# Single-cell RNA-Seq Resolves Cellular Heterogeneity and Transcriptional Dynamics during Spermatogonia Stem Cells Establishment and Differentiation

Jinyue Liao[1, 2], Shuk Han Ng[1, 2], Jiajie Tu[1, 2], Alfred Chun Shui Luk[1, 2], Yan Qian[1, 2], Nelson Leung Sang Tang[5], Bo Feng[1,6],Wai-Yee Chan[1-4], Pierre Fouchet [7-10], Tin-Lap Lee[1-4*]

[1]Developmental and Regenerative Biology Program, School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China

[2]The Chinese University of Hong Kong – Shandong University (CUHK-SDU) Joint Laboratory on Reproductive Genetics, Shatin, Hong Kong SAR, China

[3]Key Laboratory for Regenerative Medicine (Jinan University – The Chinese University of Hong Kong), Ministry of Education of the People's Republic of China, China

[4]Joint CUHK-UoS (University of Southampton) Joint Laboratories for Stem Cells and Regenerative Medicine, The Chinese Univesrity of Hong Kong, Shatin, Hong Kong SAR, China

[5]Department of Chemical Pathology, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China

[6]SBS Core Laboratory, CUHK Shenzhen Research Institute, Shenzhen, China

[7]Commissariat à l'Energie Atomique, Direction des Sciences du Vivant, Institut de Radiobiologie Cellulaire ET Moléculaire, Laboratoire Gamétogenèse, Apoptose, Genotoxicité, 92265 Fontenay-aux-Roses, Paris, France

[8]Institut National de la Santé ET de la Recherche Médicale, Unité 967, Fontenay-aux-Roses, France

[9]Université Paris-Diderot, Paris 7, Fontenay-aux-roses, France

[10]Université Paris-Sud, Paris 11, Fontenay-aux-roses, France

[*]To whom correspondence should be addressed. Tel: +852 39434436; Fax: +852 26035139; Email: leetl@cuhk.edu.hk

## SUMMARY

In mammals, the transition of gonocytes to spermatogonia and subsequent differentiation provide the foundation of spermatogenesis. Disturbance in this process has been suggested to cause infertility and testicular germ cell tumor. However, systematic understanding on the cellular and molecular basis of this process is still limited, mainly impeded by the asynchrony in development and the lack of stage-specific markers. Using single-cell RNA sequencing on Oct4-GFP+/KIT- cells isolated from PND5.5 mice testes, we dissected the cellular heterogeneity of early germ cells and established molecular regulations during the gonocyte-spermatogonial transition and early spermatogonia differentiation. We demonstrated that gonocyte-spermatogonial transition was characterized by gene expression change related to apoptosis, cell cycle progression, and regulation of migration processes. Pseudotime analysis reconstructed developmental dynamics of the spermatogonial populations and unraveled sequential cellular and molecular transitions. We also uncovered an unexpected subpopulation of spermatogonia primed to differentiation within the undifferentiated compartment, which is characterized by the lack of self-renewal genes and enhanced *Oct4* expression and RA responsiveness. Lastly, we revealed a novel cellular state during the germ cell KIT transition, in which cells have initiated differentiation but still retains the expression of molecular determinants of self-renewal. Our study thus provides a novel understanding of cellular and molecular changes during spermatogonia establishment and a comprehensive resource for studying early male germ development and related diseases.

## KEYWORDS

Single cell, transcriptomes, spermatogonia, differentiation, development

## 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 SSC is derived from the transition of postnatal gonocytes (Kluin and Derooij, 1981). Disturbances in cell fate decision during this process can lead to significant clinical consequences. In cryptorchidism, gonocyte transition into SSC is thought to be inhibited, which may result in infertility after puberty (Kamisawa et al., 2012). Defective SSC formation from gonocytes might be the cause of testicular germ cell tumors (TGCT) (Rajpert-de Meyts and Hoei-Hansen, 2007). SSC self-renewal and differentiation must also be sophistically balanced to maintain normal spermatogenesis and avoid tumorigenesis (Ishii et al., 2012).

The precursors of SSC pool arise from primordial germ cells (PGCs) during embryogenesis. The PGCs then differentiate to gonocytes, and quickly enter quiescent state after E13.5. Shortly after birth, gonocytes resume mitotic activity and begin migrating to the basal lamina of the testicular cords. Once they become resident at the basement membrane, the cells become spermatogonia including self-renewing SSCs (Culty, 2009; Manku and Culty, 2015; Saitou and Yamaji, 2012). A major challenge in deciphering the cellular and molecular mechanisms during this process is that the germ cells in neonatal testes are not developmentally synchronized, resulting in the coexistence of germ cells at different developmental stages (Nagano et al., 2000). It is impossible to visually distinguish the co-existing gonocytes and spermatogonia at various stages as they share similar morphology (Culty, 2009; Manku and Culty, 2015) and there is significant heterogeneity regarding marker expression among the germ cell populations (Hermann et al., 2015; Nakagawa et al., 2007; Suzuki et al., 2009; Yoshida et al., 2006a). This makes the population-based approach for molecular characterization infeasible, thereby hampering the pace of research in this field.

To resolve cellular heterogeneity and capture developmental dynamics, we have undertaken single-cell RNA-Seq to tackling this long-standing question through revealing biologically intermediate states of germ cell development from a single chronological time point (Li et al., 2016; Llorens-Bobadilla et al., 2015; Shin et al., 2015). We profiled 71 single-cell transcriptomes of germ cells from mouse testes on postnatal day (PND) 5.5, which constitute a mixture of gonocytes undergoing transition into spermatogonia and undifferentiated spermatogonia, including a subset of stem cells and committed progenitors.

This has allowed us to assign single cells to distinct cell states and reconstruct their developmental chronology. Our analysis further delineated the molecular cascades governing gonocyte-spermatogonial transition (GST) and the balance of SSC self-renewal and differentiation. Moreover, we characterized the transcriptional dynamics during germ cell KIT- to KIT+ transition and uncovered a unique cell state during spermatogonia differentiation, in which self-renewal and differentiation signatures co-expressed.

## RESULTS

### Single-cell transcriptioanal profiling of germ cells from mouse testes at PND5.5

GST occurs between PND3 and PND6, during which gonocytes either give rise to the undifferentiated spermatogonia or directly into differentiating spermatogonia that enter the first wave of spermatogenesis(de Rooij, 1998; Yoshida et al., 2006b). It's proposed that the primary SSCs first appear at PND5 and become enriched at PND7–PND10(Culty, 2009, 2013). Therefore, we decided to investigate PND5.5 because germ cells at this single developmental time point should represent a continuum of differentiation from gonocytes to differentiating spermatogonia and single-cell transcriptome profiling of the cells would allow reconstruction of developmental process (Figure S1A). Oct4-GFP transgenic mouse line is a great tool to study early germ cells development as it allows visualization of endogenous *Oct4* expression, which is restricted to undifferentiated germ cells at prepubertal stages(Ohbo et al., 2003) . Since this line of mice has not been evaluated at PND5.5, we first characterized the Oct4-GFP-expressing cells in the testis at this stage. We found that majority of the Oct4-GFP+ cells were localized at the basement membrane, where a small portion of cells resided within the center of testicular cords or positioned adjacent to the basement membrane with protrusions towards the periphery (Figure 1), implying the existence of transitioning gonocytes yet to finish the migratory phase. We then used several known markers to further define Oct4-GFP+ cells. We found that Oct4-GFP+ cells exhibit significant overlap with PLZF, a marker for the undifferentiated state (Figure 1A, D). In addition, coexpression of Oct4-GFP with the important self-renewal factor GFRA1 was also observed (Figure 1B, D). We also found that significant numbers of Oct4-GFP+ cells begin to express the differentiation marker KIT, suggesting that PND5.5 testis contains differentiating germ cells (Figure 1C). Therefore, at PND5.5, Oct4-GFP+ cells primarily label most, if not all, germ cells at different developmental stages, making this line suitable for studying the GST and subsequent differentiation *in vivo*.

To maximize the representation of the most undifferentiated germ cells, we isolated Oct4-GFP+/KIT- cells from PND5.5 testis using fluorescence-activated cell sorting (FACS) (Figure 1E, S1B-E). Single cells were then captured by Fluidigm C1 IFC system, which were subjected to automated RNA isolation, cDNA synthesis, library construction and RNA-sequencing (Figure S2A). 71 cells passed a number of stringent quality control assessments that ensured high-quality single-cell transcriptome data and were included in subsequent analyses (Figure S2B-F, Table S1). Although *Oct4* transcripts were absent in 19 of the 71 sequenced cells, likely because the expression was below the threshold of detection at this sequencing depth, *Plzf* transcripts were abundantly expressed in all cells, indicating their germ cell nature (Figure S3). Furthermore, a significant portion of germ cell-related genes such as *Bcl6b*, *Egr2*, *Sohlh1*, and *Csf3r* showed substantial variability, while housekeeping genes (*Actb*, *Gapdh*, *Rpl7*, and *Rps2*) showed uniform expression across single cells, illustrating the significant gene expression heterogeneity in early male germ cell populations as previous reported(Hermann et al., 2015) (Figure S3).

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

We next sought to interrogate the heterogeneous cellular composition. Genes with the highest loadings in principal component analysis (PCA) were analyzed by unsupervised hierarchical clustering, which identified four different subpopulations (Figure 2A, B), including 2 smaller clusters (I and IV) and 2 larger clusters (II and III). Aformention characterization implies that majority of cells captured should be spermatogonia at PND5.5 and may be linked to Cluster II and III. Visual inspection of the heatmap suggests that distinctions among Cluster II and III appeared to predominantly reflect a gradient of expression in mRNA levels, suggesting the two subsets of spermatogonia could represent temporally distinct populations. Thus, Cluster II cells with higher expression of genes associated with SSC self-renewal, such as *Etv5* and *Ccnd2* (Figure 2C), are likely to be the primitive stem cells. In contrast, Cluster III cells, which display higher expression of differentiation regulator *Sohlh1*, are the immediate progenitors (Figure 2C). The expression level of *bona fide* adult stem cell markers in Cluster II and III also reflected their cellular identities (Abid et al., 2014; Aloisio et al., 2014; Oatley et al., 2011). The majority of cells with high *Id4* expression fell into Cluster II (15 out of 23 cells, log2(TPM)>10)), consistent with the recent observation that cells with high level of *Id4* represent true SSCs ((Helsel et al., 2017)). *Erbb3* and *Pax7* were also preferentially expressed in this primitive stem cell pool

(Figure 2C). We further defined Cluster IV cells as differentiation-primed spermatogonia, based on their close proximity in the PCA space to the Cluster III progenitors and the absence of *Gfra1* (Figuer 2B and C), which are likely to be primed to differentiation and lose self-renewal potential (Meng et al., 2000). The last cell subset (Cluster I) identified in the present experiment corresponds to the gonocyte, based on the PC2 distance between them and the spermatogonia (Cluster II, III and IV) and the fact that existence of gonocyte population should be reflected in single-cell transcriptomes of the captured cells (Figure 2B). It's also notable that *CD81*, which is reported to mediate migratory gonocytes movement (Tres and Kierszenbaum, 2005), showed higher expression in Cluster I (Mean TPM: Cluster I=62, cluster II=18, Cluster III=22, and Cluster IV=10) (Figure 2C). A single-cell characterization of germ cells in the neonatal testis was recently published using a different reporter ID4-GFP (Song et al., 2016). This provides a unique opportunity to explore the relationship among germ cells at different developmental stages. When we projected cells from our study onto this reference dataset, we observed that Cluster I cells bear a property intermediate between that of P3 gonocytes and P5.5 spermatognoa (Cluster II), while Cluster III and IV cells more closely resemble the P7 spermatogonia, affirming the gonocyte nature of Cluster I cells (Figure 2D). We further reconstructed the transitional dynamics and ontogeny relationship of the cell populations by Monocle (Trapnell et al., 2014b). The result supported our proposed assignments, where the trajectory path began with transitional gonocytes (Cluster I), passing through undifferentiated spermatogonia (Cluster II and III) and ending at differentiation-primed spermatogonia (Cluster IV) (Figure2E, left). Monocle also revealed three major development stages that might correspond to GST (State 1), SSS self-renewal (State 2), and differentiation commitment (State 3) (Figure 2E, right).

