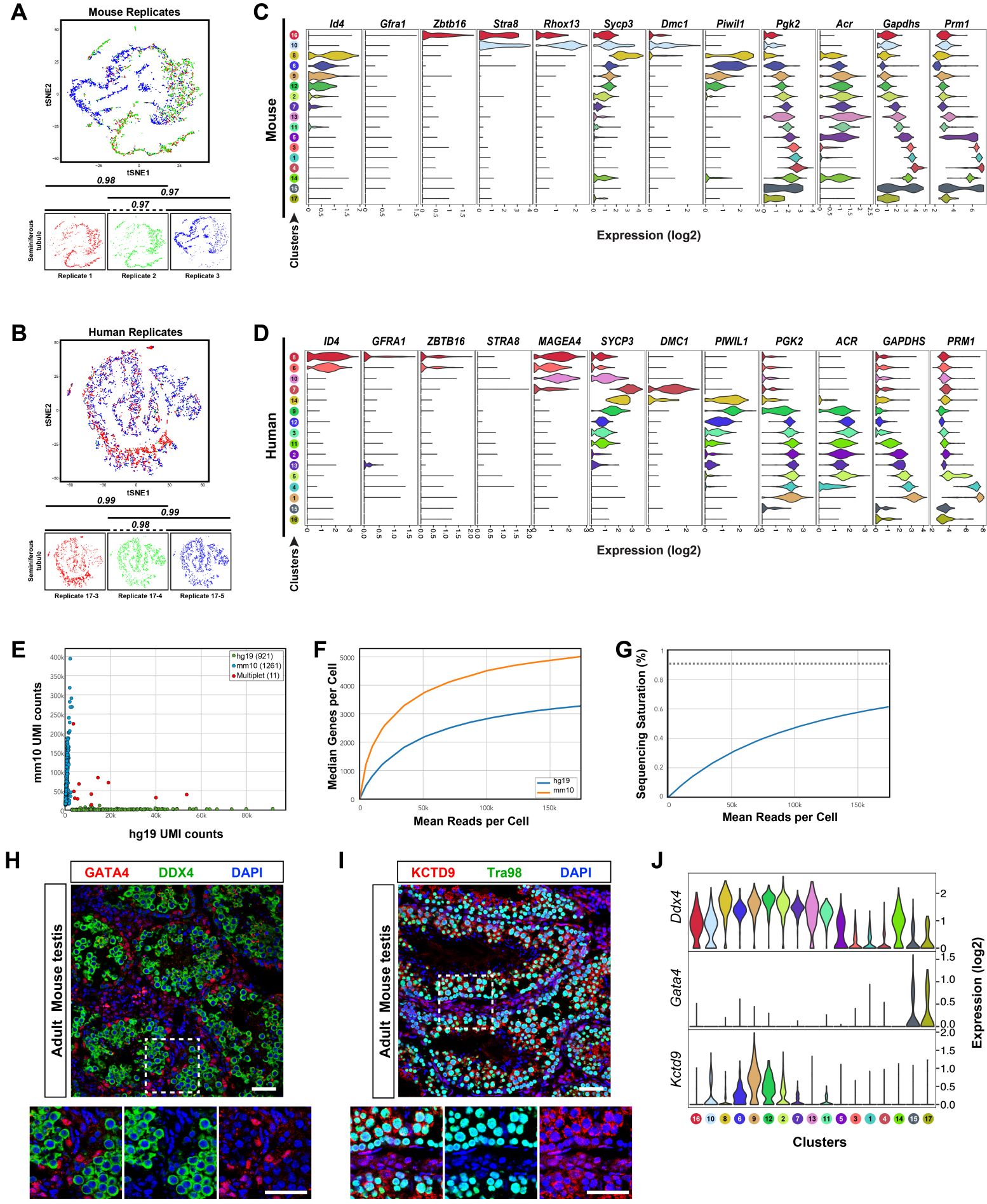
Cell Reports, Volume *25*

Supplemental Information

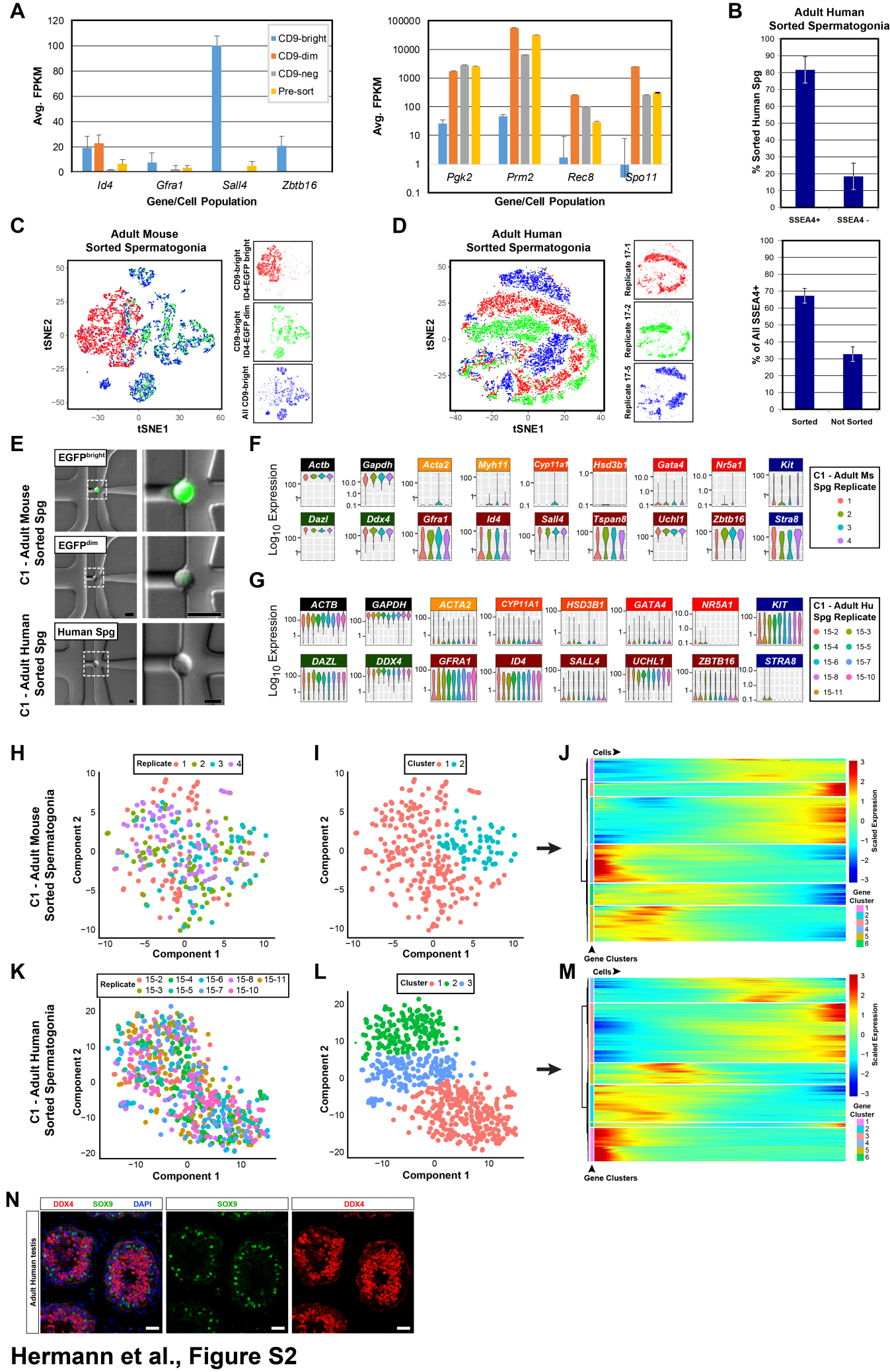
The Mammalian Spermatogenesis Single-Cell Transcriptome, from Spermatogonial Stem Cells to Spermatids

