## SUPPLEMENTAL INFORMATION

### Supplemental Experimental Procedures

*Generation of Id4 conditional overexpression mouse line*

To establish the human ubiquitin C (UBC) promoter - ATG - 5' *loxP* - TAG -polyA - 3' *loxP* - Id4 - SV40 splice & poly A transgene, the mouse *Id4* cDNA was amplified from pI-Id4 cDNA clone (Riechmann et al., 1994), changing the ATG start codon to ATC. Amplified Id4 cDNA fragment was ligated to a SmaI-ApaI fragment containing an ATG start codon in-frame with the 5' *loxP* open reading frame, the 3' end of neomycin (containing translational (TAG) and transcriptional stop (polyA) sequences) and a 3' *loxP* sequence (Zhang et al., 1996). A SalI-XhoI fragment containing ATG - 5' *loxP* - TAG - poly A - 3' *loxP* -Id4 cDNA was isolated and cloned into the pUC18 - UBI promoter - junB cDNA vector containing SV40 splice and polyA sequences replacing the junB cDNA (Schorpp et al., 1996). Transgenic mice carrying a single copy integration were established through pronuclear injection. Cre-mediated deletion of the *loxP*-flanked TAG - polyA sequences (375 bp) will result in in-frame fusion of ATG - *loxP* - *Id4* sequences and expression of N-terminally tagged *Id4*.

*qRT PCR analyses*

Total RNA was isolated using Trizol reagent from ID4-EGFP$^{Bright}$ and ID4-EGFP$^{Dim}$ spermatogonia isolated by FACS from P7 pups and treated with DNase I. For each sample, 0.1-4μg of RNA was reverse transcribed using oligo d(T) priming and Reverse Transcriptase III (Invitrogen). Quantitative PCR was then performed using an ABI 7500 Fast Sequence Detection system (Applied Biosystems) and validated Taqman assays for the endogenous *Id4* transcript and ribosomal protein S2 (*Rps2*) transcript for normalization. The assay involved a probe to exon 1 of the endogenous *Id4* transcript where the *eGfp* coding sequencing is inserted for the *Id4-eGfp* transgene. Relative *Id4* transcript abundance was calculated using the formula for $2^{-\Delta\Delta CT}$.

*Testis histology and immunofluorescent staining*

Testes were fixed in either 4% PFA or Bouin's solution, embedded in paraffin, and 5-7 μm cross-sections were mounted on glass slides for hematoxylin and eosin staining or immunostaining for ZBTB16, STRA8, TRA98 or EGFP. Primary antibodies and the dilutions used are listed in Table S2. Immunostained cross-sections were visualized by fluorescent microscopy and digital images capture with a DP72 digital camera using CellSense software (Olympus Inc.).

*RNA sequencing analysis*

Populations of ID4-EGFP$^{Bright}$ and ID4-EGFP$^{Dim}$ spermatogonia were isolated by FACS from P8 mice as described above. Total RNA was extracted from 1-5x$10^5$ cells using Trizol reagent and treated with DNase I. Next, cDNA libraries were generated using oligo d(T) priming and sequencing was performed using an Illumina HiSeq2500 by the Genomics Core Service at WSU. For each sample (n=3 of both cell populations), 30-46 million reads of 100bp in length were generated and aligned to the mouse genome (mm9 build) using TopHat version 2.1.1. All confidently mapped reads were used to generate fragments per kilobase of exon per million fragments mapped (FPKM) values for each transcript using Cufflinks. To determine significant differences in gene expression between the spermatogonial subsets, stringent cutoff of $p<0.05$ and a false discovery rate (q-value) of <0.01 was applied using cuff.diff (One Array, Phalanx Biotech Group). Gene ontology analysis was conducted on list of differentially expressed genes using the Database for Annotation, Visualization and Integrated Discovery (DAVID).