Importantly, to exclude that any population was merely the result of biological or technical artifacts, we validated our findings in a second set of PND 5.5 germ cells. We chose a limited number of genes to identify each subpopulation and sorted an additional 96 cells and quantitatively measured expression of these genes utilizing Biomark microfluidic quantitative PCR platform (Figure S4A-C, Table S2). Our analysis confirmed that the pattern of PCA and hierarchical clustering followed the same structure as observed in the RNA-Seq experiment (Figure S4D and S4E). This means that the cell populations and their relative relationships are conserved in this biological replicate. Similarly, the gene expression generally recapitulated patterns we identified in the initial RNA-Seq experiment (Figure S4F).

**Transcriptional regulation of gonocyte-spermatogonial transition**

The hallmarks of perinatal gonocytes are their relocation from the center to the periphery of seminiferous cords and their cell cycle reentry (Nagano et al., 2000). To directly link Cluster I subpopulation identified in our data to gonocyte, we first assessed expression of migration-related genes within subpopulations. We focused on the cells resided in gonocyte-spermatogonial transition state as revealed by Monocle (State 1) (Figure 2E) and compared gene expression between the first half (9 cells from Cluster I) and the second half (9 cells from Cluster II) in the trajectory. Top-ranked differential expressed genes identified were postulated to regulate cell migration (Figure 3A). For example, this analysis revealed *CD81* as one of the top genes. Cluster I also expressed higher *Crk*, an adapter protein of PDGFR, which might be involved in PDGF-mediated gonocyte survival and migration (Basciani et al., 2008). Moreover, among genes up-regulated in the gonocytes, the top GO categories enriched was "Integrin", which contained genes *Nf2*, *Itgb1*, *Utrn*, *Mfge8*, and *P4hb*. *Itgb1* (β1-integrin) has been implicated as an essential adhesion receptor required for both gonocytes migration and SSC homing (Kanatsu-Shinohara et al., 2008; Tres and Kierszenbaum, 2005). 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).

To further confirm the gonocyte nature of Cluster I, we perform Gene Set Enrichment Analysis (GSEA) and found that gene sets associated with cell cycle and protein synthesis (aminoacyl-tRNA biosynthesis) related pathways were underrepresented in the gonocyte population, implying a resting phase of these gonocytes prior to their transition to spermatogonia (Figure 3B). Interestingly, gonocytes were shown to express lower levels of basal transcriptional factors (TFIIB) like *Taf4b*, which encodes a chromatin-associated factor and has been demonstrated to regulate early germ stem cell specification (Falender et al., 2005). To experimentally validate the existence of this population of "cell cycle-low" cells, we examined the proliferation marker Ki-67 in Oct4-GFP+/KIT- cells. In line with the presence of 12.6% (9 in 71 cells) cells (Cluster I) within the cells captured, we observed a similar fraction of Oct4-GFP+/KIT- cells that were negative for the Ki-67 expression (Figure 3C). Immunostaining of the testis revealed that the centrally located cells contain greater proportion of Ki-67-negative cells than peripherally located cells (Figure 3D and 3E). GO analysis identified significant up-regulation of genes associated with death receptor pathway in gonocytes, which contained *Fadd*, *Tnfrfs1b*, *Icam1*, *Madd*, *Stx4a*, and *Skil*. We also found elevated expression of genes involved in intrinsic apoptosis signaling (*Itm2b*) (Figure 3F)

(Fleischer et al., 2002). Genes associated with DNA damage/repair pathway were shown to be repressed in gonocytes, including *Ercc1*, *Trp53*, and *Bre* (Figure 3F). These results are in line with the notion that 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 germ cells during GST.

**Characterization of stem cell and progenitor spermatogonia by pseudotime ordering and network analysis**

To investigate the underlying molecular features of the progression from the early, stem-like stage to the progenitor stage of spermatogonia, we analyzed differentially expressed genes between Cluster II and Cluster III. We identified 511 genes differentially expressed, which included *Etv5* and *Pax7* (Table S3). GO analysis revealed significant enrichment for signatures associated with integrin complex, Rap1 signaling and lipid metabolism in primitive stem cell pool (Figure S5A), whereas TGF signaling (e.g. Smad3 and Creb1) overrepresented in progenitor pool (Figure S5B). GSEA analysis further revealed that JAK/STAT signaling, Cell adhesion molecules (CAMs), Neuroactive ligand-receptor interaction, and RIG-I-like receptor signaling pathway were upregulated in stem cell cluster, while ribosome biogenesis pathway upregulated in progenitor subset (Figure 4A and S6A). To show coordination on gene expression during spermatogonia differentiation at PND5.5, we applied Monocle to position genes with similar expression trends along differentiation. We revealed three gene regulatory events following the self-renewal phase during spermatogonia differentiation (Figure 4B). During the initiation phase, known self-renewal genes such as *Ccnd2* and *Etv5* as well as signaling pathways, including MAPK/Erk1 and VEGF were repressed, as reported previously (Figure 4C) (Caires et al., 2009; He et al., 2008). Concurrently, a series of genes highly enriched for cell cycle and cell division was activated and then repressed before commitment of differentiation, suggesting an expansion phase precedes terminal differentiation of spermatogonia. The cells passed through an intermediate phase and transiently expressed an elevated level of *Nanos3* and genes related to DNA repairing pathway (*Bre* and *Trp53*) (Figure 4C). Previous work has indicated that Nanos3 preferentially expresses in $A_{al}$ spermatogonia that are committed to differentiation in adult testis (Suzuki et al., 2009). The commitment phase featured complete repression of SSC self-renewal genes and elevation of differentiation genes, such as Sohlh1, a point we return to below. The interaction between cell cycle and differentiation promted us to dissect the cell

cycle stages by PCA-based approach (Buettner et al., 2015). We found that the majority of cells (70%, 14 cells out of 20) in G2/M phase were from Cluster II, from which we inferred that more cells in G2/M were sampled from primitive stem cell pool (Figure S6B). This implied that stem cells tend to have a relatively shorter G1 phase and proliferate faster, which is reminiscent of embryonic and adult stem cells (He et al., 2009; Lange et al., 2009).

To identify novel regulatory networks associated with SSC self-renewal and differentiation, we applied Weighted Gene Co-expression Network Analysis algorithm on transcriptomes of spermatogonia (WGCNA) (Langfelder and Horvath, 2008). In search for modules that correlate with our differentiation model, we identified the pink module that enriched in primitive stem cells and decreased across pseudotime (Figure S6C, left). GO terms of this module are associated with cell proliferation, regulation of developmental process and neuron projection. This unbiased approach highlighted key regulators as central to this network, including *Etv5*, *Ccnd2 and Gfra1* as well as the presence of GDNF signaling regulation (19 out of 107 genes are GDNF-regulated), which confirmed the current knowledge of SSC self-renewal (He et al., 2009). It would be interesting to characterize functional significance of other highly connected hub genes such as *Tcl1*, *Spry4*, *Nefm*, *Dusp6*, *Tmem59l* and *Impact* (Figure S6C, left). The second module was enriched for genes involved in mitotic cell cycle and RNA binding. It was mostly repressed in differentiation-primed cells, suggesting a general reduction in proliferation rate upon differentiation commitment (Figure S6C, right). Hub-gene-network analysis indicated *Tex13* and *Rbm35a* are the most highly connected genes, both of which might exert potential functional roles in spermatogonia differentiation (Fagoonee et al., 2013; Kwon et al., 2014).

**Identification and validation of novel surface-marker for primitive stem cells**

This analysis also enables us to identify putative markers or regulators that may be specific to the earlier, potentially self-renewing spermatogonia (Cluster II). Among the multitude of potential novel surface markers is *Tspan8*, whose expression has recently been shown to correlate with stem cell activity in neonatal testis (Figure S7A) (Mutoji et al., 2016). To ensure that novel subgroup-enriched mRNAs truly represent protein expression and to enhance the usefulness of our discovery, we screened commercially available antibodies for surface markers that could be used for histological characterization and isolation of stem cells. We identified antibodies for IL1R2, F2R, CD93, and CD87, which were specifically or predominantly expressed in Cluster II. Immunostaining signals of CD87, F2R, and IL1R2 showed that they are more restrictedly expressed when compared to Oct4-GFP signal, which

therefore establishes novel cellular heterogeneity within undifferentiated spermatogonia (Figure 5A, S8). In line with development model we reconstructed, FACS-sorted cells positive for CD87 and CD93 showed higher expression of *Etv5* or *Ccnd2* and reduced expression of *Sohlh1* and *Ngn3* (Figure 5B-C, S7B-D).

We found CD87 particularly interesting as it has been implicated as a marker for a subset of Hematopoietic stem/progenitor cells (HSPCs) and functional in hematopoietic cell migration and homing (Tjwa et al., 2009). To gain more insight of CD87's function in germ cells, we examined its surface expression at postnatal stages by FACS. All Oct4-GFP+ cells expressed CD87 at early stages and the proportion gradually decreased from PND4.5 to PND7.5 and became nearly undetectable by PND14 (Figure 5D). This data indicated that CD87 participated in spermatogenesis shortly after birth and had an intimate relationship with GST or spermatogenesis initiation. We next explored whether CD87 participated in germ-cell migration. We performed matrix adhesion assays using germ cell (GS) cultured on matrigel and examine the formation of the cellular protrusion, which has been described as an essential step for cell migration(Ladoux and Nicolas, 2012). Interestingly, CD87+ cells display cellular protrusion more frequently than CD87- cells, suggesting a greater migration potential (Figure 5E). The foregoing analysis predicted that primitive stem cells tend to have a relatively shorter G1 phase and proliferate faster. Therefore, we investigated whether CD87+ cells, as a proxy to primitive stem cells, differed from CD87- cells in regard to cell-cycle activity. FACS analysis confirmed a higher frequency of cells in G2/M phase in CD87+ versus CD87- cells, likely reflecting a faster proliferation rate (Figure 5F). Indeed, we observed significantly higher growth rate for GS culture derived from CD87+ cells (Figure 5G).

