorted cells using the AllPrep DNA/RNA Mini kit (Qiagen) and subjected to low input RNA-Seq library preparation. Further details of methods and bioinformatic analysis can be found in supplementary Materials and Methods.

Single cell qPCR using Biomark

Single cell gene expression analysis was performed using the BioMark 48.48 Dynamic Array platform (Fluidigm) and custom-designed gene-specific primer pairs (Table S13). We sorted Oct4-GFP⁺/KIT⁻ single cells using FACS into 96-well PCR plates containing 5 µl of CellsDirect reaction mix and these cells were immediately stored at –80°C. Upon thawing, a mix containing 2.5 µl gene-specific primers, 1.2 µl CellsDirect RT/Taq mix and 0.3 µl TE buffer were added to each well. RT-PCR pre-amplification cycling conditions were: 50°C, 15 min; 95°C, 2 min; 22× (95°C, 15 s; 60°C, 4 min). Samples were diluted 1:5 in TE buffer and a 2.25 µl aliquot of each cDNA was mixed with 2.5 µl of 2× SsoFast EvaGreen Supermix with Low ROX (Bio-Rad) and 0.25 µl of 20× DNA Binding Dye Sample Loading Reagent (Fluidigm). The sample mix and custom-designed primer pairs were loaded separately into wells of 48.48 Gene Expression Dynamic Arrays (Fluidigm) in the presence of appropriate loading reagents. The arrays were read using a Biomark analysis system (Fluidigm). Data were analyzed using Fluidigm Real-Time PCR Analysis software (v3.0.2). Cts were recovered from BioMark. The quality threshold was set to 0.65 and a linear derivative as baseline correction was used. The limit of detection was set to a Ct of 28; Cts for qRT-PCR reactions that failed the quality threshold were converted to 28. Cts were converted to log₂ expression by subtracting Cts from 28. Expression matrix of the genes analyzed can be found in Table S4.

RA-induced differentiation

For RA-induced differentiation, 5 mM all-trans-RA (Sigma-Aldrich) stock in DMSO was diluted to 1 µM in germ cell culture medium before applying to cells; vehicle (DMSO) alone was added in the control group. For Fig. 6E, the cells were collected for FACS analysis after 16 h. For Fig. 7E, the cells were collected by FACS 4 h after treatment and sorted cells were subjected to RNA extraction and qRT-PCR analysis.

Cell cycle analysis

Cells were stained for 30 min on ice with APC-H7-conjugated anti-KIT and APC-conjugated anti-CD87 antibodies and washed with PBS. Cells were then fixed in 2% PFA and permeabilized in Perm buffer. Cells were washed and stained for 45 min on ice with APC-conjugated anti-Ki-67 antibody. The cells were washed again and stained with Hoechst 33342 (Sigma-Aldrich) at 5 µg/ml final concentration in complete medium for 30 min at 37°C. The stained cells were re-suspended in cold PBS and run on the FACS analyzer. Data were analyzed using FlowJo software (Tree Star).

Matrigel adhesion assay

Culture plates were coated with Matrigel (BD Biosciences) by the following procedure. A Matrigel bottle was thawed on ice in a 4°C refrigerator overnight until the Matrigel liquified. Matrigel was divided into 300 µl aliquots and stored at –20°C until use. Working Matrigel solution was prepared by diluting 300 µl of Matrigel with 29 ml of DMEM/F12 medium (Invitrogen) and thorough mixing. This solution was added to 24-well plates (0.5 ml per well) to cover the whole surface of the wells. The plates were allowed to sit for 1 h at room temperature. Excess Matrigel solution was then removed, and the plates were washed once with DMEM/F12. An equal number of the FACS-sorted Oct4-GFP⁺/KIT⁻/CD87⁺ or Oct4-GFP⁺/KIT⁻/CD87⁻ cells were plated on three Matrigel-coated wells. After 16 h incubation, images were taken from four random fields of each well and the proportion of cells with protrusion was calculated as the number of cells with protrusions divided by the total number of cells in all 12 images in each biological replicate.

Knockdown of *Oct4* by siRNA

Small interference RNAs (siRNAs) against *Oct4* were synthesized by GenePharma (Shanghai, China). The sequences of siRNAs are: sense, 5'-AAGGAUGUGGUUCGAGUAUGG-3'; antisense, 5'-CCATACCTCG-AACCACATCCTT-3'. The negative control sequence has no homology

with known mouse genes. The siRNA was transfected into SSC culture cells using Lipofectamine 2000. Briefly, a 100 μ l OptiMEM mixture containing 2.5 μ l Lipofectamine and 10 pmol siRNA was prepared according to the manufacturer's instructions. The mixture was added to 400 μ l of SSC culture medium containing 1.5×10^5 trypsin-dissociated SSCs and incubated at 37°C for 30 min. The SSC-Lipofectamine-siRNA mixture was then added to a 24-well of pre-seeded MEF and incubated for 6-8 h after which the medium was replaced with normal SSC culture medium. The SSCs were allowed to grow for another 48 h and subjected to direct gene expression analysis or RA-induced differentiation.

Statistical analyses

Experimental data are shown as mean \pm s.e.m. All statistical analyses were either conducted with JMP or as specified in relevant sections.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: J.L., T.L.L.; Methodology: J.L., S.H.N., T.L.L.; Software: J.L.; Validation: J.L., A.C.L., H.C.S.; Formal analysis: J.L., S.H.N.; Investigation: J.L., S.H.N., T.L.L.; Resources: J.L., B.F., T.L.L.; Data curation: J.L., S.H.N.; Writing - original draft: J.L., T.L.L.; Writing - review & editing: J.L., H.C.S., Y.Q., A.W.T.L., J.T., J.C.L.F., N.L.S.T., B.F., W.Y.C., P.F., R.M.H., T.L.L.; Visualization: J.L.; Supervision: W.Y.C., T.L.L.; Project administration: T.L.L.; Funding acquisition: T.L.L.

Funding

This work was supported by the Lo Kwee-Seong Biomedical Research Seed Fund from the School of Biomedical Sciences, The Chinese University of Hong Kong (project no. 7104687).

Data availability

All RNA-Seq data generated in this study have been deposited in Gene Expression Omnibus under accession number GSE107711.

Supplementary information

Supplementary information available online at <http://dev.biologists.org/lookup/doi/10.1242/dev.174953.supplemental>