Brian P. Hermann, Keren Cheng, Anukriti Singh, Lorena Roa-De La Cruz, Kazadi N. Mutoji, I-Chung Chen, Heidi Gildersleeve, Jake D. Lehle, Max Mayo, Birgit Westernströer, Nathan C. Law, Melissa J. Oatley, Ellen K. Velte, Bryan A. Niedenberger, Danielle Fritze, Sherman Silber, Christopher B. Geyer, Jon M. Oatley, and John R. McCarrey



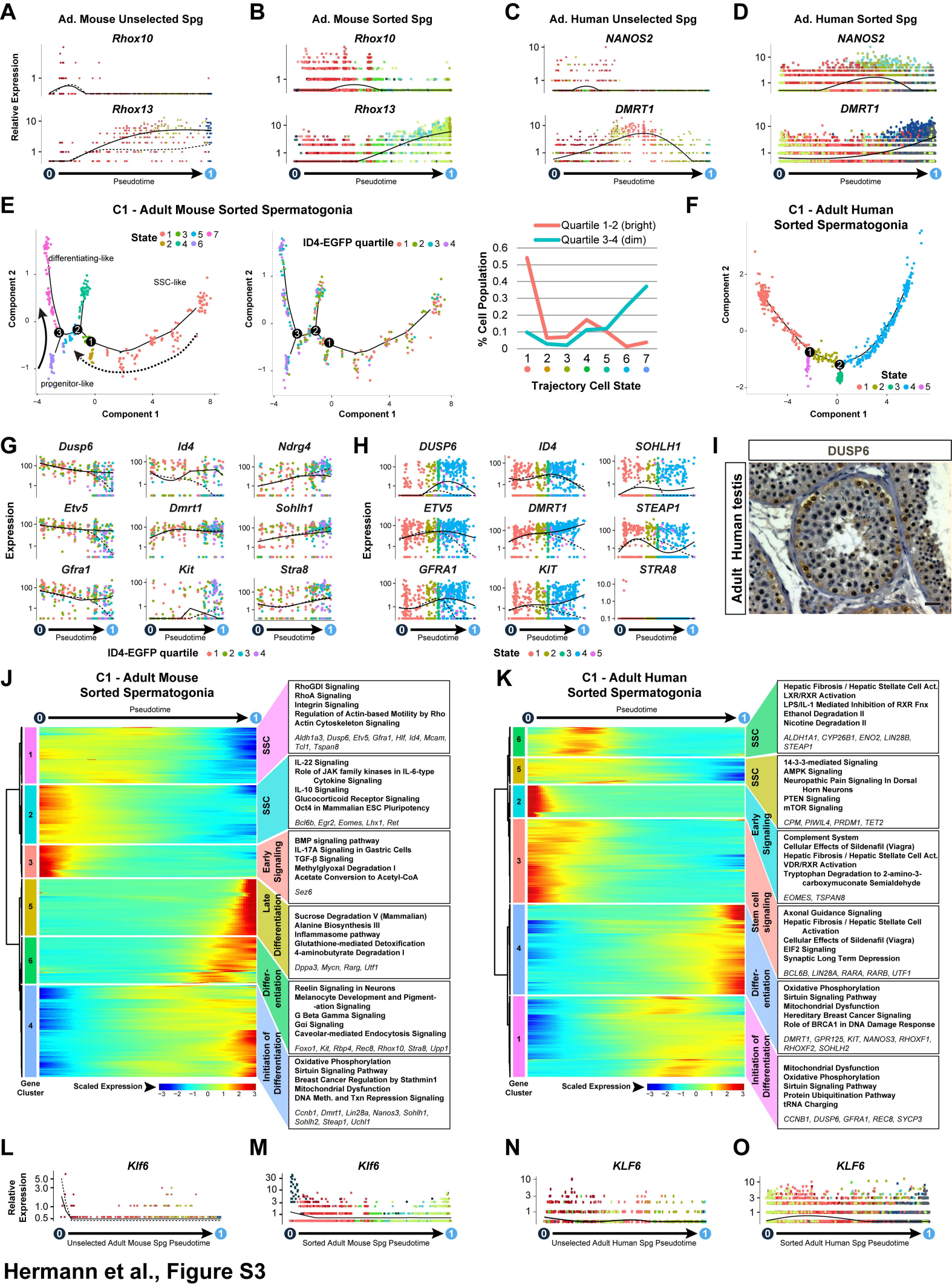
Hermann et al., Figure S1

Figure S1 – Related to Figure 1. tSNE plots show biological replicates of 10X Genomics profiling of unselected spermatogenic cells (Fig. 1A-B) from adult **(A)** mouse testes and **(B)** human testes. Modified RV coefficients (multivariate squared Pearson correlation coefficients) demonstrate a high degree of gene expression matrix pairwise similarity between replicates (Smilde et al., 2009). **(C, D)** Violin plots show mRNA level variation among clusters in Fig. 1A-B for the landmark genes shown in Fig. 1E. The horizontal axis shows normalized expression levels. **(E-G)** To ensure singularity of our 10X Genomics data, we performed a mouse-human cell mixing experiment (multiplet analysis). Equal numbers of freshly-isolated unselected spermatogenic cells from each species were mixed prior to emulsion compartmentalization and resulting data were analyzed using a combined hg19 (human) and mm10 (mouse) genome annotation and assembly. **(E)** Eleven of 2,193 detected cells contained substantial sequences corresponding to both species and were designated as mouse:human multiplets, giving a multiple rate of ~1% (presuming an equal number of mouse:mouse and human:human multiplets which were not detected in this analysis). **(F)** In human and mouse cells, we detected different median numbers of genes per cell at similar sequencing depth, but saturated gene detection at similar sequencing depths. **(G)** Sequencing saturation for this experiment was 61.4% at 174k mean reads per cell. **(H-I)** Immunostaining of adult mouse testes for proteins encoded by genes which are differentially-expressed among testicular cell types and **(J)** corresponding mRNA expression profiles for the noted genes in violin plot formats. Dotted white box in the large images indicate area enlarged in second row of images. Scale bar = 25μm. GATA4 (somatic) and DDX4 (spermatogonia) staining was mutually exclusive, matching the expectation from mRNA measurements. Likewise, KCTD9 was only detected in germ cells located at least one cell layer removed from the basement membrane, matching expectations that it is expressed by spermatocytes and early round spermatids (clusters 2, 6, 7, 8, 9, 10, and 12 – see Fig. 1E).