These results cross-verified our development reconstruction, and we successfully identify surface markers that can be readily used to study most primitive spermatogonia or SSCs.

**Gene expression changes during early spermatogonia differentiation**

To identify the molecular regulation of spermatogonia differentiation, we compared gene expression between "differentiation-primed" (*Gfra1-*/*Kit-*, Cluster IV) population and its undifferentiated counterpart (*Gfra1+*/*Kit-*, Cluster II and III). This analysis revealed down-regulation of SSC self-renewal related genes, such as *Etv5* and *Ccnd2*, as well as embryonic stem (ES) cell self-renewal regulator *Tcl1* (Figure 6A) (Table S4). GSEA analysis showed a strong enrichment of glial cell line-derived neurotrophic factor (GDNF) regulated genes in

*Gfra1*+/*Kit*- cells, which confirmed the loss the responsiveness to GDNF regulation in *Gfra1*-/*Kit*- cells (Figure 6B) (Oatley et al., 2006). On the other hand, expression of differentiation regulator *Rarg* was increased in *Gfra1*-/*Kit*- cells, which might contribute to the heterogeneous differentiation competence in response to retinoic acid (RA) that resembled steady-state spermatogenesis in adult (Ikami et al., 2015). We found genes related to Ras signaling pathway (FDR= 1.088E-3) enriched in *Gfra1*+/*Kit*-, which is consistent with the previous finding (Lee et al., 2009). Moreover, a significant enrichment for cell-cell communication and GnHR, as well as neurotrophic signaling related genes were also observed (Figure S9). Functional annotation of genes upregulated in *Gfra1*-/*Kit*- revealed enrichment for DNA methylation e.g. *Tdrd5*, *Dnmt3b*, and *Ehmt2* (*G9a*), consistent with its role in spermatogonia differentiation (Figure 6A) (Shirakawa et al., 2013). From the pseudotime analysis, it is evident that the transition from stem cells and progenitor to differentiation-primed cells entails a greater magnitude of the changes in gene expression (e.g. Complete lost of *Gfra1* and *Etv5* expression). GO analysis indicated other factors involved in mTOR pathway (*Eif4ebp1*, *Pdcd4*, *Cycs*, and *Prkab2*), which is in line with enhanced mTOR pathway activity in adult spermatogonia differentiation (Figure 4C) (Hobbs et al., 2010). Notably, *Nsbp1* and *Eno1*, which have been implicated in cell differentiation in other systems, were also elevated (Figure 4C) (Diaz-Ramos et al., 2012; Rochman et al., 2010).

**Elevated *Oct4* expression signifies early spermatogonia differentiation and enhanced RA responsiveness**

O*ct4* has previously been shown to be important for SSC maintenance, given that *Oct4* is down-regulated by RA-induced differentiation and the knockdown of *Oct4* results in differentiation (Dann et al., 2008). A priori, it might be expected that *Oct4* expression level would decrease along the differentiation process. Surprisingly, *Oct4* was higher expressed in differentiation-primed cells compared to the other subsets (Figure 6C). Since indexed FACS sorting showed that GFP intensity from the Oct4-GFP transgene positively correlates with endogenous *Oct4* expression at single cell level (Figure S4B and S4D), we sought to experimentally validate this prediction using the level of Oct4-GFP intensity (Figure 6D). As predicted by the single-cell data, germ cells sorted by FACS with higher levels of GFP fluorescence had significantly higher expression of *Oct4* and *Rarg* (Figure 6E). To further associate Oct4-GFP expression to the GFRA1 protein level, we pregated on Oct4-GFP+ cells and then examine GFRA1 staining in different subsets. Strikingly, we found that GFP-high

gated cells were largely depleted of GFRA1 expression (Figure 6D). We next investigated whether higher *Rarg* in differentiation-primed cells would contribute to different differentiation competence in response to RA *in vitro*. GS culture derived from PND5.5 Oct4-GFP+/KIT- cells contained a small fraction of cells showing KIT expression, which indicates spontaneous differentiation. Interestingly, KIT+ cells primarily reside in Oct4-GFP-high cells (Figure 6F, left). RA-induced differentiation led to an increased proportion of cells positive for KIT and this effect is more prominent in Oct4-GFP-high cells, suggesting an enhanced responsiveness to RA signaling in Oct4-GFP-high cells (Figure 6F, right). As an independent means to assess the possibility that Oct4-GFP-high cells are primed to differentiation in vivo, we examine Oct4-GFP expression through developmental stages. We observed that gonocytes expressing strong Oct4-GFP cells at PND3.5 and PND4.5 largely disappear by PND5.5, while another cell population with relatively lower Oct4-GFP level emerged at PND5.5, which represent newly formed spermatogonia. On the other hand, a small fraction of PND5.5 germ cells exhibiting higher GFP level expanded in number and this subset became more apparent at PND7.5 (Figure 6G). We investigated the two possible scenarios in which this cohort of cells either represents a very early population of gonocytes or a later differentiation-primed stage but display higher Oct4 level with unknown mechanism. We reasoned that blockage of differentiation of spermatogonia after PND4.5 should keep them in an undifferentiated state and lead to the reduction in the proportion of cells primed to differentiation but not affect early gonocyte development. We treated PND4.5 Oct4-GFP mice with WIN 18,446, which chemically inhibits RA synthesis and retains spermatogonia at their SSC state(Agrimson et al., 2017; Paik et al., 2014). As expected, KIT expression was down-regulated upon treatment, suggesting a lower level of RA activity and blockage of differentiation (Figure 6H, left, and S10). Importantly, WIN 18,446 treated PND5.5 testis not only lack differentiated Oct4-GFP-low cells that are detectable only after PND5.5 but are also essentially devoid of cells expressing the highest Oct4-GFP present in normal PND5.5 and PND7.5 testis (Figure 6H, right). These results, therefore, verified that Oct4-GFP-high population represented differentiation-primed cells but not gonocyte nor primary spermatogonia. Taken together, these results show that Oct4-GFP-high cells display molecular and functional properties as cells primed to differentiation. Elevated of Oct4 expression thus reflects a switch from self-renewing to differentiation commitment.

**Characterization of transcriptional dynamics during early germ cell KIT transition**

KIT expression has been classically indicated for distinguishing differentiating spermatogonia from their undifferentiated precursors (Rossi et al., 2000). However, our discovery of an unexpected differentiation-primed population in KIT- cells suggests a more complex model. Therefore, we extended our analysis to investigate transcriptional dynamics during KIT- to KIT+ transition at PND5.5 by bulk RNA-Seq. Key differentiation-related genes (e.g. *Kit*, *Stra8*, and *Cyp26a1*) as well as RA-induced genes showed preferential expression in KIT+ cells (Figure 7A-D) (Wang et al., 2016). In contrast, undifferentiated markers such as *Plzf* and *Ddit4* were downregulated during the KIT transition (Figure 7A-D). We also compared our data to a recent study on spermatogonia at PND7.5, when all the gonocytes have transitioned into spermatogonia (Table S5) (Kubo et al., 2015). Scatter plot of gene expression changes revealed the overall similarity and specificity of genes differentially expressed in common (R2 = 0.908) (Figure 7D). Strikingly, key niche factors of SSC (e.g. *Gfra1*, *Fgf1r*, and *Cxcr4*) decreased significantly during the KIT transition at PND7.5 but were barely changed at PND5.5 (Figure 7E). Concordantly, repression of *Etv5*, *T*, *Sox2*, *Bcl6b*, and *Eomes*, which are induced by FGF2 and GDNF, was exclusively detected during the PND7.5 KIT transition (Figure 7D) . Furthermore, systematic analysis of the genes displaying significant silencing during KIT transition only at PND7.5 (Figure 7D, shown in red) showed an over-representation of pathways associated with GDNF signaling pathway (Figure 7F) as well as the interaction between stem cells and niche (Figure 7G). These results indicated that although bulk KIT+ cells at PND5.5 have activated the differentiation pathway, they still retain the machinery at least at the transcriptional level to respond to the stem cell niche, which represents a unique phase of germ cell differentiation that has not been explored before.

## DISCUSSION

Previous single-cell gene expression analysis on neonatal spermatogonia at PND6 is restricted to only 172 genes, which limits the resolution to dissect the molecular heterogeneity (Hermann et al., 2015). In this study, we extended this analysis to whole-transcriptome profiling on Oct4-GFP+/KIT- cells at PND5.5 testes, which represent a broader undifferentiated germ cell population. In agreement with the recent study (Song et al., 2016), this approach allowed us to reveal early germ cell development trajectory in which migratory gonocytes transition into undifferentiated spermatogonia that subsequently proceed to differentiation.

Importantly, our study provides the global view of gene expression dynamics underlying GST, which permits elucidating important events that dictate the fate of gonocytes. First, we discovered several molecules that might potentially involved in gonocyte migration. Second, GST is accompanied by a global activation of transcription and change of cell cycle status. Third, our results showed that apoptosis-related genes are actively regulated during gonocyte transition. Particularly, transitional gonocytes seem to have a lower threshold for initiating apoptosis (elevated death receptor pathway) and increased susceptibility to DNA damage-induced apoptosis (reduced DNA repair pathway). This raises the possibility that transitional gonocytes possess the poised apoptotic machinery, which renders them more vulnerable to apoptotic cell death. Further studies using our resources will help shed light on the mechanistic nature of how signals regulate gonocyte cell cycle and migration during GST.