References

- Basciani, S., De Luca, G., Dolci, S., Brama, M., Arizzi, M., Mariani, S., Rosano, G., Spera, G. and Gnassi, L. (2008). Platelet-derived growth factor receptor beta-subtype regulates proliferation and migration of gonocytes. *Endocrinology* **149**, 6226-6235.
- Chen, J.-P., Luan, Y., You, C.-X., Chen, X.-H., Luo, R.-C. and Li, R. (2010). TRPM7 regulates the migration of human nasopharyngeal carcinoma cell by mediating Ca(2+) influx. *Cell Calcium* **47**, 425-432.
- Culty, M. (2009). Gonocytes, the forgotten cells of the germ cell lineage. *Birth Defects Res. C Embryo Today* **87**, 1-26.
- Dann, C. T., Alvarado, A. L., Molyneux, L. A., Denard, B. S., Garbers, D. L. and Porteus, M. H. (2008). Spermatogonial stem cell self-renewal requires OCT4, a factor downregulated during retinoic acid-induced differentiation. *Stem Cells* **26**, 2928-2937.
- de Rooij, D. G. (2017). The nature and dynamics of spermatogonial stem cells. *Development* **144**, 3022-3030.
- Falender, A. E., Freiman, R. N., Geles, K. G., Lo, K. C., Hwang, K., Lamb, D. J., Morris, P. L., Tjian, R. and Richards, J. S. (2005). Maintenance of spermatogenesis requires TAF4b, a gonad-specific subunit of TFIID. *Genes Dev.* **19**, 794-803.
- Fleischer, A., Ayllon, V. and Rebollo, A. (2002). ITM2BS regulates apoptosis by inducing loss of mitochondrial membrane potential. *Eur. J. Immunol.* **32**, 3498-3505.
- Garcia, T. and Hofmann, M.-C. (2012). Isolation of undifferentiated and early differentiating type A spermatogonia from Pou5f1-GFP reporter mice. *Methods Mol. Biol.* **825**, 31-44.
- Green, C. D., Ma, Q., Manske, G. L., Shami, A. N., Zheng, X., Marini, S., Moritz, L., Sultan, C., Gurczynski, S. J., Moore, B. B. et al. (2018). A comprehensive roadmap of murine spermatogenesis defined by single-cell RNA-seq. *Dev. Cell* **46**, 651-667.e610.
- Hasegawa, K., Namekawa, S. H. and Saga, Y. (2013). MEK/ERK signaling directly and indirectly contributes to the cyclical self-renewal of spermatogonial stem cells. *Stem Cells* **31**, 2517-2527.
- He, S., Nakada, D. and Morrison, S. J. (2009). Mechanisms of stem cell self-renewal. *Annu. Rev. Cell Dev. Biol.* **25**, 377-406.
- Helsel, A. R., Yang, Q.-E., Oatley, M. J., Lord, T., Sablitzky, F. and Oatley, J. M. (2017). ID4 levels dictate the stem cell state in mouse spermatogonia. *Development* **144**, 624-634.
- Hermann, B. P., Mutoji, K. N., Velte, E. K., Ko, D. J., Oatley, J. M., Geyer, C. B. and McCarrey, J. R. (2015). Transcriptional and translational heterogeneity among neonatal mouse spermatogonia. *Biol. Reprod.* **92**, 54.
- Hermann, B. P., Cheng, K., Singh, A., Roa-De La Cruz, L., Mutoji, K. N., Chen, I.-C., Gildersleeve, H., Lehle, J. D., Mayo, M., Westernströmer, B. et al. (2018). The mammalian spermatogenesis single-cell transcriptome, from spermatogonial stem cells to spermatids. *Cell Rep.* **25**, 1650-1667.e1658.
- Hobbs, R. M., Seandel, M., Falcatori, I., Rafii, S. and Pandolfi, P. P. (2010). Plzf regulates germline progenitor self-renewal by opposing mTORC1. *Cell* **142**, 468-479.
- Huang, Y., Mao, X., Boyce, T. and Zhu, G. Z. (2011). Dispensable role of PTEN in mouse spermatogenesis. *Cell Biol. Int.* **35**, 905-908.
- Ikami, K., Tokue, M., Sugimoto, R., Noda, C., Kobayashi, S., Hara, K. and Yoshida, S. (2015). Hierarchical differentiation competence in response to retinoic acid ensures stem cell maintenance during mouse spermatogenesis. *Development* **142**, 1582-1592.
- Ishii, K., Kanatsu-Shinohara, M., Toyokuni, S. and Shinohara, T. (2012). FGF2 mediates mouse spermatogonial stem cell self-renewal via upregulation of ETV5 and Bcl6b through MAP2K1 activation. *Development* **139**, 1734-1743.
- Ito, K. and Suda, T. (2014). Metabolic requirements for the maintenance of self-renewing stem cells. *Nat. Rev. Mol. Cell Biol.* **15**, 243-256.
- Kamisawa, H., Kojima, Y., Mizuno, K., Imura, M., Hayashi, Y. and Kohri, K. (2012). Attenuation of spermatogonial stem cell activity in cryptorchid testes. *J. Urol.* **187**, 1047-1052.
- Kanatsu-Shinohara, M., Takehashi, M., Takashima, S., Lee, J., Morimoto, H., Chuma, S., Raducanu, A., Nakatsui, N., Fässler, R. and Shinohara, T. (2008). Homing of mouse spermatogonial stem cells to germline niche depends on beta1-integrin. *Cell Stem Cell* **3**, 533-542.
- Kitadate, Y., Jorg, D. J., Tokue, M., Maruyama, A., Ichikawa, R., Tsuchiya, S., Segi-Nishida, E., Nakagawa, T., Uchida, A., Kimura-Yoshida, C. et al. (2019). Competition for mitogens regulates spermatogenic stem cell homeostasis in an open Niche. *Cell Stem Cell* **24**, 79-92.e76.
- Kluin, P. M. and Derooij, D. G. (1981). A comparison between the morphology and cell-kinetics of gonocytes and adult type undifferentiated spermatogonia in the mouse. *Int. J. Androl.* **4**, 475-493.
- Komai, Y., Tanaka, T., Tokuyama, Y., Yanai, H., Ohe, S., Omachi, T., Atsumi, N., Yoshida, N., Kumano, K., Hisha, H. et al. (2014). Bmi1 expression in long-term germ stem cells. *Sci. Rep.* **4**, 6175.
- Kumar, L. and Futschik, M. E. (2007). Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics* **2**, 5-7.
- La, H. M., Mäkelä, J.-A., Chan, A.-L., Rossello, F. J., Nefzger, C. M., Legrand, J. M. D., De Seram, M., Polo, J. M. and Hobbs, R. M. (2018). Identification of dynamic undifferentiated cell states within the male germline. *Nat. Commun.* **9**, 2819.
- Ladoux, B. and Nicolas, A. (2012). Physically based principles of cell adhesion mechanosensitivity in tissues. *Rep. Prog. Phys.* **75**, 116601.
- Langfelder, P. and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559.
- Lee, J., Kanatsu-Shinohara, M., Morimoto, H., Kazuki, Y., Takashima, S., Oshimura, M., Toyokuni, S. and Shinohara, T. (2009). Genetic reconstruction of mouse spermatogonial stem cell self-renewal in vitro by Ras-Cyclin D2 activation. *Cell Stem Cell* **5**, 76-86.
- Lord, T., Oatley, M. J. and Oatley, J. M. (2018). Testicular architecture is critical for mediation of retinoic acid responsiveness by undifferentiated spermatogonial subtypes in the mouse. *Stem Cell Rep.* **10**, 538-552.
- Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.
- Lu, B. C., Cebrían, C., Chi, X., Kuure, S., Kuo, R., Bates, C. M., Arber, S., Hassell, J., MacNeil, L., Hoshi, M. et al. (2009). ETV4 and ETV5 are required downstream of GDNF and Ret for kidney branching morphogenesis. *Nat. Genet.* **41**, 1295-1302.
- Manku, G. and Culty, M. (2015). Mammalian gonocyte and spermatogonia differentiation: recent advances and remaining challenges. *Reproduction* **149**, R139-R157.
- Nagano, R., Tabata, S., Nakanishi, Y., Ohsako, S., Kurohmaru, M. and Hayashi, Y. (2000). Reproliferation and relocation of mouse male germ cells (gonocytes) during prespermatogenesis. *Anat. Rec.* **258**, 210-220.
- Nakagawa, T., Nabeshima, Y.-I. and Yoshida, S. (2007). Functional identification of the actual and potential stem cell compartments in mouse spermatogenesis. *Dev. Cell* **12**, 195-206.
- Oatley, J. M. and Brinster, R. L. (2008). Regulation of spermatogonial stem cell self-renewal in mammals. *Annu. Rev. Cell Dev. Biol.* **24**, 263-286.
- Oatley, J. M., Avarbock, M. R., Telaranta, A. I., Fearon, D. T. and Brinster, R. L. (2006). Identifying genes important for spermatogonial stem cell self-renewal and survival. *Proc. Natl. Acad. Sci. USA* **103**, 9524-9529.
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417-419.
- Ploug, M., Ronne, E., Behrendt, N., Jensen, A. L., Blasi, F. and Dano, K. (1991). Cellular receptor for urokinase plasminogen activator. Carboxyl-terminal

- processing and membrane anchoring by glycosyl-phosphatidylinositol. *J. Biol. Chem.* **266**, 1926-1933.
- Pollen, A. A., Nowakowski, T. J., Shuga, J., Wang, X., Leyrat, A. A., Lui, J. H., Li, N., Szpankowski, L., Fowler, B., Chen, P. et al.** (2014). Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.* **32**, 1053-1058.
- Rajpert-de Meyts, E. and Hoei-Hansen, C. E.** (2007). From gonocytes to testicular cancer: the role of impaired gonadal development. *Ann. N. Y. Acad. Sci.* **1120**, 168-180.
- Sailland, J., Tribollet, V., Forcet, C., Billon, C., Barenton, B., Carnesecchi, J., Bachmann, A., Gauthier, K. C., Yu, S., Giguère, V. et al.** (2014). Estrogen-related receptor alpha decreases RHOA stability to induce orientated cell migration. *Proc. Natl. Acad. Sci. USA* **111**, 15108-15113.
- Saitou, M. and Yamaji, M.** (2012). Primordial germ cells in mice. *Cold Spring Harb. Perspect. Biol.* **4**, a008375.
- Song, H.-W., Bettgowda, A., Lake, B. B., Zhao, A. H., Skarbrevik, D., Babajanian, E., Sukhwani, M., Shum, E. Y., Phan, M. H., Plank, T.-D. M. et al.** (2016). The homeobox transcription factor RHOX10 drives mouse spermatogonial stem cell establishment. *Cell Rep.* **17**, 149-164.
- Suzuki, H., Sada, A., Yoshida, S. and Saga, Y.** (2009). The heterogeneity of spermatogonia is revealed by their topology and expression of marker proteins including the germ cell-specific proteins Nanos2 and Nanos3. *Dev. Biol.* **336**, 222-231.
- Tang, Y.-A., Chen, C.-H., Sun, H. S., Cheng, C.-P., Tseng, V. S., Hsu, H.-S., Su, W.-C., Lai, W.-W. and Wang, Y.-C.** (2015). Global Oct4 target gene analysis reveals novel downstream PTEN and TNC genes required for drug-resistance and metastasis in lung cancer. *Nucleic Acids Res.* **43**, 1593-1608.
- Tjwa, M., Sidenius, N., Moura, R., Jansen, S., Theunissen, K., Andolfo, A., De Mol, M., Dewerchin, M., Moons, L., Blasi, F. et al.** (2009). Membrane-anchored uPAR regulates the proliferation, marrow pool size, engraftment, and mobilization of mouse hematopoietic stem/progenitor cells. *J. Clin. Invest.* **119**, 1008-1018.
- Tourtellotte, W. G., Nagarajan, R., Auyeung, A., Mueller, C. and Milbrandt, J.** (1999). Infertility associated with incomplete spermatogenic arrest and oligozoospermia in Egr4-deficient mice. *Development* **126**, 5061-5071.
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S. and Rinn, J. L.** (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381-386.
- Tres, L. L. and Kierszenbaum, A. L.** (2005). The ADAM-integrin-tetraspanin complex in fetal and postnatal testicular cords. *Birth Defects Res. C Embryo Today* **75**, 130-141.
- Wang, R.-A., Nakane, P. K. and Koji, T.** (1998). Autonomous cell death of mouse male germ cells during fetal and postnatal period. *Biol. Reprod.* **58**, 1250-1256.
- Yoshida, S., Sukeo, M., Nakagawa, T., Ohbo, K., Nagamatsu, G., Suda, T. and Nabeshima, Y.** (2006). The first round of mouse spermatogenesis is a distinctive program that lacks the self-renewing spermatogonia stage. *Development* **133**, 1495-1505.
- Zeineddine, D., Hammoud, A. A., Mortada, M. and Boeuf, H.** (2014). The Oct4 protein: more than a magic stemness marker. *Am. J. Stem Cells* **3**, 74-82.

