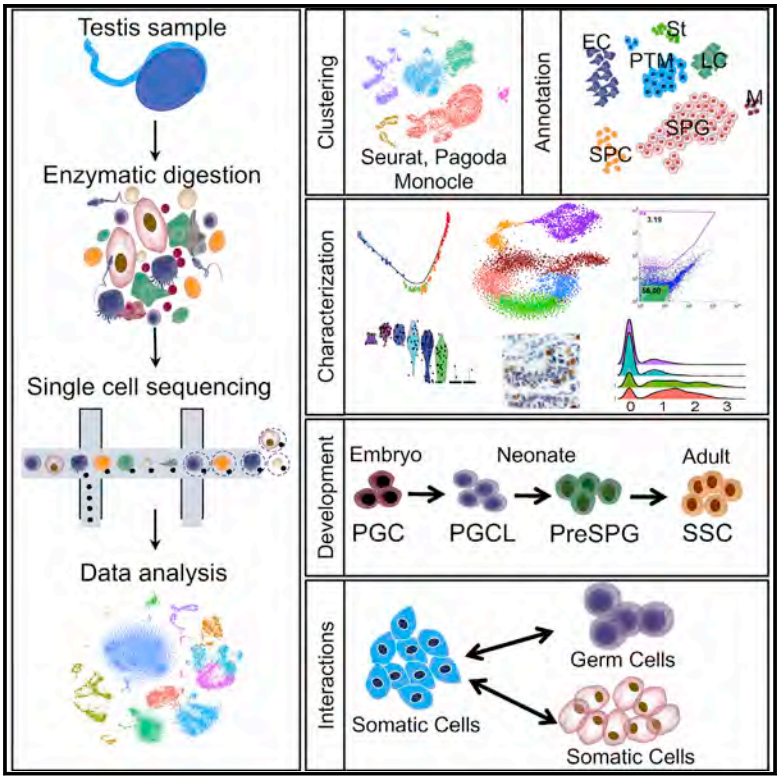
Article

Cell Reports

The Neonatal and Adult Human Testis Defined at the Single-Cell Level

Graphical Abstract



Authors

Abhishek Sohni, Kun Tan, Hye-Won Song, ..., Saher Sue Hammoud, Elena Vicini, Miles F. Wilkinson

Correspondence

mfwilkinson@ucsd.edu

In Brief

Sohni et al. use scRNA-seq analysis to define cell subsets in the human testis. Highlights include the identification of primordial germ cell- and spermatogonial stem cell-like cell subsets in neonatal testes, numerous undifferentiated spermatogonial cell states in adult testes, and somatic cell subsets in both neonatal and adult testes.

Highlights

- Neonatal and adult human testicular cell subsets are defined by scRNA-seq analysis
- The neonate contains both primordial germ cell- and SSC-like cells
- Numerous adult spermatogonial states with distinct markers are defined
- Neonatal and adult testicular "niche" cells and factors are defined

Sohni et al., 2019, Cell Reports *26*, 1501–1517
February 5, 2019
https://doi.org/10.1016/j.celrep.2019.01.045

CellPress

Cell Reports
Article

The Neonatal and Adult Human Testis Defined at the Single-Cell Level

Abhishek Sohni,[1,8] Kun Tan,[1,8] Hye-Won Song,[1,8] Dana Burow,[1] Dirk G. de Rooij,[2] Louise Laurent,[1] Tung-Chin Hsieh,[3] Raja Rabah,[4] Saher Sue Hammoud,[5] Elena Vicini,[6] and Miles F. Wilkinson[1,7,9,*]

[1]Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California, San Diego, La Jolla, CA 92093, USA
[2]Reproductive Biology Group, Division of Developmental Biology, Department of Biology, Faculty of Science, Utrecht University, 3584 CH Utrecht, the Netherlands
[3]Department of Urology, University of California, San Diego, La Jolla, CA 92103, USA
[4]Pediatric and Perinatal Pathology, Michigan Medicine, CS Mott and VonVoigtlander Women's Hospitals, Ann Arbor, MI 48109-4272, USA
[5]Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109, USA
[6]Department of Anatomy, Histology, Forensic Medicine and Orthopedic, Section of Histology Sapienza University of Rome, 00161 Rome, Italy
[7]Institute of Genomic Medicine, University of California, San Diego, La Jolla, CA 92093, USA
[8]These authors contributed equally
[9]Lead Contact
*Correspondence: mfwilkinson@ucsd.edu
https://doi.org/10.1016/j.celrep.2019.01.045

SUMMARY

Spermatogenesis has been intensely studied in rodents but remains poorly understood in humans. Here, we used single-cell RNA sequencing to analyze human testes. Clustering analysis of neonatal testes reveals several cell subsets, including cell populations with characteristics of primordial germ cells (PGCs) and spermatogonial stem cells (SSCs). In adult testes, we identify four undifferentiated spermatogonia (SPG) clusters, each of which expresses specific marker genes. We identify protein markers for the most primitive SPG state, allowing us to purify this likely SSC-enriched cell subset. We map the timeline of male germ cell development from PGCs through fetal germ cells to differentiating adult SPG stages. We also define somatic cell subsets in both neonatal and adult testes and trace their developmental trajectories. Our data provide a blueprint of the developing human male germline and supporting somatic cells. The PGC-like and SSC markers are candidates to be used for SSC therapy to treat infertility.

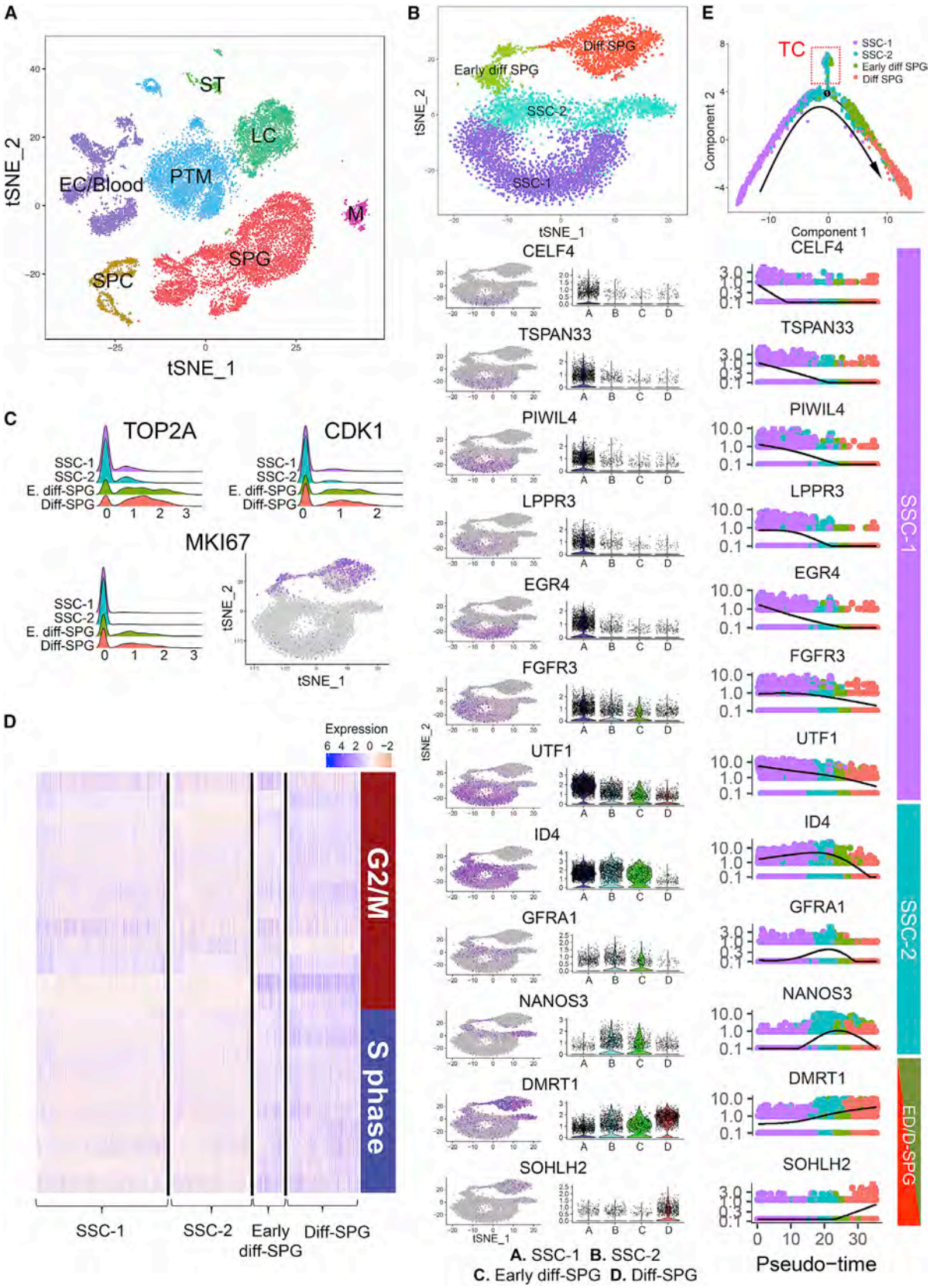
INTRODUCTION

Spermatogenesis is the process by which sperm are generated from male germ cell precursor cells. Spermatogenesis depends on an orchestrated series of events in germ cells first initiated in undifferentiated spermatogonia (SPG). A subset of undifferentiated SPG—called spermatogonial stem cells (SSCs)—have the ability to continuously self-renew and, thus, are responsible for maintaining the male germline throughout life. When not self-renewing, SSCs form progenitors, which proliferate and differentiate to form more advanced SPG cell types. The most differentiated SPGs give rise to spermatocytes (SPCs), which go through meiosis to become haploid cells known as spermatids (STs), which ultimately become sperm.

Germ cell differentiation requires the support of specialized somatic cells. This includes Sertoli cells (SCs), the nurse cells in direct contact with all germ cells in the seminiferous epithelium; peritubular myoid cells (PTMs), which are factor-secreting muscle cells surrounding the seminiferous tubule; and Leydig cells (LCs), which reside outside of the seminiferous epithelium and secrete androgens and other factors critical for spermatogenesis (Oatley and Brinster, 2012).

Most of what we know about spermatogenesis comes from investigations in rodents (Kanatsu-Shinohara and Shinohara, 2013). Although some of this information is likely to bear on human spermatogenesis, it is clear that human spermatogenesis is significantly different from rodent spermatogenesis, including seminiferous epithelium organization, the pattern of SPG development, and sperm output per gram of tissue (Fayomi and Orwig, 2018).

Given the differences between rodent and human spermatogenesis, there has been increasing interest in conducting studies on spermatogensis in humans. A major focus has been human SSCs, as these cells have the potential to be used clinically to treat infertility (Valli et al., 2014a). An active area of investigation has been the identification of protein markers that label cells with the morphology of human SSCs. However, many of these markers—including ENO2, LIN28, PLZF, SALL4, SSEA4, UCHL1, and UTF1—recognize not only undifferentiated SPG but also differentiating SPG (Dym et al., 2009; Fayomi and Orwig, 2018). Others—such as ID4 and FGFR3—are relatively specific for undifferentiated SPG (Guo et al., 2017; Sachs et al., 2014), but their relative selectivity for human SSCs is unclear. As another approach to identify SSCs and SSC markers, Guo et al. (2017) used single-cell RNA sequencing (scRNA-seq) to identify 4 SPG "states" and define markers that label the state most likely to be enriched for SSCs. Although this study was an important advance, a marker of unclear specificity—SSEA4—was used to enrich undifferentiated SPG, which introduced potential bias and, thus, most SSCs may not have been



(legend on next page)

included in their analysis. The purified populations used in this study also precluded an analysis of other testicular subsets, including other germ and all somatic cell subsets.

In this communication, we used scRNA-seq to analyze all cells in the human testis. This allowed us to define all major germ and somatic cell subsets, including a specific undifferentiated SPG subset exhibiting the characteristics of highly enriched SSCs. Using immunofluorescence (IF), immunohistochemistry (IHC), and fluorescence-activated cell sorting (FACS), marker proteins were identified that labeled this cell subset and allowed for its purification. We also addressed the events that lead to the initial establishment of human SSCs. In mice, primordial germ cells (PGCs) undergo epigenetic reprogramming and convert into SSC precursor cells—called ProSPG or gonocytes—that progress through distinct proliferative and quiescent stages, leading to mitotically active SSCs soon after birth (de Rooij, 2017). In contrast, we know little about SSC formation in humans. Germ cells have been identified in late-stage human male fetuses and young male children (Sadri-Ardekani et al., 2011; Wu et al., 2009), and yet, these cells are not well characterized and their relationship with SSCs is unknown. Here, we performed scRNA-seq on neonatal human testes, allowing us to define the germ and somatic cell subsets at this critical stage of development. By comparing gene expression profiles of cells from neonatal and adult testes, we also defined the developmental trajectories of germ and somatic cell types. Together, our results provide a roadmap of germ and somatic cell development in the human testis.

RESULTS

Cell Subsets in the Adult Human Testis Defined by scRNA-Seq

To define the cell types in the adult human testis, we used scRNA-seq to analyze cells from freshly isolated whole testes samples from two fertile adults (aged 37 and 42 years). After filtering out poor-quality cells, 18,723 cells remained for subsequent analysis. Using pathway and gene set overdispersion analysis (PAGODA2), we identified cell clusters corresponding to known cell types in the testis, including SPG, SPCs, STs, PTMs, LCs, blood and endothelial cells (ECs), and macrophages (Ms) (Figures 1A and S6C). Figure S1A shows some of the gene markers we used to identify these cell subsets (other markers are listed in Table S1). Few whole SCs were observed, probably because adult SCs are extremely large branched cells; indeed, we obtained many likely adult SC fragments that expressed SC markers, none of which were included in our analysis (data not shown).

Identification of SPG Cell Clusters

Our initial clustering analysis suggested that the SPG cluster is itself comprised of several cell subsets (Figure 1A). To further assess this possibility, we performed clustering analysis on only SPG (Figure 1B). This revealed 4 distinct cell clusters, which we defined as (1) SSC-1, (2) SSC cluster-2 (SSC-2), (3) Early differentiating (diff) SPG (Early diff-SPG), and (4) Diff-SPG, based on several lines of evidence, as described below. The pan-germ-cell markers *DDX4* and *MAGEA4* (Kossack et al., 2013) labeled cells in all 4 clusters (Figure S1B), as expected. Table S2 lists genes enriched in each of these 4 SPG subsets.

The SSC-1 subset is likely to be SSC enriched, as most cells in this subset expressed previously recognized human SSC marker genes (Di Persio et al., 2017), including *UTF1* (Figure 1B). The SSC-2 subset shares the expression of several SSC genes with the SSC-1 subset but also selectively expresses some genes, including *NANOS3* (Figure 1B). Most cells in both the SSC-1 and SSC-2 subsets have a cell cycle gene expression profile indicative of them being in gap 0 (G0) or gap 1 (G1) phase, with only a few cells in synthesis (S) or gap 2 (G2)-mitosis (M) phase (Figures 1C and 1D). This suggests that most SSC-1 and SSC-2 cells are not proliferating and/or undergoing slow proliferation, consistent with past studies that have directly examined the proliferation of undifferentiated human SPG; i.e., A-SPG (Di Persio et al., 2017). We note that while we refer to the SSC-1 and SSC-2 cell clusters as "SSCs," scRNA-seq is not a functional assay and thus we do not know the percentage of cells in these clusters with SSC activity. These subsets almost certainly contain other A-SPG cells, including SPG progenitors that have committed to differentiate.

