PNAS
www.pnas.org

Supplementary Information for

Transcriptome profiling reveals signaling conditions dictating human spermatogonia fate *in vitro*

Kun Tan, Hye-Won Song, Merlin Thompson, Sarah Munyoki, Meena Sukhwani, Tung-Chin Hsieh, Kyle E. Orwig, Miles F. Wilkinson

Miles F. Wilkinson
Email: mfwilkinson@health.ucsd.edu

This PDF file includes:

- Supplementary text
- Figures S1 to S3
- Tables S1 to S3
- SI References

Other supplementary materials for this manuscript include the following:

- Dataset S1

www.pnas.org/cgi/doi/10.1073/pnas.2000362117

Supplementary Information Text

SI Materials and Methods

Human testis samples preparation

The experiments with human material were approved by the UCSD Human Research Protections Program (HRPP) council. Informed consent was obtained from all the human subjects. Testicular biopsies were obtained from 29 fertile men aged between 30- and 50-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 + 40% αMEM + 50% FBS under controlled cooling conditions in a freezing container (Thermo Fisher Scientific) at -80 °C. The samples were subsequently transferred to liquid nitrogen storage until use.

Single testicular cells were isolated using a two-step enzymatic digestion protocol described previously (1, 2). In brief, testicular tissue was mechanically disrupted and enzymatically digested with 1 mg ml^{-1} collagenase type IV (Worthington Biochemical) in Hanks 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 and 40 μm strainers (ThermoFisher) and pelleted by centrifugation at 300 g for 5 min.

Xenograft transplantation Assay

For xenograft transplantation, human materials were approved by the Oregon National Primate Research Center, Oregon Health and Sciences University (assurance A3304-01). Healthy adult deidentified human testes were procured through the University of Pittsburgh Health Sciences Tissue bank and Center for Organ recovery under the University of Pittsburgh CORID 686.

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 (PLPPR3 [HPA057034, Atlas Antibodies], washed with staining buffer, incubated with secondary antibodies for 20 mins on ice, washed and resuspended in staining buffer and sorted by FACS. Small debris and doublets were gated out based on size. Unstained cells and cells stained with secondary antibody only were served as negative controls to gate out unstained and false-positive stained cells respectively.

Transplantation analysis was performed in busulfan-treated nude (immunodeficient) mice, as previous described (3). Briefly, 6-week-old nude mice were treated with a single dose of busulfan

(40 mg/kg, Sigma-Aldrich) to eliminate all germ cells. Six weeks after busulfan treatment, ~7 μl of cell suspension containing 10% trypan blue (Invitrogen) was injected into the seminiferous tubules of the testes via cannulation of the efferent ducts (at 100-200 × 10^6 cells per milliliter). Eight weeks post xenograft transplantation, the testes were collected for quantitative analysis of donor-derived colonies, as described previously (3). Briefly, after fixation, tubules were incubated with blocking buffer (PBS, 0.02g/ml blotto dry milk powder, 10% Triton-X100) and stained with a rabbit anti-primate testis cell primary antibody (3) at 4°C overnight. Goat anti-rabbit AlexaFlour488 was used to detect the primary antibody. Tubules were mounted on slides with Vectashield mounting medium containing DAPI and raised cover slip and imaged with fluorescent microscopy. SSCs colonies were counted if colonies contained at least 4 cells in a continuous area (≤ 100μm between cells), located on the basement membrane of the seminiferous tubules, were ovoid shaped, had a high nuclear to cytoplasmic ratio. Average number of stained cells was determined by dividing the total number of tubules counted (at least 100 tubules per sample for 3 replicate samples).

Immunofluorescence analysis

Human testis tissues were fixed in 1× PBS containing 4% paraformaldehyde (sigma) overnight at 4 °C with agitation on a nutator. Fixed samples were then washed three times in cold PBS. Paraffin embedding were performed at the UCSD Tissue Technology Shared Resource (supported by a National Cancer Institute Cancer Center Support Grant [CCSG Grant P30CA23100]). Paraffin 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. The following primary antibodies were used: anti-PLPPR3 (HPA057034, Atlas Antibodies), anti-KIT (AF332; R&D Systems), anti-KIT (18696-1-AP, Proteintech), anti-phospho-AKT (AF887, R&D Systems), anti-GFRA1 (AF714, R&D Systems), and anti-NANOS3 (21679-1-AP, Proteintech). After wash three times with PBS, sections were 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).