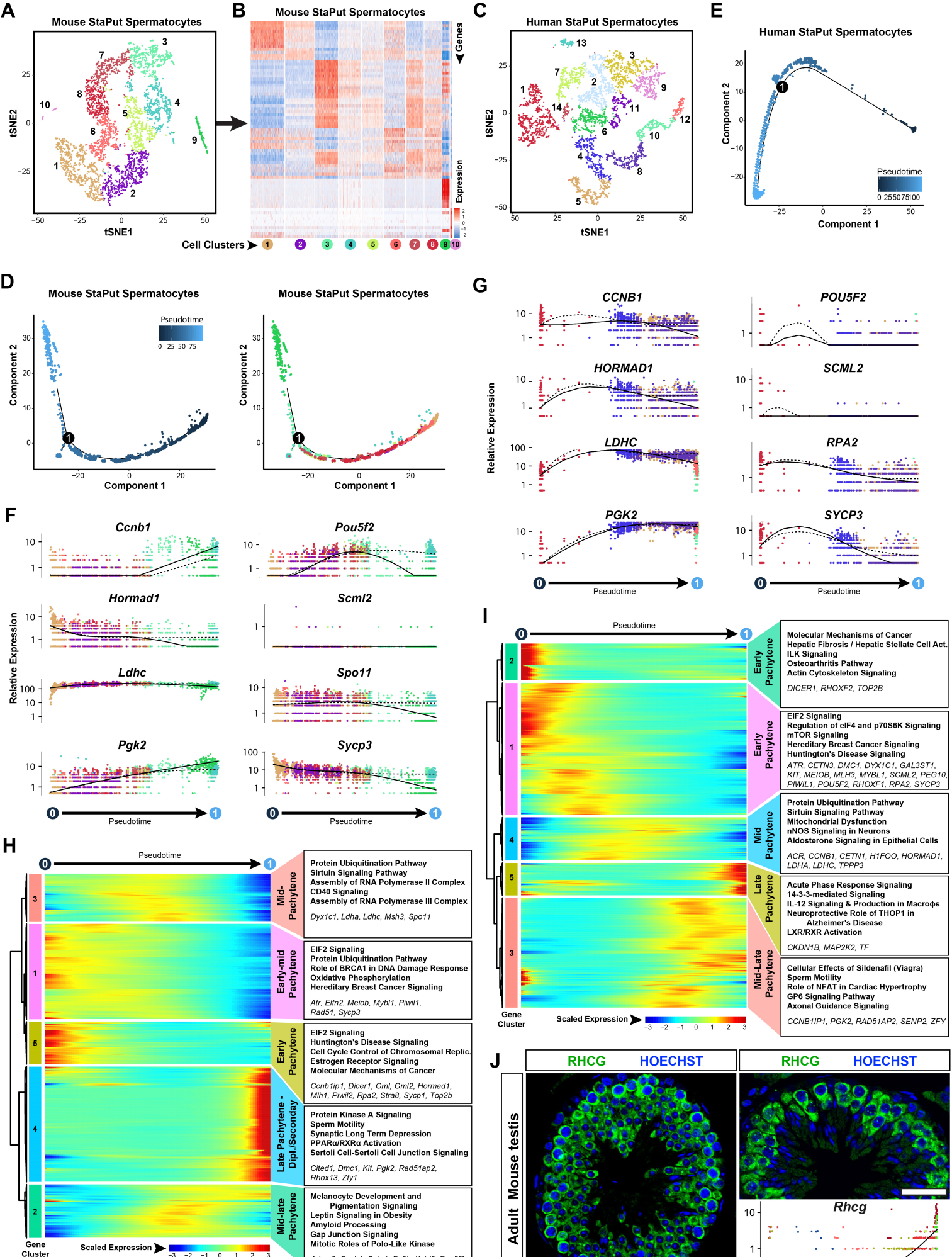
Hermann et al., Figure S2

Figure S2 – Related to Figure 2. (A) Low-pass RNA-seq assessment of gene expression in pre-sort and sorted subpopulations of adult mouse ID4-GFP+ cells exhibiting different staining intensities for CD9 (bright, dim, negative). Levels of mRNA for key genes that are known to be expressed by (left) spermatogonia or (right) spermatocytes demonstrated that the ID4-EGFP+/CD9^{bright} population contained spermatogonia and were used for subsequent experiments. **(B)** Bar graphs show the proportion of sorted human spermatogonia (Fig. 2F) that were SSEA4+ (top) and the proportion of SSEA4+ that were sorted in the spermatogonia used for experiments (bottom). tSNE plots show the independent samples of 10X Genomics profiled sorted spermatogonia from adult **(C)** mouse testes (Fig. 2G) and **(D)** human testes (Fig. 2I). It was notable that while some clusters from sorted human spermatogonia contained significant contributions from all three replicates (clusters 4, 7, 8), others were derived primarily from one replicate (clusters 2, 3, 5, 6, 9, 10) or two replicates (cluster 1) (compare Figs. 2I to S2D). However, replicate-biased clusters from different replicates always overlapped in pseudotime (see Fig. 2R). Therefore, batch effects arising from the relatively few transcriptome differences among human spermatogonia are muted by pseudotime trajectory analyses. **(E)** Images of sorted spermatogonia captured in Fluidigm C1 IFC chips (right column shows enlargements of boxed area in left column images), (top) mouse ID4-EGFP^{bright} spermatogonium, (middle) mouse ID4-EGFP^{dim} spermatogonium, (bottom) human spermatogonium. Violin plots show distribution of mRNA levels among individual **(F)** C1 mouse adult sorted spermatogonia or **(G)** C1 human adult sorted spermatogonia (each violin shows an individual biological replicate) for genes considered to mark all testis cells (black), peritubular myoid cells (yellow), Leydig cells (orange), Sertoli cells (red), spermatogonia (green), undifferentiated spermatogonia (maroon), and differentiating spermaotogonia (blue). tSNE plots show heterogeneity of single-cell transcriptomes from C1-captured **(H)** adult mouse sorted spermatogonia (300 cells) or **(K)** adult human sorted spermatogonia (635 cells), in which individual cells are colored by biological replicate, and **(I, L)** unsupervised clustering which defined sub-groups of cells in each dataset. Heat maps show six clusters of significantly differentially expressed genes (DEGs) among **(J)** mouse and **(M)** human C1-captured sorted spermatogonia. **(N)** To determine concordance between our mRNA expression data and protein expression, we performed immunostaining for proteins encoded by genes which are differentially-expressed among human testicular cell types including for DDX4 (red, germ cells) and SOX9 (green, Sertoli cells) and counterstained with DAPI (blue). Scale bar = 25μm.