Figure S1

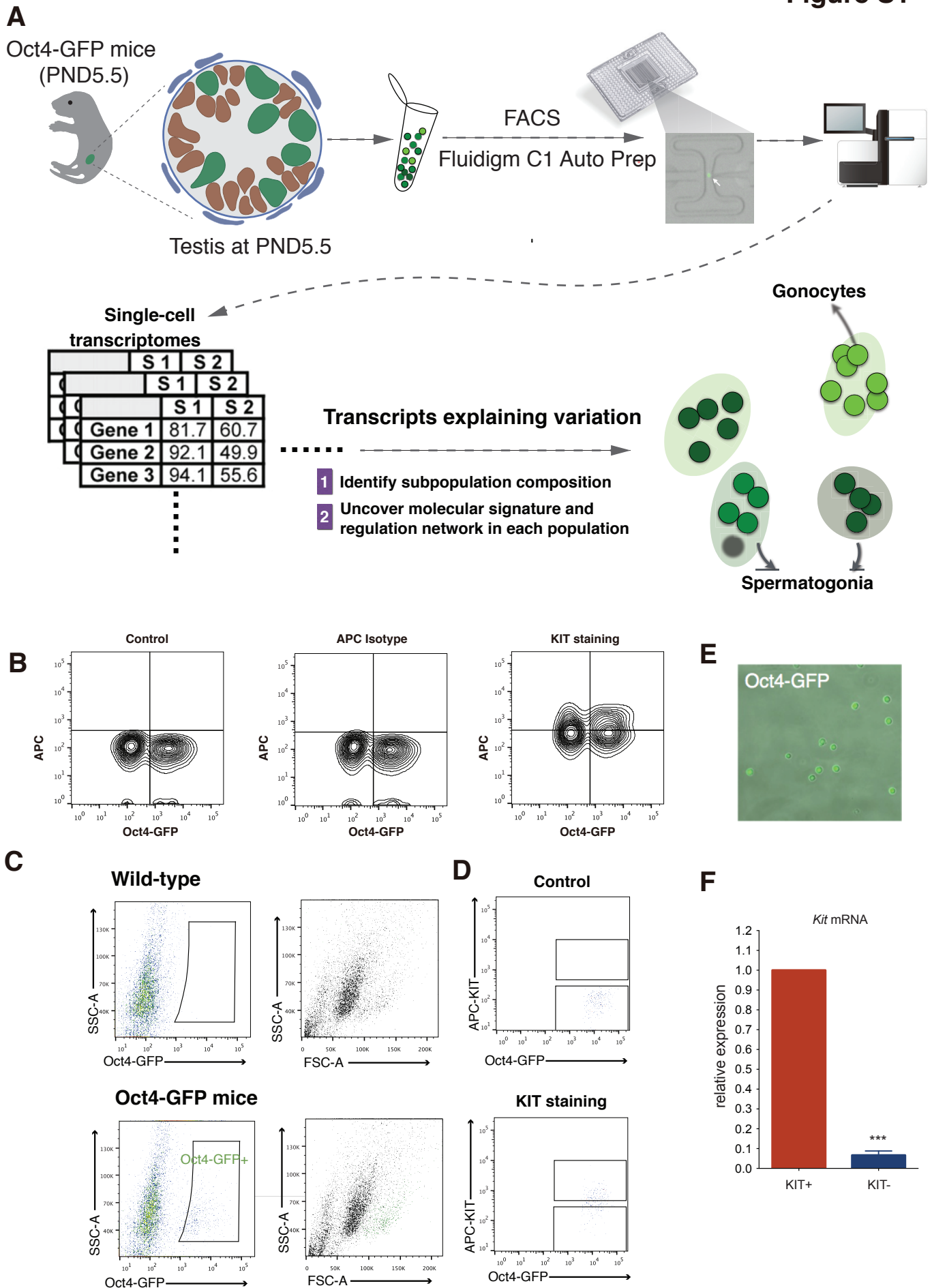


Figure S1. Workflow of the single-cell RNA-Seq experiment (Related to Figure 1).

(A) A schematic diagram of the single-cell RNA-Seq experimental strategy. Germ cells at PND5.5 are a mixture of gonocyte and spermatogonia. After dissection of testes of the Oct4-GFP male mice, Oct4-GFP+/KIT- cells were collected by FACS. Single cells were captured by Fluidigm C1 Auto Prep system. cDNA was generated, and libraries were amplified for deep sequencing. The resulting 71 transcriptomes were subjected to downstream analysis. See Supplemental experimental procedure for further details on the isolation procedures. (B) Flow cytometry analysis of testicular cells without staining (control), labeled with APC-conjugated rat IgG2b isotype control antibody (APC isotype) or APC-KIT antibody (KIT staining). (C) Positive Oct4-EGFP was determined by comparison to testis cells from PND5.5 wild-type littermates. As noted in the FSC-SSC scatter plot, Oct4-GFP+ cells are only located in the FSC-A-high area (green dots). (D) The KIT- and KIT+ populations were gated using APC-KIT labeled sample compared with unstained control (FMO). FSC indicates forward scatter; SSC, side scatter. (E) GFP signal examination confirms isolation of Oct4-GFP+ cells. (F) Differential *Kit* expression revealed by qRT-PCR confirm successful isolation of Oct4-GFP+/KIT- germ cells.

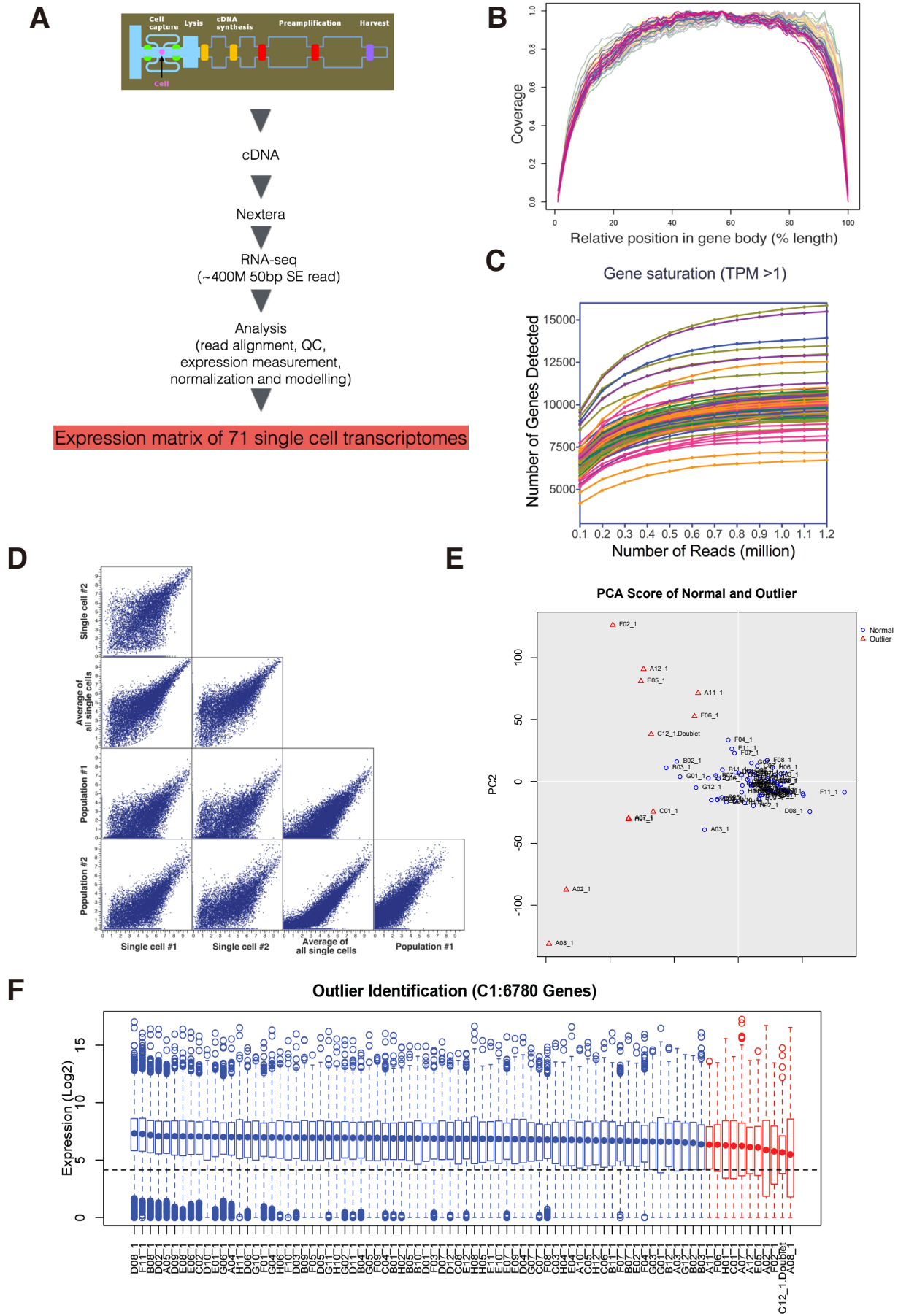
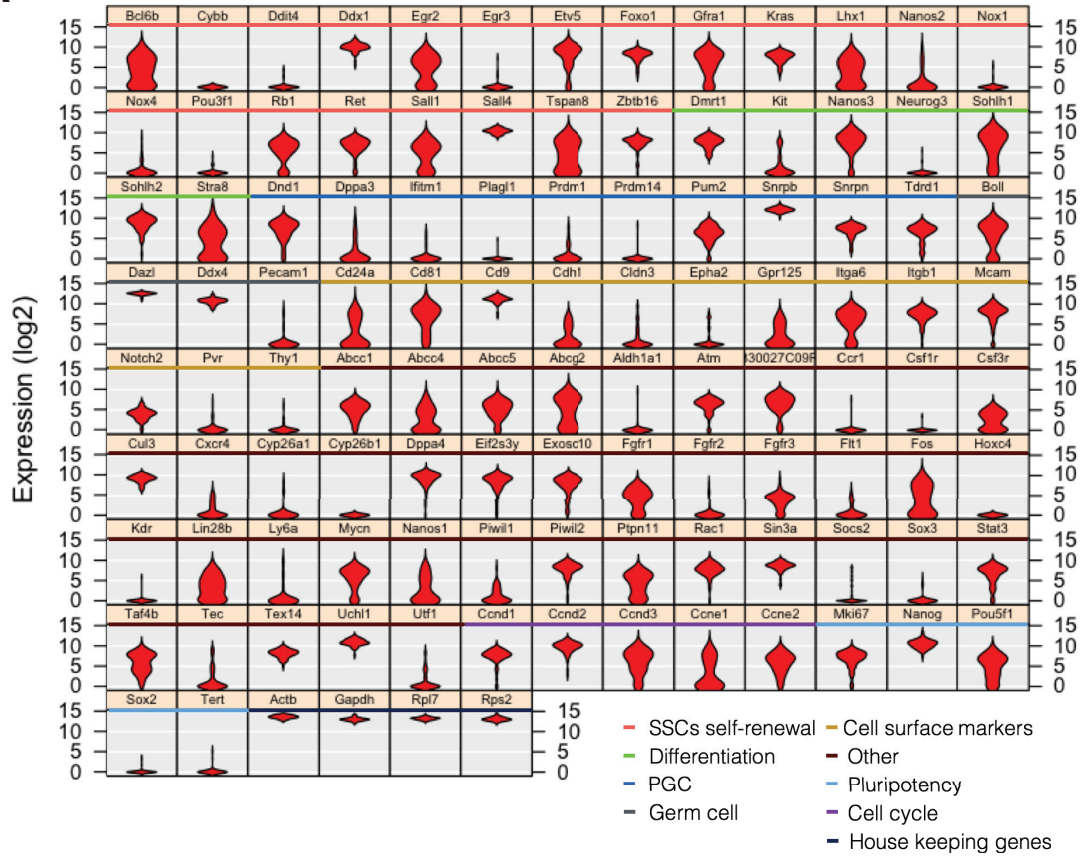


Figure S2. Quality assessment of single-cell RNA-Seq data.

(A) Schematic overview of the C1 workflow. (B) Mean of read coverage over gene body of transcripts. Transcripts are normalized for length and the average plots show uniform read coverage from 5' to 3' position. (C) Saturation curves of RNA-seq data with TPM>1, which confirmed that the sequencing depth was sufficient to detect most of the genes expressed in individual cells. (D) Pairwise comparison of gene expression levels (\log_2 TPM) for 2 single cells, the 200-cell pool and the 'average' single-cell. The aggregate of gene expression from individual cells recapitulated the population samples well. (E-F) Results from SINGuLAR outlier analysis, presented as PCA and box plots. Each box plot represents an individual cell and shows the distribution of expression of the most stably expressed genes across all the cells. Dashed horizontal line indicates the 15th percentile of the gene expression distribution, and cells whose median expression level falls below the threshold are red.