Bulk RNA Sequencing

RNAseq was performed on PLPPR3+ and KIT+ cells FACS purified from 4 independent human testicular biopsy samples (from 30- to 41-year old individuals). The antisera used for FACS purification were anti-PLPPR3 (HPA057034, Atlas Antibodies) and anti-KIT (550412, BD Biosciences). Total RNA was extracted using the RNeasy Plus Micro Kit (Qiagen), following the manufacturer's protocol. 10 ng of total RNA was used to make the library using the SMARTer®

stranded total RNAseq kit v2 - pico input system (Takara Bio USA), as per manufacturer's instructions. Libraries were sequenced (single-end reads) with an Illumina HiSeq 4000 platform for 75 cycles at the UCSD institute for Genomic Medicine (IGM) core. The average number of reads per sample ranged from approximately 29 to 36 million, with an average of approximately 33 million reads. Reads were filtered for quality and aligned with STAR (2.5.2b) (4) against Homo sapiens, release-95, Ensembl genome (GRCh38). The exon counts were aggregated for each gene to build a read count table using SubRead function featureCounts (5). DEGs were defined using DESeq2 (6) using the following threshold: log2FC ≥ 1 or ≤ –1, *q*-val < 0.01. The R package program "pheatmap" was used for clustering and to generate heatmap plots. The database for annotation, visualization and integrated discovery (DAVID) v6.8 was used for GO analysis. The Ingenuity pathway analysis (IPA) was used for signaling pathway and genetic regulatory network analysis.

Human SPG culture

Integrin-α6 (ITGA6)+ cells were purified from dissociated human testicular cells using the magnetic-activated cell sorting (MACS) system. Cells were re-suspended in MACS buffer (PBS, 0.5% BSA, 2 mM EDTA) + Deoxyribonuclease I, incubated with ITGA6 antisera (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). Cells were then eluted with MACS buffer and resuspended in culture medium. 1×10^5 MACS-sorted cells/well were plated in 24-well plate pre-coated with laminin (20 μg/ml). Cells were passaged by incubation in StemPro accutase solution (Thermo Fisher Scientific) for 5 min.

The ITGA6+ cells were cultured in basal culture medium (IMDM/SFM) previously used to culture mouse germline stem cells under feeder-free conditions (7). This basal medium contained Iscove modified Eagle medium (IMDM) supplemented with 25 μg/ml insulin (Sigma), 100 μg/ml apo-Transferrin (sigma), 200 μg/ml sodium pyruvate (Thermo Fisher Scientific), 60 μM putrescine (Sigma), 30 nM sodium selenite (Sigma), 6 mg/ml D-(+)-glucose (Sigma), 1 μl/ml DL-lactic acid (sigma), 5 mg/ml bovine albumin (Sigma), 2 mM GlutaMAX™ supplement (Thermo Fisher Scientific), 50 μM 2-mercaptoethanol (Thermo Fisher Scientific), 1 × MEM vitamin solution (Thermo Fisher Scientific), 1 × non-essential amino acids (Thermo Fisher Scientific), 100 μM ascorbic acid (Sigma), 10 μg/ml d-Biotin (Sigma), 30 ng/ml β-Estradiol (sigma), 60 ng/ml progesterone (Sigma), 1 mg/ml fetuin (Sigma), 10 μl/ml CD Lipid concentrate (Thermo Fisher Scientific), 2 μl/ml cholesterol solution (Sigma), and 50 μl/ml knockout serum replacement (Thermo Fisher Scientific). To this basal medium, we added 10 ng/ml bFGF (Sigma) with one or more of the following: 15 ng/ml recombinant human GDNF protein (R&D Systems) and/10 ng/ml recombinant human/mouse/rat activin A protein (R&D Systems), 5 ng/ml human recombinant M-CSF (STEMCELL Technologies), 10 ng/ml human recombinant BMP8B (Sigma), and 100 nM of the AKT1/2/3 inhibitor MK-2206 2HCl (Selleck Chemicals).