Human A-SPG differentiate into actively mitotic B-SPG (Aponte et al., 2005). We identified two subsets—Early diff-SPG and Diff-SPG—that likely correspond to early and late stages, respectively, of B-SPG. Evidence for this includes (1) most cells in both subsets predominantly express S or G2-M genes consistent with their undergoing active mitosis (Figures 1C and 1D) and (2) most cells in these 2 subsets express known B-SPG markers, such as *KIT*, *SOHLH1/2*, *STRA8*, *MAGEA4*, and *DMRT1* (He et al., 2010; Suzuki et al., 2012) but lack or weakly express meiotic genes (e.g., *SYCP3* and *PRDM9*) (Figure 1B

Figure 1. Identification of Adult Human SPG Subsets

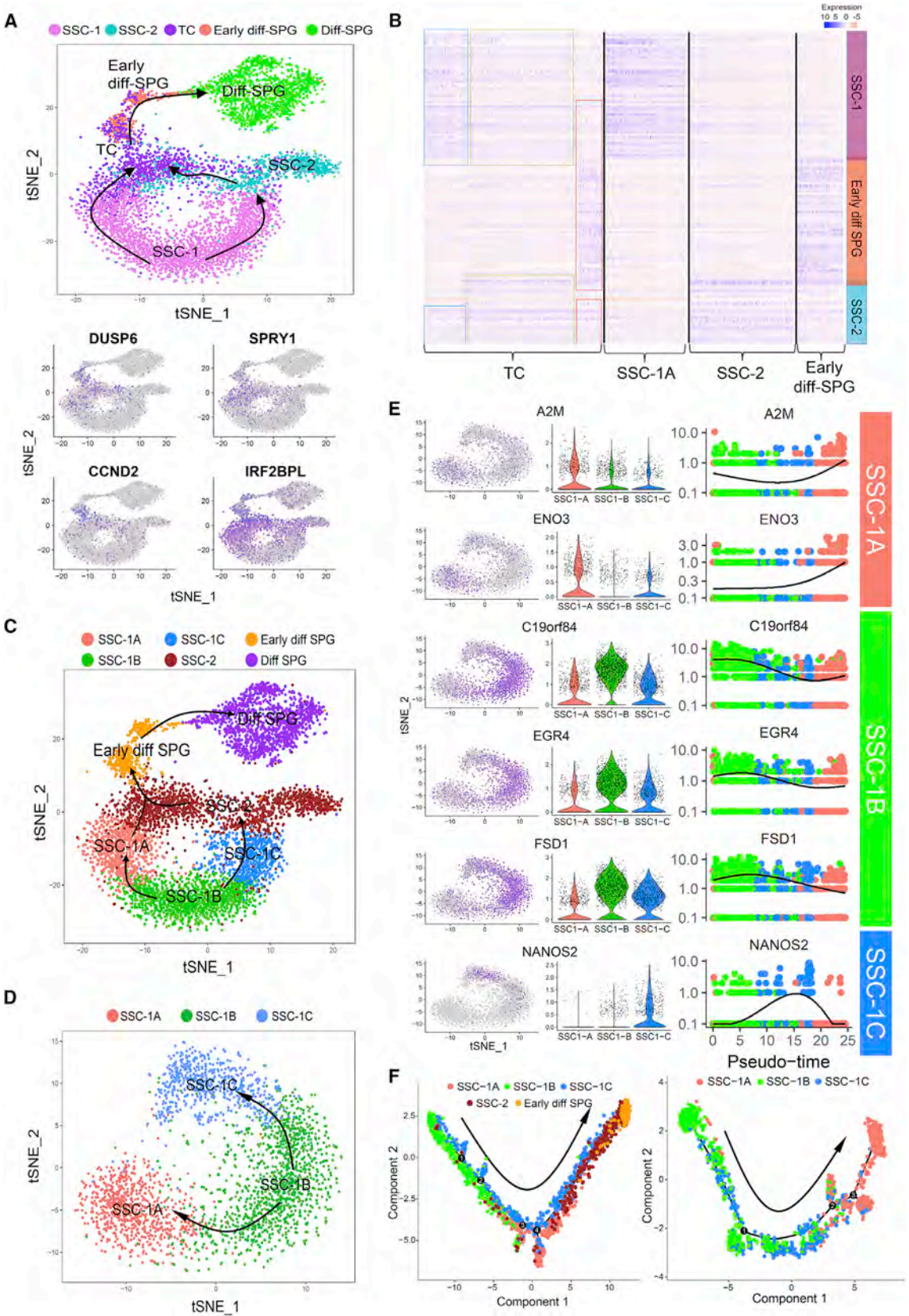
(A) tSNE plot of adult human testes cell clusters defined by scRNA-seq analysis. The fraction of the total cells in each subset is: spermatogonia (SPG) (39%), spermatocyte (SPC) (6%), spermatid (ST) (1%), Leydig cell (LC) (2%), peritubular myoid cell (PTM) (24%), endothelial cell (EC) and blood (27%), and macrophages (M) (1%).

(B) SPG subsets defined by clustering analysis. The tSNE plot (top) was generated from the cells in the SPG subset defined in (A). The feature plots (bottom left) and violin plots (bottom right) show the expression pattern of markers (both known and new) that chart the progression of human SPG.

(C) Cell cycle gene expression in SPG subsets. The ridge plots show the expression of S, G2-M, and S-G2-M phase genes (*TOP2A*, *CDK1*, and *MKI67*, respectively). The *MKI67* tSNE features plot corroborates that Early diff-SPG and Diff-SPG cells tend to express higher levels of proliferation genes than spermatogonial stem cell 1 (SSC-1) and SSC-2 cells.

(D) Heatmap of G2-M and S phase genes in the SPG subsets.

(E) Developmental timeline of SPG cell subsets. The top shows Monocle pseudotime trajectory analysis of the SPG cell subsets defined in (B). The arrow indicates the developmental direction, based on the SSC-1 and Diff-SPG subsets being the least and most advanced, respectively. The red box labels a branch of cells that evidence suggests are TCs (see text). The expression pattern of marker genes is plotted along the pseudotime axis (bottom; same genes as those shown in B). The marker genes are predominantly expressed in the SPG subsets indicated in the vertical bar on the right. ED, early differentiating; D, differentiating.



(legend on next page)

and Table S2). The Early diff-SPG subset is positioned—as defined by PAGODA2 clustering analysis—between the SSC-2 and Diff-SPG subsets (Figure 1B). Consistent with this, most Early diff-SPG cells express the *NANOS3*, *L1TD1*, and *ASB9* genes, which are also expressed by most SSC-2 cells (Figures 1B and S1B). Similarly, most Early diff-SPG also express several genes expressed by Diff-SPG, including *DMRT1*, *TUBA3D*, *DNMT1*, and *CALR* (Figures 1B and S1B). Diff-SPG are selectively labeled by *SOHLH2* (Figure 1B), consistent with the fact that *SOHLH2* encodes a transcription factor that promotes SPG differentiation in mice (Suzuki et al., 2012).

To examine the developmental relationship of these 4 SPG subsets, we performed Monocle pseudotime trajectory analysis, which aligns individual cells along a developmental trajectory, based on differentially expressed genes (Trapnell et al., 2014). This analysis indicated that these cell subsets had the following developmental order: SSC-1 → SSC-2 → Early diff-SPG → Diff-SPG (Figure 1E, top). Figures 1E (bottom) and S1C show the expression of stage-specific markers along this developmental trajectory. Using the likelihood ratio test (Trapnell et al., 2014), we found that the differentially expressed genes (DEGs) most enriched for the SSC-1 subset are *CELF4*, *EGR4*, *FGFR3*, *FSD1*, *LPPR3*, *PIWIL4*, and *TSPAN33* (Figures 1E and S1C), which we propose is a molecular signature of SSCs.

Gene Ontology (GO) analysis revealed distinct categories of genes enriched in each of the 4 SPG subsets (Figure S1D; Table S2). SSC-1 are enriched for gene regulatory pathways; SSC-2 express genes involved in transcription and cellular differentiation; and Early diff-SPG and Diff-SPG are both enriched for functions related to cell proliferation, which is consistent with their high proliferation rate (Di Persio et al., 2017) (Figures 1C and 1D).

Transition Cells Bridging SSCs with Differentiating SPG

Monocle analysis revealed the existence of a group of cells that branched from the linear path of the pseudotime trajectory (Figure 1E, top). These Monocle-defined cells had characteristics of three distinct cell clusters—SSC-1, SSC-2, and Early diff-SPG—all of which also converge when defined by Pagoda clustering analysis (Figure 1B, top). We postulated that this group of cells is comprised of transition cells (TCs) that bridge the SSC and differentiating SPG stages (Figure 2A). A prediction of this hypothesis is these cells actively express genes of the later stage but still retain expression of genes corresponding to the earlier stage. In agreement with this, two groups of TCs co-expressed SSC-1 and SSC-2 marker genes (Figure 2B, left and middle boxes), suggesting they are transitioning between the SSC-1 and SSC-2 stages. A third group of cells co-express marker genes from all three subsets (Figure 2B, right box), suggesting they are transitioning to become Early diff-SPG. While it is possible that some of these "TCs" are cell doublets, the average mitochondrial content, number of genes expressed, and unique molecular indices (UMIs) per cell in the TC subset were comparable with other cell subsets (see STAR Methods); in addition, these TCs expressed markers largely not expressed in the related cell clusters (Figure 2A, bottom), suggesting doublets are rare. TCs may represent the transition between infrequent and active cell proliferation. The preceding stages—SSC-1 and SSC-2—have a cell cycle gene expression patterns consistent with infrequent cell proliferation, whereas the stage that immediately follows the TC stage—Early diff-SPG—has an active proliferation expression pattern (Figures 1C and 1D).

Identification of 3 Undifferentiated SPG States Expressing Distinct SSC Marker Genes

We noticed that several genes were expressed in distinct "zones" within the SSC-1 cluster (Figure 1B, bottom). This raised the possibility that there are distinct cell states within the SSC-1 subset, which we tested by performing PAGODA clustering analysis on only SSC-1 cells. This revealed the existence of three sub-clusters—SSC-1A, SSC-1B, and SSC-1C (Figures 2C and 2D)—each of which selectively expressed specific genes (Figures 2E and S2). These 3 sub-clusters, as well as the other SPG sub-clusters in and above the "doughnut," were robust, as they were also observed when clustering parameters (such as n.oddgenes), and sub-sampling iterations were tested (data not shown). We also validated these sub-clusters using DoKMeans (Seurat function) and non-negative matrix factorization clustering analyses (data not shown). Monocle trajectory analysis of these 3 subsets—alone or in combination with other germ cell subsets—indicated that SSC-1B is the least developmentally advanced (Figure 2F), which is consistent with the notion that SSC-1B is the most SSC enriched. Figure 2C presents a model—supported by both our Monocle and Pagoda analyses—which posits that SSC-1B cells convert into either SSC-1A or SSC-1C cells (travel left or right around the doughnut, respectively) to ultimately become differentiating SPG. However, our data do not exclude other possibilities, including that SSC-1A and/or SSC-1C are alternative stem cell states (see Discussion). Table S2 lists genes enriched in these 3 undifferentiated cell states.

Figure 2. Identification and Characterization of Human Undifferentiated SPG Subsets

(A) tSNE plot from Figure 1B showing that TCs are predominantly at the juxtaposition point of SSC-1, SSC-2, and Early diff-SPG subsets (top). The feature plot shows the expression of some gene markers enriched in TCs (bottom). The arrows indicate the developmental directionality of the cell subsets, based on our model (see text).

(B) Heatmap of top differentially expressed genes (DEGs) from the three cell subsets related to TCs compared to their expression in TCs themselves. DEGs that specifically mark the SSC-1, SSC-2, and Early diff-SPG subsets were defined as genes enriched in these 3 respective cell clusters when TCs were excluded. The blue and green boxes denote TCs expressing SSC-1A- and SSC-2-enriched genes. The red box marks TCs expressing genes enriched in SSC-1A, SSC-2, and Early diff-SPG.

(C) tSNE plot from Figure 1B showing the three SSC-1 subsets identified using PAGODA2.

(D) tSNE plot of only the 3 SSC-1 subsets.

(E) Feature plot, violin plot, and Monocle pseudotime trajectory expression pattern of SSC-1A, -1B, and -1C marker genes.

(F) Pseudotime trajectories of the adult SPG subsets shown in the key.

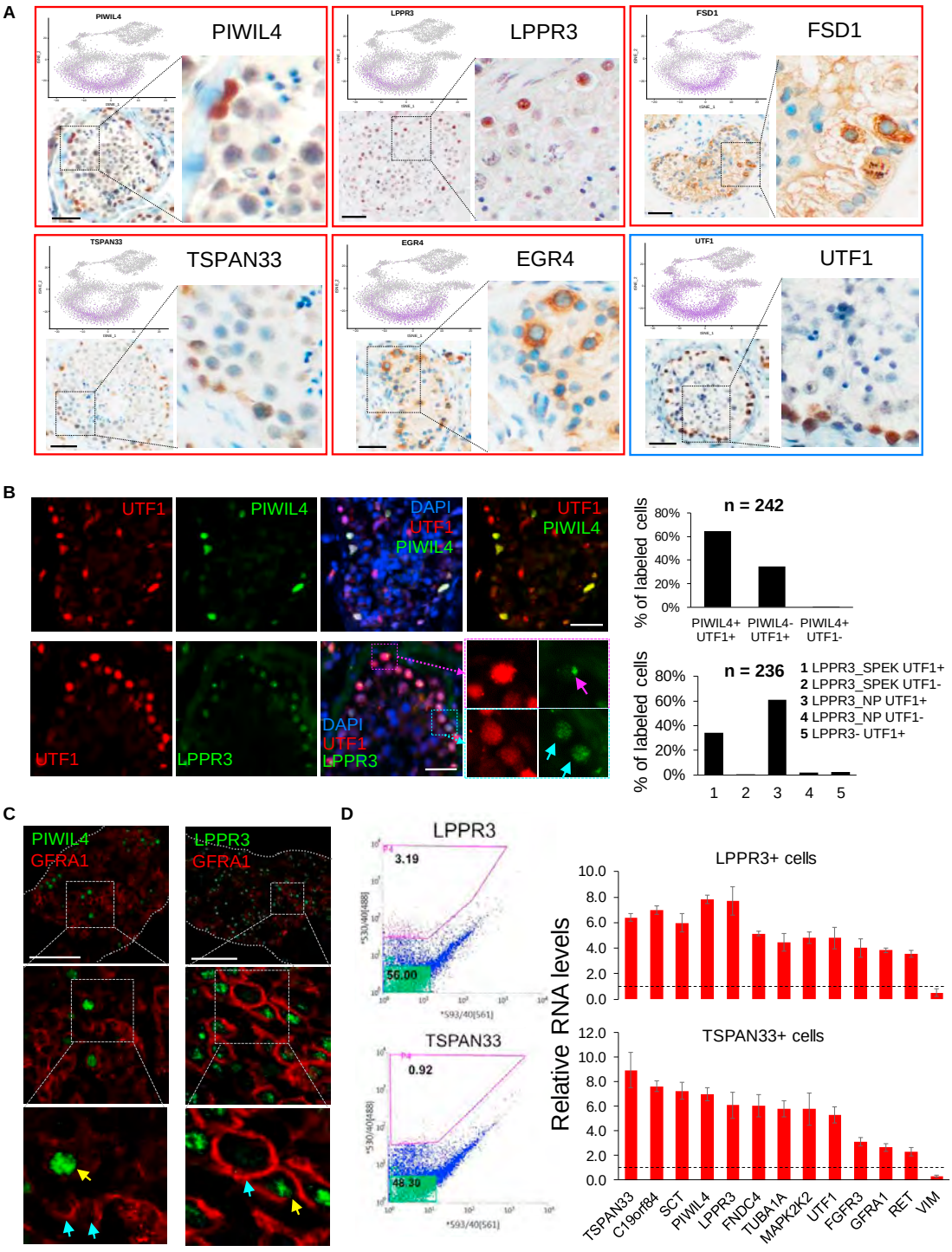


Figure 3. Identification and Characterization of Primitive Undifferentiated SPG Markers

(A) IHC analysis of 5 candidate SSC marker proteins (representative images of 2 biological replicates). The genes encoding these proteins are predominantly expressed in the SSC-1B region, as shown by the tSNE plot. As a control, we included UTF1, a previously defined undifferentiated SPG marker (Di Persio et al., 2017) whose gene is broadly expressed in SPG, as shown in its tSNE plot.

(B) IF analysis of adult testis sections stained with the antibodies shown. Anti-LPPR3 produced two different staining patterns: speckled (SPEK [pink arrow]) and whole nucleoplasm (NP [light blue arrow]). The cells were co-stained with an antibody against UTF1. Quantification of two biological replicates showed that both PIWIL4 and LPPR3 are expressed in a more selective set of cells than is UTF1. Scale bar: 50 μM.

(legend continued on next page)

Identification of Candidate SSC Marker Proteins

The identification of human SSC protein markers is critical for studying human SSCs, including generating methods to expand these stem cells for clinical use. Our discovery of a small subset of highly undifferentiated SPG provided an opportunity to potentially identify more selective SSC marker proteins than those currently known. As a positive control, we first performed IHC analysis with the well-established undifferentiated SPG markers FGFR3 and UTF1 (Di Persio et al., 2017; Guo et al., 2017) (Figures 3A and S3A). Antibodies against these 2 proteins marked cells at the periphery of the seminiferous tubule (Figures 3A and S3A), where SSCs are known to selectively reside. We then performed IHC analysis with antibodies against 5 proteins encoded by genes expressed in the least advanced SSC subset—SSC-1B—and found that all 5 are also localized in the tubule periphery (Figure 3A). We focused our subsequent analysis on 2 of these candidate SSC marker proteins: PIWIL4 and LPPR3 (also known as PLPPR3). Co-staining cells with PIWIL4 and UTF1 antibodies showed that PIWIL4 labeled ~64% of UTF1-positive cells, suggesting that the PIWIL4 protein is a more specific marker than the UTF1 protein (Figure 3B). This is consistent with the *PIWIL4* gene exhibiting a more specific pattern of expression in undifferentiated SPG subsets than the *UTF1* gene (Figures 1B and 1E). LPPR3 exhibited two patterns of expression: whole nucleoplasm (LPPR3-NP+ cells) and speckled nuclei (LPPR3-SPEK+ cells) (Figure 3B). Co-staining with UTF1 showed that ~61% of UTF1+ cells were LPPR3-NP+ and ~34% of UTF1+ cells were LPPR3-SPEK+ (Figure 3B, right), suggested that, similar to PIWIL4, LPPR3 labels a highly specific undifferentiated SPG subset.

We next examined the expression profile of selected SSC markers by whole mount IF staining, which allows for an analysis of intact SSCs in their natural topographical distribution (Di Persio et al., 2017). By co-staining with the undifferentiated SPG marker GFRA1 (Di Persio et al., 2017), we found that PIWIL4 was expressed in only ~38% of GFRA1-positive cells (Figures 3C and S3B), providing further evidence for the specificity of PIWIL4. Whole-mount IF of LPPR3 co-staining with GFRA1 shows that LPPR3-NP+ cells tended to express higher levels of GFRA1 than LPPR3-SPEK+ cells (Figure S3C). Together, this protein analysis, coupled with our scRNA-seq analysis, suggested that PIWIL4 and LPPR3 label a specific subset of undifferentiated SPG with the characteristics of SSCs.