Figure S3

A



B

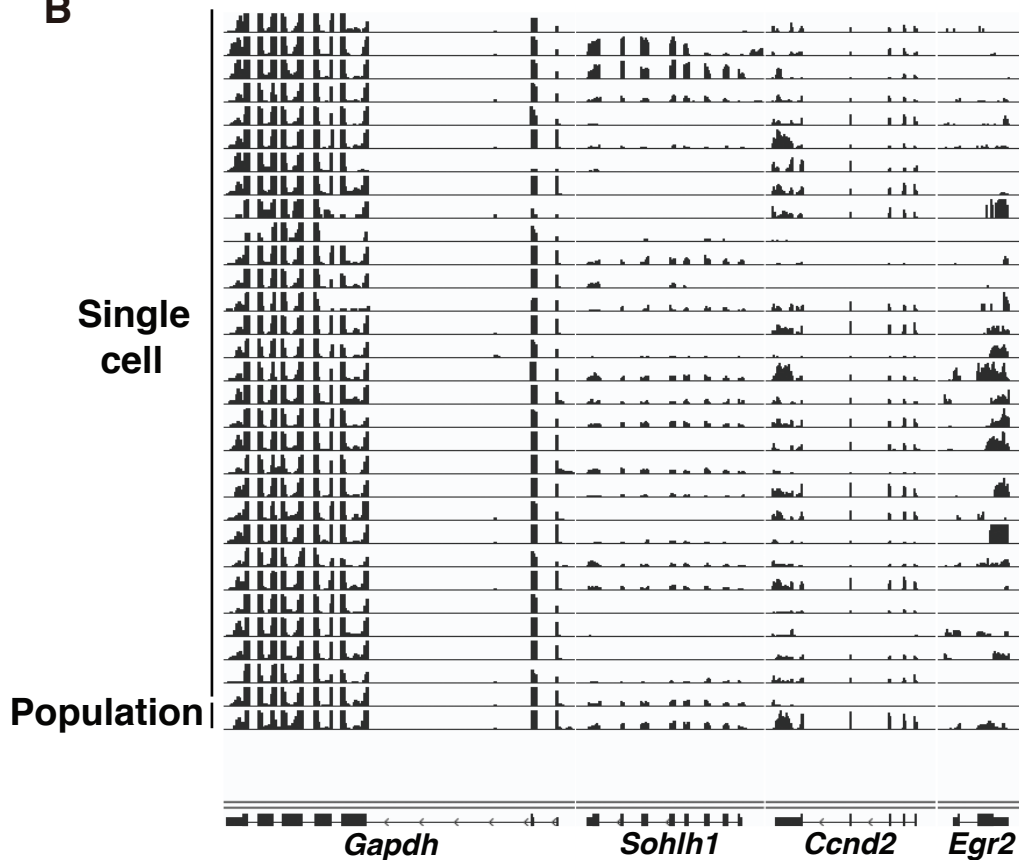


Figure S3. Expression of selected genes across the single cells.

(A) Violin plots showing the expression level a panel of genes associated with spermatogonia development in all the single-cell samples. The plots showed substantial variability of the 103 expressed genes, which could be summarized as uniform distribution (n=48), bimodal distribution (n=21) and sporadic distribution (n=34). Genes exhibited uniform and high expression included early germ cell markers *Dazl* and *Ddx4*, confirming that our data included only germ cells. *Bcl6b*, *Tcl1*, *Lhx1* and *Egr2* exhibited bimodal expression pattern, while *Nanos3* and *Nanos2* displayed sporadic expression. (B) Representative coverage profiles of single-cell transcriptomes and bulk sample at selected genomic loci, including housekeeping gene *Gapdh* and known genes related to germ cell development.

Figure S4

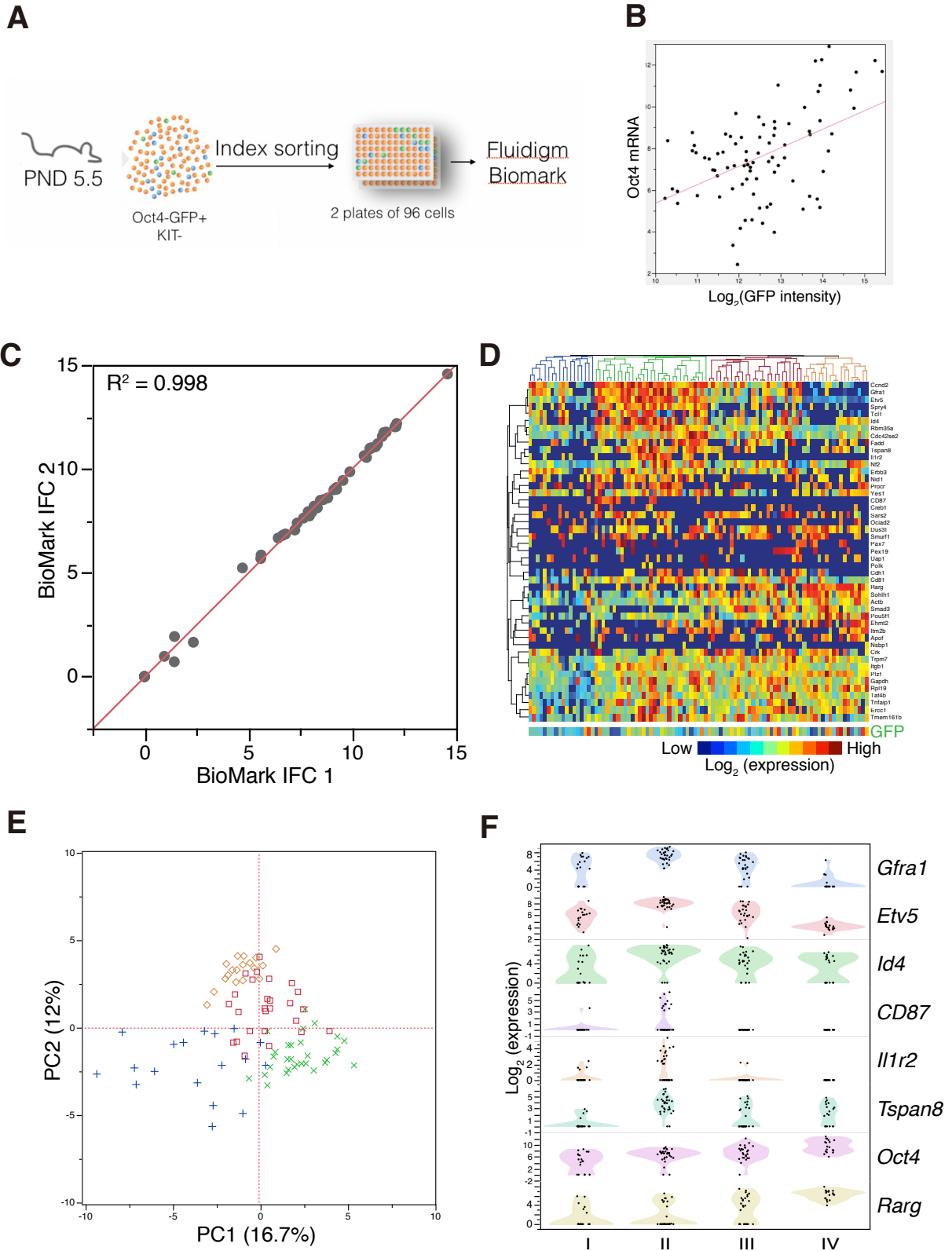
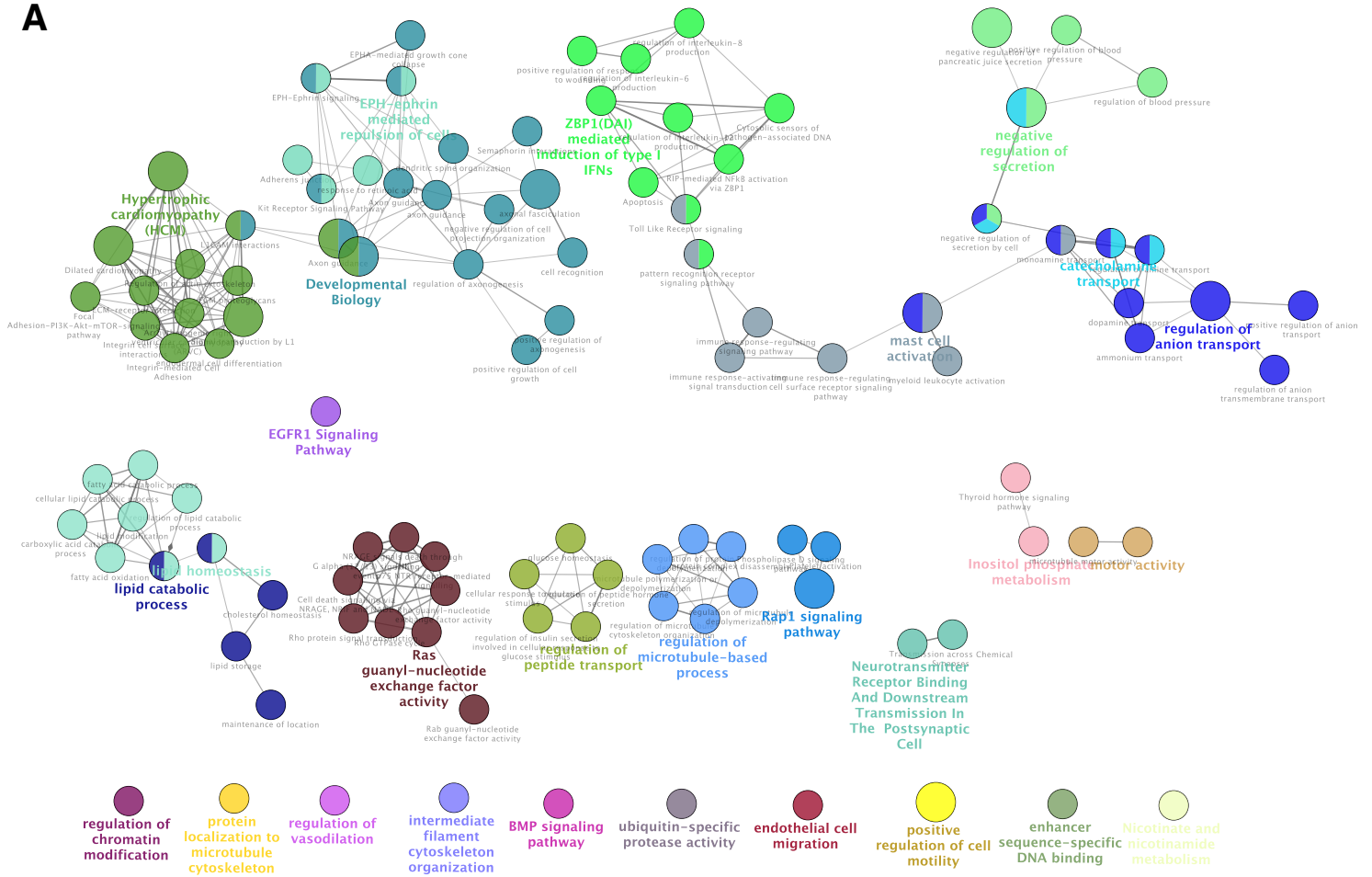


Figure S4. Single-cell qPCR validates 4 subpopulations in PND5.5 Oct4-GFP+/KIT- germ cells.