Annexin V/PI staining and flow cytometry

Apoptosis and necrosis was assessed using the Annexin V-FITC/PI apoptosis detection kit (Sigma) coupled with flow cytometry, following the manufacturer's instructions. Briefly, cells were collected and washed twice with PBS and simultaneously stained with Annexin V-fluorescein isothiocyanate (FITC) and the non-vital dye propidium iodide (PI), which allowed the discrimination of intact cells (FITC−PI−), early apoptotic cells (FITC+PI−), and necrotic cells (FITC+PI+). For FACS, gating was set based on unstained cells.

qRT-PCR analysis

cDNAs were generated using the *Iscript* reverse transcriptase (RT) kit, according to the manufacturer's protocol (Bio-Rad). The RT product and primer pairs 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.

10X Genomics library preparation

Testicular cells from biopsies obtained from two fertile men aged 32- and 37-years, respectively, were cultured under different conditions for two weeks and then collected for single cell RNA sequencing (scRNAseq). Dead cells were removed using the ClioCell Dead Cell Removal kit (Amsbio) following the manufacturers' instructions. 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 3′ Library Construction Kit v3). In brief, 5,000 cells were targeted for capture per sample, after cDNA synthesis, 11 cycles were used for library amplification. The resultant libraries were size selected, pooled and sequenced using 2 x 100 paired-end sequencing protocol on an Illumina NovaSeq 6000 instrument. All sequencing was performed at the UCSD IGM core.

Mapping, cell identification and clustering analysis

Demultiplexed raw sequencing reads were processed and mapped to the human genome (GRCh38) using Cell Ranger software (v3.0.2) with default parameters. Filtered count matrices for each library were tagged with a library batch ID and combined across independent experiments using the Seurat package (version 3.1.0) (8) 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 as used previously (2) with more stringent: number of detected features (nFeature_RNA) per cell, number of UMIs expressed per cell (nCount_RNA) and mitochondrial content, using the following

threshold parameters: nFeature_RNA (500 to 10000), nCount_RNA (1000 to 50000), and percentage of mitochondrial genes expressed (<0.2%). In addition, we used known lineage marker profile to rule out that same barcode was not assigned to two cells of different lineages (multiplets). Gene expression values were log normalized and regressed by nFeature_RNA, nCount_RNA, the mitochondrial expression ("percent.mt") and cell cycle ("S.Score" and "G2M.Score") using the *SCTransform* function.

To identify cell clusters, we employed Pathway and Geneset OverDispersion Analysis (PAGODA2) (9). 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 (2). 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 (10). The *DoHeatmap* function was used to generate an expression heatmap for given cells and features. Pearson correlation analysis of cell clusters was performed using the *CellScatter* function. GO analysis (DAVID v6.8) was done using top differentially (positively) expressed genes with a *p*-adjusted cut-off of 0.01. Lists of differentially expressed genes were analyzed by Ingenuity Pathway Analysis to identify biological pathways that are significantly over-represented among the genes in each list.

Cell trajectory analysis

Single-cell pseudotime trajectories were constructed with the Monocle 2 package (v2.10.1) (11). UMI counts were modeled as a negative binomial distribution. The 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. Differentially expressed genes were identified and used for dynamic trajectory analysis (false discovery rate [FDR] < 0.01), which ordered cells in pseudotime. The plot_pseudotime_heatmap function was used to generate heatmaps.

Statistical analysis

The details of the statistical method used for identifying the differential gene expression and pseudotime trajectory analysis are provided in the detailed methods above. 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.