Because it is essential that an SSC marker be present on the cell surface to be useful for SSC isolation, we tested the efficacy of candidate SSC markers encoding cell surface proteins for this purpose. We selected *LPPR3* and *TSPAN33*, as both encode cell membrane proteins (Luu et al., 2013; Yu et al., 2015) and both are selectively expressed in primitive undifferentiated SPG (Figures 1B and 1E), with relatively infrequent expression in other adult testicular cells (Figure S3D). Using FACS, we found that antibodies against both of these cell-surface proteins highly enriched cells expressing undifferentiated SPG and candidate SSC marker genes (Figure 3D). GFRA1, which is widely used as both a mouse and human SSC marker (Garbuzov et al., 2018; He et al., 2012), was less enriched than most other SSC markers (Figure 3D), consistent with our evidence that *GFRA1* is infrequently expressed at levels detected by scRNA-seq in primitive undifferentiated SPG (Figure 1B). As more evidence for selectivity, LPPR3 and TSPAN33 marked a more selective subset of testicular cells (~1% to 3% of total testicular cells) (Figure 3D) than FGFR3 (~6% of total testicular cells) (Figure S3A). FGFR3 is a previously identified cell-surface protein enriched on undifferentiated SPG (Guo et al., 2017) that is encoded by a gene that we find is broadly expressed across all undifferentiated SPG states (Figures 1B and 1E).

Identification of Cell Subsets in Testes from Human Neonates

To assess the developmental origins of the primitive undifferentiated SPG in adult testes, we performed scRNA-seq analysis on human neonatal testis. After pre-filtering, 14,862 cells were obtained from two newborn human whole testes (2 days and 7 days old). Clustering using PAGODA2 allowed us to classify these neonatal testes cells into 5 major cell clusters—germ cells, SCs, PTMs, LCs, and ECs (Figures 4A and S6C)—based on known marker genes (Figure S4A; Table S2). A striking difference between the neonatal testes with the adult testes is the former have a much lower percentage of germ cells (Figure 4A). This is consistent with the fact that spermatogenesis has yet to initiate at birth, as well as past studies quantifying germ cells by morphology from human newborns and infants (Cortes et al., 1995).

Germ Cell Subsets in Neonatal Human Testes

To elucidate the nature of the germ cells in the neonatal testis, we performed clustering analysis specifically only on the germ cells. PAGODA2 clustering analysis revealed 2 neonatal germ cell clusters (Figure 4B). One cell cluster has an expression profile highly reminiscent of PGCs (e.g., expression of pluripotency genes; e.g., *POU5F1* and *NANOG* [Figure 4B, bottom]), and thus, we named this cluster "PGC-like" (PGCL). The second cell cluster is dominated by cells that have largely extinguished the expression of PGC genes and instead express several genes known to mark human pre-SPG (Hayashi et al., 2012; Song et al., 2013), including *MAGEA4* and *RHOXF1* (Figure 4B, bottom), and thus, we refer to this cluster as "PreSPGs." Monocle clustering divided the PreSPG into two distinct clusters that we named "PreSPG-1" and "PreSPG-2" (Figure 4B). We validated these 2 PreSPG clusters and the PGCL cluster using the DoKmean

(C) IF analysis of intact human seminiferous tubules (whole mount) stained with the indicated antibodies (representative images of 2 biological replicates). Dashed lines outline the seminiferous tubules. The yellow arrows label PIWIL4+GFRA1+ and LPPR3+GFRA1+ cells, whereas the light blue arrows label GFRA1 single-positive cells. Scale bars: 100 (top) and 10 μM (bottom).

(D) FACS plot of adult human testicular cells stained with the antibodies shown (left). The percentage of positive and negative cells are indicated in upper and lower regions, respectively; <0.1% positive cells stained with secondary antibody only (see Figure S3A). qPCR analysis was performed on sorted positive cells and unfractionated cells. The values shown are from FACS-purified cells relative to unfractionated cells (mean ± SD from two biological replicates), the latter of which was given a value of 1 (dotted line). The 10 genes listed first are known undifferentiated SPG markers or candidate SSC markers identified herein. The last gene *VIM* is a somatic cell-expressed gene and, thus, serves as a negative control.

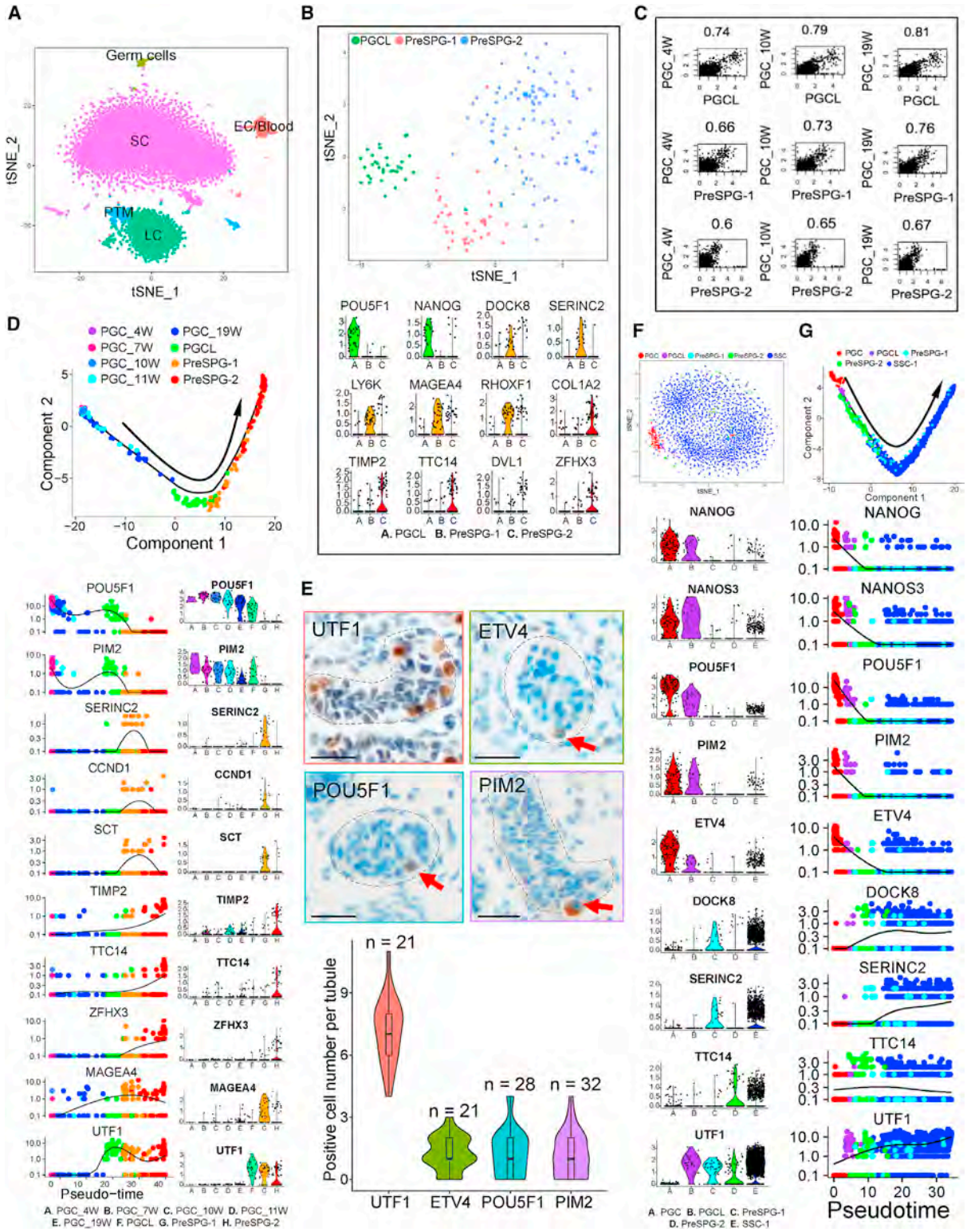


Figure 4. Identification and Characterization of Neonatal Germ Cells

(A) tSNE plot of neonatal human testicular cells analyzed by scRNA-seq analysis. Cell subsets were identified based on the expression pattern of known marker genes (Figure S4A). The fraction of the total cells in each subset is: germ cells (1%), Sertoli cell (SC) (59%), LC (32%), PTM (5%), and EC and blood (4%).

(legend continued on next page)

function in Seurat (data not shown). The PreSPG-1 subset expresses several genes that most PreSPG-2 cells only weakly express, including *DOCK8*, *SERINC2*, *LY6K*, *MAGEA4*, and *RHOXF1* (Figure 4B, bottom). Conversely, other genes selectively mark the PreSPG-2 subset, including *COL1A2*, *TIMP2*, *TTC14*, *DVL1*, and *ZFHX3* (Figure 4B, bottom, and data not shown). Table S2 lists genes enriched in the 3 neonatal germ cell clusters we identified.

To investigate the nature of these 3 neonatal germ cell subsets, we compared their expression profiles with the published datasets from Guo et al., who performed scRNA-seq analysis on gonadal PGCs from human fetuses between 4 weeks and 19 weeks old (Guo et al., 2015). Pearson correlation analysis showed that the average expression of genes in PGCs (at all 5 gestational time points) was more similar to that of the PGCL subset than it was to PreSPGs (Figure 4C and data not shown), confirming the similarity of the neonatal PGCL subset with fetal PGCs. The correlation coefficient increased with PGC gestational age, consistent with the maturation of PGCs as fetuses age. As further evidence for the similarity of neonatal PGCLs with embryonic PGCs, t-distributed stochastic neighbor embedding (tSNE) analysis showed that the PGCL subset clustered near cells from the last PGC time point (19 weeks) (Figure S4B). Together, these data support the notion that the PGCL subset in the neonate are similar to (and perhaps directly derived from) the PGCs in the developing embryonic gonad.

To further examine the developmental relationship of PGCs and the neonatal germ cell subsets, we performed Monocle trajectory analysis. The least differentiated cells—the 4-week PGCs—clustered on one end of the pseudotime axis (the left side) and the later stage PGC samples aligned according to their gestational age (Figure 4D, top). Directly following the oldest PGC sample (19 weeks) were cells from the PGCL cluster (Figure 4D, top), confirming that PGCs and PGCLs are highly related. Following the PGCL cluster along the pseudotime trajectory were PreSPG-1 and PreSPG-2 cells (Figure 4D, top). Together, these data are consistent with the following developmental order: PGC → PGCL → PreSPGs (PreSPG-1 and PreSPG-2).

To identify neonatal germ cell marker genes, we generated a list of genes differentially expressed in neonatal germ cell subsets (Table S2). This allowed us to identify several genes co-expressed in the fetal PGCs and the neonatal PGCL subset but rarely in cells in the neonatal PreSG subsets (Figure 4D, bottom). This not only provided a useful set of markers but was further evidence that PGCs and PGCLs are similar to each other but distinct from PreSPG. We identified other marker genes differentially expressed between PGCs and PGCLs (Figure 4D), which will be useful to specifically mark the neonatal stage of germ cell development. We found that *MAGEA4*—which has widely been used to mark both PreSPG and adult SPG (Hayashi et al., 2012)—is a good gene marker to discriminate between PreSPG and PGCL, as it is rarely expressed by the latter (Figure S4B, bottom). Together, these genes and other genes listed in Table S2 have the potential to serve as valuable markers to study both embryonic and neonatal male germ cell development.

Identification of PGCL Marker Proteins

Our data suggest that the PGCL subset is a key transition stage between embryonic PGCs and adult SPG. To screen for PGCL-specific protein markers, we identified genes differentially expressed between PGCLs and other germ cells (Table S2) and examined the expression of their encoded proteins in neonatal testes by IHC. We found that antibodies against 3 proteins—ETV4, PIM2, and POU5F1—selectively labeled neonatal cells with large nuclei (Figure 4E, top), a characteristic typical of germ cells (Gaskell et al., 2004). To determine the selectivity of these markers, we also stained with an antibody against UTF1, which is encoded by a gene highly expressed in all 3 neonatal germ subsets (Figure 4D). There was an average ~7 UTF1+ cells/tubule section, whereas there was an average of ~2 ETV4+, PIM2+, and POU5F1+ cells per tubule section (Figure 4E, bottom). This demonstrates the specificity of these markers and suggests that they selectively label PGCLs. We note that although the POU5F1 and PIM2 antibodies only appeared to stain germ cells, the antibody we used against ETV4 also stained cells outside of the seminiferous tubule (Figure S4C), suggesting that ETV4 is also expressed in LCs. Together, these PGCL protein markers will be valuable for analyzing this key transition stage of human germ cell development.

Neonatal Germ Cells Progress to a Near-Adult SSC State

To determine the relationship of neonatal germ cells with adult SSCs, we performed PAGODA2 analysis on neonatal germ cells merged with the adult SSC-1 subset. SSC-1 cells formed a large

(B) Neonatal germ cell subsets and markers defined by clustering analysis using Monocle. The tSNE plot (top) was generated from the cells in the germ subset defined in (A). Violin plots (bottom) show the expression pattern of selected PGCL and PreSPG gene markers.

(C) Pearson correlation analysis of the cumulative expression of genes in PGCs at the gestational ages indicated with neonatal cell clusters. The values at top are the R^2 values.

(D) Developmental timeline of PGCs and neonatal germ cell subsets. The top shows Monocle pseudotime trajectory analysis of the neonatal germ cell subsets defined in (B), as well as PGCs at the gestational ages indicated from Guo et al. (2015). The arrow indicates the developmental direction, based on the 4-week PGC and PreSPG subsets being the least and most differentiated, respectively. The expression pattern of marker genes is plotted along the pseudotime axis (bottom left) and with violin plots (bottom right).

(E) Representative images of the expression pattern of 3 PGCL protein markers in a single neonatal (D7) human testis, as determined by IHC analysis. Red arrows mark positive cells. Gray dotted lines mark the approximate margins of the seminiferous tubules. As a positive control, cells were stained with an antibody against UTF1, which is encoded by a gene expressed in most neonatal germ cells (Figure 4D). The violin plot below shows quantification of positively stained cells. Scale bar: 50 μM.

(F) Relationship of neonatal germ cells with PGCs and adult SSCs. tSNE analysis and violin plots of PGCs (from Guo et al., 2015), neonatal germ cell subsets (defined in B, above) and the SSC-1 subset. The violin plots show stage-specific marker genes.

(G) The neonatal testis has adult SSC-like cells. The top shows Monocle pseudotime trajectory analysis of the indicated germ cell subsets. The arrow indicates the developmental direction. The expression pattern of marker genes is plotted along the pseudotime axis (bottom).

cluster that also harbored PreSPG-1 and PreSPG-2 cells (Figure 4F, top), consistent with these neonatal germ cell subsets exhibiting a similar gene expression profile as adult SSCs. Indeed, the PreSPG and SSC-1 subsets co-expressed many genes, including *ID4*, *MAGEA4*, and *UTF1* (Figure 4F, bottom; Table S2). In contrast to PreSPGs, PGCLs clustered with embryonic PGCs on the periphery of the SSC-1 cluster (Figure 4F, top), further supporting our earlier analyses indicating that the PGCL and PreSPG neonatal subsets are distinct and that the latter is more advanced (Figure 4D). As further support, Pearson correlation analysis of cumulative gene expression showed a stronger correlation between the SSC-1 subset and the PreSPG-1 and PreSPG-2 subsets than the PGCL subset (Figure S4D). Monocle analysis confirmed that PGCL cells are the most undifferentiated neonatal subset (migrating just after embryonic PGCs), followed by PreSPG-2, PreSPG-1, and SSC-1 (Figure 4G). Although most PreSPG-1 cells preceded SSC-1 cells along the pseudotime trajectory, some co-migrated with the SSC-1 cells, suggesting that the neonate has a small cohort of germ cells with SSC characteristics.

Somatic Cells in Developing and Adult Human Testes

To define somatic cells in neonatal and adult testes, we first performed additional filtering followed by PAGODA2 clustering on our entire dataset described above to obtain 25,669 human testicular cells. This analysis revealed discrete cell clusters corresponding to neonatal ECs, LCs, SCs, and PTMs (Figures S5A and S5B). Adult ECs, LCs, and PTM clusters were also identified; their position in the tSNE plots differed from the corresponding neonatal cell clusters (Figures S5A and S5B), indicative of significant differences in gene expression between the neonatal and adult stage. As discussed above, we were not able to identify an adult SC cluster, likely due to the very large size of adult SCs. Cluster identities were assigned based on the expression of established mammalian somatic cell markers (Table S1). Re-clustering of only somatic cells from neonatal and adult testis corroborated these assignments (Figures 5A and 5B). An analysis of cell cycle gene expression suggests somatic cells are seldom cycling (Figure 5C), consistent with an earlier study (Berensztein et al., 2002).

To identify marker genes enriched in the SC, LC, or PTM populations, we performed differential gene expression analysis, comparing each somatic cell population with all other cells from the entire dataset. These data corroborated previously known mammalian somatic cell markers (Figures 5D and S5C–S5E) and defined new markers labeling LCs (*LHX9*, *IGF2*, *DCN*, and *PTGDS*) (Figure S5C), PTMs (*AEBP1* and *PTCH1*) (Figure S5D), and SCs (*BEX1* and *CITED*) (Figure S5E).