(A) Schematic illustrating the methods used to isolate cells for Biomark single-cell qPCR experiment. Oct4-GFP+/KIT- cells were index sorted from Oct4-GFP mice into 96-well PCR plate. Specific Target Amplification (STA) was carried out on the same plate and the resulted cDNA was subjected to Biomark single-cell qRT-PCR analysis. (B) Correlation between the expression of *Oct4* mRNA (y-axis) and the GFP intensity (x-axis) across the single cell analyzed. *Oct4* mRNA expression was reported by Singular analysis toolset and GFP intensity was defined by the log₂-transformed FITC signal. (C) Comparison of gene expression in single-cell qPCR data between 2 BioMark IFCs. cDNA from the same cell was profiled on each BioMark IFC to test for variability in the qPCR system. The plot shows strong correlations (Pearson's *r*) between the expression level measured by 2 different IFC chips, indicating that data collected on different IFC chip are reproducible. (D, E) Plots show hierarchical clustering (D) and PCA (E) of single cells based on qPCR-based detection of 48 genes. Four major clusters were identified. *n* = 96 cells combined from two IFC chips. See Figure 2 for cell group designations. (F) Violin plots illustrating the distribution of representative genes for subpopulations.

A



B

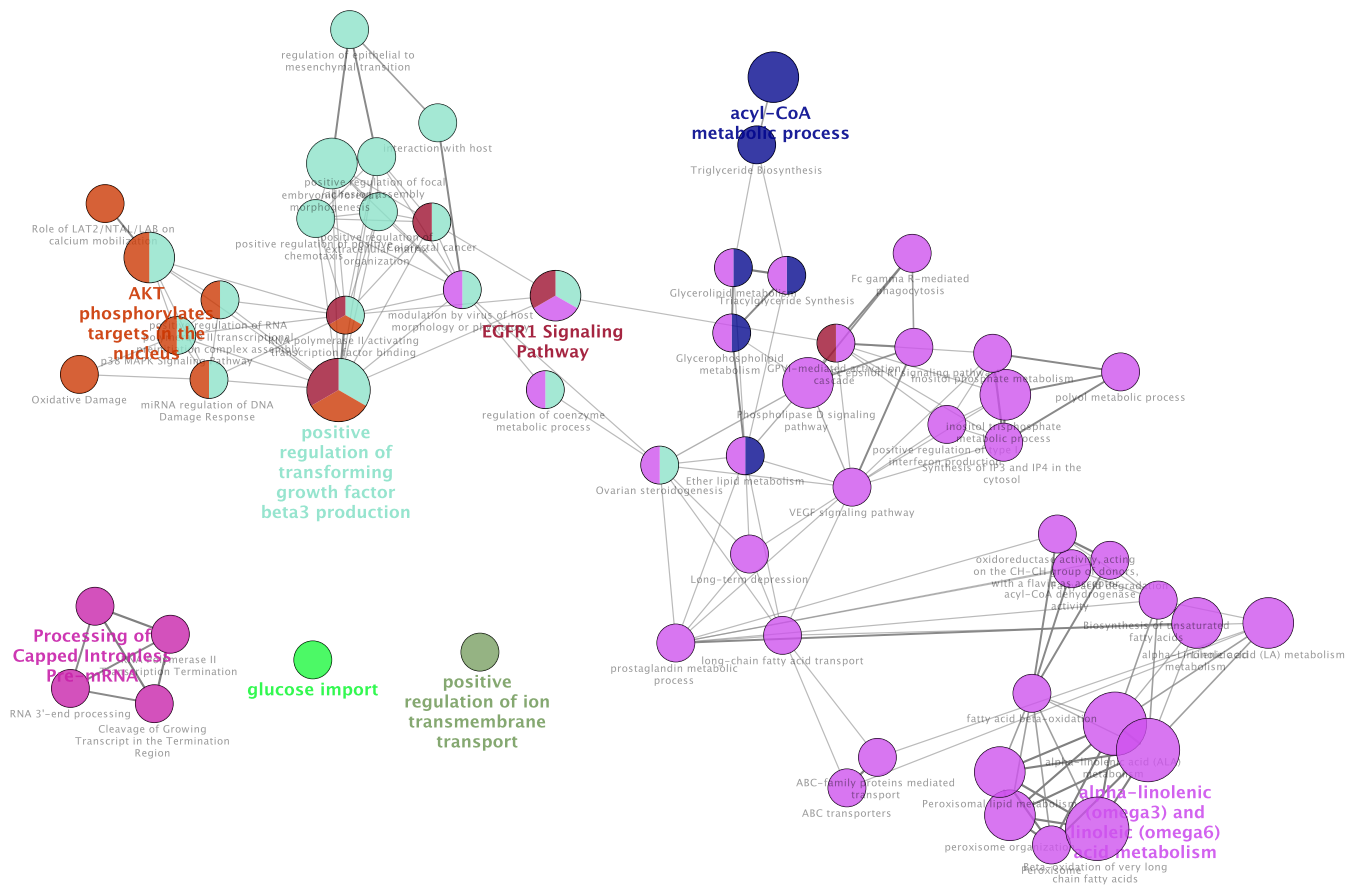


Figure S5. Gene Ontology Enrichment Analysis comparing Cluster II and Cluster III.

GO analysis for genes enriched in primitive stem cells (A) and GO analysis for genes enriched in progenitors (B). The size of each point represents the number of genes present. All the data points have adjusted p-value less than 0.05.

Figure S6

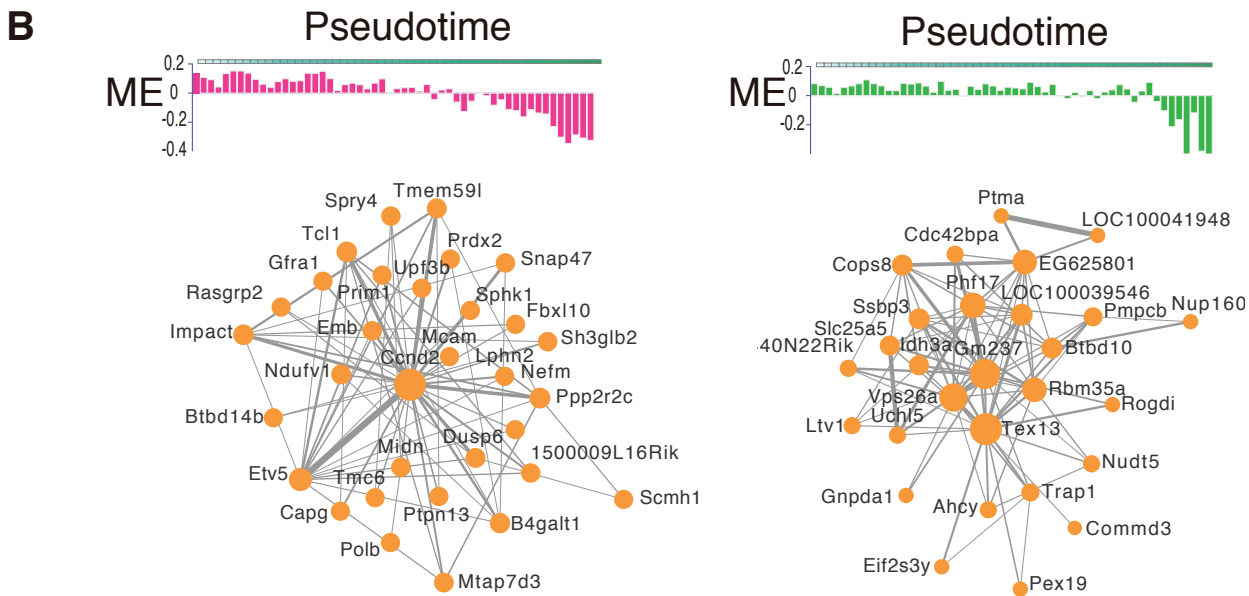
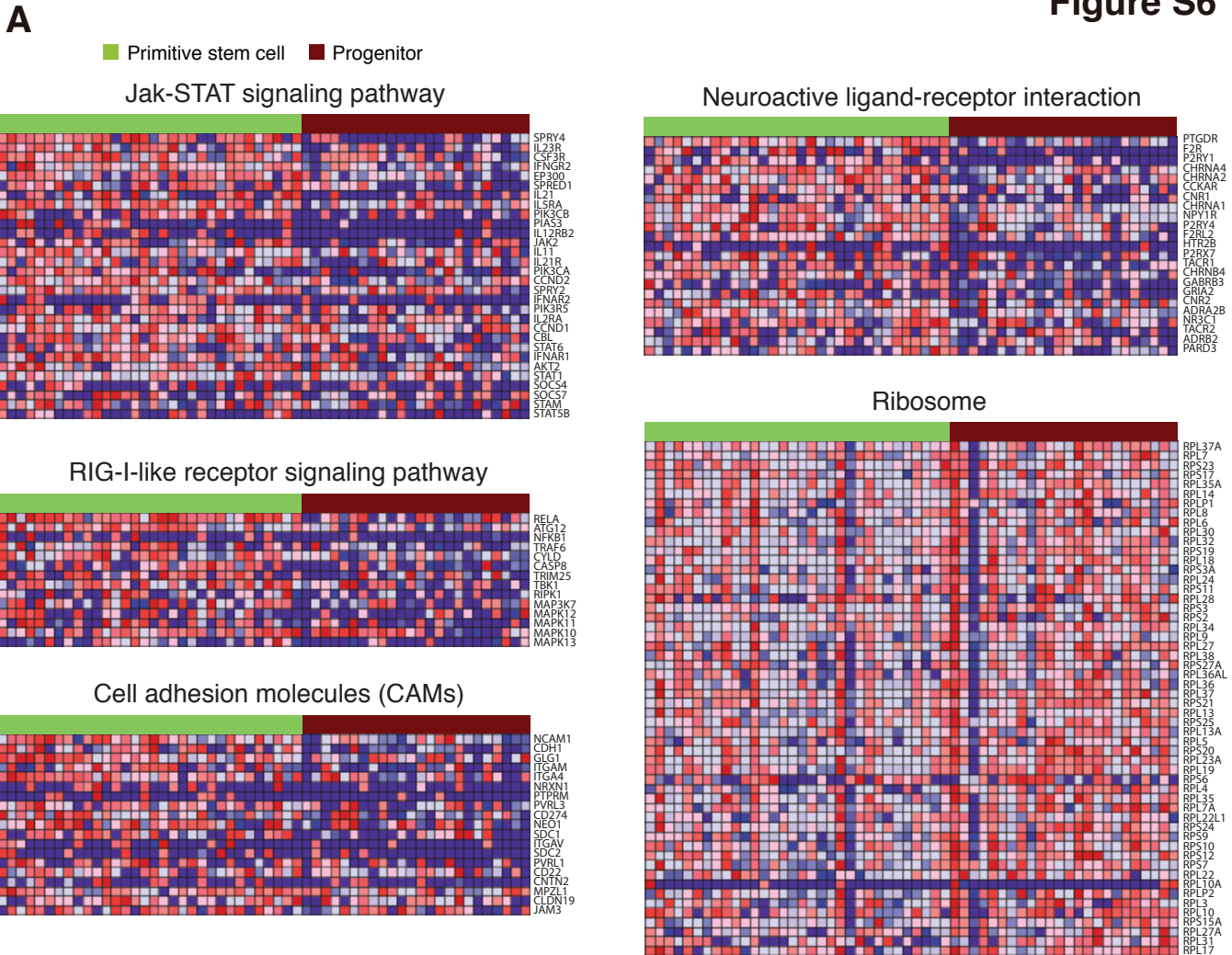


Figure S6. Transcriptional signatures of primitive stem cells and progenitors (related to Figure 4).