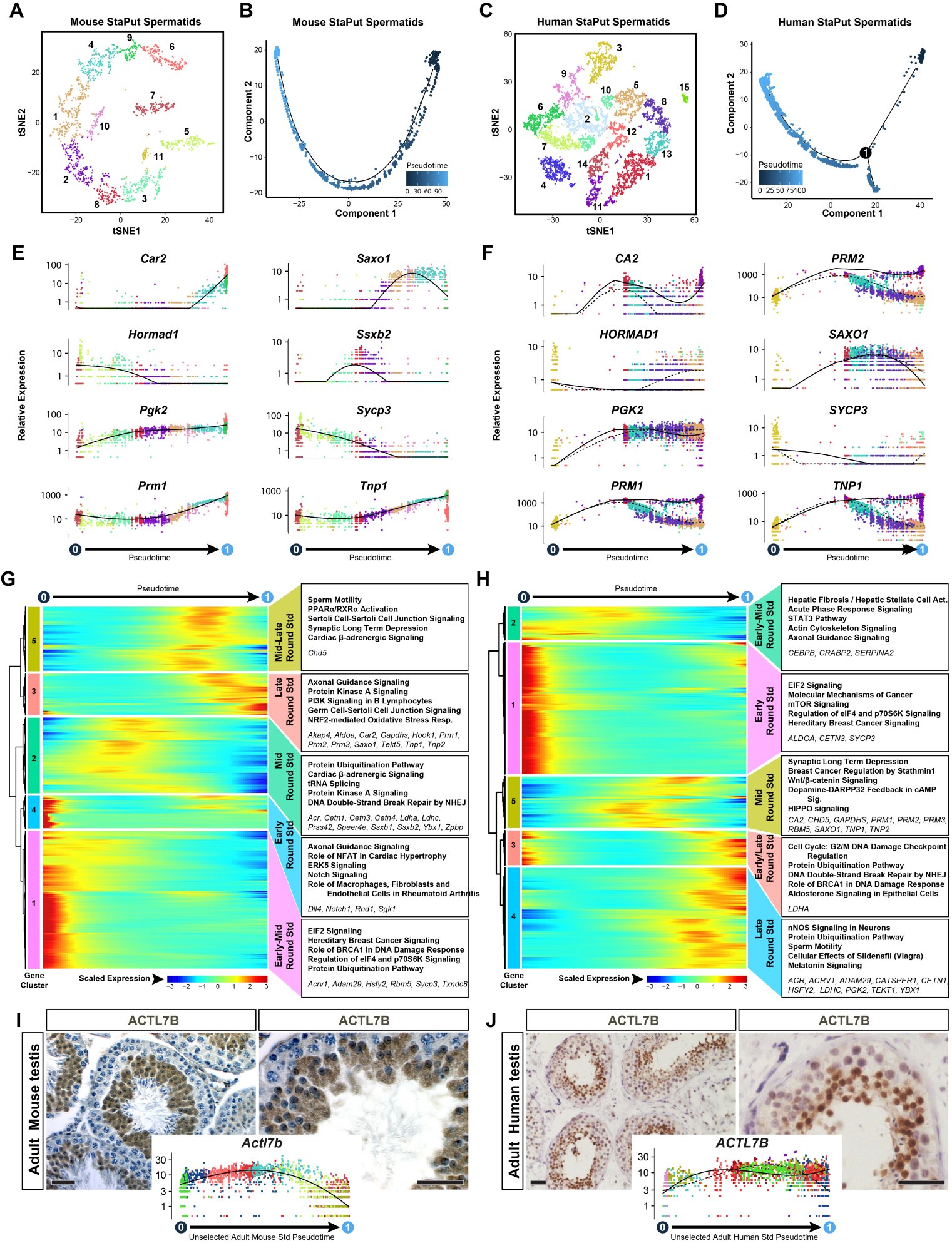
Hermann et al., Figure S3

Figure S3– Related to Figure 3. (A-D) Expression of additional key landmark genes from 10X Genomics spermatogonial datasets (Fig. 3A-D) are shown in pseudotime. Single-cell transcriptomes from C1-captured sorted spermatogonia (Fig. S2E-M) from **(E)** adult mouse testes and **(F)** adult human testes were used for unbiased dynamic lineage analysis using Monocle, which produced trajectories with cells ordered in pseudotime. Plots are shown as trajectories formatted with cell state. For adult mice, trajectories were also projected retrospectively with ID4-EGFP epifluorescence quartile (1=brightest, 4=dimmest). The graph shows the proportion of ID4-EGFP^{bright} (quartiles 1-2) and ID4-EGFP^{dim} (quartiles 3-4) distributed in each spermatogonial trajectory state. Branchpoints in the single-cell trajectories are noted by black numbered circles. **(G-H)** Expression levels (vertical axis) of key landmark genes among spermatogonia over pseudotime. Mouse spermatogonia are colored according to ID4-EGFP epifluorescence quartiles and human spermatogonia are colored by cell state. Pseudotime (scaled) is indicated below each gene plot column. **(I)** Antibody staining for DUSP6 in a human testis histological section where brown staining indicates immunoreactivity in spermatogonia. Scale bar = 25µm. **(J-K)** Differentially-expressed gene clusters and their hierarchical relationship across pseudotime from C1 sorted spermatogonial datasets are shown in heat map format (see expression level legend). Enriched biological pathways (GO Analysis) within each cluster are noted in the boxes to the right (Table S3) and key genes are identified. **(L-O)** Pseudotime expression profile of a novel SSC marker arising from the Hepatic Stellate Cell Activation pathway, *Klf6/KLF6*, in 10X Genomics spermatogonial datasets (Fig. 3A-D).