Figure S1

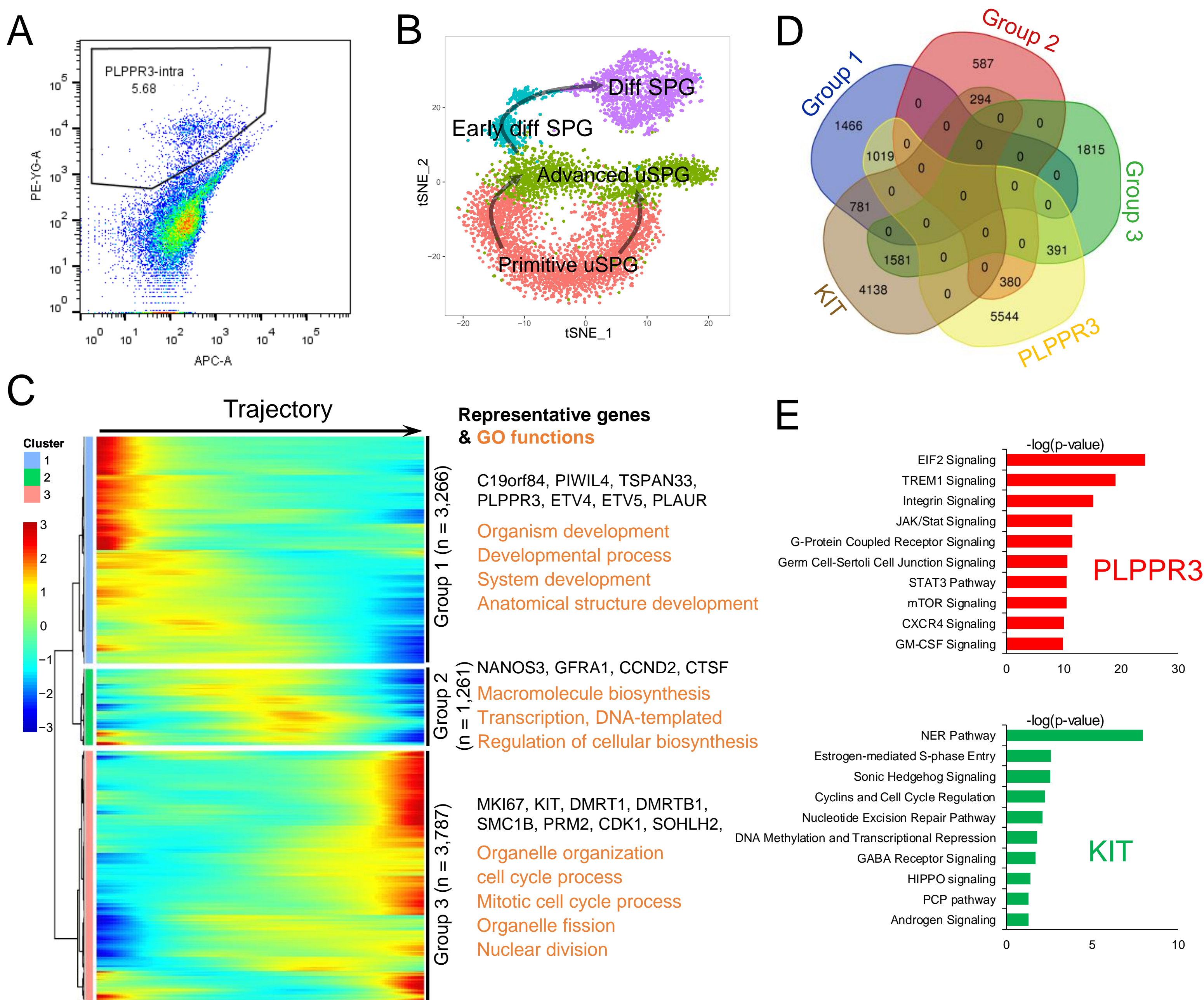

Figure S1. Transcriptomic analysis of human SPG development. (*A*) FACS plot of permeabilized adult human testicular cells stained with a polyclonal antisera against human PLPPR3. (*B*) tSNE plot of different SPG cell subsets in adult human testes, as defined in Sohni A *et al.* (15). (*C*) Heatmap of differentially expressed genes (DEGs) from different human SPG subsets following the trajectory timeline shown in (15). Top, pseudotime directions; Right, the number of DEGs and the representative biological processes. (*D*) Venn plot showing the overlapping DEGs identified from scRNAseq (*C*) and from RNAseq (Fig. 1*F*). (*E*) The most significantly enriched signaling pathways in PLPPR3+ and KIT+ cells, as defined by Ingenuity pathway analysis of the DEGs in these two cell subsets.