Pseudotime analysis showed that neonatal LCs comprise one end of the developmental trajectory, whereas adult LCs are on the opposite end (Figure 5E), providing evidence that LCs undergo a discrete developmental shift between the neonatal and adult stage. The LC marker genes *INHBA*, *LHX9*, and *IGF2* are downregulated, whereas *DCN*, *PTGDS*, and *IGF1* are upregulated during this developmental transition (Figure S5C). Some classes of genes enriched for expression in fetal LCs compared to adult (Figure 5E; e.g., ribosomal genes, Table S3) suggest that fetal LCs are more metabolically active than adult LCs.

Pseudotime analysis showed that, similar to LCs, PTMs primarily align according to their age of origin, although we identified a subset of neonatal-like adult PTMs (PTM-NL) (Figure 4F). This PTM-NL subset has a gene expression pattern similar with neonatal PTMs and resides in a distinct cluster alongside neonatal PTMs (Figures 5B and 5F). Our data suggest the PTM marker genes *TAGLN* and *MYH11* are downregulated as PTMs develop, whereas *AEBP1* and *PTCH1* are upregulated during this developmental transition (Figure S5D).

Ligand and Receptor Gene Expression in Human Neonatal Testes

It is well established that spermatogenesis depends on signaling between somatic and germ cells (Rossi and Dolci, 2013). However, the signaling events occurring prior to spermatogenesis—including during SSC establishment—are largely unknown. To investigate signaling events that may occur between cells at the neonatal stage, we identified ligand and receptor gene pairs expressed in neonatal testicular cell populations (Table S3). This revealed potential instances of NOTCH, KIT, HEDGEHOG (Hh), and WNT signaling (Figure 6). Other examples of ligand-receptor pairs expressed in germ and somatic cell types of the neonatal and/or adult testis are provided in Table S3. Although many of the signaling genes appear to be expressed in a relatively small subset of germ cells, this may be an underestimate due to the low sensitivity of scRNA-seq analysis. The expression of these signaling genes may also be transient in germ cells. See the Discussion for implications of these findings.

DISCUSSION

Male germ cell development is a complex and intricate pathway that propels embryonic germ cells—PGCs—to progress through a series of states involving cell proliferation, apoptosis, and differentiation to become SSCs, which, through self-renewal and differentiation, continuously generate sperm through the process of spermatogenesis. Although much has been learned about these various steps in rodents, we are only just beginning to unravel male germ cell development and spermatogenesis in humans. In this communication, we shed light on these processes through scRNA-seq analysis of neonatal and adult human testis.

Neonatal Male Germ Cell Development

Classical studies using morphological analysis and protein markers have suggested that only a single germ cell type exits in human newborn testes (Fukuda et al., 1975; Gaskell et al., 2004; Vilar, 1970). Our discovery of 3 distinct germ cell states at the newborn stage—which we have named PCGL, PreSPG-1, and PreSPG-2—enriches our view of this stage of human male germ cell development. Our analyses support a model in which human fetal PGCs differentiate into PGCLs and, subsequently, PreSPGs, both of which populate the human testes at birth (Figure 7A). Although neonatal PGCLs are similar to embryonic PGCs, they express different gene and protein markers, which will be critical for future studies to isolate, culture, and characterize PGCLs. It will be intriguing to know whether

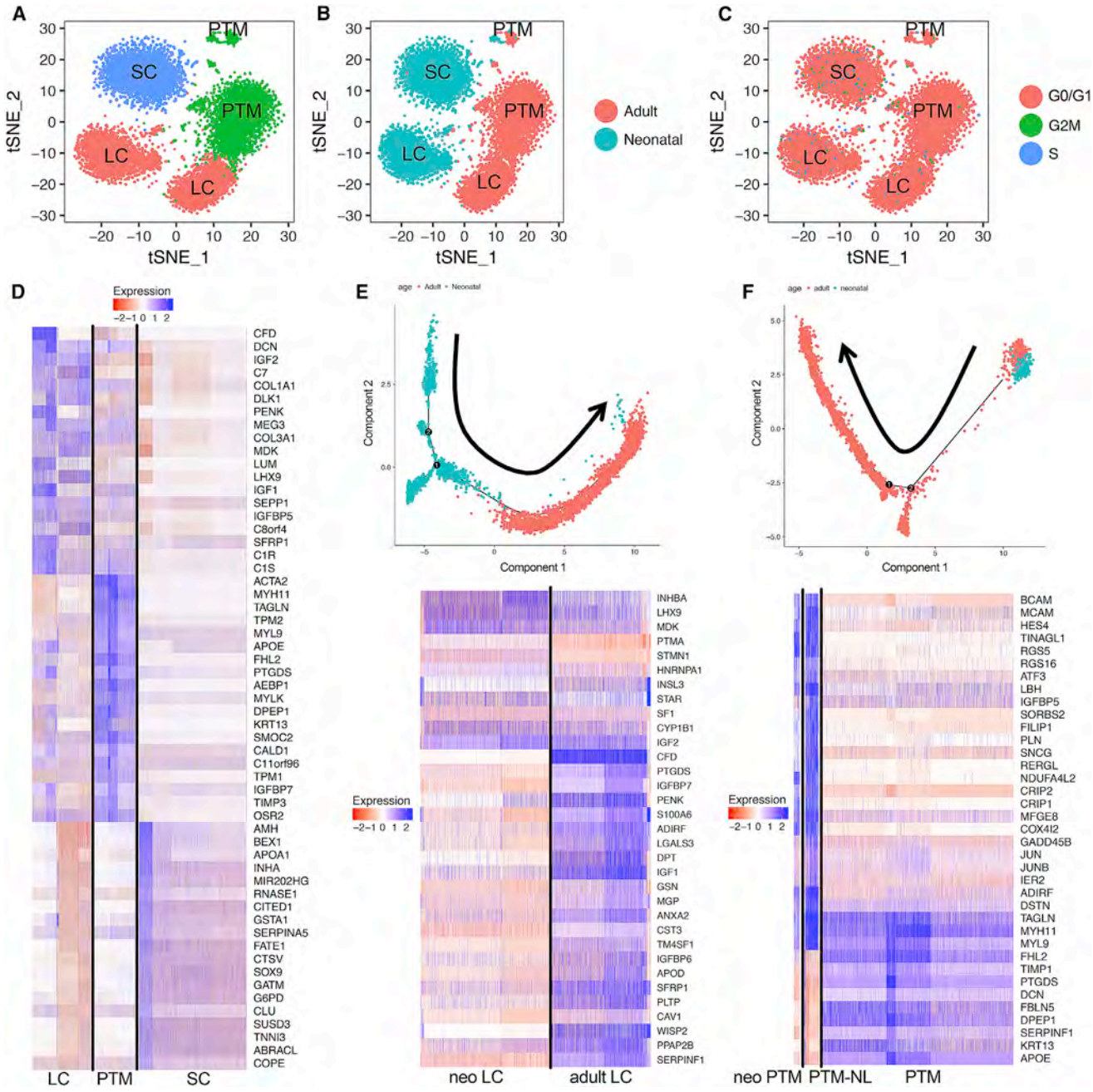


Figure 5. Characterization of Somatic Cell Populations in Human Testes

(A) tSNE plot of adult and neonatal testicular somatic cells. Clusters were annotated based on known markers shown in (D)–(F). The percentage of the total in each subset is neonatal SC (20.0%), neonatal LC (31.3%), neonatal PTM (1.3%), adult SC (0.2%), adult LC (23.0%), and adult PTM (24.2%).

(B) tSNE plot of the same clusters as in (A) with original sample source identities (neonatal and adult) indicated.

(C) tSNE plot of the same clusters as in (A) marked by phase of cell cycle, as determined by average expression of G2-M and S phase gene expression (Kowalczyk et al., 2015) in each cell cluster relative to each other.

(D) Heatmap showing the expression profile of neonatal and adult somatic cell marker genes.

(E) LC markers used for cluster annotation. Monocle pseudotime trajectory of neonatal and adult LC subsets (top). Heatmap shows the expression profile of neonatal and adult LC markers (bottom).

(F) PTM markers used for cluster annotation. Monocle pseudotime trajectory of neonatal and adult PTM subsets. Heatmap shows the expression profile of neonatal and adult PTM markers (bottom).

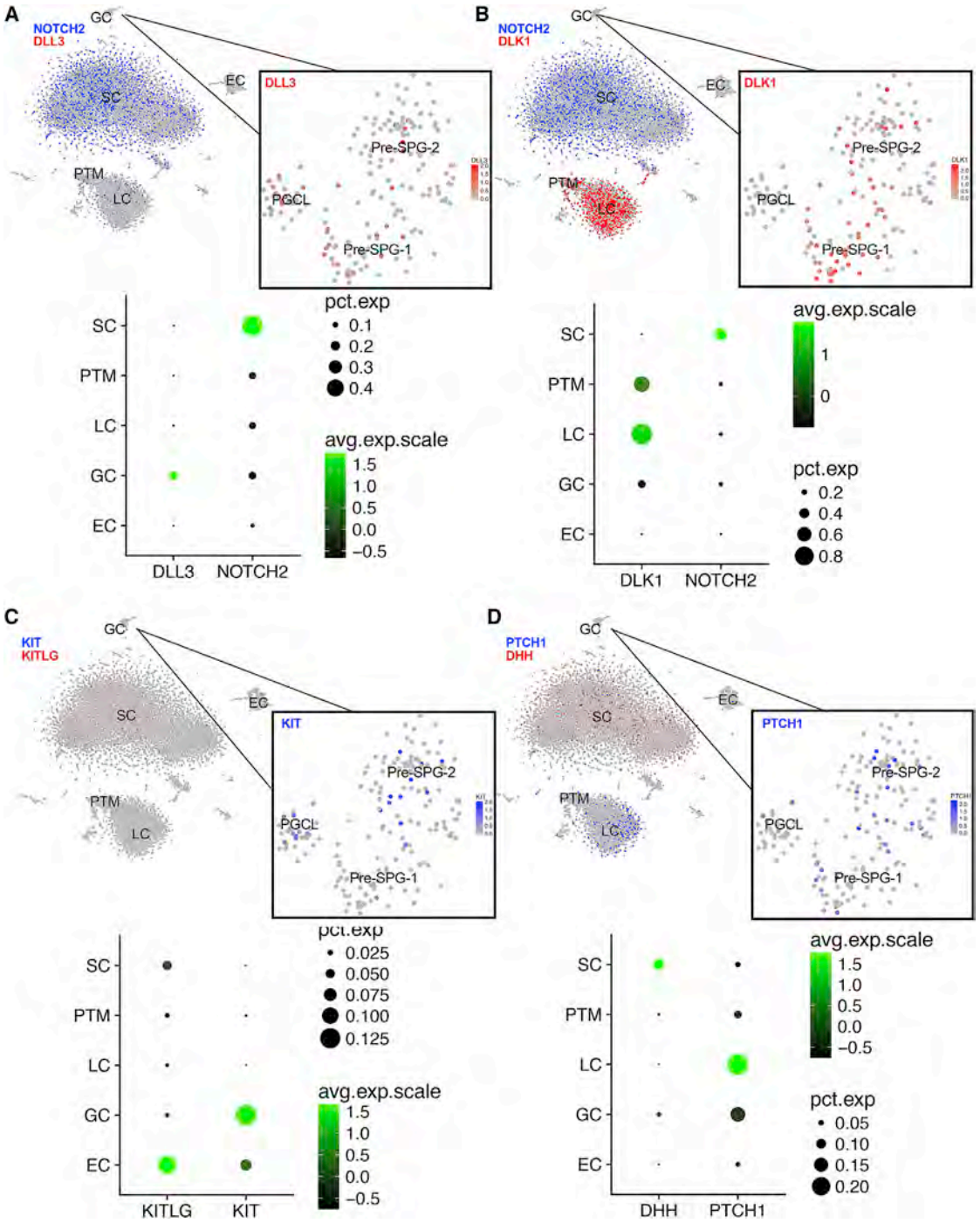


Figure 6. Receptor-Ligand Signaling Interactions in Human Testes

(A and B) Notch ligand-receptor co-expression in the neonatal testis. tSNE plot of all neonatal cells (from Figure 4A) showing expression of the receptor NOTCH2 (in neonatal SC) and the ligands DLL3 (in neonatal germ cells) (A) and DLK1 (in neonatal PreSPG, LC, and PTM) (B). The inset shows tSNE plot of expression of only the ligand genes in neonatal germ cells (from Figure 4B). The dotplot shows expression of the indicated genes in the cell subsets shown. Note that percentage of positive cells indicated is a conservative estimate from scRNA-seq, an insensitive technique.

(C) Kit ligand-receptor co-expression in the neonatal testis. KIT is expressed in neonatal germ cells and the ligand KITLG is expressed in neonatal SC and EC.

(D) Hh ligand-receptor co-expression in the neonatal testis. The Hh receptor PTCH1 is expressed in neonatal PreSPG, LC, and PTM. The Hh ligand DHH is expressed in neonatal SC.

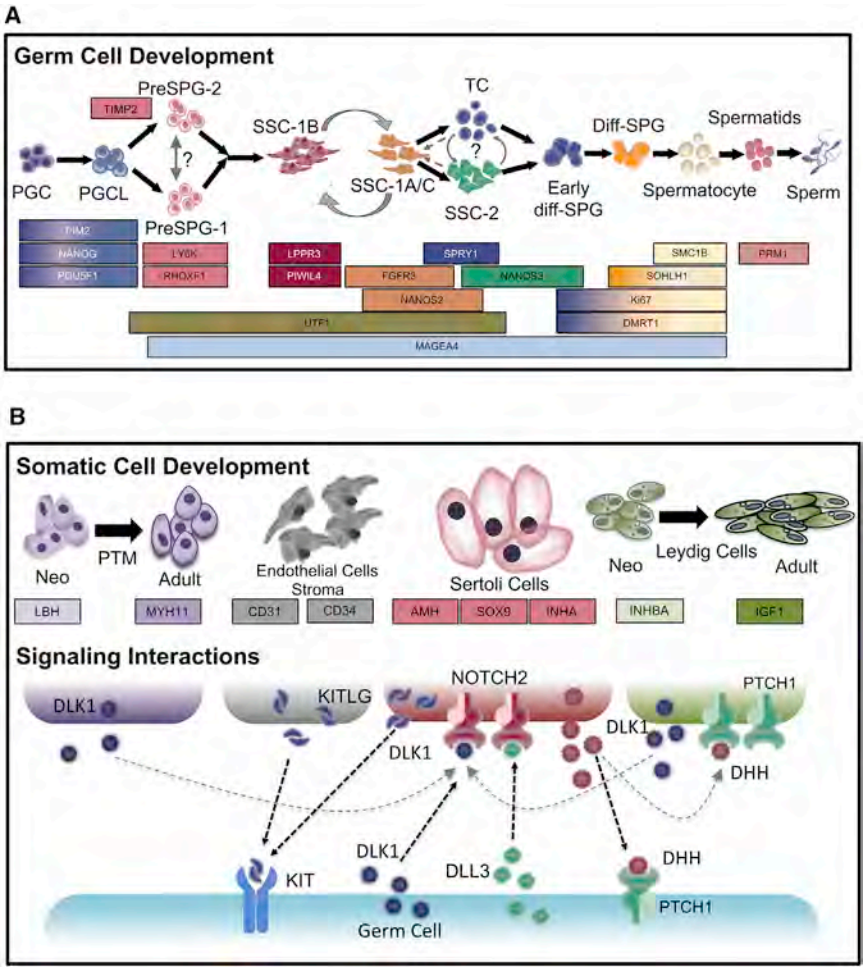


Figure 7. Human Male Gametogenesis Model and Neonatal Signaling

(A) Model. Black arrows indicate the direction of differentiation. Grey arrows indicate transitions between different cell states. The expression pattern of key marker genes (both previously known and discovered in this study) are indicated by boxes (other marker genes are listed in Table S2).

(B) Top: neonatal and adult somatic cells and selected marker genes. Bottom: signaling interactions between cells in the neonatal testes predicted from the scRNA-seq data shown in Figure 6. Gray arrows, somatic cell-somatic cell signaling. Black arrows, somatic cell-germ cell signaling.

epigenetic reprogramming events occur in PGCLs as a means to prepare them to ultimately become SSCs.

It will also be intriguing to functionally define the other two neonatal germ cell states we identified: PreSPG-1 and PreSPG-2. Another future question will be to determine their relationship: one may give rise to the other, they may be interconvertible, or they may be part of two parallel differentiation pathways in the neonatal testis. We identified some PreSPG cells that have a SSC-like gene expression pattern. These may be emergent SSCs that have begun to seed the human testis, by analogy with what occurs in mice soon after birth (Niedenberger et al., 2015). This idea is also consistent with classical studies showing that neonatal germ cells begin being replaced by cells harboring the morphology of SSCs in the first year of human life (Culty, 2009; Paniagua and Nistal, 1984). Further investigation is needed to assess whether these SSC-like neonatal cells are functional stem cells. The only currently known approach to measure human SSC activity is xenotransplantation of human testicular cells into the testes of immunocompromised mice, followed by an assessment of colony number (Nickkholgh et al., 2014; Valli et al., 2014b). Although transplanted human SPG are unable to fully progress through spermatogenesis, this xenotransplantation assay allows one to at least infer human SSC activity (Fayomi and Orwig, 2018).