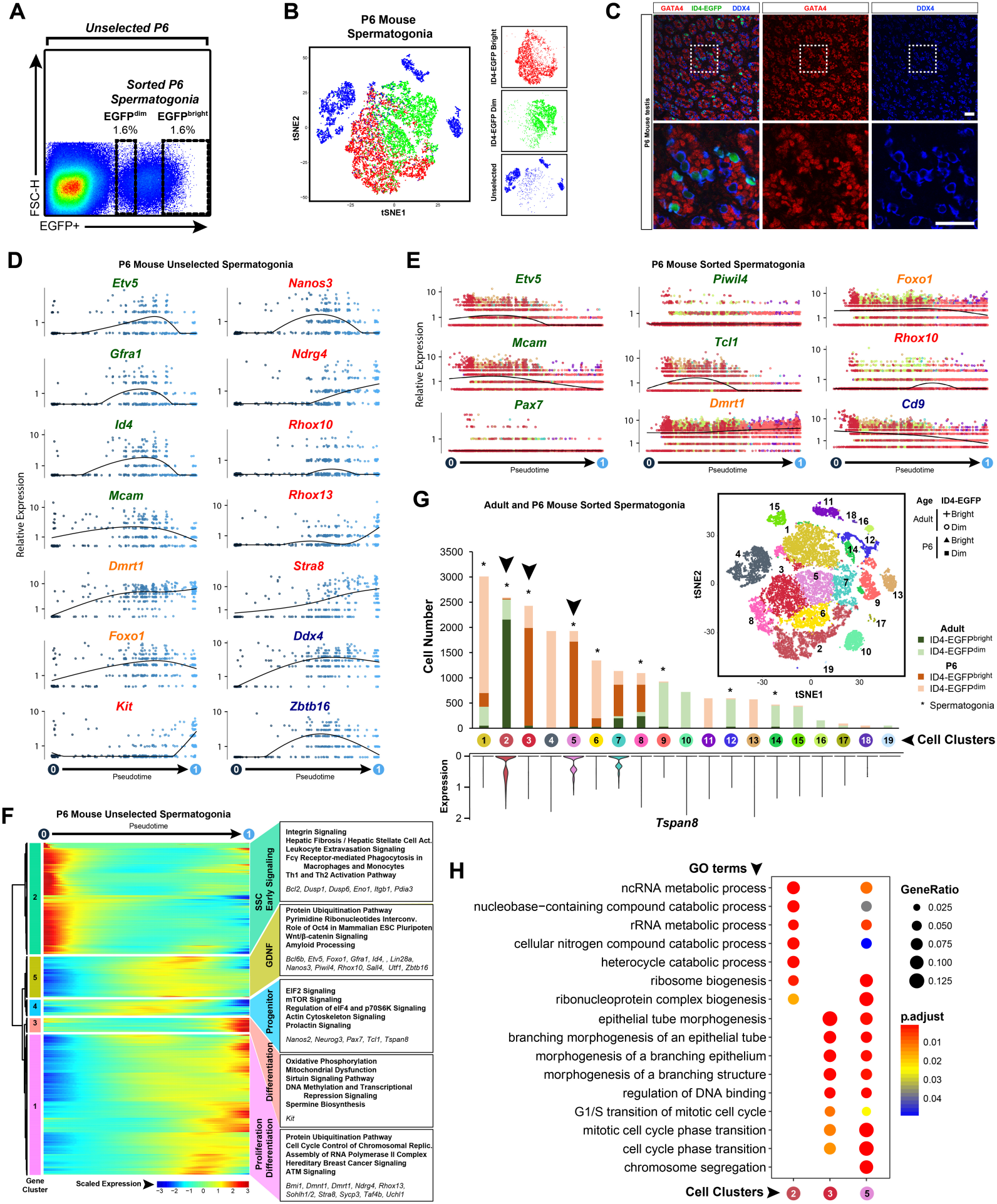
Hermann et al., Figure S4

Figure S4 – Related to Figure 4. 10X Genomics single-cell transcriptomes from StaPut (density)-enriched pachytene spermatocytes from **(A)** adult mouse and **(C)** adult human testes were subject to unbiased clustering as shown on the tSNE plots. Heat map shows the top 10 significantly differentially expressed genes (DEGs) between each cell cluster for **(B)** mouse StaPut spermatocytes (and human data not shown). Colors and numbering of circles below the heatmap matches the corresponding tSNE plot. **(D-E)** StaPut spermatocyte transcriptomes were subsequently used for unbiased dynamic lineage analysis producing single-cell trajectories in pseudotime and ordering cells from each cell cluster. Mouse StaPut spermatocyte trajectory is also displayed with labeling according to cell cluser. Branchpoints in the single-cell trajectories are noted by black numbered circles. **(F-G)** Expression levels (vertical axis) of key genes among StaPut spermatocytes across pseudotime (cell coloring is according to tSNE cluster). Dotted trend lines reflect expression across the minor branch, while the solid trend line reflects expression across the trunk. **(H-I)** Heat maps show clusters of genes (and their hierarchical relationship) that were differentially-expressed across pseudotime from StaPut spermatocytes (expression color code noted at the bottom, Table S2). The top five over-represented biological pathways from GO analyses of each cluster are noted at the right in bold (Table S3) and key genes are italicized. **(J)** Immunostaining of adult mouse testes for RHCG (green) and counterstained with Hoechst 33342. Inset shows *Rhcg* mRNA levels among unselected mouse spermatocytes across pseudotime. Scale bars = 25μm.



Hermann et al., Figure S5

Figure S5 – Related to Figure 5. StaPut-enriched round spermatids from **(A)** adult mouse and **(C)** adult human testes were used to generate single-cell transcriptomes by 10X Genomics analysis and data were subjected to unbiased clustering as shown on the tSNE plots. **(B, D)** Single-cell trajectories resulting from unbiased dynamic lineage analysis ordered StaPut spermatids in pseudotime (left). Branchpoints in the single-cell trajectories are noted by black numbered circles. **(E-F)** Expression levels (vertical axis) of key genes across pseudotime among StaPut spermatids (horizontal axis) is shown for cells colored according to tSNE cluster. Dotted trend lines reflect expression across the minor branch, while the solid trend line reflects expression across the trajectory trunk. **(G-H)** Heat maps show hierarchical relationship between clusters of genes that were differentially-expressed across pseudotime from StaPut spermatids (scaled expression as shown in legend, Table S2). The top five over-represented biological pathways from GO analyses of each cluster are noted at the right in bold (Table S3) and key genes are italicized. **(I-J)** Immunostaining of adult mouse testis and adult human testis for ACTL7B and counterstained with hematoxylin. Insets show *Actl7b/ACTL7B* mRNA levels among unselected mouse or human spermatids across pseudotime (cells colored as in Fig. 5G-H). Scale bars = 25μm.