Figure S2

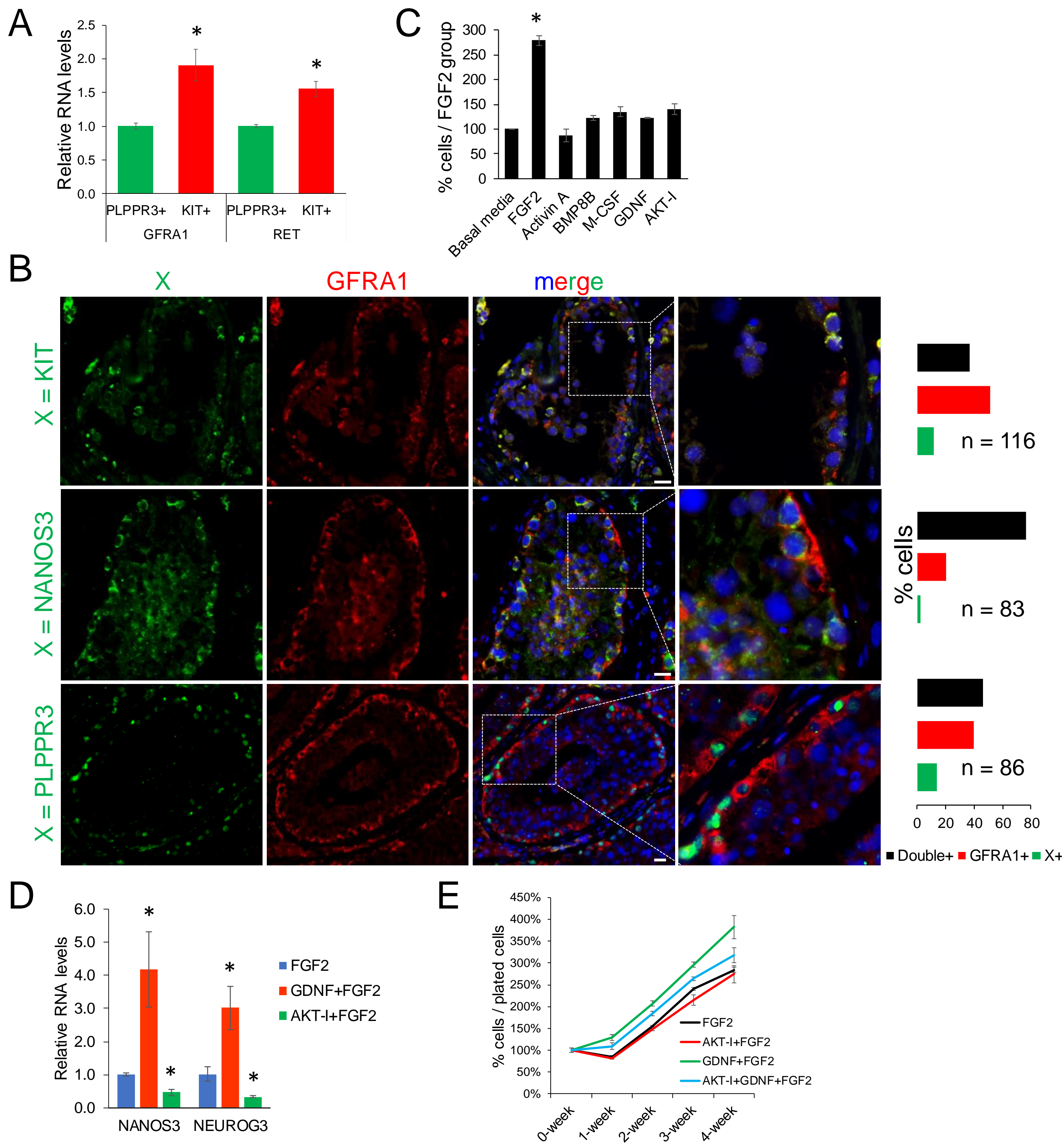

Figure S2. (*A*) *GFRA1* and *RET* expression in PLPPR3+ and KIT+ cells, as determined by qPCR. (*B*) Left, IF analysis of human adult testes sections co-stained with antisera against GFRA1 and the other proteins shown. Scale bar, 20 µm. Right, proportion of single- and double-positive cells. n, number of cells counted. (*C*) Percentage of cells after two-week culture, relative to cells incubated with basal media condition. (*D*) qPCR analysis of NANOS3 and NEUROG3 gene expression in MACS-purified ITGA6+ cells cultured for 2 weeks under the conditions indicated. The values shown are relative to our standard culture media that contains FGF2 (mean ± SD from three biological replicates). (*E*) Percentage of cells during four-week culture, relative to the number of cells initially plated (0-week). *, $p<0.05$.