Globally, our data explicitly pinpointed many developmental hallmarks in the literature previously assumed to exist in neonates based on studies in adult spermatogenesis (Figure S11A). More importantly, we clearly *in silico* recaptured the cellular states and temporal gene expression pattern with a high-resolution and our observation is largely in agreement with the recent single-cell analysis in human adult SSCs(Guo et al., 2017). In corroboration of a recent 'hierarchinal model' of adult SSC, we showed that neonatal primitive stem cell and progenitor populations are distinguished by a gradient expression of SSC maintenance genes (*Id4*, *Gfra1*, *Etv5* and *Tspan*8) and differentiaoin-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 and would be enriched by cell surface markers that are specific to this population. We described a panel of cell-surface markers and characterized CD87 in more details as a proof of principle. Although it's tempting to predicte the biological relevance of these genes in controlling SSC maintenance, future studies by transplantation studies or exclusive lineage tracing are needed to assess their functional role. Adding to this picture, the transition from Gfra1+/Rarg$^{low}$ to Gfra1-/Rarg$^{high}$ (cluster IV) cells signify the early spermatogonia differentiation, which closely resembles the recently described differentiation model in adult spermatogenesis (Ikami et al., 2015). This subset of differentiation-primed cells notably display a surprisingly higher level of Oct4 but lacked expression of signature self-renewal promoting genes *Gfra1, Etv5*. Given the known role of *Oct4* in SSC maintenance, our observation thus raised the possibility that it's crucial that SSCs maintain *Oct4* at an appropriate level in order to preserve self-renewal capacity, which mirrors its function in ESC (Niwa et al., 2000). Future studies will still be

required to determine whether *Oct4*'s level is able to specify self-renewal and differentiation cell autonomously or dependently of differential extrinsic stimuli, such as RA signaling. Nevertheless, our finding suggests Oct4-GFP can be used as an excellent reporter system for real-time visualization of changes in cellular state during SSC differentiation. In the combination of bulk RNA-Seq experiment, our study further highlighted the complex dynamics of spermatogonial differentiation regarding KIT transition during early development. Contrary to later neonatal (PND7.5) and adult stage, KIT+ cells at PND5.5 still retain part of the molecular determinants of self-renewal as suggested by expression of niche receptors and their dependent factors. It's possible that germ cells at PDN5.5 are undergoing active cycling to populate the niche and exempted from turning off of the self-renewal program, which is similar to phenotypic plasticity that has been previously reported in the situation of germ cell transplantation or in vitro culture (Morimoto et al., 2009).

At the same time, this high-resolution data set provides a more comprehensive characterization of molecular cascades underlying early spermatogonia differentiation (Figure S11B). Our analysis showed that the regulation of energy metabolism is tightly coupled to spermatogonia differentiation. Genes related to lipid metabolism are highly enriched in 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 the ribosome biogenesis increased in progenitors compared to stem cell spermatogonia. Because translation rate is correlated positively with ribosome biogenesis, our results implied an enhanced need for global protein synthesis during differentiation. Interestingly, increased protein synthesis has been observed in GSC development in Drosophila (Sanchez et al., 2016). Investigating the cell cycle heterogeneity further demonstrated that differentiation of spermatogonia is intertwined with cell cycle states. Interestingly, cell cycle regulator *Ccnd2* was identified as the hub gene in self-renewal module (Figure S6C), which strongly suggested that it might serve as an important modulator for the coupling of spermatogonia differentiation and cell cycle progression. Our results, therefore, strengthened the hypothesis that faster cell cycle progression and a shorter G1 phase in stem cell pool could be accomplished by induction of cyclin D expression as reminiscent of the interconnection of cell fate decision and cell cycle demonstrated in other stem cell systems (Becker et al., 2006).

This study also provided an opportunity to improve our understanding of developmental disease processes, further attesting the importance of our work. Aberrant expression of *Ccnd2* is believed to contribute directly to the formation of testicular germ cell tumors (TGCTs) (Rodriguez et al., 2003). Similarly, another hub gene *Spry4* has been

proposed as one of the susceptibility genes for TGCTs (Karlsson et al., 2013). Moreover, the expression of the CD87 was significantly increased in TGCT, which may be relevant in testicular cancer progression (Ulisse et al., 2010). On the other hand, low level of ERCC1 or recruitment of FADD is likely to be key determinants of cisplatin sensitivity in testicular germ cell tumor (Spierings et al., 2003; Wood, 2011). Because defective gonocyte differentiation might be at the origin of testicular germ cell tumors, it is intriguing to speculate that intrinsic gene expression phenotype ($Ercc1^{low}$ and $Fadd^{high}$) of the transitional gonocytes may be retained in TGCTs and play a significant role in determining the cisplatin sensitivity phenotype.

In conclusion, we envisaged that our comprehensive data set and strategy would serve as a powerful tool to further dissect the heterogeneity in adult human and mouse spermatogonia, paving the way for identification of molecular mechanisms underlying adult male germ cell development and associated diseases in an unbiased and precise manner.

## MATERIALS AND METHODS

### Immunofluorescence Staining

Immunofluorescence staining of testes sections were performed as described previously (Oatley et al., 2011) in conjunction with the antibodies against KIT, GFRA1, and PLZF. Images were acquired using Leica SP2 confocal microscope.

### RNA Sequencing

Undifferentiated and differentiating germ cells from PND5.5 testes were purified using the method described previously (Garcia and Hofmann, 2012) in conjunction with KIT antibody. For single-cell RNA-Seq, isolated cell were captured using the Fluidigm C1 Auto Prep system. cDNAs were prepared using the SMARTer Ultra Low RNA Kit for Illumina (Clontech). Libraries were generated using the Nextera XT (Illumina) protocol. $1x10^6$ cells of each population in 2 independent biological experiments were FACS sorted for bulk RNA-Seq. Total RNAs were isolated from collected cells using the QIAGEN AllPrep DNA/RNA Micro Kit. Sequencing libraries were prepared with ribosomal depletion using Ribo-Zero Gold Kit followed by Directional RNA-Seq library preparation (Wafergen). RNA-Seq libraries were sequenced as 100-bp paired-end reads on an Illumina HiSeq 2000. Data is available in the public repository GEO (GSE107711).

### Analysis of RNA-Seq data

Raw reads were pre-processed with trim galore for trimming Illumina adaptor sequences and mapped to mouse mm9 genome. We estimated expression level for all RefSeq genes using Salmon (Patro et al., 2017) and selected the 500 top-ranked PCA genes for clustering and PCA analysis. Pseudotime ordering were performed with Monocle (v1.2.0) as previously described (Trapnell et al., 2014a). Differential gene expression analysis was performed with the Bayesian approach to single-cell differential expression (SCDE) (Kharchenko et al., 2014) or DESeq2 method (Love et al., 2014). R package "WGCNA" was used to construct the weighted gene co-expression network. GO enrichment was performed using DAVID (Huang da et al., 2009), ToppGene Suite (Chen et al., 2009) or ClueGO (Bindea et al., 2009).

## AUTHOR CONTRIBUTIONS

JL and TL designed the study, performed experiments, analyzed and interpreted data, and prepared the manuscript. SN participated in experimental design, single-cell capture experiment and manuscript writing. JT, AL, and YQ provided intellectual inputs. NT, WC and PF advised on data interpretation. All authors read and approved the final manuscript.

## ACKNOWLEDGEMENT

The authors acknowledge the support of Core laboratory in School of Biomedical Sciences, The Chinese University of Hong Kong and Prof. Bo Feng (School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong) for providing Oct4-GFP transgenic mice. This work was supported by General Research Funds from Hong Kong Research Grant Council (Project no. 468312, 469713) and Lo Kwee-Seong Biomedical Research Seed Fund from School of Biomedical Sciences, The Chinese University of Hong Kong (Project no. 7104687).

## TABLES AND FIGURES LEGENDS

**Figure 1. Characterization and isolation of Oct4-GFP+ germ cells from mouse testes at PND5.5.**

(A-C) Representative images of 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 undifferentiated spermatogonial marker GFRA1 and PLZF (A and B). Oct4-GFP expressing cells can be classified as KIT- and KIT+ (C). Transitional gonocytes reside in the center of the testicular cord (arrow) or

near the basement membrane with cytosolic projection towards the periphery (arrow head). (D) Overlap of Oct4-GFP with PLZF or GFRA1 in testis section immunostaining result as shown in (A) or (B). Results are expressed as means ± SE of 4 testis sections. (E) Sorting strategy for isolation of undifferentiated (Oct4-GFP+/KIT-) at PND5.5. Scale bar = 50um.

**Figure 2. Single-cell transcriptom analysis reveals subpopulations and cellular transitions.**

(A) Hierarchical clustering analysis of 71 single-cell transcriptomes by Ward's method uncovered 4 subpopulations. Germ cell subpopulation identities are resolved by known marker genes as shown in (C). (B) Principal component analysis of 71 single-cell RNA-seq transcriptomes. Each point represents a single cell and labelled according to clustering result as shown in (A). Data were plotted along the first and second principal components. (C) Violin plots illustrating the distribution of representative genes for subpopulations (top) as well as adult $A_s$ marker genes (bottom). (D) Comparison of FACS sorting schemes implemented in our study and in the Song et al. study was shown at the top. PCA on all ID4-GFP+ transcriptome from Song et al. study using the expression [log2(TPM+1)] of the top500 PCA genes in our study was shown at the bottom. 71 single-cell transcriptomes in our study are projected onto the resulting principal component space. Cells colored by developmental identity indicated at the top. (E) Reconstructing continuous trajectory and assigning pseudotime for germ cell development using Monocle. Coloring of cells based on the group assignment from hierarchical clustering reveals stepwise development from gonocytes (Cluster I) to undifferentiated spermatogonia (Cluster II and III) to differentiation-primed spermatogonia (Cluster IV) (left). Developmental states revealed by Monocle (right).

**Figure 3. Characterization of gene expression and signaling pathways in 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 Figure 2E, right. (B) Signalling pathways enriched in primitive stem cell as compared to transitional gonocytes by GSEA/KEGG analysis. Left, GSEA enrichment plot of KEGG signalling pathways. Right, heat maps of genes enriched in representative signaling pathways. The values indicated on individual plots are the Normalised 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. (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, showing a significantly higher percentage of Ki-67- cells in the centrally located population. (E) Cross-sectional images of Ki-67 (red) immunostaining of testis from Oct4-GFP mice. The outlines of seminiferous tubules are indicated by dashed lines. The asterisk denotes the cell resides at the center of the tubules and lacks Ki-67 expression. (F) Box plots for apoptosis-related genes identified as differentially expressed between transitional gonocytes and stem cell spermatogonia.

**Figure 4. Molecular cascades regulating spermatogonia self-renewal and differentiation.**

(A) Signalling pathways enriched in stem cell spermatogonia as compared to progenitors by GSEA/KEGG analysis. The values indicated on GSEA enrichment plots are the Normalised Enrichment Score (NES) and p-value of the enrichment. (B) Hierarchical clustering of genes that significantly vary over pseudotime in spermatogonia differentiation process using expression kinetic curves calculated by Monocle. The bar on the right shows 4 distinct kinetic trends recovered. (C) Pseudotime profiles of representative genes showing different kinetics of expression.