Hermann et al., Figure S6

Figure S6 – Related to Figure 6 – (A) FACS plot shows testis cells from postnatal day (P) 6 *Id4-eGfp* transgenic mice with unselected cells noted (encompassing all cells) along with sort gates used to select ID4-EGFP^{bright} and ID4-EGFP^{dim} spermatogonia (each comprising roughly 1.6% of starting testis cells). **(B)** tSNE shows the samples of unsorted P6 testis cells (red) and sorted P6 spermatogonia (green and blue) profiled by 10X Genomics analysis. **(C)** To validate mRNA expression data, we performed immunostaining in sections of P6 *Id4-eGfp* mouse testes for proteins encoded by genes which are differentially-expressed among testicular cell types including GATA4 (red) and DDX4 (blue) with ID4-EGFP epifluorescence (green). Scale bars = 25μm. **(D)** Expression levels (vertical axis) of key landmark genes from unselected P6 mouse spermatogonia (Fig. 6E) are shown in pseudotime. **(E)** Expression levels (vertical axis) of additional key landmark genes from the P6 sorted spermatogonial trajectory (Fig. 6F) are shown in pseudotime. **(F)** The heat map shows the hierarchical relationship between clusters of genes that were differentially-expressed across pseudotime from unselected P6 mouse spermatogonia (scaled expression as shown in legend, Table S2). The top five over-represented biological pathways from GO analyses of each cluster are noted at the right in bold (Table S3) and key genes are italicized. **(G)** Adult and P6 sorted spermatogonia datasets were merged and subjected to unbiased analysis together and tSNE shows unsupervised clustering of the merged data. Graph shows distribution of cells among all cell clusters by their origin (P6 or Adult, ID4-EGFP^{bright} or ID4-EGFP^{dim} spermatogonia) according to the key. Each bar represents one cell cluster and bars from clusters containing spermatogonia are noted with asterisks. Arrows denote three clusters containing predominantly P6 or Adult ID4-EGFP^{bright} which were used for GO analyses. The inverted violin plot below the graph depicts single-cell mRNA levels for the SSC cell surface marker *Tspan8* among the 19 cell clusters from the tSNE plot. Each violin is colored according to the noted cell cluster appearing immediately above. Populations of spermatogonia expressing *Tspan8* were only found in clusters containing ID4-EGFP^{bright} spermatogonia. Interestingly, only clusters 5 and **(H)** GO term analysis results from Bioconductor – clusterProfiler of DEGs between the noted cell clusters.