(A) Signaling pathways enriched in stem cell spermatogonia as compared to progenitors by GSEA/KEGG analysis (related to Figure 4). Heatmaps shows the expression of gene enriched in representative signaling pathways. (B) WGCNA analysis uncovered the existence of different gene modules. Module eigengene values (ME) (y-axis) of two selected modules among the 62 single-cell spermatogonia samples (Cluster II, III, and IV) across pseudotime (x-axis) was shown. Expression of pink module is enriched in primitive stem cell population while green module is expressed in both primitive stem cells and progenitors (Upper panel). Relationship of the top hub genes identified by network analysis is shown in lower panel.

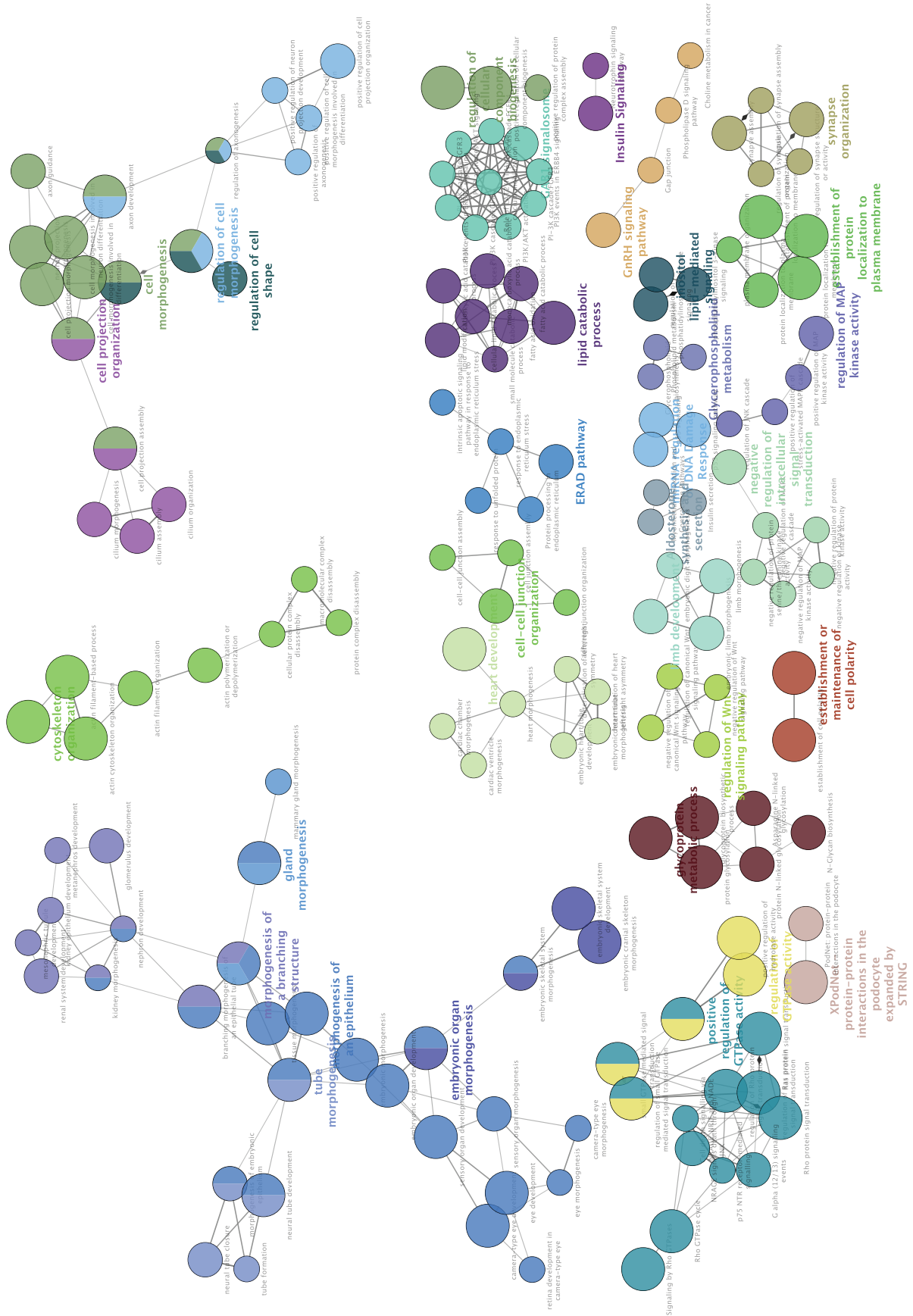
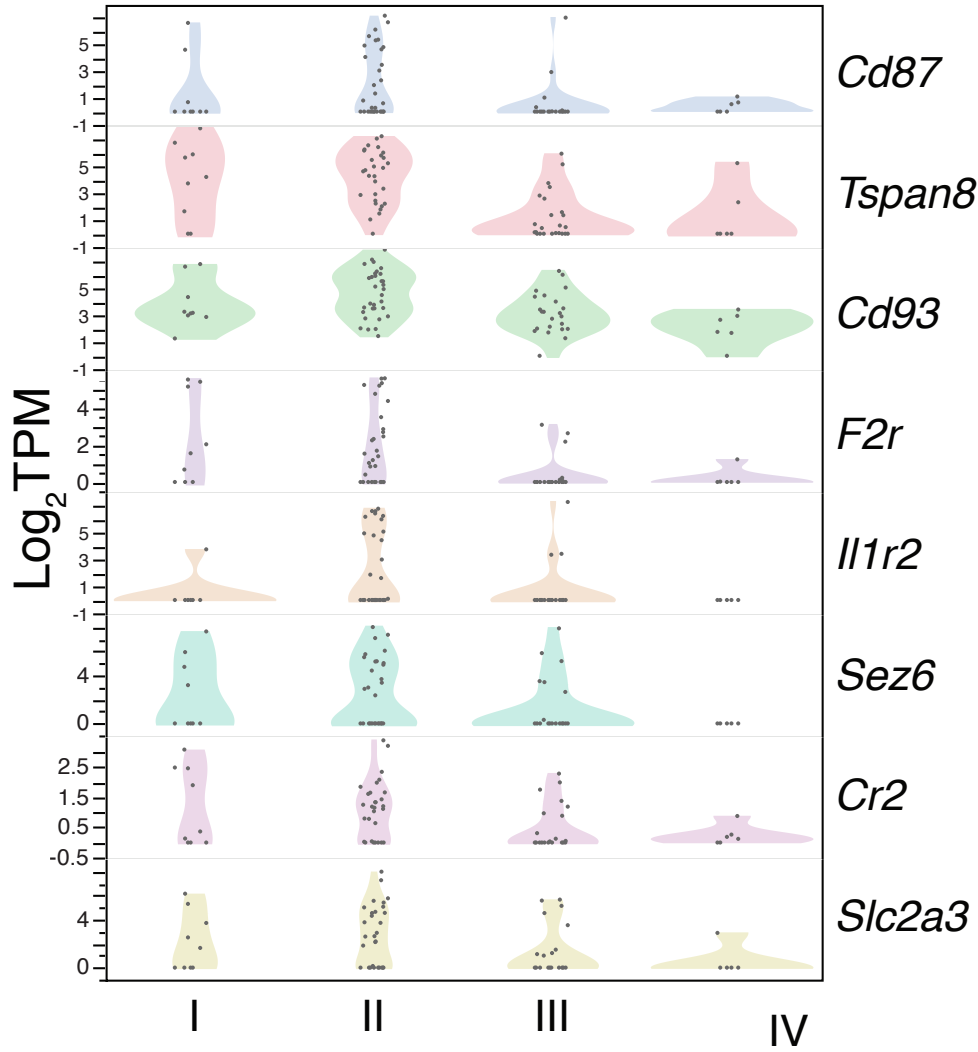


Figure S7. GO analysis for genes enriched in *Gfra1*⁺ cells as compared with *Gfra1*⁻ cells.

The size of each point represents the number of genes present. All the data points have adjusted p-value less than 0.05.

Figure S8

A



B

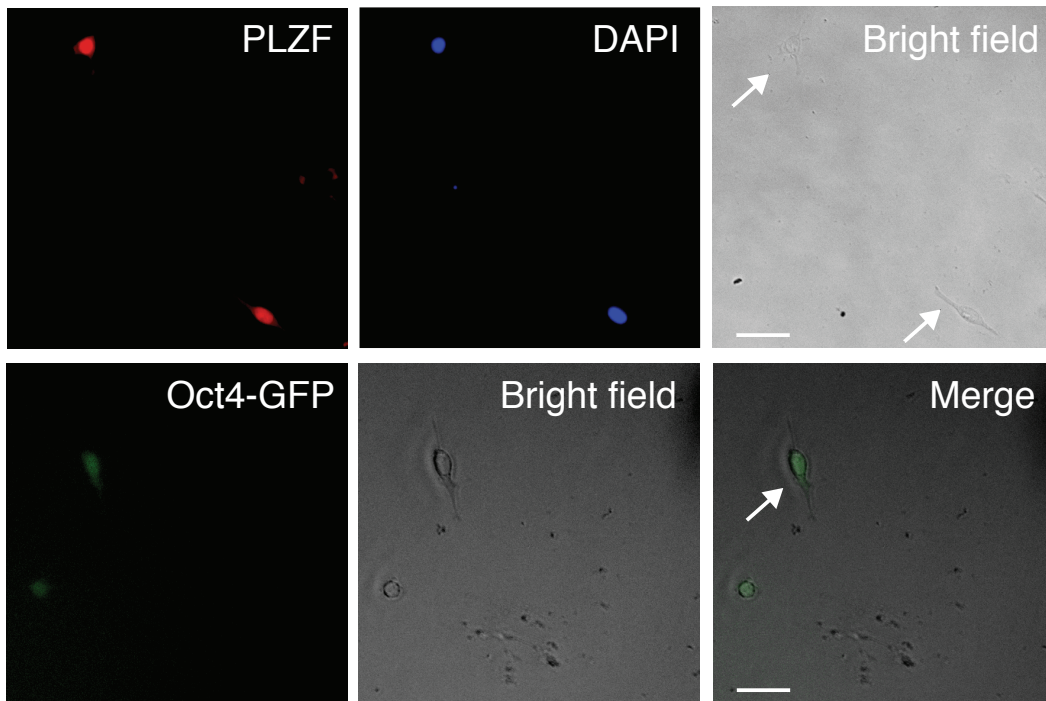


Figure S8. Identification of candidate surface markers in the primitive stem cell population (related to Figure 5).