**Figure 5. CD87 is a novel surface marker for primitive stem cells.**

(A) Confocal image of testis section 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. (B) The computationally reconstructed temporal gene expression pattern of genes tested. Each point corresponds to one cell and the color on each cell corresponds the scaled expression value of the respective gene. (C) Validation of gene expression patterns in CD87+ and CD87- subpopulations by FACS and qRT-PCR. (D) The CD87 expression is dynamic throughout the spermatogenesis initiation. Oct4-GFP cells were isolated at 3.5–14 days after birth and stained with APC-conjugated anti-CD87 antibody. The histogram plots show the expression of surface CD87 at different time points. The dashed line represents the cutoff for the determination of CD87+ v.s. CD87- based on staining with isotype controls. (E) Brightfield images showing that CD87+ cells display more spreading and lamellipodia formation on Matrigel. Quantification of cell protrusion comparing CD87+ and CD87- cells is shown on the right. (F) The representative result of cell cycle distribution analysis comparing the CD87+ and CD87-cells. Distribution of cells in G1, S and G2/M phases of the cell cycle as determined using FlowJo software. The percentages of cell populations of different cell cycle phases are shown in the upper-right corner. Quantification on the right

showing a significantly higher percentage of G2/M cells in the CD87+ population. (G) The appearance of GS cell cultured derived from freshly isolated CD87+ and CD87- cells. 10000 FACS-sorted cells were seeded in each well. The representative brightfield images taken 3 days after seeding were shown on the left. The quantification shows significantly higher proliferation rate of CD87+ cells.

**Figure 6. Differentiation-primed cells display enhanced RA responsiveness.**

(A) Box plots for representative genes identified as differentially expressed (FDR<0.05) between *Gfra1*+ (Cluster II and III) and *Gfra1*- (Cluster IV) spermatogonia. (B) Enrichment plots from GSEA analysis using GDNF-responsive gene sets. Note the significant upregulation of GDNF-responsive genes in *Gfra1*+ cells. (C) Both *Oct4* and *Rarg* mRNA were higher expressed in Cluster IV. (D) FACS sorting scheme to isolate subpopulations of germ cells expressing various levels of Oct4-GFP. (E) Patterns of gene expression measured by RT-qPCR in subpopulations of Oct4-GFP+/KIT- cells sorted by level of Oct4-GFP fluorescence. The population with high Oct4-GFP display significant higher expression of *Rarg*. (F) Significantly enhanced RA response in Oct4-GFP-high cells. Left panel, the representative plot showing KIT staining in GS culture derived from Oct4-GFP+/KIT- cells. Spontaneously differentiating KIT+ cells mainly reside in Oct4-GFP-high subpopulation. Right panel, the representative plot showing KIT staining in GS culture after RA-induced differentiation. RA treatment induced KIT expression, which is more profound in the Oct4-GFP-high subpopulation. (G) The Oct4-GFP expression is dynamic throughout the spermatogenesis initiation. Oct4-GFP cells were isolated 3.5-7.5 days after birth and analyzed by FACS. The dashed line represents the cutoff for the determination of Oct4-GFP+ v.s. Oct4-GFP- based on wild-type animal controls. (H) The effect of retinoid signalling on Oct4-GFP expression *in vivo*. Histograms of measured intensity showing KIT expression (left panel) and Oct4-GFP (right panel) in WIN18,466-treated and control testis. KIT+ cells, Oct4-GFP-low, and Oct4-GFP-high cells were significantly reduced upon WIN 18,446 treatment.

**Figure 7. Gene expression dynamics during KIT transition revealed by bulk RNA-Seq analysis.**

(A) The MA plot shows the logarithm of the ratio between expression levels in Oct4-GFP+/KIT+ and Oct4-GFP+/KIT- populations versus the average expression of individual genes. The red dots depict the differentially expressed genes (FDR< 0.05) with top 3 most significant genes shown. (B) Screenshots from the UCSC genome browser, displaying the mRNA-Seq reads from KIT- and KIT+ samples mapping to *Ddit4* and *Stra8* locus. (C)Venn

diagram showing the overlap in the number of genes upregulated in KIT+ cells identified in bulk RNA-seq and the number of genes upregulated upon RA treatment in GS culture (Wang et al., 2016). The enrichment was tested by Fisher's exact test. (D) Scatter plot comparing KIT+/KIT- gene expression ratios at PND5.5 (x-axis) to ratios at PND7.5 (y-axis). Dots represent genes with significant expression difference in either condition. Blue dots represent genes with significant expression difference in both conditions. Red denotes genes showing discrepant changes of gene expression (FDR<0.05, $\log_2$FC<-1 at PND7.5 and $\log_2$FC >-0.3 at PND5.5). (E) Relative expression (KIT+/KIT-) of genes encoding niche receptors in KIT transition at both PND5.5 and PND7.5. (F) Venn diagram showing the overlap in the number of genes regulated in KIT transition at either developmental time points, the number of genes regulated only at PND7.5, and the number of GDNF-responsive genes. The enrichment was tested by Fisher's exact test. (G) Bar plot showing enrichment of biological processes for genes only regulated during KIT transition at PND7.5 as denoted in red in (C).

# REFERENCES

Abid, S.N., Richardson, T.E., Powell, H.M., Jaichander, P., Chaudhary, J., Chapman, K.M., and Hamra, F.K. (2014). A-single spermatogonia heterogeneity and cell cycles synchronize with rat seminiferous epithelium stages VIII-IX. Biol Reprod *90*, 32.

Agrimson, K.S., Oatley, M.J., Mitchell, D., Oatley, J.M., Griswold, M.D., and Hogarth, C.A. (2017). Retinoic acid deficiency leads to an increase in spermatogonial stem number in the neonatal mouse testis, but excess retinoic acid results in no change. Dev Biol.

Aloisio, G.M., Nakada, Y., Saatcioglu, H.D., Pena, C.G., Baker, M.D., Tarnawa, E.D., Mukherjee, J., Manjunath, H., Bugde, A., Sengupta, A.L., *et al.* (2014). PAX7 expression defines germline stem cells in the adult testis. J Clin Invest *124*, 3929-3944.

Basciani, S., De Luca, G., Dolci, S., Brama, M., Arizzi, M., Mariani, S., Rosano, G., Spera, G., and Gnessi, L. (2008). Platelet-derived growth factor receptor beta-subtype regulates proliferation and migration of gonocytes. Endocrinology *149*, 6226-6235.

Becker, K.A., Ghule, P.N., Therrien, J.A., Lian, J.B., Stein, J.L., van Wijnen, A.J., and Stein, G.S. (2006). Self-renewal of human embryonic stem cells is supported by a shortened G1 cell cycle phase. J Cell Physiol *209*, 883-893.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pages, F., Trajanoski, Z., and Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics *25*, 1091-1093.

Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C., and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nat Biotechnol *33*, 155-160.

Caires, K.C., de Avila, J., and McLean, D.J. (2009). Vascular endothelial growth factor regulates germ cell survival during establishment of spermatogenesis in the bovine testis. Reproduction *138*, 667-677.

Chen, J., Bardes, E.E., Aronow, B.J., and Jegga, A.G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res *37*, W305-311.

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.

Culty, M. (2013). Gonocytes, from the fifties to the present: is there a reason to change the name? Biol Reprod *89*, 46.

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. (1998). Stem cells in the testis. Int J Exp Pathol *79*, 67-80.

de Rooij, D.G. (2017). The nature and dynamics of spermatogonial stem cells. Development *144*, 3022-3030.

Diaz-Ramos, A., Roig-Borrellas, A., Garcia-Melero, A., and Lopez-Alemany, R. (2012). alpha-Enolase, a multifunctional protein: its role on pathophysiological situations. J Biomed Biotechnol *2012*, 156795.

Fagoonee, S., Bearzi, C., Di Cunto, F., Clohessy, J.G., Rizzi, R., Reschke, M., Tolosano, E., Provero, P., Pandolfi, P.P., Silengo, L., *et al.* (2013). The RNA Binding Protein ESRP1 Fine-Tunes the Expression of Pluripotency-Related Factors in Mouse Embryonic Stem Cells. Plos One *8*.

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.

Guo, J., Grow, E.J., Yi, C., Mlcochova, H., Maher, G.J., Lindskog, C., Murphy, P.J., Wike, C.L., Carrell, D.T., Goriely, A., *et al.* (2017). Chromatin and Single-Cell RNA-Seq Profiling Reveal Dynamic Signaling and Metabolic Transitions during Human Spermatogonial Stem Cell Development. Cell Stem Cell *21*, 533-546 e536.

He, S., Nakada, D., and Morrison, S.J. (2009). Mechanisms of stem cell self-renewal. Annu Rev Cell Dev Biol *25*, 377-406.

He, Z.P., Jiang, J.J., Kokkinaki, M., Golestaneh, N., Hofmann, M.C., and Dym, M. (2008). GDNF upregulates c-fos transcription via the Ras/ERK1/2 pathway to promote mouse spermatogonial stem cell proliferation. Stem Cells *26*, 266-278.

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. Biology of Reproduction *92*.

Hobbs, R.M., Seandel, M., Falciatori, I., Rafii, S., and Pandolfi, P.P. (2010). Plzf regulates germline progenitor self-renewal by opposing mTORC1. Cell *142*, 468-479.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

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., Nakatsuji, N., Fassler, R., and Shinohara, T. (2008). Homing of mouse spermatogonial stem cells to germline niche depends on beta1-integrin. Cell Stem Cell *3*, 533-542.

Karlsson, R., Andreassen, K.E., Kristiansen, W., Aschim, E.L., Bremnes, R.M., Dahl, O., Fossa, S.D., Klepp, O., Langberg, C.W., Solberg, A., *et al.* (2013). Investigation of six testicular germ cell tumor susceptibility genes suggests a parent-of-origin effect in SPRY4. Hum Mol Genet *22*, 3373-3380.

Kharchenko, P.V., Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to single-cell differential expression analysis. Nature Methods *11*, 740-U184.

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. International Journal of Andrology *4*, 475-493.

Kubo, N., Toh, H., Shirane, K., Shirakawa, T., Kobayashi, H., Sato, T., Sone, H., Sato, Y., Tomizawa, S.I., Tsurusaki, Y., *et al.* (2015). DNA methylation and gene expression dynamics during spermatogonial stem cell differentiation in the early postnatal mouse testis. Bmc Genomics *16*.

Kwon, J.T., Jin, S., Choi, H., Kim, J., Jeong, J., Kim, J., Kim, Y., Cho, B.N., and Cho, C. (2014). Identification and Characterization of Germ Cell Genes Expressed in the F9 Testicular Teratoma Stem Cell Line. Plos One *9*.

Ladoux, B., and Nicolas, A. (2012). Physically based principles of cell adhesion mechanosensitivity in tissues. Rep Prog Phys *75*, 116601.

Lange, C., Huttner, W.B., and Calegari, F. (2009). Cdk4/CyclinD1 Overexpression in Neural Stem Cells Shortens G1, Delays Neurogenesis, and Promotes the Generation and Expansion of Basal Progenitors. Cell Stem Cell *5*, 320-331.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. Bmc Bioinformatics *9*.

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.

Li, J., Luo, H., Wang, R., Lang, J., Zhu, S., Zhang, Z., Fang, J., Qu, K., Lin, Y., Long, H., *et al.* (2016). Systematic Reconstruction of Molecular Cascades Regulating GP Development Using Single-Cell RNA-Seq. Cell Rep *15*, 1467-1480.