Hermann et al., Figure S7

Figure S7 – Related to Figure 7 – We identified genes which were significantly differentially-expressed in pseudotime among spermatogonia, spermatocytes and spermatids in both mice and men. Those germ cell expressed genes which were found to be germ cell specific in their expression pattern were included in the Figure 7 analysis. Shown are violin plots that depict single-cell mRNA levels for the noted genes among the unselected cells shown in clusters overlaid on the tSNE plots from Fig. 1A-B. These represent the expression profile of the noted genes in adult **(A)** mouse testes and **(B)** human testes. Each violin corresponds to the noted cell cluster from Figure 1 (designated with the colored/numbered circle below the axis). In addition to the spermatogenic cell-type specific genes, expression profiles are also shown for genes known to be expressed by various testicular somatic cell types, pan germ cell markers, and ubiquitously (housekeeping). The vertical axis shows log-normalized expression levels.

Table S5 – Human testicular tissue sources (related to Figs. 1-5, 7)

Sample ID	***Experiment and/or Replicate****	***Age** (years)*	***Source***	***Ischemic time** (hr:min)*
15-1	qRT-PCR	48	Biopsy	19:45
15-2	C1, qRT-PCR	50	Biopsy	17:05
15-3	C1, qRT-PCR	40	Biopsy	17:25
15-4	C1, qRT-PCR	38	Biopsy	17:10
15-5	C1, qRT-PCR	46	Biopsy	17:15
15-6	C1, qRT-PCR	35	Biopsy	17:05
15-7	C1, qRT-PCR	54	Biopsy	17:30
15-8	C1, qRT-PCR	53	Biopsy	17:05
15-10	C1, qRT-PCR	30	Organ Donor	18:36
15-11	C1, qRT-PCR	40	Biopsy	18:00
16-1	qRT-PCR	41	Biopsy	44:10
16-2	qRT-PCR	36	Biopsy	20:10
17-1	10X Spg Rep1, qRT-PCR	37	Biopsy	17:26

17-2	10X Spg Rep2	38	Biopsy	17:22
17-3	10X Rep 1 10X Ms-Hu mixing	34	Biopsy	18:25
17-4	10X Rep 2, qRT-PCR	36	Biopsy	21:55
17-5	10X Rep 3 10X Spg Rep3, qRT-PCR	49	Biopsy	22:45
17-6	10X StaPut Sct/Std Rep 1, bulk StaPut Sct/Std, qRT-PCR	43	Biopsy	41:15
17-7	qRT-PCR	52	Biopsy	22:00
17-8	qRT-PCR	52	Biopsy	18:43
17-9	qRT-PCR	44	Biopsy	17:07
17-10	qRT-PCR	44	Biopsy	20:18
17-11	10X StaPut Sct/Std Rep 2, bulk StaPut Sct/Std, qRT-PCR	43	Biopsy	23:05
17-12	Spg bulk, qRT-PCR	43	Biopsy	21:45
17-13	Spg bulk, qRT-PCR	28	Biopsy	22:09
17-16	qRT-PCR	44	Biopsy	21.67
17-17	qRT-PCR	43	Biopsy	26.00
17-23	qRT-PCR	38	Biopsy	18:04
17-24	qRT-PCR	27	Organ Donor	24:00
17-25	qRT-PCR	42	Biopsy	23:00

* qRT-PCR = Figure 7, C1=single-cell RNA-seq with Fluidigm C1, 10X=10X Genomics method.

Table S7: Oligodeoxynucleotide primers (related to Fig. 7)