**Table S1.** Donor-derived colonies of spermatogenesis derived by transplanting 10 ID4-EGFP$^{Bright}$ cells per recipient testis.

| Donor Cell Preparation | Recipient Mouse ID | Cells Injected | Colonies |
|---|---|---|---|
| 1 | 517 | 10 | 0 |
| | 519 | 10 | 0 |
| | 521 | 10 | 0 |
| | 522 | 10 | 0 |
| 2 | 650 | 10 | 2 |
| | 650 | 10 | 0 |
| | 651 | 10 | 0 |
| | 656 | 10 | 0 |
| 3 | 712 | 10 | 3 |
| | 716 | 10 | 0 |
| | 718 | 10 | 1 |
| Total | 11 | 110 | 6 |

**Table S2.** List of primary antibodies used for immunostaining analyses.

| Antibody | Manufacturer | Dilution Used |
|---|---|---|
| Rabbit anti Human ZBTB16 | Santa Cruz Biotechnology (sc-22839) | 1:200 |
| Rabbit anti Mouse STRA8 | Dr. Michael Griswold Lab | 1:200 |
| Rat anti Mouse TRA98 | Abcam (ab82527) | 1:200 |
| Goat anti EGFP | Abcam (ab6662) | 1:200 |

**Table S3.** Lists of overlapping genes up-regulated in different spermatogonial subpopulations isolated from testes of pre-pubertal mice.

| **Genes Up-Regulated in ID4-EGFP$^{Bright}$, ID4-EGFP+/TSPAN8-Hi, and THY1+ populations** |
|---|
| *Lhx1, Mpzl2, Nefm, Etv5, Tcl1, Tubg2, Ret, Mmp9, Slc9a3r1, Capg, Sirpa, Morc1* |

| **Genes Up-Regulated in ID4-EGFP$^{Bright}$ and ID4-EGFP+/TSPAN8-Hi Spermatogonia** |
|---|
| *9930012K11Rik, Agpat4, Akr1b8, Aldh1a3 ,Ap1m2, Arrdc3 ,Atad2, Atp6v1b2, Bcl6b, Cbln2, Ccdc3, Cdc42ep3, Chn2, Cited2, Clip4, Cnot6l, Cpq, Dmrt2, Dnajc10, Dppa4, Dpysl2, Dusp6, Ecm1, Egr2, Esrp1, Fgf12, Fgf9, Fst, Fstl1, Fyn, G2e3, Gfpt2, Gfra1, Gjb2, Glb1l3, Gmpr, Gramd1a, Gstp1, H2afy2, Hdc, Hhex, Ica1, Id4, Ifnar2, Il1r2, Jdp2, Klf6, Krt18, Ldha, Lonp2, Lpar3, Lpcat1, Mcam, Med14, Mid1ip1, Mtap7d3, Mtus1, Nckap1, Ndufv1, Nlrp4f, Nrd1, Ocln, Odc1, Oit1, Optn, Parl, Parm1, Parp8, Pced1b, Pcp4l1, Pdha1, Phyh, Pip4k2a, Plb1, Plvap, Plxdc2, Pmaip1, Ppfibp2, Pqlc1, Prkab1, Prrx2, Pygl, Rasl11a, Rhebl1, Rpap3, Rspo2, Sdcbp2, Slc25a4, Slc38a1, Smoc2, Snap25, Sphk1, Stx3, T, Tagln2, Tgif1, Thnsl2, Tlr3, Tmem176b, Tnfsf4, Tpm1, Tspan8, Tuba1a, Ubald2, Upf3b, Usp3, Usp4, Usp44, Usp9x, Vsx1, Wif1* |