Llorens-Bobadilla, E., Zhao, S., Baser, A., Saiz-Castro, G., Zwadlo, K., and Martin-Villalba, A. (2015). Single-Cell Transcriptomics Reveals a Population of Dormant Neural Stem Cells that Become Activated upon Brain Injury. Cell Stem Cell *17*, 329-340.

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.

Manku, G., and Culty, M. (2015). Mammalian gonocyte and spermatogonia differentiation: recent advances and remaining challenges. Reproduction *149*, R139-157.

Meng, X., Lindahl, M., Hyvonen, M.E., Parvinen, M., de Rooij, D.G., Hess, M.W., Raatikainen-Ahokas, A., Sainio, K., Rauvala, H., Lakso, M., *et al.* (2000). Regulation of cell fate decision of undifferentiated spermatogonia by GDNF. Science *287*, 1489-1493.

Morimoto, H., Kanatsu-Shinohara, M., Takashima, S., Chuma, S., Nakatsuji, N., Takehashi, M., and Shinohara, T. (2009). Phenotypic plasticity of mouse spermatogonial stem cells. PLoS One *4*, e7909.

Mutoji, K., Singh, A., Nguyen, T., Gildersleeve, H., Kaucher, A.V., Oatley, M.J., Oatley, J.M., Velte, E.K., Geyer, C.B., Cheng, K., *et al.* (2016). TSPAN8 Expression Distinguishes Spermatogonial Stem Cells in the Prepubertal Mouse Testis. Biol Reprod *95*, 117.

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., and Yoshida, S. (2007). Functional identification of the actual and potential stem cell compartments in mouse spermatogenesis. Dev Cell *12*, 195-206.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet *24*, 372-376.

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 U S A *103*, 9524-9529.

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, M.J., Kaucher, A.V., Racicot, K.E., and Oatley, J.M. (2011). Inhibitor of DNA binding 4 is expressed selectively by single spermatogonia in the male germline and regulates the self-renewal of spermatogonial stem cells in mice. Biol Reprod *85*, 347-356.

Ohbo, K., Yoshida, S., Ohmura, M., Ohneda, O., Ogawa, T., Tsuchiya, H., Kuwana, T., Kehler, J., Abe, K., Scholer, H.R., *et al.* (2003). Identification and characterization of stem cells in prepubertal spermatogenesis in mice. Dev Biol *258*, 209-225.

Paik, J., Haenisch, M., Muller, C.H., Goldstein, A.S., Arnold, S., Isoherranen, N., Brabb, T., Treuting, P.M., and Amory, J.K. (2014). Inhibition of retinoic acid biosynthesis by the bisdichloroacetyldiamine WIN 18,446 markedly suppresses spermatogenesis and alters retinoid metabolism in mice. J Biol Chem *289*, 15104-15117.

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.

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.

Rochman, M., Malicet, C., and Bustin, M. (2010). HMGN5/NSBP1: A new member of the HMGN protein family that affects chromatin structure and function. Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms *1799*, 86-92.

Rodriguez, S., Jafer, O., Goker, H., Summersgill, B.M., Zafarana, G., Gillis, A.J., van Gurp, R.J., Oosterhuis, J.W., Lu, Y.J., Huddart, R., *et al.* (2003). Expression profile of genes from 12p in testicular germ cell tumors of adolescents and adults associated with i(12p) and amplification at 12p11.2-p12.1. Oncogene *22*, 1880-1891.

Rossi, P., Sette, C., Dolci, S., and Geremia, R. (2000). Role of c-kit in mammalian spermatogenesis. J Endocrinol Invest *23*, 609-615.

Saitou, M., and Yamaji, M. (2012). Primordial germ cells in mice. Cold Spring Harb Perspect Biol *4*.

Sanchez, C.G., Teixeira, F.K., Czech, B., Preall, J.B., Zamparini, A.L., Seifert, J.R., Malone, C.D., Hannon, G.J., and Lehmann, R. (2016). Regulation of Ribosome Biogenesis and Protein Synthesis Controls Germline Stem Cell Differentiation. Cell Stem Cell *18*, 276-290.

Shin, J., Berg, D.A., Zhu, Y., Shin, J.Y., Song, J., Bonaguidi, M.A., Enikolopov, G., Nauen, D.W., Christian, K.M., Ming, G.L., *et al.* (2015). Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis. Cell Stem Cell *17*, 360-372.

Shirakawa, T., Yaman-Deveci, R., Tomizawa, S., Kamizato, Y., Nakajima, K., Sone, H., Sato, Y., Sharif, J., Yamashita, A., Takada-Horisawa, Y., *et al.* (2013). An epigenetic switch is crucial for spermatogonia to exit the undifferentiated state toward a Kit-positive identity. Development *140*, 3565-3576.

Song, H.W., Bettegowda, A., Lake, B.B., Zhao, A.H., Skarbrevik, D., Babajanian, E., Sukhwani, M., Shum, E.Y., Phan, M.H., Plank, T.D., *et al.* (2016). The Homeobox Transcription Factor RHOX10 Drives Mouse Spermatogonial Stem Cell Establishment. Cell Rep *17*, 149-164.

Spierings, D.C., de Vries, E.G., Vellenga, E., and de Jong, S. (2003). Loss of drug-induced activation of the CD95 apoptotic pathway in a cisplatin-resistant testicular germ cell tumor cell line. Cell Death Differ *10*, 808-822.

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.

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.

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. (2014a). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol *32*, 381-386.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S.Q., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014b). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nature Biotechnology *32*, 381-U251.

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.

Ulisse, S., Baldini, E., Mottolese, M., Sentinelli, S., Gargiulo, P., Valentina, B., Sorrenti, S., Di Benedetto, A., De Antoni, E., and D'Armiento, M. (2010). Increased expression of urokinase plasminogen activator and its cognate receptor in human seminomas. BMC Cancer *10*, 151.

Wang, R.A., Nakane, P.K., and Koji, T. (1998). Autonomous cell death of mouse male germ cells during fetal and postnatal period. Biology of Reproduction *58*, 1250-1256.

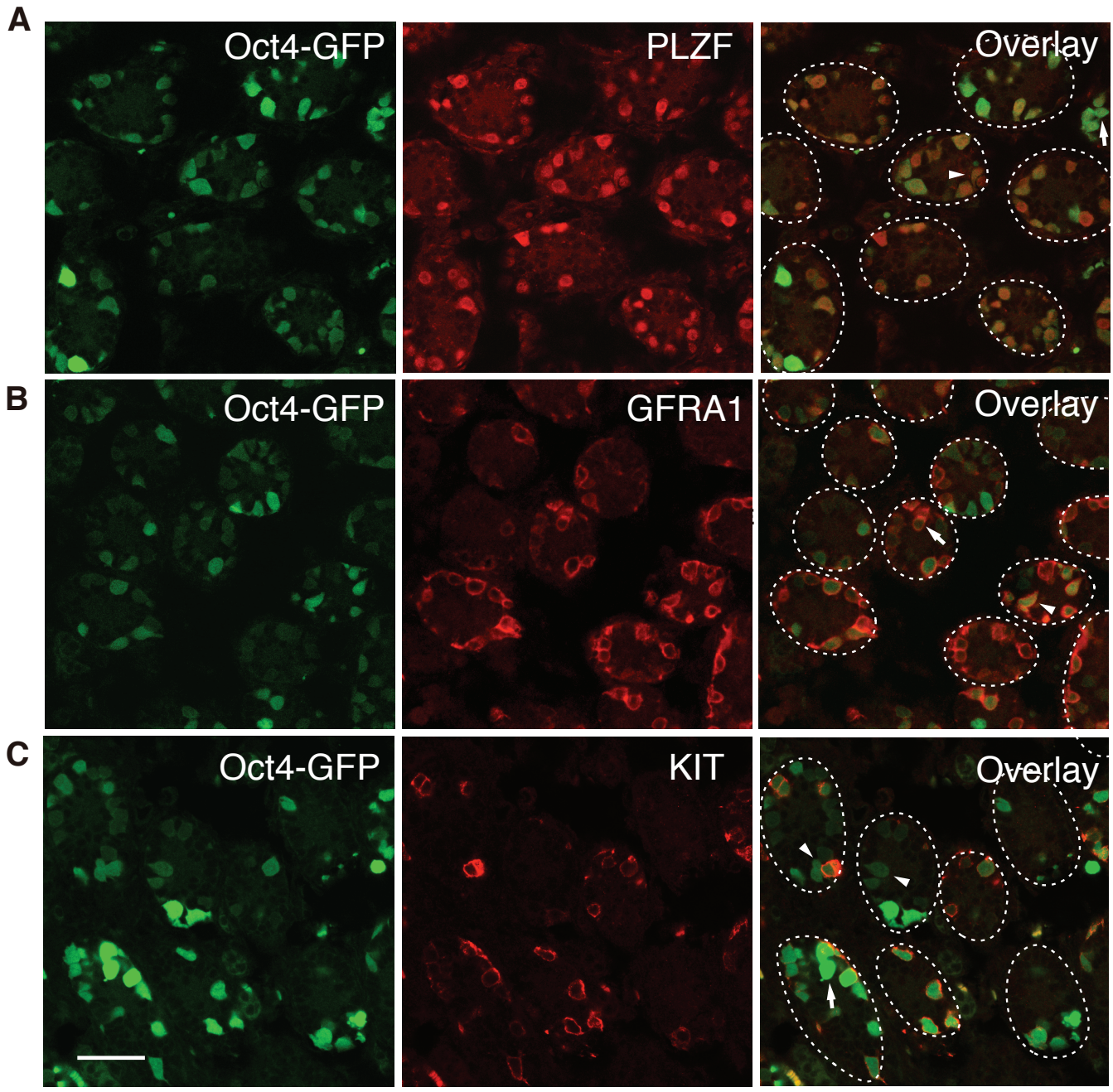
Wang, S., Wang, X., Ma, L., Lin, X., Zhang, D., Li, Z., Wu, Y., Zheng, C., Feng, X., Liao, S., *et al.* (2016). Retinoic Acid Is Sufficient for the In Vitro Induction of Mouse Spermatocytes. Stem Cell Reports *7*, 80-94.

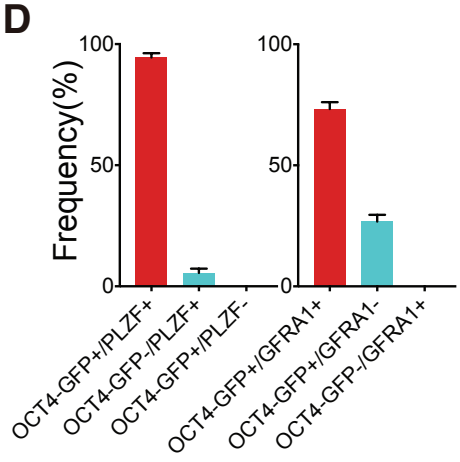
Wood, D.P. (2011). Cisplatin Sensitivity of Testis Tumour Cells is Due to Deficiency in Interstrand-Crosslink Repair and Low ERCC1-XPF Expression Editorial Comment. Journal of Urology *186*, 457-457.