Cell Signaling during Neonatal Testicular Development

Although cell-cell signaling events have been shown to be important for developing male embryonic germ cells and during spermatogenesis (Rossi and Dolci, 2013), we know little about signaling mechanisms that act on male germ cells between these two stages. In this study, we begin to fill this gap by identifying several examples of cell-cell signaling events that may occur during the neonatal period (Figure 7B). First, NOTCH ligand and receptor genes are expressed in human neonatal testis in a pattern suggesting that germ cells signal to SCs through the NOTCH pathway. This is of interest given that NOTCH signaling is critical for germ cells and SCs at other stages of development (Barrionuevo et al., 2006). Second, KIT and its ligand are expressed in neonatal germ cells and somatic cells, respectively. It will be interesting to determine what role KIT signaling might have in neonatal cells. It has been shown that the KIT pathway is essential for PGC survival and proliferation during fetal development and postnatal SPG differentiation in mice (Rossi, 2013). Third, the Hh ligand gene *DHH* and its receptor gene *PTCH1* are expressed in the neonatal testes in a pattern that suggests that neonatal SCs signal through the Hh pathway to several other neonatal cell types, including PreSPGs. This is intriguing given that *Dhh* is essential for normal testis formation and germ cell development in mice (Bitgood et al., 1996), and mutations in *DHH* are associated with human gonadal disgenesis (Canto et al., 2004). Finally, we found that neonatal germ cells and SCs express the WNT activator ligand and receptor genes *RSPO2* and *LGR4*, respectively, the latter of which is known to be critical for fertility in mice (Mendive et al., 2006). WNT signaling may also be important for somatic-somatic cell communication, as we found that

neonatal SCs and PTMs express the WNT ligand *WNT6* and corresponding receptor *FRZB*, respectively.

We also identified receptor-ligand pairs expressed in adult human testes. For example, adult germ cells express the WNT activator *RSPO2*, whereas adult PTMs express the WNT receptor *LGR4*, suggesting a role for WNT signaling in spermatogenesis. Hh signaling may also have a role in spermatogenesis, based on our finding that the Hh ligand/receptor pair *DHH* and *PTCH1* is expressed in adult germ and somatic cell subsets. Together, the receptor-ligand datasets we generated (Table S3) provide a potentially important resource for future investigation into the signaling events critical for both early human male germ cell development and human spermatogenesis.

SPG Subsets

Our scRNA-seq and protein marker analyses of adult human testes revealed the existence of 4 human SPG subsets: SSC-1, SSC-2, Early diff-SPG, and Diff-SPG (Figure 7A). Our evidence indicates that the SSC-1 subset is the most primitive, suggesting that this subset harbors all or most human SSCs. Similar to the SSC-1 subset, the SSC-2 subset is comprised of undifferentiated SPG but appears to be more advanced developmentally. The Early diff-SPG and Diff-SPG likely correspond to early and late B-SPG stages, respectively, based on the expression of several marker genes.

We obtained several lines of evidence for the existence of an intermediate cell state between the undifferentiated and Early diff-SPG stages that we named the "TC" subset (Figure 7A). We posit that TCs represent the transition between infrequent and active cell proliferation. The preceding stage, SSC-1, has a cell cycle gene expression pattern consistent with infrequent cell proliferation, whereas the stage that immediately follows the TC stage, Early diff-SPG, has an active proliferation expression pattern. Several genes exhibit elevated expression in TCs relative to the less advanced SPG subsets. Two of these TC-enriched genes—*CCND2* and *SPRY1*—encode cell proliferation-promoting proteins (Glickstein et al., 2009; Shea et al., 2010), raising the possibility that these genes coerce TCs to enter the proliferation program characteristic of differentiating B-SPG. We note that Wang et al. (2018) also obtained some evidence for cells intermediate between undifferentiated and differentiating SPG, but the gene markers expressed in these cells (e.g., *DMRT1* and *KIT*) suggest a later stage of differentiation than the TC state we identified.

We also obtained several lines of evidence for the existence of 3 states within the SSC-1 subset, each of which is selectively labeled with different gene markers (Figures 2C and 7A). Clustering and trajectory analyses suggested that the most primitive state—SSC-1B—harbor cells that can traverse two different pathways to generate differentiating SPG. There are various explanations for this bifurcation. The possibility we favor comes from the classical notion that there are two types of stem cells in the testes: reserve SSCs that function only in response to injury and active SSCs that divide slowly on a regular basis to maintain homeostasis (Oakberg, 1971). The SSC-1B and SSC-2 subsets may represent these two stem cell states. Both ultimately give rise to cells that differentiate to form B-SPG, but cells in the SSC-1B state must go through the SSC-1A and TC states to become B-SPG (clockwise around the doughnut in Figure 2C), whereas cells in the SSC-2 state can go directly to the TC state to become B-SPG (anti-clockwise around the doughnut in Figure 2C). We suggest that the reserve and active SSCs each serve a different purpose; the former provide a stem-cell reserve (e.g., as a reservoir in the case of insult) and the former divide regularly (albeit infrequently to minimize mutation rate) to provide a continuous supply of progenitors under steady-state conditions. By analogy, the olfactory system depends on reserve and active stem cells (Schwob, 2002), and the hematopoietic system harbors long-term and short-term stem cells (Wilson et al., 2009).

We further speculate that the cells in the different states in the doughnut in Figure 2C are in equilibrium, such that they can interconvert. Although this concept is somewhat controversial in the field,several studies have obtained evidence that some SPG states in mice can interconvert (e.g., Hara et al., 2014; Nakagawa et al., 2010). We suspect that a particular direction will be favored (e.g., SSC-1B converting to the SSC-1A and -1C states), but the ability to reverse direction at some frequency explains why we found many cells in the different states, particularly because SSCs only rarely divide.

During preparation and review of our manuscript, three papers were published that also used scRNA-seq analysis to define human SPG and other human testicular cell types (Guo et al., 2018; Hermann et al., 2018; Wang et al., 2018). Wang et al. (2018) and Hermann et al. (2018) identified SPG subsets that likely correspond—based on gene markers—to the SSC-1 and SSC-2 subsets we defined. Guo et al. (2018) performed detailed analysis of 3 undifferentiated SPG states that they named "State 0-2," which likely correspond to our SSC-1B, SSC-1A and SSC-1C, and SSC-2 subsets, respectively, based on gene markers expressed. Together, our four studies reinforce each other by defining common cell states marked by specific genes. Our studies also complement each other. By analyzing more cells and using additional approaches, our study identified additional SPG subsets, obtained evidence as to their developmental lineage, and defined new protein markers that are strong candidates to be highly selective for human SSCs.

SSC Markers

A focal point of interest in the field has been SSCs, as these stem cells maintain spermatogenesis and have clinical potential for treating male infertility. Despite the intense interest in SSCs, it is not yet possible to identify SSCs with any level of certainty because no SSC-specific markers have yet been defined. Although considerable progress has been made in identifying SSC markers in the mouse, no single unambiguous SSC marker has been identified that only labels mouse SSCs (Kubota and Brinster, 2018). Some effort has been made to identify SSC markers in humans, but only a few human SSC markers have been identified and their specificity is unclear (von Kopylow and Spiess, 2017). In our study, we leveraged our scRNA-seq analysis to identify genes selectively expressed in the most primitive undifferentiated SPG subset: SSC-1B. We verified the specificity of five of these SSC-1B-enriched genes by performing IF, IHC, and FACS analyses on their encoded proteins. Using an initial scRNA-seq screen similar to ours, Guo et al. (2018) identified some of the same candidate SSC protein markers.

Our analyses suggested that several markers shown to enrich for mouse SSCs—such as *Id4*, *Eomes*, and *Zbtb16* (Fayomi and Orwig, 2018; Helsel et al., 2017)—may also be useful for enriching human SSCs. However, we found that human *ID4* and *ZBTB16* are broadly expressed in human SPG, whereas *EOMES* is only expressed in a small proportion of primitive undifferentiated human SPG, at least as detected by scRNA-seq. We found that several markers previously suggested to enrich for human SSCs—such as *GFRA1*, *FGFR3*, and *UTF1* (Di Persio et al., 2017; Guo et al., 2017)—are not only expressed in the primitive SSC-1B subset but also more advanced SPG subsets, including SSC-2 and/or Early diff-SPG. Although these previously defined markers are less selective than the SSC-1B markers we have defined, they will potentially still be useful, as they will likely mark SPG progenitors and may also mark alternative SSC-enriched cell subsets. The expression level of broadly expressed SPG markers can also be useful for staging.

We used antibodies against two of the cell-surface SSC-1B protein markers we identified—LPPR3 and TSPAN33—to highly enrich for primitive undifferentiated SPG, based on qPCR analysis of a plethora of markers. We were able to use unfractionated adult testicular cells for this enrichment, as scRNA-seq analysis of all human testicular cells showed that the genes encoding these two cell-surface proteins are relatively selective for primitive SPG. Together, this provides a proof of principal for future clinical studies to use SSC therapy as a means to treat human male infertility. SSC therapy will also require the development of methods to expand human SSCs in culture (Vij et al., 2018). Our identification of genes expressed by testicular somatic cells, including those encoding cell-cell signaling factors, may prove useful in developing cocktails containing "testes niche" factors that allow for the propagation and expansion of human SSCs *in vitro* for clinical application *in vivo*.

STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Subject details
  - Human testis samples preparation
- METHODS DETAILS
  - 10X Genomics library preparation
  - Mapping, cell identification and clustering analysis
  - Cell trajectory analysis
  - FACS and qRT-PCR analysis
  - Immunohistochemical analysis
  - Immunofluorescence analysis
  - Whole mount staining
- QUANTIFICATION AND STASTISTICAL ANALYSIS
  - Single Cell Seq analysis
  - Immuno-staining quantification
  - qRT-PCR
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and four tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2019.01.045.

ACKNOWLEDGMENTS

We thank Blue Lake, Song Chen, and Kun Zhang for technical advice for our initial scRNA-seq experiments. We also thank Cuong To and the UCSD Institute for Genomic Medicine for technical support, as well as the San Diego Supercomputer Center for providing data analysis resources. This work was supported by NIH grants R01 GM119128 (M.F.W.) and T32 HD007203 (D.B.), as well as the Lalor Institute (K.T.).

AUTHOR CONTRIBUTIONS

The experiments were initially conceived and performed by H.-W.S. with help from A.S. and M.F.W. The bulk of experiments were performed by H.-W.S., K.T., A.S., D.B., and E.V. Testes samples were provided by T.-C.H., R.R., and S.H. The manuscript was primarily written by A.S., K.T., D.B., and M.F.W. Critical thinking was provided by D.G.dR., S.H. and L.L.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 24, 2018
Revised: November 21, 2018
Accepted: January 10, 2019
Published February 5, 2019

REFERENCES

Aponte, P.M., van Bragt, M.P., de Rooij, D.G., and van Pelt, A.M. (2005). Spermatogonial stem cells: characteristics and experimental possibilities. APMIS *113*, 727–742.

Barrionuevo, F., Bagheri-Fam, S., Klattig, J., Kist, R., Taketo, M.M., Englert, C., and Scherer, G. (2006). Homozygous inactivation of Sox9 causes complete XY sex reversal in mice. Biol. Reprod. *74*, 195–201.

Berensztein, E.B., Sciara, M.I., Rivarola, M.A., and Belgorosky, A. (2002). Apoptosis and proliferation of human testicular somatic and germ cells during prepuberty: high rate of testicular growth in newborns mediated by decreased apoptosis. J. Clin. Endocrinol. Metab. *87*, 5113–5118.

Bitgood, M.J., Shen, L., and McMahon, A.P. (1996). Sertoli cell signaling by Desert hedgehog regulates the male germline. Curr. Biol. *6*, 298–304.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Canto, P., Söderlund, D., Reyes, E., and Méndez, J.P. (2004). Mutations in the desert hedgehog (DHH) gene in patients with 46,XY complete pure gonadal dysgenesis. J. Clin. Endocrinol. Metab. *89*, 4480–4483.

Cortes, D., Thorup, J.M., and Beck, B.L. (1995). Quantitative histology of germ cells in the undescended testes of human fetuses, neonates and infants. J. Urol. *154*, 1188–1192.

Culty, M. (2009). Gonocytes, the forgotten cells of the germ cell lineage. Birth Defects Res. C Embryo Today *87*, 1–26.

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

Di Persio, S., Saracino, R., Fera, S., Muciaccia, B., Esposito, V., Boitani, C., Berloco, B.P., Nudo, F., Spadetta, G., Stefanini, M., et al. (2017). Spermatogonial kinetics in humans. Development *144*, 3430–3439.

Durcova-Hills, G., Adams, I.R., Barton, S.C., Surani, M.A., and McLaren, A. (2006). The role of exogenous fibroblast growth factor-2 on the reprogramming of primordial germ cells into pluripotent stem cells. Stem Cells *24*, 1441–1449.

Dym, M., Kokkinaki, M., and He, Z. (2009). Spermatogonial stem cells: mouse and human comparisons. Birth Defects Res. C Embryo Today *87*, 27–34.

Fan, J., Salathia, N., Liu, R., Kaeser, G.E., Yung, Y.C., Herman, J.L., Kaper, F., Fan, J.B., Zhang, K., Chun, J., and Kharchenko, P.V. (2016). Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. Nat. Methods *13*, 241–244.

Fayomi, A.P., and Orwig, K.E. (2018). Spermatogonial stem cells and spermatogenesis in mice, monkeys and men. Stem Cell Res. (Amst.) *29*, 207–214.

Fukuda, T., Hedinger, C., and Groscurth, P. (1975). Ultrastructure of developing germ cells in the fetal human testis. Cell Tissue Res. *161*, 55–70.

Garbuzov, A., Pech, M.F., Hasegawa, K., Sukhwani, M., Zhang, R.J., Orwig, K.E., and Artandi, S.E. (2018). Purification of GFRα1+ and GFRα1- spermatogonial stem cells reveals a niche-dependent mechanism for fate determination. Stem Cell Reports *10*, 553–567.

Gaskell, T.L., Esnal, A., Robinson, L.L., Anderson, R.A., and Saunders, P.T. (2004). Immunohistochemical profiling of germ cells within the human fetal testis: identification of three subpopulations. Biol. Reprod. *71*, 2012–2021.

Glickstein, S.B., Monaghan, J.A., Koeller, H.B., Jones, T.K., and Ross, M.E. (2009). Cyclin D2 is critical for intermediate progenitor cell proliferation in the embryonic cortex. J. Neurosci. *29*, 9614–9624.

Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., Hu, Y., Wang, X., Wei, Y., et al. (2015). The transcriptome and DNA methylome landscapes of human primordial germ cells. Cell *161*, 1437–1452.

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.

Guo, J., Grow, E.J., Mlcochova, H., Maher, G.J., Lindskog, C., Nie, X., Guo, Y., Takei, Y., Yun, J., Cai, L., et al. (2018). The adult human testis transcriptional cell atlas. Cell Res. *28*, 1141–1157.

Hadziselimovic, F., Hadziselimovic, N.O., Demougin, P., and Oakeley, E.J. (2011). Testicular gene expression in cryptorchid boys at risk of azoospermia. Sex Dev. *5*, 49–59.

Hara, K., Nakagawa, T., Enomoto, H., Suzuki, M., Yamamoto, M., Simons, B.D., and Yoshida, S. (2014). Mouse spermatogenic stem cells continually interconvert between equipotent singly isolated and syncytial states. Cell Stem Cell *14*, 658–672.

Hayashi, Y., Saitou, M., and Yamanaka, S. (2012). Germline development from human pluripotent stem cells toward disease modeling of infertility. Fertil. Steril. *97*, 1250–1259.

He, Z., Kokkinaki, M., Jiang, J., Dobrinski, I., and Dym, M. (2010). Isolation, characterization, and culture of human spermatogonia. Biol. Reprod. *82*, 363–372.

He, Z., Kokkinaki, M., Jiang, J., Zeng, W., Dobrinski, I., and Dym, M. (2012). Isolation of human male germ-line stem cells using enzymatic digestion and magnetic-activated cell sorting. Methods Mol. Biol. *825*, 45–57.

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., Cheng, K., Singh, A., Roa-De La Cruz, L., Mutoji, K.N., Chen, I.C., Gildersleeve, H., Lehle, J.D., Mayo, M., Westernstroer, B., et al. (2018). The mammalian spermatogenesis single-cell transcriptome, from spermatogonial stem cells to spermatids. Cell Rep. *25*, 1650–1667.e1658.

Kanatsu-Shinohara, M., and Shinohara, T. (2013). Spermatogonial stem cell self-renewal and development. Annu. Rev. Cell Dev. Biol. *29*, 163–187.

Kossack, N., Terwort, N., Wistuba, J., Ehmcke, J., Schlatt, S., Schöler, H., Kliesch, S., and Gromoll, J. (2013). A combined approach facilitates the reliable detection of human spermatogonia in vitro. Hum. Reprod. *28*, 3012–3025.

Kowalczyk, M.S., Tirosh, I., Heckl, D., Rao, T.N., Dixit, A., Haas, B.J., Schneider, R.K., Wagers, A.J., Ebert, B.L., and Regev, A. (2015). Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. Genome Res. *25*, 1860–1872.

Kubota, H., and Brinster, R.L. (2018). Spermatogonial stem cells. Biol. Reprod. *99*, 52–74.