Figure S3

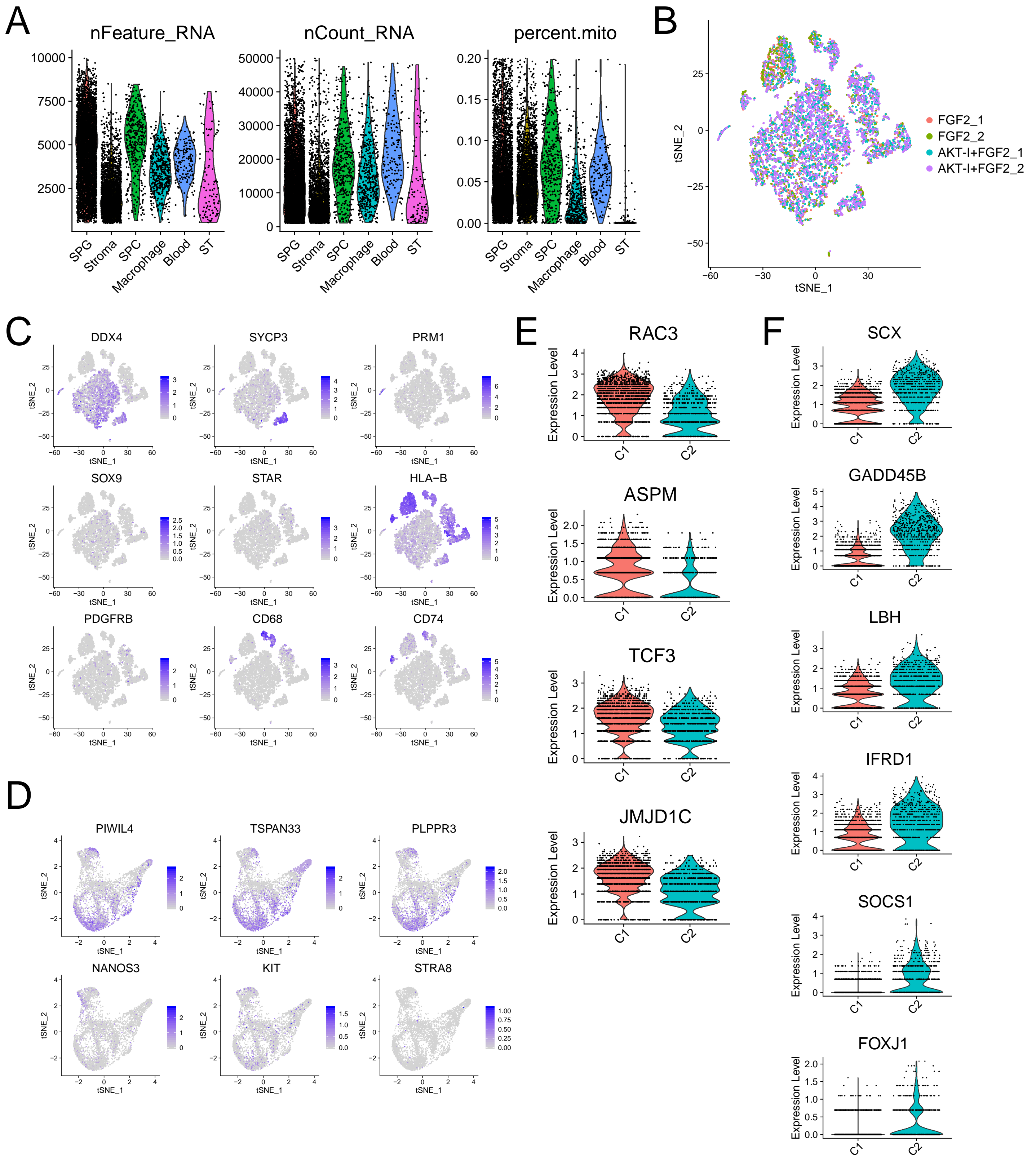

Figure S3. scRNAseq analysis of the major cell types and key SPG genes expressed in ITAG6+ testicular cell cultures. (*A*) Violin plots of the features indicated in the cell types in the cultures described in Fig. 4*A*. (*B*) tSNE plot of cells from different samples and biological replicates, clusters were defined in Fig. 4*A*. (*C*) Examples of gene marker genes used to annotate the cell clusters in Fig. 4*A*. (*D*) Examples of gene markers used to annotate the cell sub-clusters defined in Fig. 4*B*. (*E*) Violin plots showing the expression of genes associated with stem cell self-renewal or maintenance in C1 and C2 cells. (*F*) Violin plots showing the expression of genes associated with cell differentiation in C1 and C2 cells.