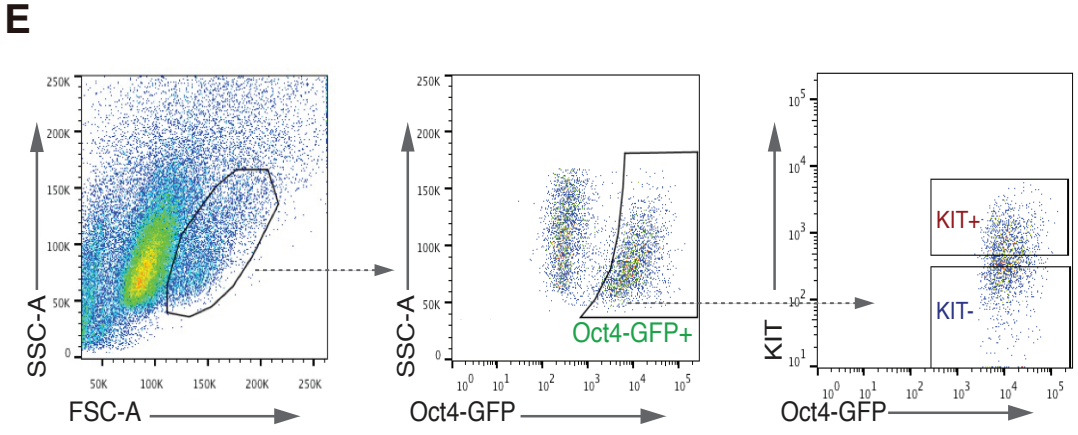
Yoshida, S., Sukeno, M., Nakagawa, T., Ohbo, K., Nagamatsu, G., Suda, T., and Nabeshima, Y. (2006a). The first round of mouse spermatogenesis is a distinctive program that lacks the self-renewing spermatogonia stage. Development *133*, 1495-1505.

Yoshida, S., Sukeno, M., Nakagawa, T., Ohbo, K., Nagamatsu, G., Suda, T., and Nabeshima, Y. (2006b). The first round of mouse spermatogenesis is a distinctive program that lacks the self-renewing spermatogonia stage. Development *133*, 1495-1505.