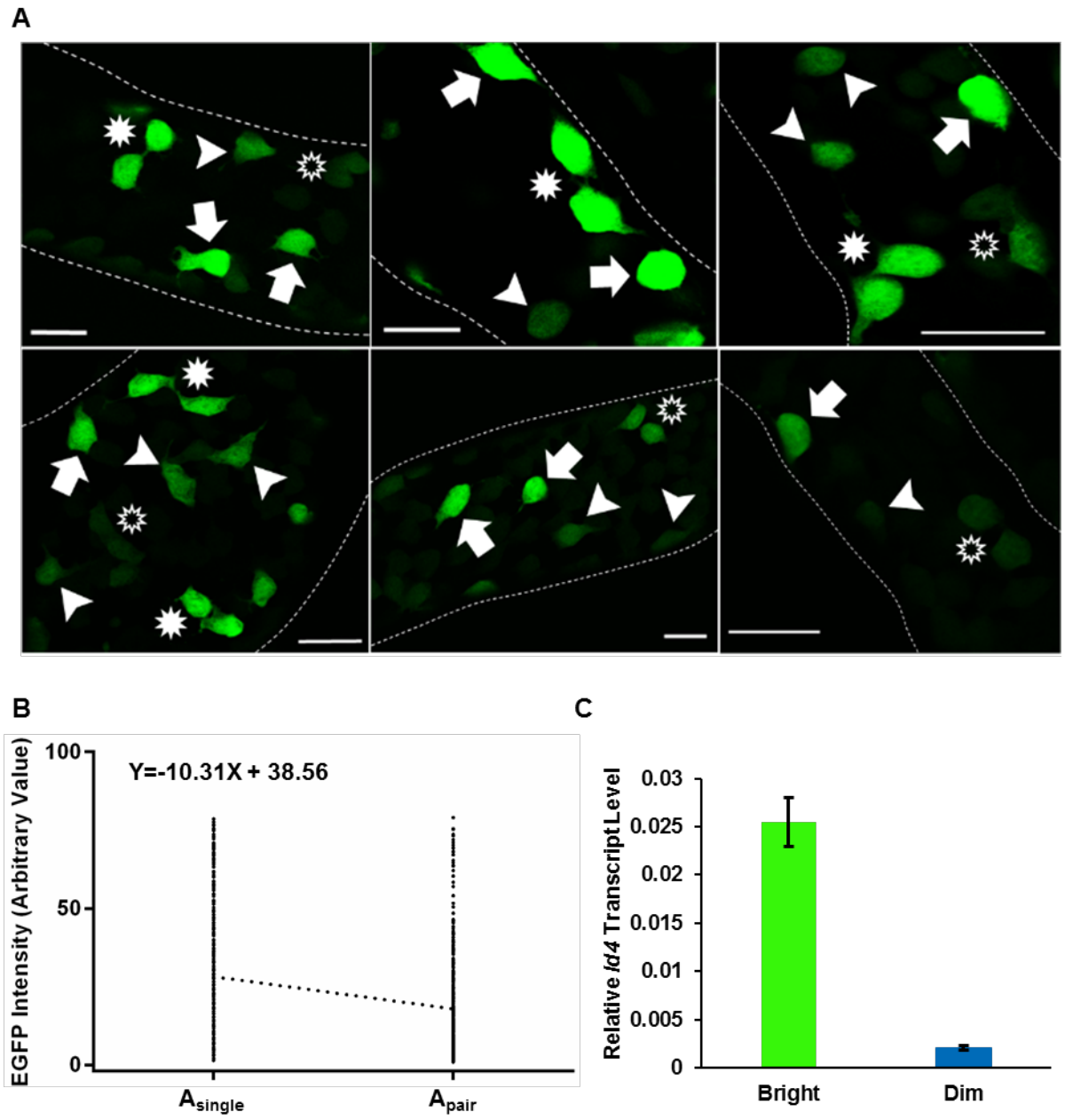


**Figure S1. Identification of spermatogonial subsets based on ID4-EGFP expression.** (A) Whole mount confocal images of live seminiferous from *Id4-eGfp* transgenic mice at postnatal day 8. EGFP+ cells are ID4 expressing spermatogonia. Arrows indicate $A_{single}$ with bright EGFP intensity (ID4-EGFP$^{Bright}$). Arrowheads indicate $A_{single}$ with dim EGFP intensity (ID4-EGFP$^{Dim}$). Closed stars indicate $A_{pair}$ with bright EGFP intensity. Open stars indicate $A_{pair}$ with dim EGFP intensity. Scale bars are 25 or 50 μm. (B) Linear regression analysis for EGFP intensity in $A_{single}$ and $A_{pair}$ spermatogonia. (C) qPCR analysis of endogenous Id4 transcript levels in ID4-EGFP$^{Bright}$ and ID4-EGFP$^{Dim}$ spermatogonial subsets.