Luu, V.P., Hevezi, P., Vences-Catalan, F., Maravillas-Montero, J.L., White, C.A., Casali, P., Llorente, L., Jakez-Ocampo, J., Lima, G., Vilches-Cisneros, N., et al. (2013). TSPAN33 is a novel marker of activated and malignant B cells. Clin. Immunol. *149*, 388–399.

Mendive, F., Laurent, P., Van Schoore, G., Skarnes, W., Pochet, R., and Vassart, G. (2006). Defective postnatal development of the male reproductive tract in LGR4 knockout mice. Dev. Biol. *290*, 421–434.

Nakagawa, T., Sharma, M., Nabeshima, Y., Braun, R.E., and Yoshida, S. (2010). Functional hierarchy and reversibility within the murine spermatogenic stem cell compartment. Science *328*, 62–67.

Nickkholgh, B., Mizrak, S.C., Korver, C.M., van Daalen, S.K., Meissner, A., Repping, S., and van Pelt, A.M. (2014). Enrichment of spermatogonial stem cells from long-term cultured human testicular cells. Fertil. Steril *102*, 558–565.e555.

Niedenberger, B.A., Busada, J.T., and Geyer, C.B. (2015). Marker expression reveals heterogeneity of spermatogonia in the neonatal mouse testis. Reproduction *149*, 329–338.

Oakberg, E.F. (1971). Spermatogonial stem-cell renewal in the mouse. Anat. Rec. *169*, 515–531.

Oatley, J.M., and Brinster, R.L. (2012). The germline stem cell niche unit in mammalian testes. Physiol. Rev. *92*, 577–595.

Paniagua, R., and Nistal, M. (1984). Morphological and histometric study of human spermatogonia from birth to the onset of puberty. J. Anat. *139*, 535–552.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods *14*, 979–982.

Rossi, P. (2013). Transcriptional control of KIT gene expression during germ cell development. Int. J. Dev. Biol. *57*, 179–184.

Rossi, P., and Dolci, S. (2013). Paracrine mechanisms involved in the control of early stages of Mammalian spermatogenesis. Front. Endocrinol. (Lausanne) *4*, 181.

Sachs, C., Robinson, B.D., Andres Martin, L., Webster, T., Gilbert, M., Lo, H.Y., Rafii, S., Ng, C.K., and Seandel, M. (2014). Evaluation of candidate spermatogonial markers ID4 and GPR125 in testes of adult human cadaveric organ donors. Andrology *2*, 607–614.

Sadri-Ardekani, H., Akhondi, M.A., van der Veen, F., Repping, S., and van Pelt, A.M. (2011). In vitro propagation of human prepubertal spermatogonial stem cells. JAMA *305*, 2416–2418.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

Schwob, J.E. (2002). Neural regeneration and the peripheral olfactory system. Anat. Rec. *269*, 33–49.

Semon, J.A., Nagy, L.H., Llamas, C.B., Tucker, H.A., Lee, R.H., and Prockop, D.J. (2010). Integrin expression and integrin-mediated adhesion in vitro of human multipotent stromal cells (MSCs) to endothelial cells from various blood vessels. Cell Tissue Res. *341*, 147–158.

Shea, K.L., Xiang, W., LaPorta, V.S., Licht, J.D., Keller, C., Basson, M.A., and Brack, A.S. (2010). Sprouty1 regulates reversible quiescence of a self-renewing adult muscle stem cell pool during regeneration. Cell Stem Cell *6*, 117–129.

Song, H.W., Anderson, R.A., Bayne, R.A., Gromoll, J., Shimasaki, S., Chang, R.J., Parast, M.M., Laurent, L.C., de Rooij, D.G., Hsieh, T.C., and Wilkinson, M.F. (2013). The RHOX homeobox gene cluster is selectively expressed in human oocytes and male germ cells. Hum. Reprod. *28*, 1635–1646.

Suzuki, H., Ahn, H.W., Chu, T., Bowden, W., Gassei, K., Orwig, K., and Rajkovic, A. (2012). SOHLH1 and SOHLH2 coordinate spermatogonial differentiation. Dev. Biol. *361*, 301–312.

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.

Valli, H., Phillips, B.T., Shetty, G., Byrne, J.A., Clark, A.T., Meistrich, M.L., and Orwig, K.E. (2014a). Germline stem cells: toward the regeneration of spermatogenesis. Fertil. Steril. *101*, 3–13.

Valli, H., Sukhwani, M., Dovey, S.L., Peters, K.A., Donohue, J., Castro, C.A., Chu, T., Marshall, G.R., and Orwig, K.E. (2014b). Fluorescence- and magnetic-activated cell sorting strategies to isolate and enrich human spermatogonial stem cells. Fertil Steril. *102*, 566–580.e567.

Vij, S.C., Sabanegh, E., Jr., and Agarwal, A. (2018). Biological therapy for non-obstructive azoospermia. Expert Opin. Biol. Ther. *18*, 19–23.

Vilar, O. (1970). Histology of the human testis from neonatal period to adolescence (Plenum Press), pp. 95–111.

von Kopylow, K., and Spiess, A.N. (2017). Human spermatogonial markers. Stem Cell Res. (Amst.) *25*, 300–309.

Wang, M., Liu, X., Chang, G., Chen, Y., An, G., Yan, L., Gao, S., Xu, Y., Cui, Y., Dong, J., et al. (2018). Single-cell RNA sequencing analysis reveals sequential cell fate transition during human spermatogenesis. Cell Stem Cell *23*, 599–614.e594.

Wilson, A., Laurenti, E., and Trumpp, A. (2009). Balancing dormant and self-renewing hematopoietic stem cells. Curr. Opin. Genet. Dev. *19*, 461–468.

Wu, X., Schmidt, J.A., Avarbock, M.R., Tobias, J.W., Carlson, C.A., Kolon, T.F., Ginsberg, J.P., and Brinster, R.L. (2009). Prepubertal human spermatogonia and mouse gonocytes share conserved gene expression of germline stem cell regulatory molecules. Proc. Natl. Acad. Sci. USA *106*, 21672–21677.

Yu, P., Agbaegbu, C., Malide, D.A., Wu, X., Katagiri, Y., Hammer, J.A., and Geller, H.M. (2015). Cooperative interactions of LPPR family members in membrane localization and alteration of cellular morphology. J. Cell Sci. *128*, 3210–3222.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibody		
Rabbit polyclonal anti-PLPPR3 antibody against human phospholipid phosphatase related 3.	Atlas Antibodies	Cat # HPA057034; RRID:AB_2683316
Monoclonal mouse IgG2B against TSPAN33	R&D Systems	Cat # MAB8405; RRID:AB_10629311
Rabbit polyclonal antibody to PIWIL4 (HIWI2) from human, mouse and rat.	LifeSpan BioSciences	Cat # LS-C482396; RRID:AB_2782961
Rabbit polyclonal antibody against FSD2	GeneTex	Cat # GTX131769; RRID:AB_2782962
Rabbit polyclonal antibody to EGR-4 (EGR4) from human and mouse.	LifeSpan BioSciences	Cat # LS-C402257; RRID:AB_2782963
Monoclonal mouse antibody against human FGFR3	Novus Biologicals	Cat # NBP2-52468; RRID:AB_2782964
Rabbit Polyclonal antibody against PIM2	GeneTex	Cat # GTX113928; RRID:AB_2037694
Rabbit polyclonal IgG antiboday against POU5F1	Novus Biologicals	Cat # NB100-2379; RRID:AB_2167565
Rabbit polyclonal antibody to ETV4 (PEA3) from human and mouse	LifeSpan Biosciences	Cat # LS-C335603; RRID:AB_2782965
Monoclonal Antibody against UTF1	EMD Millipore	Cat # MAB4337; RRID:AB_827541
Biological Samples		
Neonatal Day 2	Day 2 infant testis	NA
Neonatal Day 7	Day 7 infant testis	NA
Adult 1 Age 37 years	Adult testis biopsy	NA
Adult 2 Age 42 years	Adult testis biopsy	NA
Critical Commercial Assays		
10x Chromium Single Cell A Chip Kit	10x genomics	Cat# 1000009
iScript cDNA synthesis Kit	BioRad	Cat# 170-8891
SsoAdvanceD Universal SYBR Green Supermix	BioRad	Cat# 172-5274
Deposited Data		
Filtered gene matrix, Gene list and cell barcodes	Cell Ranger output from this paper	NCBI/GEO data base accession GEO: GSE124263
Experimental Models: Organisms/Strains		
Human testis samples	Homo Sapiens	Informed Consent and IRB approved
Oligonucleotides		
See Table S4 for list of primers used for qRT-PCR	This Paper	NA
Software and Algorithms		
Cell Ranger Version 2.0	10x genomics	NA
Seurat R package Version 2.0	Butler et al., 2018, Satija et al., 2015	https://satijalab.org/seurat/
Monocle R package Version 2.4.0	Qiu et al., 2017	http://cole-trapnell-lab.github.io/monocle-release/
R Version 3.4 or higher	https://www.r-project.org	NA
R studio Version 1.0.143	https://www.rstudio.com	NA
PAGODA Version 2.0	Peter Kharchenko and Nikolas Barkas (NA). pagoda2: Single CellAnalysis and Differential Expression. R package version 0.0.0.9002.	https://github.com/hms-dbmi/pagoda2

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Miles F. Wilkinson (mfwilkinson@ucsd.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Subject details

The experiments with human material were approved by the UCSD Human Research Protections Program (HRPP) council and the HRPP council coordinated by the University of Michigan Office of Research. Informed consent was obtained from all the human subjects.

Testicular biopsies were obtained from two fertile men aged 37- and 42-years, undergoing vasectomy reversal at the UCSD Medical Center, following IRB-approved protocol #120471. The biopsies were transported to the research laboratory on ice in Minimum Essential Medium Alpha Medium (αMEM) + 10% FBS. The samples were then immediately cut into smaller portions and cryopreserved using freezing media composed of 10% DMSO + 90% FBS under controlled cooling conditions in a freezing container (Thermo Fisher Scientific cat # 5100-0001) at −80°C. The samples were subsequently transferred to liquid nitrogen storage until use. The neonatal testes we used for our scRNaseq analysis were obtained from the University of Michigan Medical Center, following IRB-approved protocol HUM00097035. These testes were from unrelated day 2 and day 7 neonates who died as a result of non-testicular-related medical dysfunction; they were cryopreserved in vials in liquid nitrogen and later transported to UCSD on dry ice.

Human testis samples preparation

Single testicular cells were isolated using a two-step enzymatic digestion protocol described previously (Valli et al., 2014b). In brief, testicular tissue was mechanically disrupted and enzymatically digested with 1 mg ml^{-1} collagenase type IV (Worthington Biochemical) in Hank's Balanced Salt Solution (HBSS; GIBCO) at 37°C. The tubules were sedimented and washed with HBSS and digested in 0.25% Trypsin-EDTA (ThermoFisher) and Deoxyribonuclease I (Worthington Biochemical). The suspension was triturated vigorously ten times, incubated at 37°C for 5 min, followed by repeat trituration and incubation. The digestion was stopped by adding the same volume of αMEM + 10% FBS medium and the cells were size-filtered through 70 μm (ThermoFisher) and 40 μm strainers (ThermoFisher) and pelleted by centrifugation at 300 g for 5 min.

Integrin-α6 (ITGA6)+ cells were purified from the isolated testicular cell preparations described above using the MACS system. The cells were re-suspended in MACS buffer (PBS, 0.5% BSA, 2mM EDTA) + Deoxyribonuclease I, incubated with ITGA6 antibody (BD PharMingen; 1:50) for 20 mins on ice, washed with MACS buffer, incubated with anti-rat IgG microbeads for 20 mins on ice, washed in MACS buffer, re-suspended in 500μl MACS buffer, and run through a MACS MS column (Miltenyi Biotec, cat # 130-042-201). The cells were then eluted with MACS buffer and resuspended in PBS + 0.25% FBS. The dead cells were removed using the ClioCell Dead Cell Removal kit (Amsbio) following the manufacturers' instructions. The samples from the 4 testes biopsies (neonatal + adult) were processed in exactly the same manner to minimize technical bias.

METHODS DETAILS

10X Genomics library preparation

Viable cells were washed once in PBS and resuspended in 0.04% BSA in PBS for loading on the 10x Chromium chip. Cell capturing, and library preparation was carried as per kit instructions (Chromium Single Cell Kit [v2 chemistry]). In brief, 10,000 cells were targeted for capture per sample, after cDNA synthesis,12-14 cycles were used for library amplification. The resultant libraries were size selected, pooled and sequenced using 2 × 100 paired-end sequencing protocol on an Illumina HiSeq 4000 instrument. The libraries initially underwent shallow sequencing to access quality and to adjust subsequent sequencing depth based on the capture rate and unique molecular indices (UMI) detected. All sequencing was performed at the Institute of Genomic Medicine at UCSD.

Mapping, cell identification and clustering analysis

Demultiplexed raw sequencing reads were processed and mapped to the human genome (Hg19) using Cell Ranger software (v2.2.0) with default parameters (Figure S6A). Filtered count matrices for each library were tagged with a library batch ID and combined across independent experiments using the Seurat package (Butler et al., 2018) in R. To check the quality of the single cell data and to remove any multiplets, we performed Seurat based filtering of cells based on three criteria: number of detected genes per cell, number of UMIs expressed per cell and mitochondrial content, using the following threshold parameters: nGene (200 to 9000), nUMI (between -inf and 40,000), and percentage of mitochondrial genes expressed (< 0.5%) (see Figure S6B for violin plots of these parameters). In addition, we used known lineage marker profile to rule out that same barcode was not assigned to two cells of different lineages (multiplets). Normalization was performed as described in the package manual (https://satijalab.org/seurat/h) (Satija et al., 2015). Batch correction was performed using the *JackStraw* and/or *RunCCA* functions in the Seurat package. The *RunCCA* function was used to correct for platform differences when comparing 10x Genomics datasets with Fluidigm C1 datasets, as previously described (Butler et al., 2018; Satija et al., 2015).

After filtering out poor quality cells, we captured 7,974 ITGA6+ and 10,749 unfractionated cells from adult testes (total from 2 individuals); and 6,086 ITGA6+ cells and 8,776 unfractionated cells from neonatal testes (total from 2 individuals). Because we found that ITGA6+ cells not only enriched for germ cells, but also some somatic cell types (in alignment with previous studies [He et al., 2012; Semon et al., 2010]), it was not useful for germ cell enrichment, so most of our clustering analysis was done on unfractionated cells and ITGA6+ cells combined. For our somatic cell analysis, we first combined both adult and neonatal cells to generate a dataset of 33,585 total cells. We then removed the germ cells and applied stringent filtering parameters (i.e., mitochondrial content percent cut-off was set to 0.2), which left 25,699 cells. These cells were then re-clustered as described below to generate somatic cell clusters.

To identify cell clusters, we employed Pathway and Geneset OverDispersion Analysis (PAGODA2) (Fan et al., 2016). Parameters such as perplexity, number of overdispersed genes and K-nearest neighbor were adjusted to identify recognizable cell clusters. We then used igraph-based community prediction methods such as infograph, walktrap, and multilevel to identify clusters. The resulting tSNE cell embeddings were imported into Seurat for cluster annotation, based on established gene expression markers (Table S1). Analysis of the 2 neonatal testicular samples showed a similar distribution of cells in the cell clusters (Figure S6C). The same was observed for the 2 adult testicular samples (Figure S6C). Additional evidence for similarity of the biological replicates was the strong correlation of average gene expression ($R^2 > 0.95$; Figure S6C). Annotated germ cells derived from either ITGA6+ adult testicular cells or unfractionated adult testicular cells exhibited an overlapping distribution in germ cells clusters defined on tSNE plots (Figure S6D, top left), indicating that ITGA6+ enrichment did not significantly bias the cell population to any specific cluster. As confirmation, average gene expression between annotated germ cells derived from ITGA6+ versus unfractionated adult testicular cells was strongly correlated ($R^2 > 0.99$; Figure S6D, bottom left). Similar results were obtained with annotated SPG derived from ITGA6+ adult testicular cells versus unfractionated adult testicular cells (Figure S6D, middle), as well as annotated neonatal germ cells derived from ITGA6+ neonatal testicular cells versus unfractionated neonatal testicular cells (Figure S6D, right).

The *FindMarkers* function (Wilcoxon rank sum test) was used to determine differential gene expression between clusters (set at minimum expression in 25% of cells). The *CellCycleScoring* function was used to infer cell cycle phase, as this program determines relative expression of a large set of G2-M- and S-phase genes (Kowalczyk et al., 2015). Pearson correlation between two clusters was performed using the *CellPlot* function.

GO analysis was done using top differentially (positively) expressed genes with a p adjusted cut off of 0.05.

Cell trajectory analysis

Single-cell pseudotime trajectories were constructed with the Monocle 2 package (v2.8.0) (Qiu et al., 2017) according to the provided documentation (http://cole-trapnell-lab.github.io/monocle-release/). UMI counts were modeled as a negative binomial distribution; ordering genes were identified as having high dispersion across cells (mean_expression > = 0.01; dispersion_empirical > = 1). The discriminative dimensionality reduction with trees (DDRTree) method was used to reduce data to two dimensions. Gene sets identified from the destiny analysis were clustered and visualized using the plot_genes_in_pseudotime function.