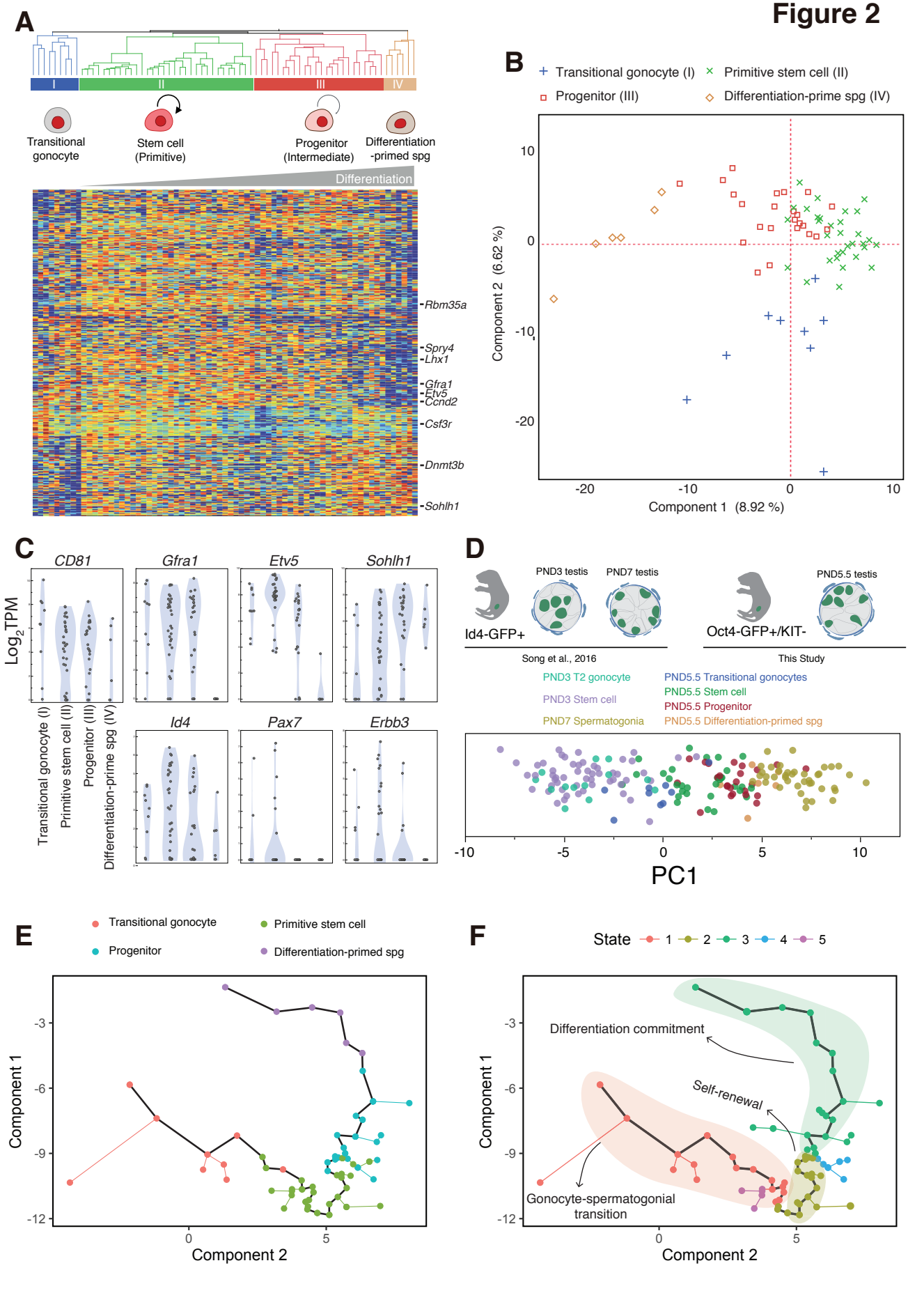
# Figure 1



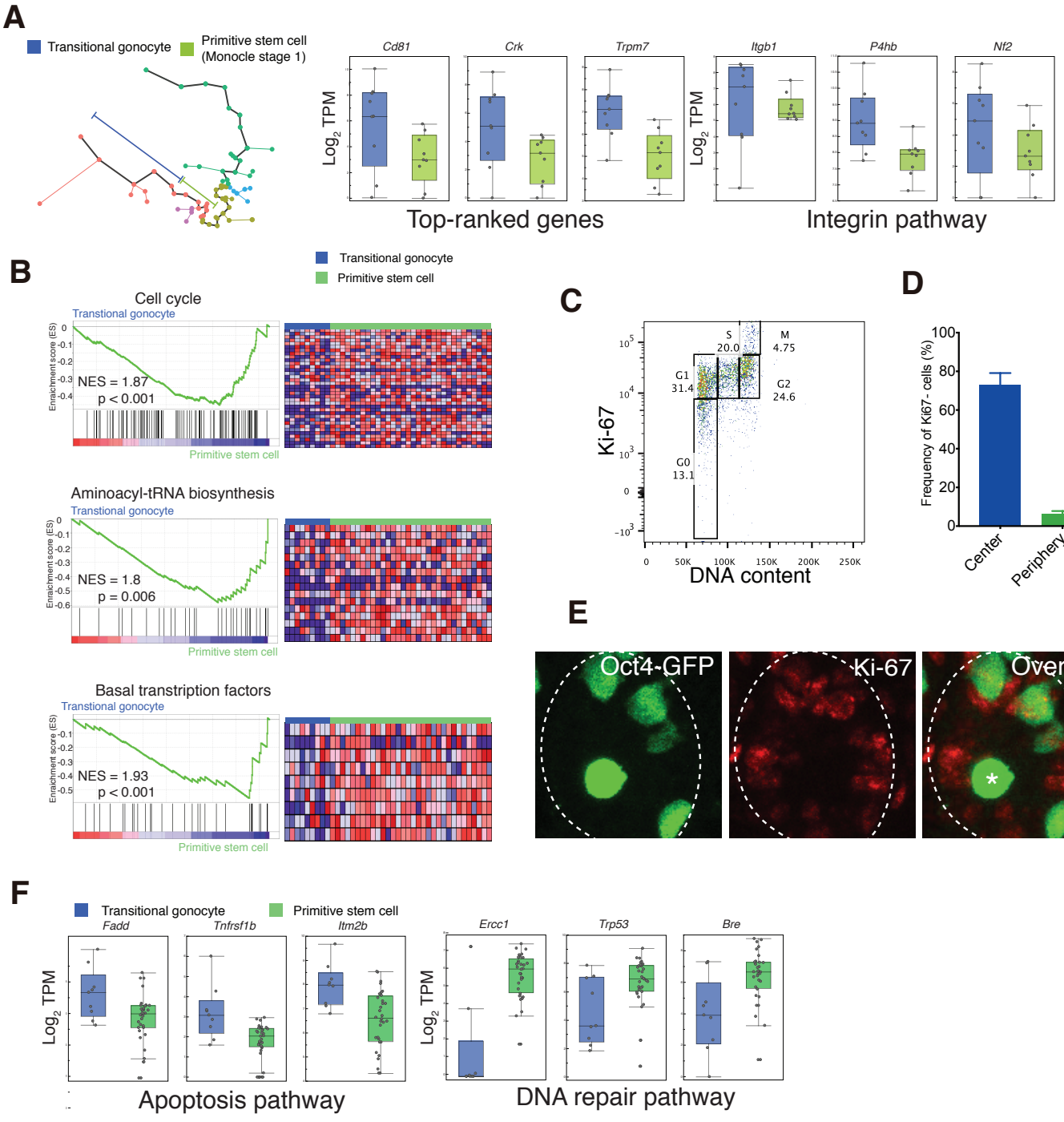



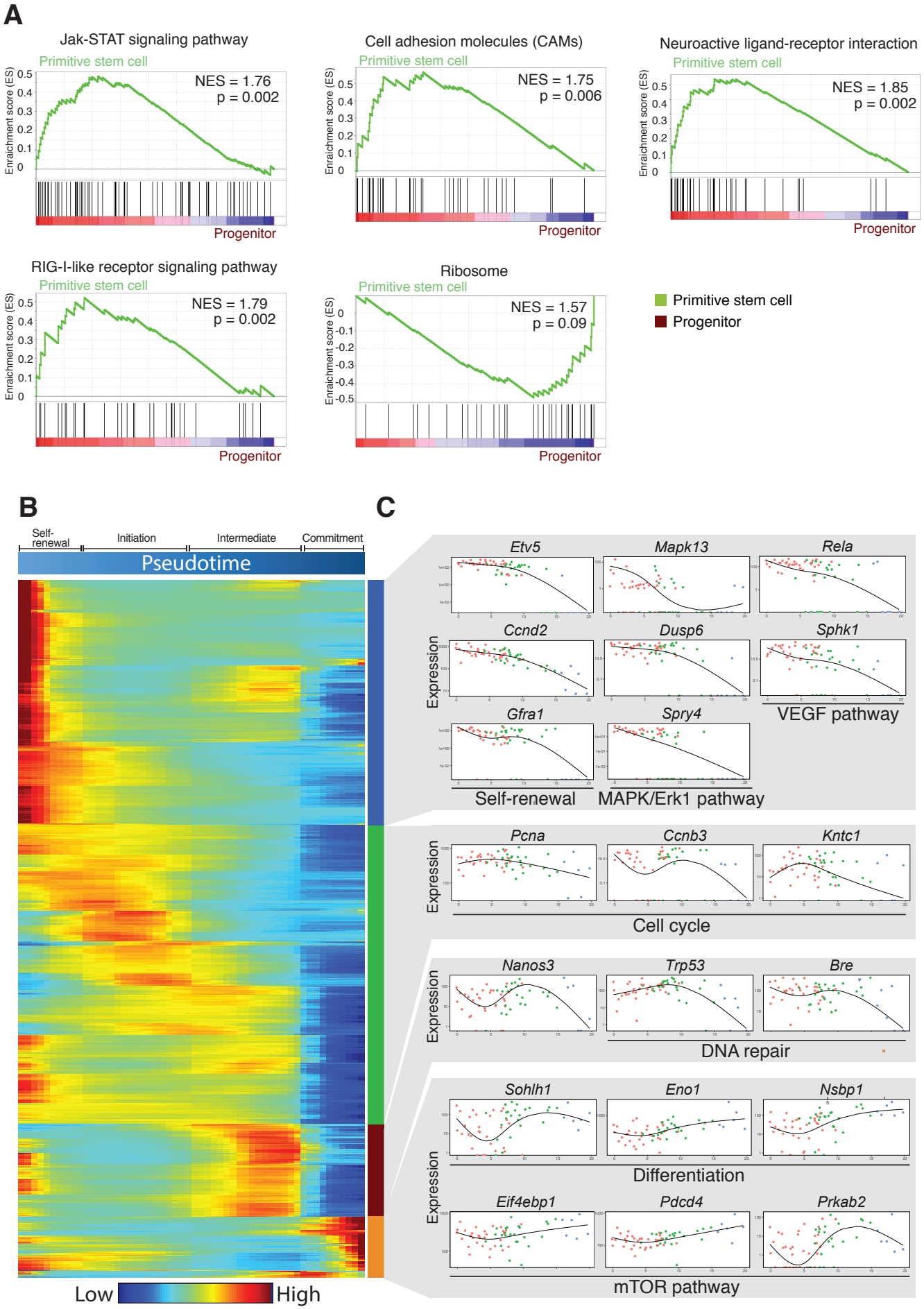
# Figure 2



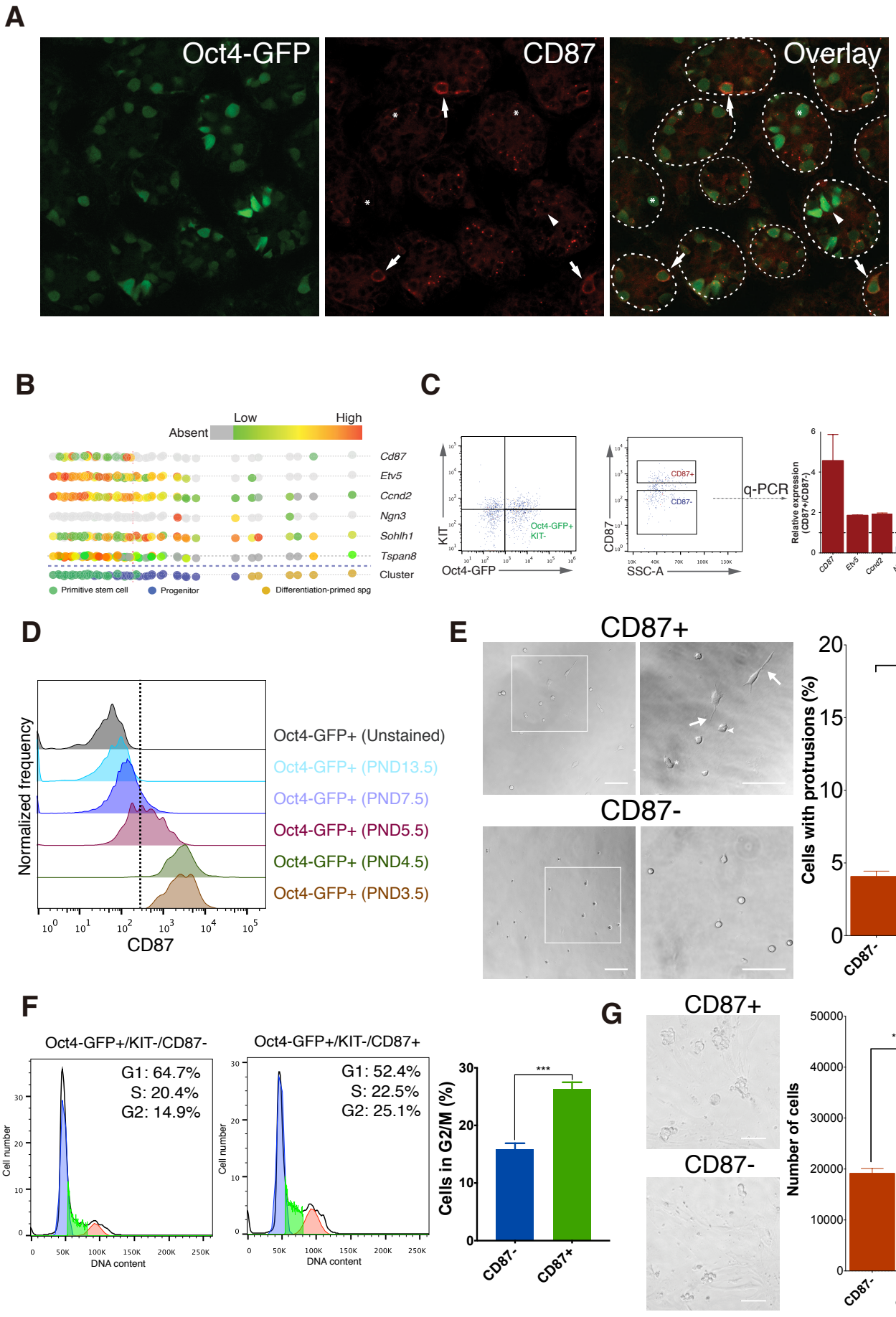
# Figure 3



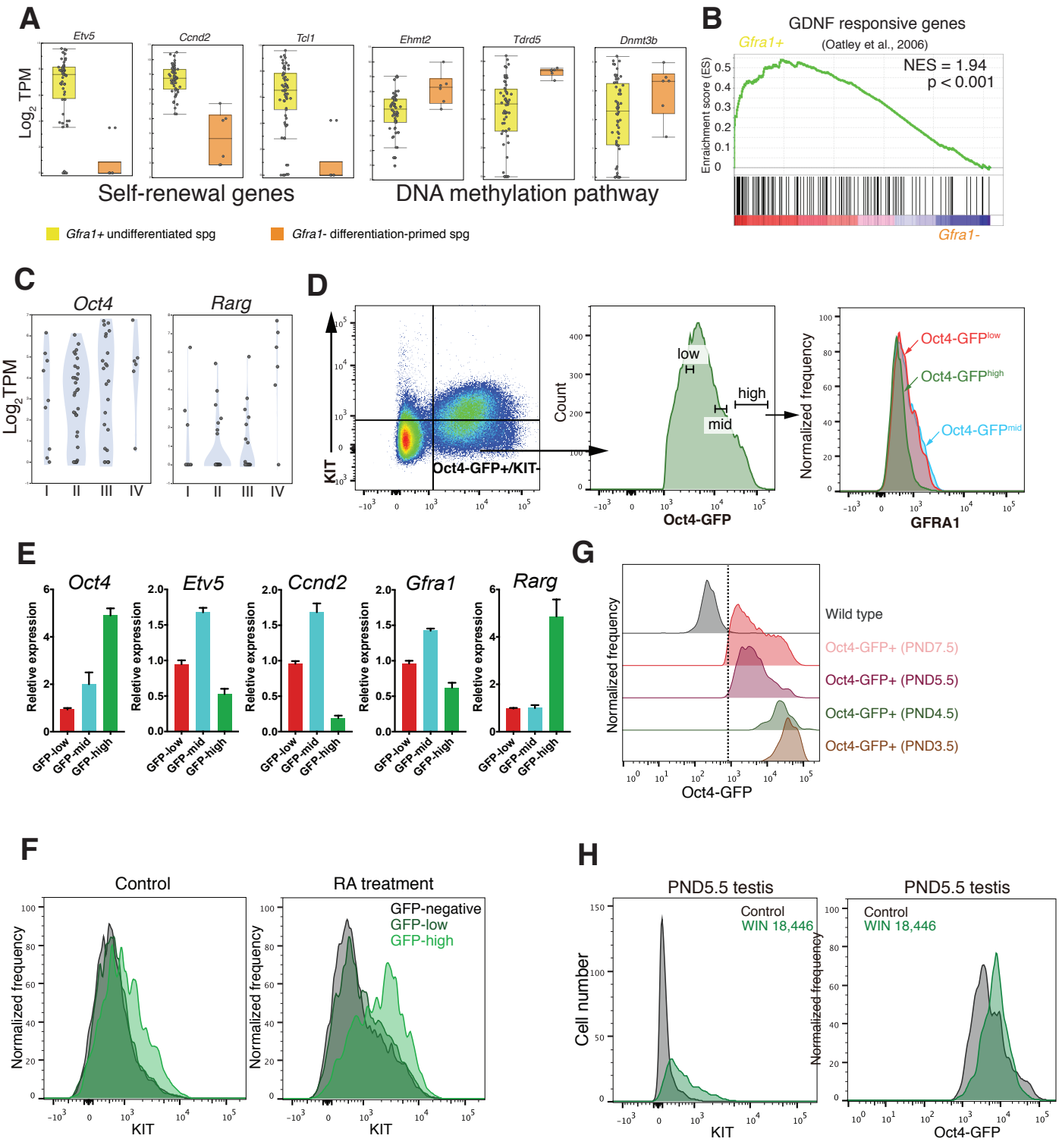
# Figure 4



# Figure 5



# Figure S16



# Figure 7