Table S1. Mapping results (from CellRanger) of testes samples analyzed by scRNAseq in this study

Samples	Total reads	Mean reads / cell	Median gene / cell	# genes detected	# cell captured
FGF2_1	263,641,733	103,959	2,418	30,648	2,536
FGF2_2	250,137,689	117,767	2,414	30,207	2,124
FGF2+Akt-I_1	218,159,661	65,790	2,901	30,719	3,316
FGF2+Akt-I_2	229,354,682	72,056	3,058	30,552	3,183

Table S2. The number of the cells in each cell type from each sample in this study

Samples	SPG	Stroma	SPC	Macrophage	Blood	ST
FGF2_1	1021	729	63	123	35	24
FGF2_2	859	608	75	98	34	18
FGF2+AktI_1	1553	764	118	129	41	39
FGF2+AktI_2	1568	689	149	115	44	20

Table S3. Percentage of cells in the cell-cycle phases indicated.

Source	Clusters	Groups	Cell-cycle phase		
			G1	G2M	S
Cultured SG	C1	FGF2	12.8%	62.9%	24.3%
		AKT-I+FGF2	12.1%	57.9%	30.0%
	C2	FGF2	12.7%	58.5%	28.8%
		AKT-I+FGF2	14.4%	57.2%	28.4%
	C3	FGF2	23.4%	50.0%	26.6%
		AKT-I+FGF2	28.3%	43.7%	28.0%
	C4	FGF2	29.8%	48.9%	21.4%
		AKT-I+FGF2	39.0%	36.5%	24.5%
	C5	FGF2	52.4%	16.7%	30.9%
		AKT-I+FGF2	53.3%	17.0%	29.6%
Fresh	Primitive uSPG	-	75.6%	20.2%	4.2%
	Advanced uSPG	-	78.5%	11.2%	10.2%

Dataset S1 (separate file).

Differentially expressed genes identified from PLPPR3+ *vs* KIT+ cells (q<0.01, Log2FC>1 or <-1).

Pseudotemporal significantly changed genes identified from scRNAseq dataset from Ref 15.

Genes encoding components in cell-surface signaling pathways enriched in PLPPR3+ cells.

Differentially expressed genes in each cluster relative to all other culture SPG subsets identified from scRNAseq.

SI References

1. Valli H*, et al.* (2014) Fluorescence- and magnetic-activated cell sorting strategies to isolate and enrich human spermatogonial stem cells. *Fertil Steril* 102(2):566-580 e567.
2. Sohni A*, et al.* (2019) The neonatal and adult human testis defined at the single-cell level. *Cell Rep* 26(6):1501-1517 e1504.
3. Hermann BP*, et al.* (2007) Characterization, cryopreservation, and ablation of spermatogonial stem cells in adult rhesus macaques. *Stem Cells* 25(9):2330-2338.
4. Dobin A & Gingeras TR (2015) Mapping RNA-seq Reads with STAR. *Curr Protoc Bioinformatics* 51:11 14 11-19.
5. Liao Y, Smyth GK, & Shi W (2013) The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* 41(10):e108.
6. Love MI, Huber W, & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550.
7. Kanatsu-Shinohara M*, et al.* (2014) Improved serum- and feeder-free culture of mouse germline stem cells. *Biol Reprod* 91(4):88.
8. Butler A, Hoffman P, Smibert P, Papalexi E, & Satija R (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 36(5):411-420.
9. Fan J*, et al.* (2016) Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nat Methods* 13(3):241-244.
10. Kowalczyk MS*, et al.* (2015) Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res* 25(12):1860-1872.
11. Qiu X*, et al.* (2017) Single-cell mRNA quantification and differential analysis with Census. *Nat Methods* 14(3):309-315.