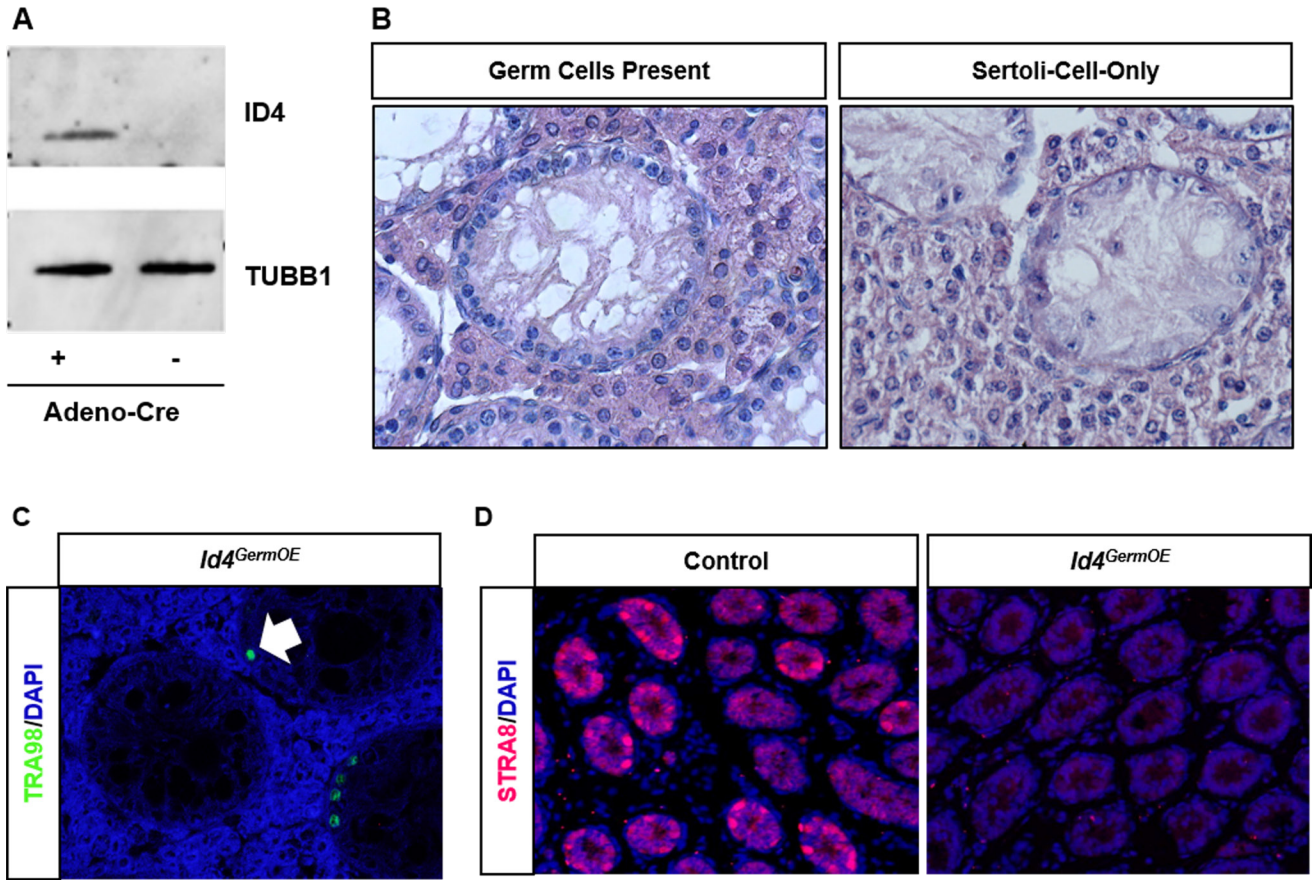


**Figure S2. Phenotype of male mice with constitutive overexpression of Id4 in the germline.** (A) Images of immunoblot analysis for ID4 expression in cultures of tail-tip fibroblasts from mice harboring an Id4 conditional overexpression transgene that were treat with or without Adenovirus expressing Cre recombinase. Tubulin-b (TUBB1) was used as a loading control. (B) Images of hematoxylin and eosin stained cross-sections from testes of a 6 month old male mouse with constitutive overexpression of Id4 in the germline. (C) Image of immunofluorescent staining for the pan germ cell marker TRA98 in a cross-section of seminiferous tubules from a 6 month old mouse with constitutive overexpression of Id4 in the germline. Arrow denotes a TRA98+ germ cell. (D) Images of immunofluorescence staining for the differentiating spermatogonial marker STRA8 in cross-sections of seminiferous tubules form a 10 day old mouse with constitutive overexpression of Id4 in the germline.