Gene mRNA*	Forward primer 5'to 3'	Reverse primer 5' to 3'
1700027A15Rik	GGACTGATGGACTTGGGGTG	TGGACACCAACACGTACAGG
1700080E11Rik	GGGAGTGTCCACTTTCAGCA	CAGGCCCAAGACCTCTAGC
4930511A02Rik	CACAGTCATAGCACCCAGTAAG	CCTCTCTGTGTGGATGTGATAC
4930513O06Rik	TACTTTGCAGGCACTGTGAAT	GTCTCAGCTTTCTCCTCACTTTC
4933414I15Rik	GTGGATCATCACTTCTCAACCT	GTCTGTGACCATTCGCTTACT
Acot10	CTCACGCTCACACTCTTCTGCT	TGTTCAAGGTCCAAAGCCTCAT
Adad2	TCTCTGCCCTTCAGTACATC	GTGGGTCAGGATAGTCTCTATG
Adam29	CCTTCCAGTGTTCACCTACTC	CTGAGTCACCCTCCACATAAC
B3galnt1	GAAAGCCAGACAAGCCATTAG	CCTGCTGGCCTAGTAAGAAA
Ccnb1ip1	CCCACCAGGGAATAACTCAAAG	GCTGTGGGAGAACACACAAA
Cyp2a12	GCGATTCTGCTTGGGAGACAG	AAGTTCTGCAAGATGGTGGTGA
Dmc1	CCAGGAGCAACTATGACCTTTC	AATCTTGGCGATCCTCAGTTC
Dyx1c1	CCGGGTGTTGATAAAGAGATGA	GTATCTCTGGTCTTCTCGCTTG
Esrp1	GTAGGGACTTCCTTCTGTCTCT	TCAGGCAGTAACACATTCTTCTT
Galnt12	CTACAGGAGGATGGCACTTTAG	GTTGGTACAGTCCCGTAAGTATG
Gm10354	CAAAGAGATCCAGCTCACTATGGA	GGTAGGGCCTCTTGTTCATGGAT
Gm35584	GCCATTCCTTTCCAACTTCAATAA	TAGGACTGAAGGGCTGTAAGA
Lmo1	CTGGACAAGTACTGGCATGAG	GGTTGGCCTTGGTGTAGAG
Loxl2	TTGGAGAACAAGGCATCACC	GGGTTAATGACAACCTGGAACA
Meiob	CTGATCCTTTCTATGGCATCCT	ACCACAGCTGGAACATCTATT
Nanos2	GGCACTATGTGTGTCCTCTATG	GACTGCTGACTGCTGTTGA
Nanos3	GACTTTCAATCTTTGGACAGATTACC	GTTTGGGATCCAGCCTTACA
Prss42	TGCATGTCTGTGGAGGTTC	TGTACTGGATTCGGCTGTAAAT
Pth1r	TCGGGAACGGGAGGTATTT	ACATGTGCATGTGGATGTAGTT
Rad51ap2	GGAACGATCACTCCATTCTACTC	AATGACTGCCTCCCTTCAATAA
Rhcg	CTGACTTCTGTGTGGCATCTT	CCTTTGCCTCTATCAGGTTCAG
Rps2 †	CGCGCTTCTTGGAGCACTATA	TGCACCGGCGTCATCC
Snx16	AGGGCTTTCTGTGAGACTTTAG	ACAATGTTCTGATTCTGGTCTCT
Sox3	CACCCCCAGTCGTATTGCTT	ACACGCACACCTGGCTATAA
Speer4e	GAGGCCAAAGAGACCAAGCA	CTTGTAAAGGCCTATTCACCCTA
Ssxb2	CTGTGGCAATCCTCTCACCA	ACAAAACCAAATGTCCTGCGT
Tbx1	AAGGCAGGCAGACGAATG	GTCATCTACGGGCACAAAGT
Upp1	GACGTGAAGTTTGTGTGTGTTG	CCTGCACAGATGTTGGGATATT
Zic1	CATGAAGGTCCATGAGTCCTCTTC	TGGTCGGGTTGTCTGTTGT
ARMC1	AGCTCTGGAGCGAATTTAAGA	CTACCGATAGAGCGTCAGGC
ASB9	ATTGATGGGCGATGCTGT	TGATCTGCCGTGATGATGTT
C17orf74	CAACGTGGCAACTATGATGTG	GGACTTTCACTGGCCTGAAT
C17orf98	TGATCAGGATCATGGCGGTA	CCCAATTTCTCCTGTTGAGGTATC
C9orf57	TGGGACGTGTCTTTCCTTTCA	CCCAGGTCACCTAAGCGG
CDT1	GGAGCGTCTTTGTGTCCGAA	ATTTCCCCAGGGCTCATGATAG
CETN3	TCATGAATTAAAGGTGGCAATGAG	CCTGTGGCTTCTCTGTCATAAT
CT55	CCATAGCCATTGTTTCTGAAGATTT	ATACAACGGATGGGCTTCAC
DHRS13	CATCCACAATGCCGGTATCA	CAGCAGAAAGGGACCGATATG

EGR4	AGGCACTACCCTGGGATTCA	AGCCTGTCTCTGGGGGTTAT
ENPP2	TTATGTGTGATCTCCTGGGATTG	ATGGTTGGCCTGAAGGTATTAG
FSCN3	TCTGAGCGCTTAAACCGAATG	TGGCGCCTCTGGGATAG
HIST3H3	GTGTCATCCATGCCAAACGG	GTGGCGAGATAGCCCTCCTA
HSD17B14	GAGTGGTTATCTGCGACAAGG	CCAGGGTCTTCACATCATCTTC
ID4	TCCCGCCCAACAAGAAAGTC	CTGCAGGTCCAGGATGTAGTC
ISOC1	CCAGAAGTAGAAGCGGCATTAG	TGGATGCACACATGAGTTTCT
LRRC3B	AGCATGCCTTCAAAGGAGTAG	CCCTGGCCTTCAGGTTATTG
MEIOB	CTTTGGGCTGCACGGTACAT	GCTTCTTTCCAAGAGAAATTGCC
MGAT4D	TGACGCAGGAGAAGATCTTAGAA	ACCCACATGCTGGAAAAGAGA
NANOS2	TACTCCTCACACCAGCTGAA	GGCAGTACTTGAGCGTATGG
NANOS3	ACGCTTCTGCCCACTTAC	TTCTTGCCTGCCGAGTTT
NMT2	TCATCCTGGCTAAATCGAAAGG	GCCATCTCCTATACCAAACTTGA
PHOSPHO1	GGGCTTCTACAACGAGTACAT	CAAAGGGATGGCTTCGTAGAT
PHOSPHO2	TGGGAGATAAGGGTGTAAGAGA	AGTTGAAGAGTTCCACCATCC
PLPPR5	GGACTCAACAGAGTAGCAGAAT	CACGCACACAACCAGAAATAC
PRSS37	GCTCCCTATTTGGTGTACCTC	CAGATTTGGTAAATAGCAGTGAGC
PRSS58	CAACCTGCCCTACCAAACTATC	GTGAATCGGGCTCTTTGTAGAT
RPL7	CTGTGCCAGAAACCCTTAAGAA	CCTTGCCTTTCGAAGCATCT
TCN2	GATCACCATGGCCATCAGAA	GGGAAGTCATGAGGAACTGTAAT
TEX101	CCACATCAGCTGCTCACTCA	CAGCAGTAGCTGTAACCCCC
TMEM55A	CACTCTGGCAAAATGCCCAC	ATCTGGGGTGCCAACAGTTA
TP53TG5	AGGGTTTCCAAGATGCAAGAT	AGACGGTTCCGCTCAATTAC
TPPP3	CTGGCGACCAAGAGATTCAAG	GCTTTAGTGACGCCCACATT
TSPAN33	AGCCCGCTGGTGAAATACCT	AGGGCTGCTTCTGCATGCTT

*Note: Per nomenclature convention, mouse genes are noted with first letter capitalization (e.g., *Gfra1*) while human genes are all caps (e.g., *GFRA1*) using the official gene ID as the gene name.

† From From (Hermann et al., 2015)