(A) Violin plots illustrating the distribution of candidate surface marker genes expressed specifically or predominantly in Cluster II subpopulation. (B) Cultured CD87+ cells with protrusions are Oct4-GFP and PLZF positive (related to Figure 5D).

Figure S9

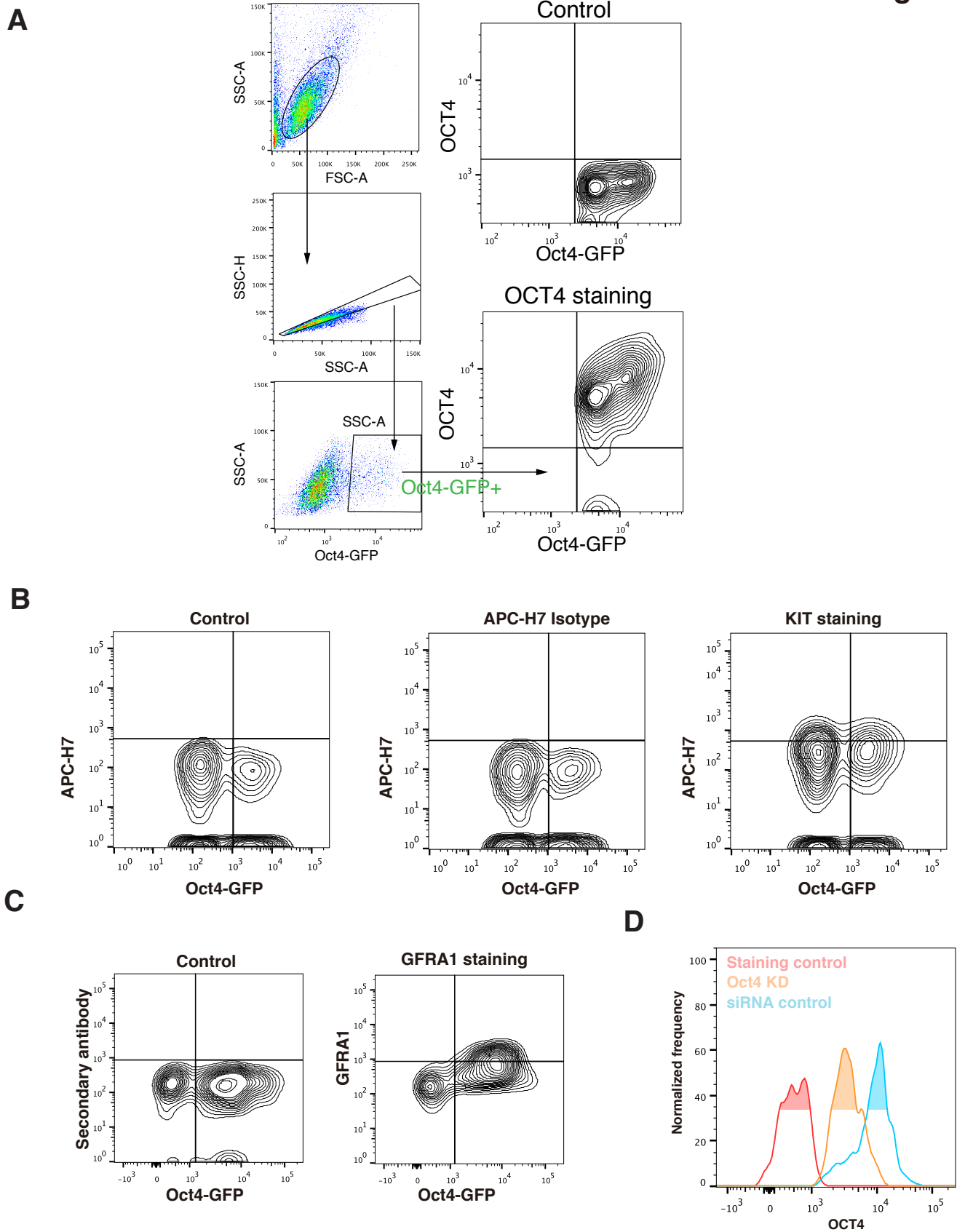


Figure S9. Characterization of Oct4-GFP-high germ cells at PND 5.5.

(A) Cells were isolated from the neonatal Oct4-GFP transgenic strain by FACS, stained with the OCT4 antibody, and analyzed by FACS. The results revealed a strong correlation between the GFP relative fluorescence intensity and the level of OCT4 immunostaining in individual cells. (B) Flow cytometry analysis of testicular cells without staining (control), labeled with APC-H7-conjugated rat IgG2b isotype control antibody (APC-H7 isotype) or APC-H7-KIT antibody (KIT staining). (C) Representative FACS results showing the control for GFRA1 gating (related to Figure 6D). (D) Representative FACS plot of cells with reduced OCT4 protein expression after the siRNA-mediated knockdown.

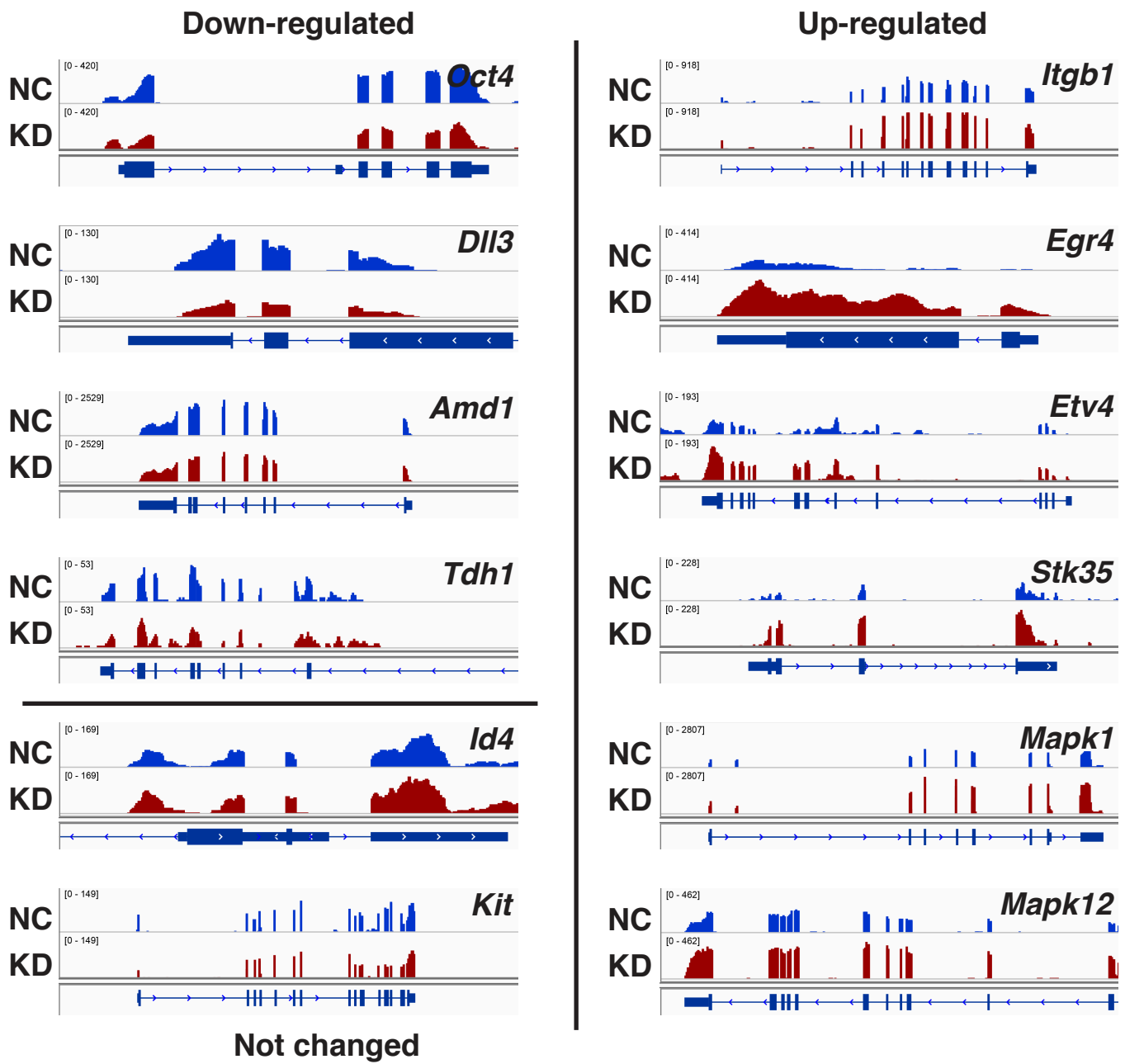


Figure S10. Genome browser representation of gene expression after *Oct4* knockdown.
Examples of individual genes showing the increased or decreased expression.

Supplementary Materials and Methods

Processing and analysis of single-cell RNA-seq data

- Read trimming and mapping

Raw reads were pre-processed with trim galore for trimming Illumina adaptor sequences (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). For the visualization and quantification of transcript coverage (related to Figure S2B and S3B), trimmed reads were mapped with Tophat2 to the mouse reference genome (mm9, Ensembl release 67) (Kim et al., 2013). Sequence read alignment BAM files were coordinate-sorted and indexed using Samtools (Li et al., 2009). Genome mapping results were visualized by using Integrative Genome Viewer (related to Figure S3B) (www.broadinstitute.org/igv/). Read coverage over gene body was calculated and plotted using `geneBody_coverage.py` in RSeQC (related to Figure S2B) (Wang et al., 2012). RNA-seq data quality metrics of each cell, including total reads, transcriptome mapped reads and transcriptome mapped rate was shown in Table S1.