FACS and qRT-PCR analysis

After dissecting single testicular cells, the cells were resuspended in staining buffer (PBS + 3% FBS) for 20 mins on ice, stained with the primary antibodies, washed with staining buffer, incubated with secondary antibodies for 20 mins on ice, resuspended in staining buffer and sorted by FACS.

For FACS, gating was set based on size (FSC) to remove small debris and doublets. We then used unstained and secondary antibody only (primary omitted) stained as negative controls for gating unstained and false positive stained cells respectively (Figure S3A, middle).

cDNAs were generated using the Iscript reverse transcriptase (RT) kit, according to the manufacturer's protocol (Bio-Rad). The RT product and primer pairs (Table S4) were mixed with iQ SYBR Green supermix (Bio-Rad) and PCR was performed using an iCycler real-time PCR machine according to the manufacturer's protocol (Bio-Rad). The production of the amplicon was measured by SYBR green fluorescence and the threshold cycle (C_t) values were calculated. C_t values obtained were normalized to C_t values for the ribosomal protein RPL19 (L19) gene.

Immunohistochemical analysis

Testicular biopsies were fixed for 6 h in 10% neutral-buffered formalin, dehydrated in a series of ethanol washes of different concentration and embedded in paraffin using standard procedures. Sections were deparaffinized two times in xylene, followed by serial dilutions of ethanol. Unmasking was performed with 10 mM sodium citrate buffer, pH 6.0, using a steamer (IHCWORLD) for 40 min. Endogenous peroxidase activity was inactivated by incubation in 0.3% hydroperoxide in methanol for 15 min. Blocking was performed by incubating in 3% serum (from the species that the secondary antibody was raised in) for 30 min at room temperature. The sections were then incubated overnight with the primary antibody at 4°C, incubated with secondary antibody for 1 h at room temperature, and peroxidase activity was detected using the Vectastain ABC kit (Vector Laboratory, Inc., Burlingame, CA, USA) and 3,3′-diaminobenzidine tetrahydrochloride (DAB) solution. Nuclei were counterstained with hematoxylin. After dehydration, a coverslip was placed over the sections with mounting medium. The sections were viewed with a Leica DMI6000 B inverted microscope (Leica). We noted that FGFR3 although a cell surface marker was found to be nuclear as reported previously

(Durcova-Hills et al., 2006). Similarly, EGR4 a DNA-binding protein was reported to be cytoplasmic (Hadziselimovic et al., 2011). Although LPPR3 displayed different nucleic staining patterns, it is a lipid phosphate phosphatase that is known to localizes to the plasma membrane (Yu et al., 2015).

Immunofluorescence analysis

Sections were deparaffinized two times in xylene, followed by serial dilutions of ethanol. Unmasking was performed with 10 mM sodium citrate buffer, pH 6.0, using a steamer (IHCWORLD) for 40 min. Blocking was performed by incubating with 5% serum (from the species that the secondary antibody was raised in) for 1 hour at room temperature. The sections were then incubated overnight with the primary antibody at 4°C and incubated with secondary antibody for 1 h at room temperature. The nuclei were counterstained with DAPI, a coverslip was placed over the sections with mounting medium, and the images were viewed using a Leica DMI4000 B fluorescence microscope (Leica).

Whole mount staining

Human seminiferous tubules whole mount immunofluorescence was performed as previously described (Di Persio et al., 2017). Briefly, seminiferous tubules were disentangled from testicular biopsies and immediately fixed in 4% PFA at 4°C for 4 h. After fixation, the seminiferous tubules were permeabilized with 0.5% Triton X-100 in PBS and treated with 1%BSA and 5% Normal Donkey Serum in PBS overnight at 4°C. Following three 30 min washes, the seminiferous tubules were incubated overnight at 4°C with primary antibody. Seminiferous tubules were washed three times for 30 min and incubated 2 h at room temperature with species-specific secondary antibodies conjugated to Alexa488, Cy3 or Cy5 fluorochromes. Following three 30 min washes, the nuclei were stained with TO-PRO-3. The primary and secondary antibody incubations and all washes were performed in 1% BSA 0.1% Triton X-100 in PBS. Seminiferous tubules were mounted on slides with Vectashield mounting medium and observed with a Leica TCS SP2 confocal microscope with 40x oil immersion objective. Quantification of relative expression levels of selected markers was performed using LAS AF Software. Mean fluorescence or mean amplitude was assessed in selected ROI of a single Z or on Z-stacks, respectively.

QUANTIFICATION AND STASTISTICAL ANALYSIS

Single Cell Seq analysis

7,974 ITGA6+ and 10,749 unfractionated cells from adult testes (total from 2 individuals) and 6,086 ITGA6+ cells and 8,776 unfractionated cells from neonatal testes (total from 2 individuals) were used for scRNaseq analysis. The statistical method used to identify DEGs and pseudotime trajectory analysis are provided in the detailed methods above.

Immuno-staining quantification

Quantification of the immunostainings was performed by counting the positively stained cells in different fields of view. The number of cells counted is indicated on the respective figure or its figure legend.

qRT-PCR

mRNA levels were normalized to the housekeeping gene, *RPL19*, and quantified using the delta-delta C_t method. The values shown are mean ± SD from two biological replicates.

DATA AND SOFTWARE AVAILABILITY

The accession number for the sequence reported in this paper is GEO: GSE124263.

Cell Reports, Volume *26*

Supplemental Information

The Neonatal and Adult Human Testis

Defined at the Single-Cell Level

Abhishek Sohni, Kun Tan, Hye-Won Song, Dana Burow, Dirk G. de Rooij, Louise Laurent, Tung-Chin Hsieh, Raja Rabah, Saher Sue Hammoud, Elena Vicini, and Miles F. Wilkinson

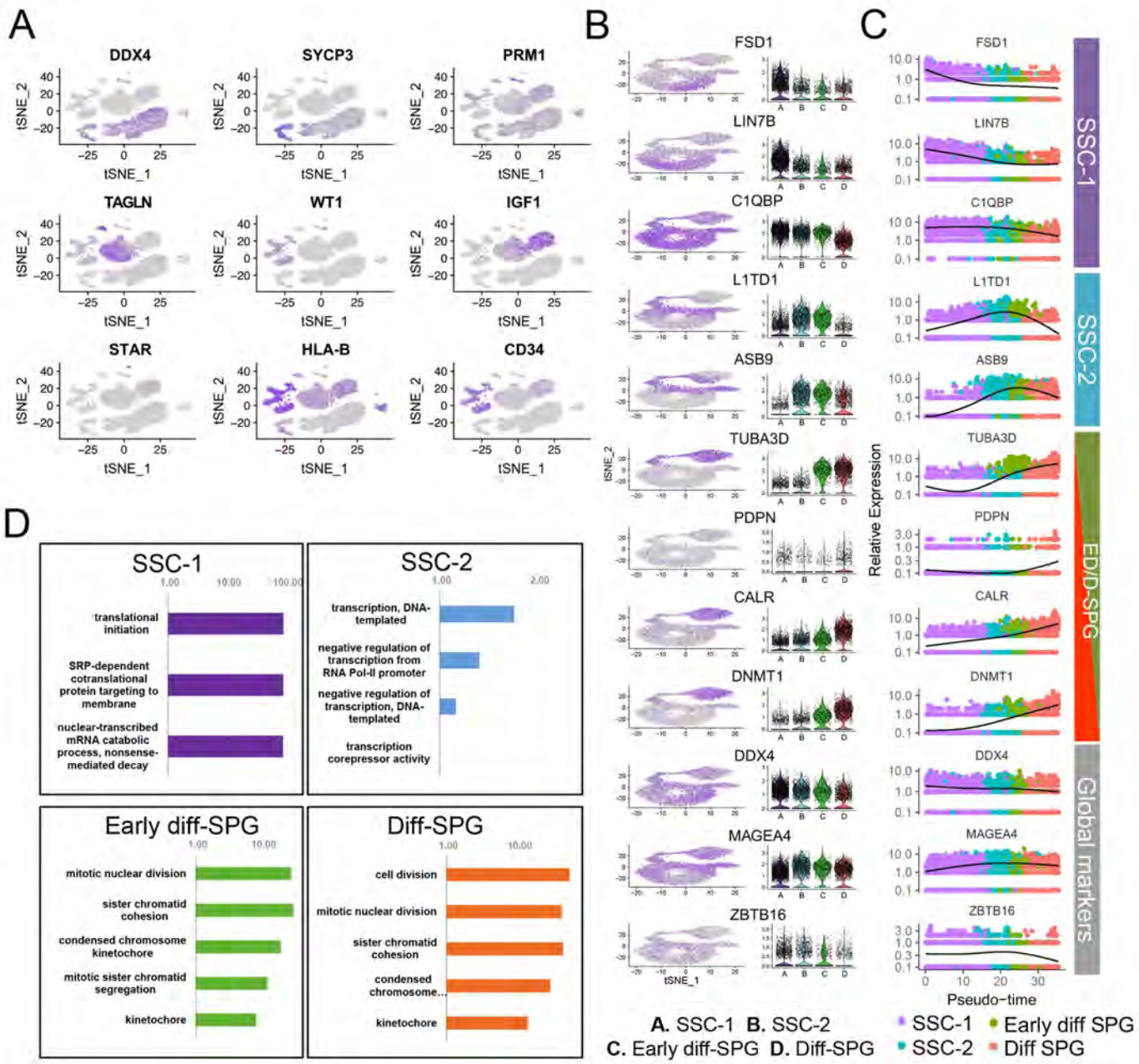


Figure S1. Identification of Adult Testicular Cell Subsets and Markers (related to Fig. 1)

(A) tSNE feature plots of some of the known marker genes used to identify the cell clusters in Fig. 1A.

(B) tSNE feature and violin plots of stage-specific SPG markers (see Fig. 1B for plots of other markers).

(C) Expression pattern of the markers shown in B across the pseudotime timeline (see Fig. 1E for the pseudotime plots of other markers).

(D) Most highly enriched GO categories corresponding to differentially expressed genes (using Padj cut off < 0.05) in the cell subsets shown (defined in Fig. 1B). -log(2) Padj values of GO terms are shown.

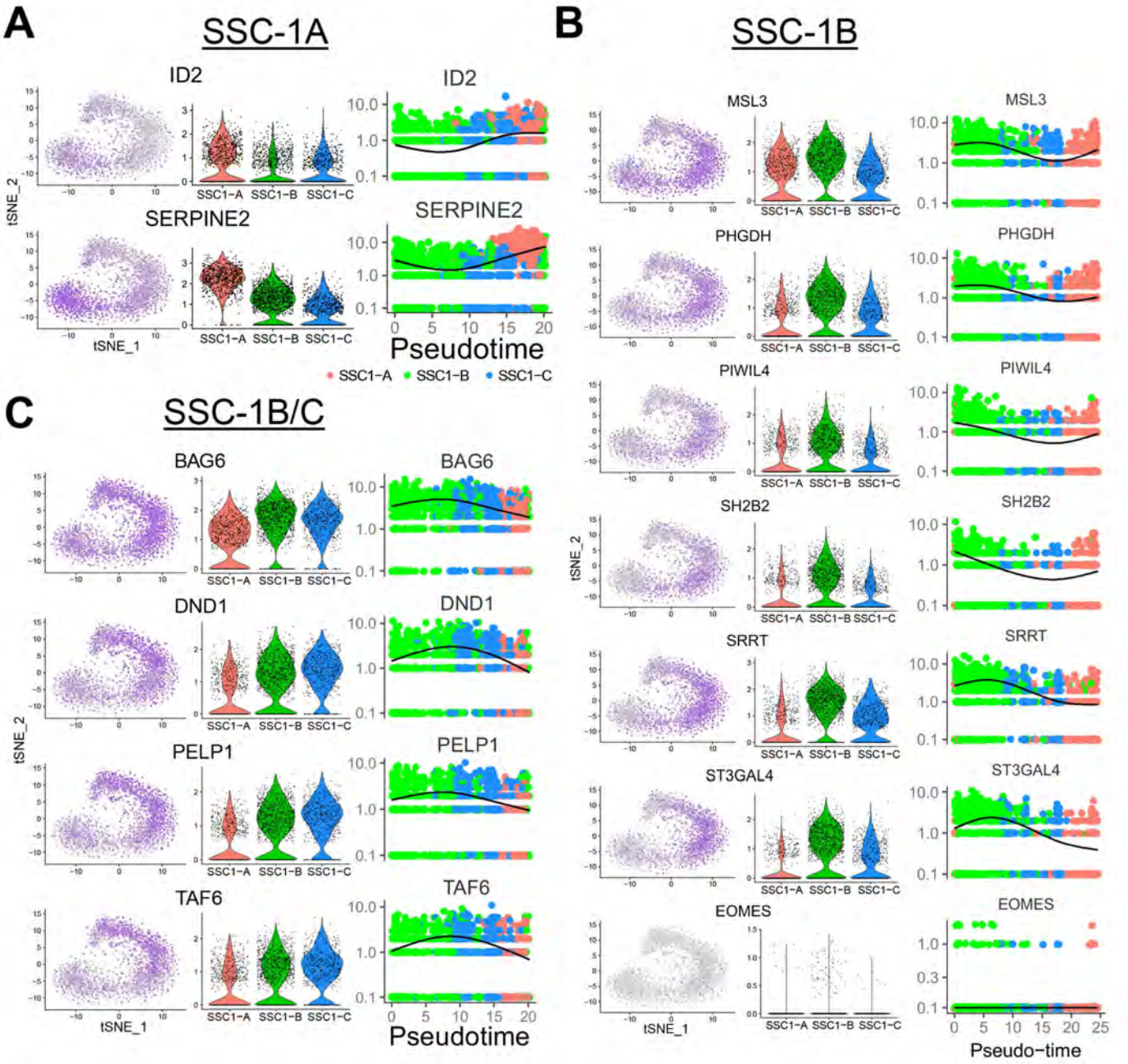


Figure S2. Undifferentiated SPG Markers (related to Fig. 2)

(A) tSNE feature, violin, and Monocle pseudotime timeline plots of genes enriched in the SSC-1A state.

(B) tSNE feature, violin, and Monocle pseudotime timeline plots of genes enriched in the SSC-1B state.

(C) tSNE feature, violin, and Monocle pseudotime timeline plots of one gene enriched in the SSC-1B and 1C states.

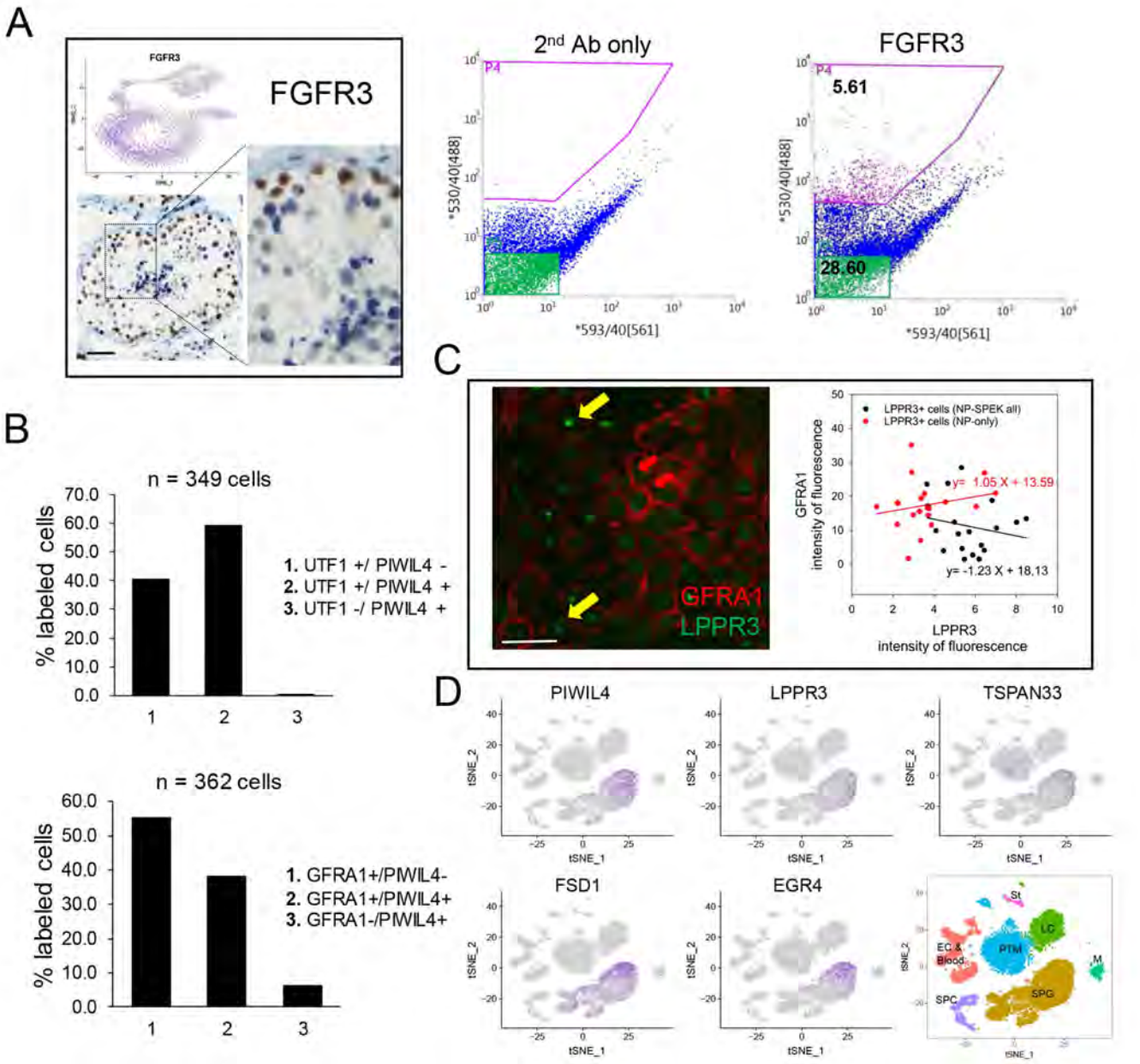


Figure S3. Identification of Candidate and Known SSC Markers (related to Fig. 3)

(A) Expression pattern of FGFR3, a previously defined SSC marker protein (Guo et al., 2017). tSNE feature plot shows the FGFR3 gene is preferential expressed in the SSC-1A and -1B states shown in Fig. 2C (top left). Representative image of FGFR3 protein expression in adult human testes determined by IHC analysis (left). FACS plot of adult human testicular cells stained with a FGFR3 antibody (right). ~6% of the cells were positive for staining with the FGFR3 antibody (representative of n=2 experiments). FACS plot of secondary antibody-only staining (primary omitted) (middle) was included as a negative control. Scale bar: 50 μM.