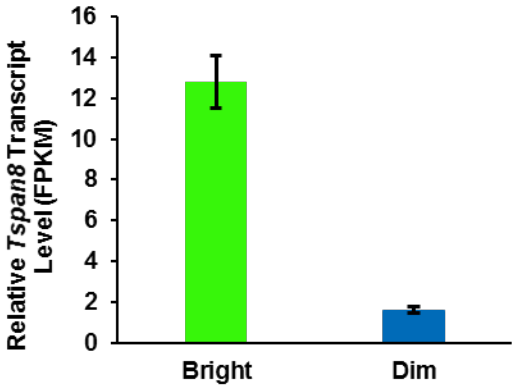


**Figure S3. Expression of Tspan8 in ID4-EGFP$^{Bright}$ and ID4-EGFP$^{Dim}$ spermatogonial subsets.** Data are FPKM values derived from RNA-seq analyses.

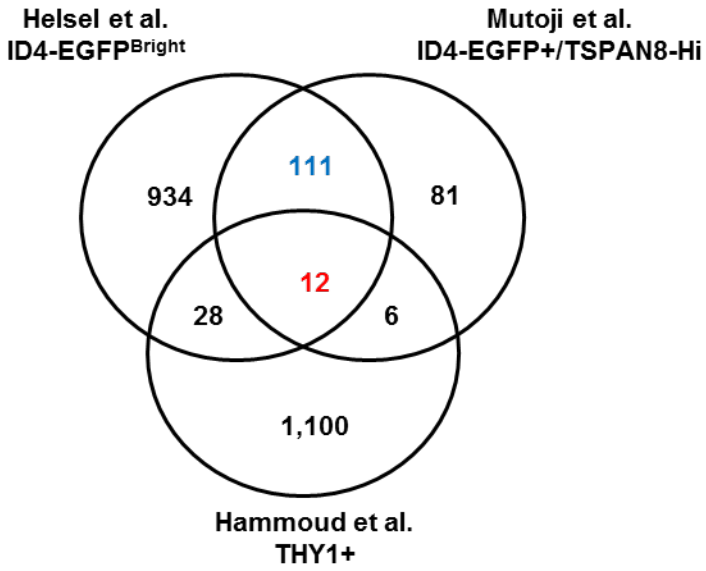


**Figure S4. Venn diagram for comparison of up-regulated genes in three different spermatogonial populations isolated from testes of pre-pubertal mice.** All data were generated by RNA-seq analysis. The data source for the ID4-EGFP$^{Bright}$ cells is the current study. The data sources for the ID4-EGFP+/TSPAN8-Hi cells and THY1+ cells are from a studies by (Mutoji et al., 2016) and (Hammoud et al., 2015), respectively.

## References

Hammoud, S. S., Low, D. H., Yi, C., Lee, C. L., Oatley, J. M., Payne, C. J., Carrell, D. T., Guccione, E. and Cairns, B. R. (2015) 'Transcription and imprinting dynamics in developing postnatal male germline stem cells', *Genes Dev* 29(21): 2312-24.

Mutoji, K., Singh, A., Nguyen, T., Gildersleeve, H., Kaucher, A. V., Oatley, M. J., Oatley, J. M., Velte, E. K., Geyer, C. B., Cheng, K. et al. (2016) 'TSPAN8 Expression Distinguishes Spermatogonial Stem Cells in the Prepubertal Mouse Testis', *Biol Reprod* 95(6): 117.

Riechmann, V., van Cruchten, I. and Sablitzky, F. (1994) 'The expression pattern of Id4, a novel dominant negative helix-loop-helix protein, is distinct from Id1, Id2 and Id3', *Nucleic Acids Res* 22(5): 749-55.

Schorpp, M., Jager, R., Schellander, K., Schenkel, J., Wagner, E. F., Weiher, H. and Angel, P. (1996) 'The human ubiquitin C promoter directs high ubiquitous expression of transgenes in mice', *Nucleic Acids Res* 24(9): 1787-8.

Zhang, Y., Riesterer, C., Ayrall, A. M., Sablitzky, F., Littlewood, T. D. and Reth, M. (1996) 'Inducible site-directed recombination in mouse embryonic stem cells', *Nucleic Acids Res* 24(4): 543-8.

## Database S1

Click here to Download Database S1

## Database S2

Click here to Download Database S2