- Gene expression matrices

Transcript quantification for RefSeq genes (mm9) was performed using Salmon (Patro et al., 2017) with Quasi-mapping-based mode using default parameter. Both the expected read count number and expression level (TPM, Transcripts Per Million) for each gene were calculated. To account for differences in total cellular RNA production across all cells, we performed an additional upper-quartile normalization on TPM using edgeR (Robinson et al., 2010).

- Quality control of single-cell RNA-seq data

Normalized TPM of the 81 libraries from lived cells as well as one doublet capture control was subjected to outlier identification analysis in R package Fluidigm SINGuLAR Analysis Toolset 3.0. 'identifyOutliers()' proceeds by trimming low-expressing genes until 95% of the genes that remain are above 1 TPM in half of the cells. A distribution of combined gene expression is created from these cells, and outliers are considered as cells whose median expression across the identified gene list is below the 15th percentile of the distribution. Using these routines, cross-contamination by capturing two cells at the same capture site can be ruled out (Figure S2E, F, sample C12_1.doublet). Additional 10 cells were excluded from downstream analysis as they may be indicative of low-quality cells that have aberrant technical properties compared to the majority of high-quality cells (Figure S2E, F). In total, 71 single cells passed quality control and were used in downstream analysis. The log-transformed gene-level in TPM (Table S2) and non-normalized expected read count matrix (GSE107711) of the remaining 71 cells

were used in the analysis throughout the whole study. To generate saturation plot, the corresponding number of millions of raw reads from each sample library were randomly selected using a python script and expression values for each gene were calculated using the same parameter. The number of genes with TPM greater than 1 was plotted.

- Clustering of single-cell RNA-seq data

Normalized TPM of the 71 single cells was imported into the Fluidigm SINGuLAR Analysis Toolset and converted into log-space. The “autoAnalysis()” command in Singular toolset was used to perform principal component analysis (PCA) to select the most variable genes. Briefly, the toolset calls the princomp R package (<http://stat.ethz.ch/R-manual/R-patched/library/stats/html/princomp.html>) and the 500 top-ranked PCA genes were selected on the basis of the maximum absolute value of each gene loading score in the first three eigenvectors (referred to as top 500 most variable genes). To detect different cell populations, we applied hierarchical clustering on cells and with the top 500 most variable genes (log transformed TPM) using Euclidean distance with Ward's method. The separation of clusters by hierarchical clustering was further confirmed by PCA analysis.

- Pseudotime ordering analysis by Monocle

Single cell expression data were analyzed with Monocle (v1.2.0) as previously described (Trapnell et al., 2014) with the default parameter. Briefly, TPM matrix of top 500 most variable genes used in aforementioned unbiased clustering was employed for pseudotime ordering. The Euclidean distance matrix among cells was calculated and projected on a high dimension space, in which each dimension represents one cell. Independent Component Analysis was applied to reduce the dimension of this space, which separates multivariate variables into additive subcomponents. Then, a minimum spanning tree (MST) of cells is constructed and the longest path within the MST is built. Finally, the cells are ordered by position for each cell on this longest path, i.e., pseudo-time ordering. To identify genes covaried significantly over pseudotime, Monocle first specified a Tobit-valued vector generalized linear model (VGLM) describing changes in each gene expression as a function of pseudotime followed by the comparison with a reduced model, in which pseudotime assignment for each cell was excluded. The p-values for each gene from the test were corrected for multiple testing by Benjamini and Hochberg's method. Resulted pseudotime were plotted using the `plot_spanning_tree()` function. Genes were plotted with respect to pseudotime using `plot_genes_in_pseudotime()`. For clustering genes by pseudotemporal expression pattern, we fit a smooth curve for each gene's expression trend as a function of pseudotime using Monocle. Heatmaps and gene

expression profile clustering were obtained using GENE-E package with the default parameter (<https://software.broadinstitute.org/GENE-E/index.html>).

- Differential gene expression analysis and GO analysis

Differential gene expression analysis was performed with the DESeq2 (Love et al., 2014). Briefly, expected read count matrix of all expressed genes 71 single cells and cell group assignment were subjected to DESeq2 analysis with the default parameter. Importantly, all expressed genes were included in this analysis. A significance p-value threshold level of 0.05 was applied. Lists of genes differentially expressed in each of the comparisons were used for gene ontology (GO) analysis using with ToppGene (Chen et al., 2009) or ClueGO app in Cytoscape 3.0.1(Shannon et al., 2003). A hypergeometric test with the Benjamini and Hochberg false discovery rate (FDR) was performed using the default parameters to adjust p-value.

- Weighted gene co-expression network analysis

R package "WGCNA" was used to construct the weighted gene co-expression network (Langfelder and Horvath, 2008). The log-transformed TPM expression matrix of all expressed genes in 71 single cells was used as input for the analysis. First, pickSoftThreshold() function was used to determine the soft-threshold power with networkType as "signed". Then the adjacency and TOM similarity matrices were generated based on soft-threshold power = 10. Hierarchical clustering of the genes was performed based on the TOM dissimilarity measure and the Dynamic Hybrid Tree Cut algorithm was used to identify the module. We summarized the expression profile of each module by representing it as the first principal component (referred to as module eigengene). The modules with eigengenes correlated with pseudotime were selected. The network connections among the most connected genes in each module were exported and visualized using Cytoscape 3.0.1.

Bulk RNA-Seq

Low input RNA-Seq library preparation was conducted according to the modified protocol of Smart-seq2 (Picelli et al., 2014). In brief, the mixture of 1 μ l of 10 mM oligo-dT30VN primer (5'-AAGCAGTGGTATCAACGCAGAGTACT30VN-3'), 1 μ l of 10mM dNTP mix and 10ng extracted RNA in 5 μ l reaction volume was incubated at 72°C for 3 min. Then, 5 μ l of the reverse transcription mixture (200U SuperScript II reverse transcriptase, 20U RNase inhibitor, 1x Superscript II first-strand buffer, 5 mM DTT, 1 M Betaine, 6mM MgCl₂, 1 μ M TSO (5' - AAGCAGTGGTATCAACGCAGAGTACATrGrG+G-3')) was added into the tube for RNA

reverse transcription. The tubes containing total 10 μ l mixture were put into ProFlex PCR System (Thermo Fisher Scientific) and the reaction condition was as follows: 42°C for 90 min, followed by ten cycles of 50°C for 2 min and 42°C for 2 min. The final enzyme inactivation was performed at 70°C for 15 min. After finishing the first-strand reaction, the 15 μ l PCR preamplification mix (1x KAPA HiFi HotStart ReadyMix, 0.1 μ M IS PCR primers (5'-AAGCAGTGGTATCAACGCAGAGT-3')) was added into the tube containing 10 μ l of the first-strand reaction. PCR preamplification reaction was performed in ProFlex PCR System by using the following program: 98°C for 3 min, 12 cycles of 98°C for 20s, 67°C for 15s and 72°C for 6 min, with the final extension cycle of 72°C for 5 min. Amplified cDNA was purified by using Ampure XP beads (1:1 ratio) according to the manufacturer's manual and was eluted in 15 μ l of ddH₂O. The yield of amplified cDNA was measure by Qubit3.0 fluorometer. Amplified cDNA with good quality were used for library construction with TruePrep DNA Library Prep Kit V2 for Illumina (Vazyme) following the instruction. The quality of all constructed libraries was assessed by Agilent Bioanalyzer 2100 with High Sensitivity DNA Kit (Agilent Technologies). Qualified libraries were sequenced on Illumina HiSeq2000 to generate 150bp pair-end reads. Raw reads were pre-processed with trim galore for trimming Illumina adaptor sequences (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). The processed reads were mapped to mouse (mm9) genome. Counts were assigned to genes defined by the Ensembl (release 67) annotation using RSEM (Li and Dewey, 2011). Differentially expressed genes were identified with edgeR package (Robinson et al., 2010). Soft clustering using differentially expressed genes was obtained using the Mfuzz package implemented in R (Kumar and M, 2007). GO enrichment was performed ToppGene Suite (Chen et al., 2009). A hypergeometric test with the Benjamini and Hochberg false discovery rate (FDR) was performed using the default parameters to adjust p-value.

Table S1. RNA-seq data quality metrics of each cell, including total reads, transcriptome mapped reads and transcriptome mapped rate.

[Click here to Download Table S1](#)

Table S2. Expression matrix in the form of log-transformed TPM of 71 single cell transcriptomes.

[Click here to Download Table S2](#)

Table S3. Log-transformed TPM values of the top 500 most variable genes.

[Click here to Download Table S3](#)

Table S4. Expression matrix of 48 genes across 96 cells in Biomark single-cell qRT-PCR experiment.

[Click here to Download Table S4](#)

Table S5. Results of DEG and GO analysis comparing gonocytes(Cluster I) with stem cell pool (Cluster II).

[Click here to Download Table S5](#)

Table S6. Results of DEG and GO analysis comparing primitive stem cell pool (Cluster II) with intermediate progenitor pool (Cluster III).

[Click here to Download Table S6](#)

Table S7. Result of DEG and GO analysis comparing undifferentiated spermatogonia pool (Cluster II and III) with differentiation-primed cells (Cluster IV).

[Click here to Download Table S7](#)

Table S8. GO analysis of pink module and green module genes identified from WGCNA analysis.

[Click here to Download Table S8](#)

Table S9. Results of DEG and GO analysis comparing CD87+ and CD87- subpopulations.

[Click here to Download Table S9](#)

Table S10. Results of Mfuzz clustering and GO analysis using the DEG among Oct4-GFP-low, Oct4-GFP-mid and Oct4-GFP-high cells.

[Click here to Download Table S10](#)

Table S11. Results of DEG and GO analysis comparing *Oct4* knockdown cells with control.

[Click here to Download Table S11](#)

Table S12. Antibodies used in immunostaining and FACS experiments.

[Click here to Download Table S12](#)

Table S13. Primers for qRT-PCR analysis.

[Click here to Download Table S13](#)

References:

- Chen, J., Bardes, E.E., Aronow, B.J., Jegga, A.G., 2009. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* 37, W305-311.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14, R36.
- Kumar, L., M, E.F., 2007. Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics* 2, 5-7.
- Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. *Bmc Bioinformatics* 9.
- Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *Bmc Bioinformatics* 12.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Genome Project Data Processing, S., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.
- Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550.
- Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14, 417-419.
- Picelli, S., Faridani, O.R., Bjorklund, A.K., Winberg, G., Sagasser, S., Sandberg, R., 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* 9, 171-181.
- Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13, 2498-2504.
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S.Q., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., Rinn, J.L., 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology* 32, 381-U251.
- Wang, L., Wang, S., Li, W., 2012. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 28, 2184-2185.