(B) IF quantification of whole mount adult testicular cells co-stained with the antibodies shown (see Figs. 3C for representative images). The number of cells counted is indicated on top of the bar graphs.

(C) Representative image of IF analysis of intact adult human seminiferous tubules (whole mount) stained with antibodies against the candidate SSC marker we defined in this study, LPPR3, and the previously recognized undifferentiated SPG marker, GFRA1 (left) (Di Persio et al., 2017). Anti-LPPR3 stained the entire nucleoplasm of some cells (red arrows show examples of these NP cells), while in other cells it produced aggregated foci or speckled pattern (yellow arrows show examples of these SPECK cells). The scatter plot shows GFRA1 and LPPR3 antibody fluorescence levels in single cells (right). Scale bar: 20 μM.

(D) tSNE feature plots of the indicated undifferentiated SPG markers in adult testicular cell subsets. Bottom right shows a tSNE plot labeled with cell types (see also Fig. 1A).

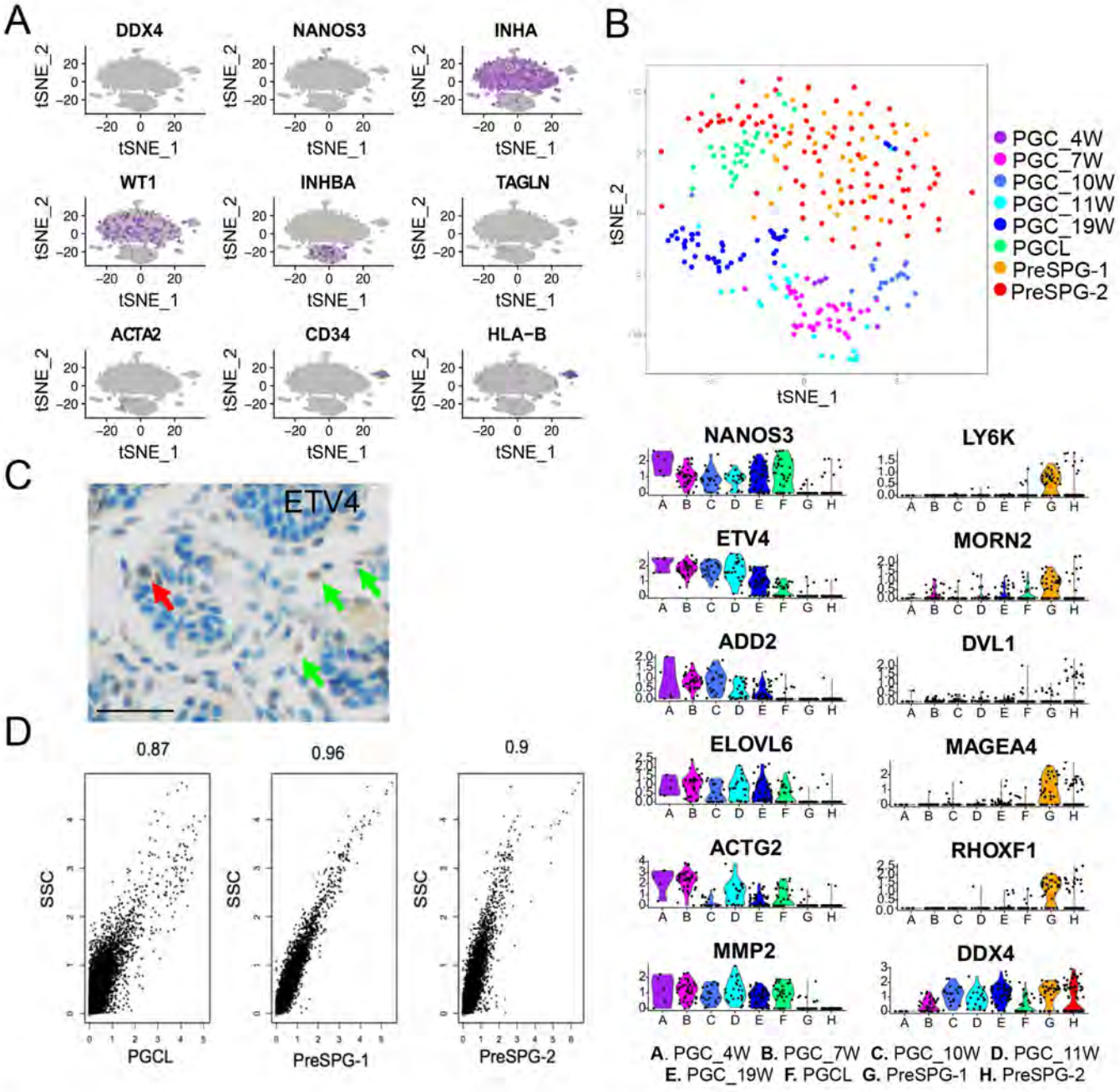


Figure S4. Characterization of Neonatal Germ Cell Subsets and Markers (related to Fig. 4)

(A) tSNE feature plots of some of the known marker genes used to identify the cell clusters in neonatal human testes in Fig. 4A.

(B) tSNE plot of neonatal germ cell subsets (defined in Fig. 4B) in combination with human PGCs of the indicated gestational time points from Guo. et. al. (Guo et al., 2015) (upper). The violin plots (lower) depict stage-specific genes markers (additional gene markers are shown in Fig. 4D and F).

(C) Representative image of ETV4 protein expression in adult human testes determined by IHC analysis (performed as in Fig. 4E). The red arrow depicts staining of a germ cell in the seminiferous tubule (also shown in Fig. 4E), while the green arrows mark staining of cells outside the seminiferous tubule. Scale bar: 50 μM.

(D) Pearson correlation analysis of the cumulative expression of the SSC-1 cluster with the neonatal cell clusters indicated. Values at the top are the R^2 value.

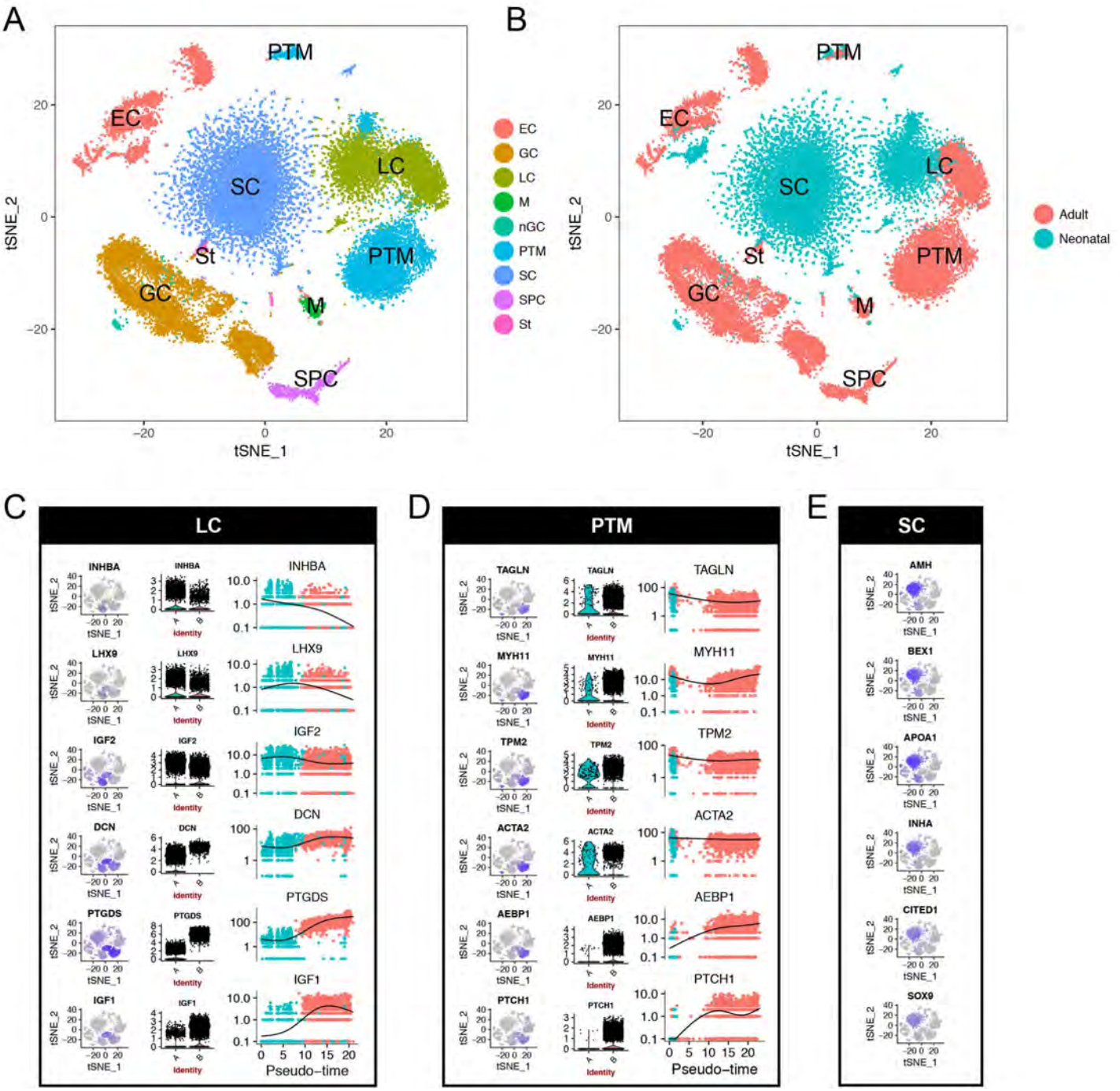


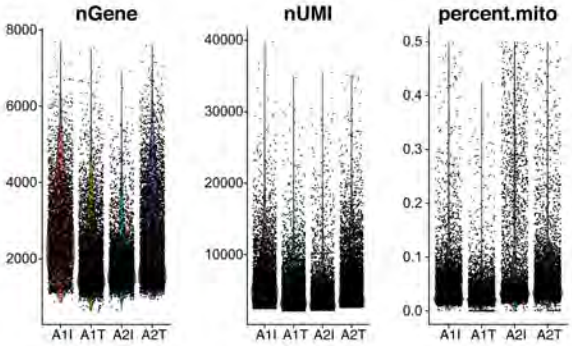
Figure S5. Characterization and Developmental Analysis of Human Testicular Somatic Cells (related to Fig. 5 and 6)

(A) tSNE plot of all neonatal and adult human testicular cells in our dataset. Cell cluster identities were determined from feature plots of known marker genes (not shown).
(B) tSNE plot of the same cells as in A, depicting the identity of the adult and neonatal cell clusters.
(C) Developmentally regulated LC genes. tSNE feature, violin, and Monocle pseudotime timeline plots depicting the expression pattern of LC marker genes in neonatal and adult LCs (5523 cells, total).
(D) Developmentally regulated PTM genes. tSNE feature, violin, and Monocle pseudotime timeline plots depicting the expression pattern of PTM marker genes in neonatal and adult PTM (4438 cells, total).
(E) Neonatal SC marker genes depicted in tSNE feature plots of all neonatal cells (14,862 total).

A

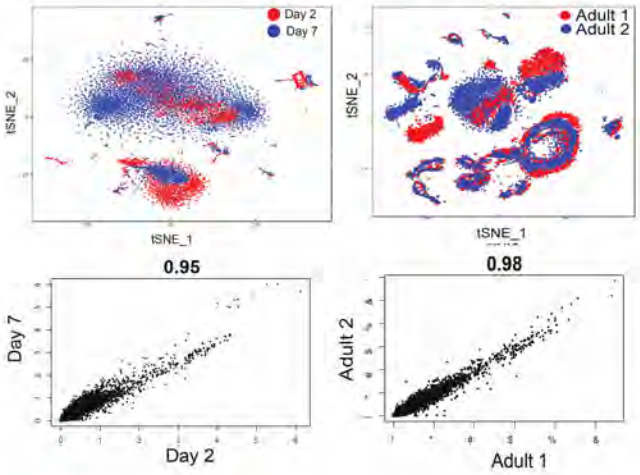
	Total Reads	Mean reads/cell	Median gene/ cell	# genes detected	# cells captured
Adult 1 –Total	97,069,723	19,590	2,008	26,935	4,955
Adult 1 –ITGA6+	130,766,252	31,253	2,669	27,161	4,184
Adult 2 –Total	191,601,744	31,236	2,420	28,356	6,134
Adult 2 – ITGA6+	162,913,699	35,679	1,792	25,819	4,566
Day2 – Total	78,866,849	21,702	1,129	22,103	3,634
Day2 – ITGA6+	73,730,641	31,753	1,365	21,043	2,322
Day 7 – Total	133,517,194	25,900	1,226	22,295	5,155
Day 7 – ITGA6+	82,352,728	21,792	1,238	20,694	3,779

B



A1I – Adult 1 ITGA6+ A2I – Adult 2 ITGA6+
A1T – Adult 1 Total A2T – Adult 2 Total

C



D

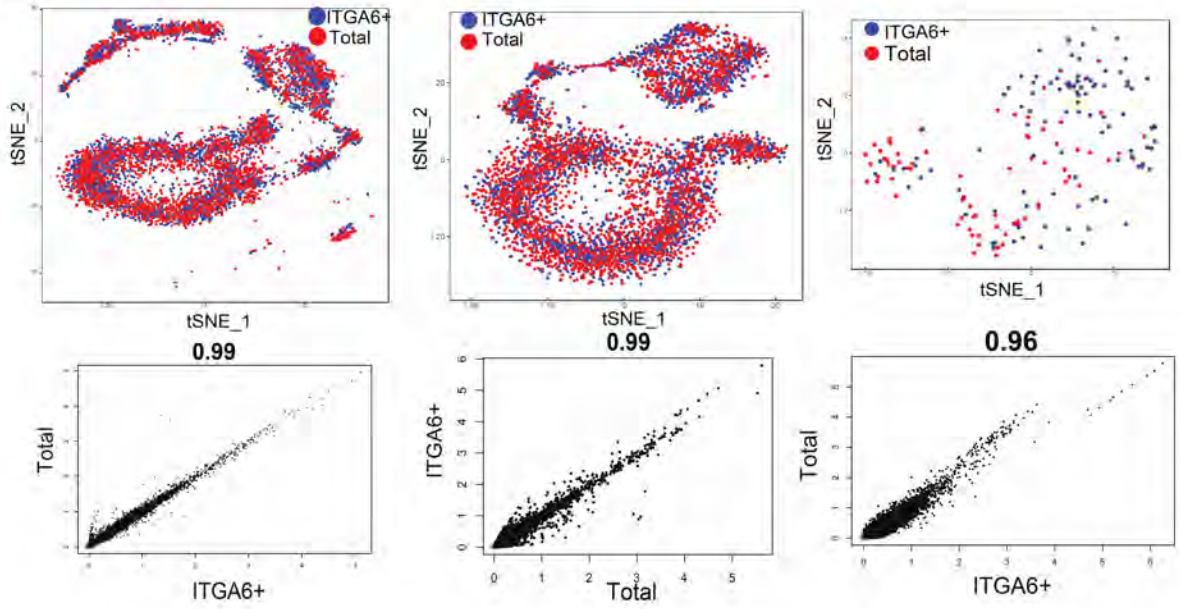


Figure S6. Mapping and quality control (related to Star Methods)
(A) Table of mapping results of testes samples analyzed by scRNAseq in this study.
(B) Violin plot of total unfractionated (T) and ITGA6+ (I) adult human testicular cells from adult 1 (A) and adult 2 (A2), showing nGene, nUMI and mitochondrial gene expression profiles—post filtering—using the parameters indicated in the Star Methods section.
(C) tSNE plots of all testicular cells from the neonatal and adult samples indicated. Pearson correlation of average gene expression between the two replicates are indicated in the plots below. R^2 values are indicated on top of the respective plots.
(D) tSNE plot of all adult germ cells (left), SPG only (middle) and neonatal germ cells (right), indicating the distribution of unsorted (Total) or ITGA6 enriched (ITGA6+) cells. Pearson correlation plot of average gene expression with R^2 values are shown below.