Identification of slow-cycling germline stem cells and their regulation by PLZF

Manju Sharma[1], Anuj Srivastava[1], Heather E. Fairfield[1], David Bergstrom[1], William F. Flynn and Robert E Braun[1#]

[1]The Jackson Laboratory, 600 Main Street, Bar Harbor, ME, 04609, USA

To whom correspondence should be addressed

bob.braun@jax.org

207-288-6841

RUNNING TITLE

PLZF protects germline stem cells from proliferative exhaustion

KEYWORDS

Spermatogonial stem cell; SSC; EOMES; GDNF; GFRA1; PLZF

ABSTRACT

Long-term maintenance of spermatogenesis in mammals is supported by GDNF, an essential growth factor required for spermatogonial stem cell (SSC) self-renewal. Exploiting a transgenic GDNF overexpression model, which expands and normalizes the pool of undifferentiated spermatogonia between *Plzf* $^{+/+}$ and *Plzf* $^{lu/lu}$ mice, we used RNAseq to identify a rare subpopulation of cells that express EOMES, a T-box transcription factor. Lineage tracing and busulfan challenge show that these are long-term SSCs that contribute to steady state spermatogenesis as well as regeneration following chemical injury. EOMES+ SSCs have a lower proliferation index than EOMES− GFRA1+ spermatogonia in wild-type but not in *Plzf* $^{lu/lu}$ mice. This comparison demonstrates that PLZF regulates their proliferative activity and suggests that EOMES+ SSCs are lost through proliferative exhaustion in *Plzf* $^{lu/lu}$ mice. Single cell RNA sequencing of EOMES+ cells from *Plzf* $^{+/+}$ and *Plzf* $^{lu/lu}$ mice support a hierarchical model of both slow- and rapid-cycling SSCs.

INTRODUCTION

Fertility in males is supported by a robust stem cell system that allows for continuous sperm production throughout the reproductive life of the individual. In humans this lasts for decades and in a mouse can last for nearly its entire lifetime. However, despite more than a half century of research, and intensive investigation by many labs over the last decade, the identity of the germline stem cell continues to be elusive and controversial.

Stem cell function clearly resides in a subpopulation of spermatogonia within the basal compartment of the seminiferous tubules. In 1971, Huckins and Oakberg proposed the "A_s" model of spermatogonial stem cells (SSC) function where A_{single} (A_s) SSCs divide in a linear and non-reversible manner to populate the spermatogenic lineage (*1-3*). In this model, A_s spermatogonia are the SSCs, dividing symmetrically and with complete cytokinesis to form two daughter A_s cells for self-renewal, or dividing with incomplete cytokinesis to form A_{paired} (A_{pr}) cells, which are irreversibly committed to differentiation. A_{pr} cells, in turn, divide to form $A_{aligned}$ (A_{al}) spermatogonia, which exist as chains of 4, 8, or 16 interconnected cells. The A_s, A_{pr}, and A_{al} spermatogonia encompass the pool of undifferentiated spermatogonia that can be identified morphologically and are thought to share important functional properties distinct from the differentiated spermatogonia; A_1-A_4, intermediate and B (*4, 5*).

Contrary to the Huckins/Oakberg A_s model, recent studies on the behavior of pulse-labeled spermatogonia populations suggest that A_{pr} and A_{al} syncytia can fragment and revert to become A_s cells following transplantation as well as during steady state spermatogenesis (*6, 7*). This suggests that undifferentiated spermatogonia are not irreversibly committed to differentiation, allowing for an alternative mechanism for SSC self-renewal. Furthermore, A_s, A_{pr}, and A_{al} spermatogonia are characterized by heterogeneous gene expression (*6-11*) including recent descriptions of specific subpopulations of A_s cells with SSC activity expressing *Id4*, *Pax7*, *Bmi1* and *T* (*12-16*) (*17*). These new data do not

easily comport to a unifying model and imply that the mode of SSC function in the testes is more complex than the original Huckins-Oakberg A_s model suggests.

A_s and A_{pr} cells express GFRA1, a GPI-anchored receptor for glial cell-derived neurotrophic factor (GDNF) (*18-22*). GDNF is secreted by neighboring somatic Sertoli (*23*) and peritubular myoid (*24*) cells and is required for establishment and self-renewal of the SSC population in a dose-dependent manner (*23*). A decrease in GDNF levels results in germ cell loss, while overexpression of GDNF promotes accumulation of SSCs due to a block in differentiation (*23, 25*). PLZF, expressed in A_s, A_{pr} and A_{al} spermatogonia, is a transcription factor required for SSC maintenance, as mutation of *Plzf* results in age-dependent germ cell loss (*26, 27*). The mechanisms by which PLZF regulates SSC maintenance are not yet known. We describe here the identification of a rare subpopulation of A_s cells, which are slow-cycling and regulated by PLZF.

RESULTS

GDNF increases the undifferentiated spermatogonial population in *Plzf* mutants

Stage-specific temporal ectopic expression of GDNF in supporting Sertoli cells results in the accumulation of large clusters of tightly-packed PLZF+ undifferentiated spermatogonia (*25*). To determine whether overexpression of GDNF could rescue germ cell loss in the *luxoid* (*lu*) mutant, we generated

Tg(Ctsl-Gdnf)^{1Reb}; *Plzf* $^{lu/lu}$ mice (referred to as *Tg(Gdnf);lu/lu*). While *lu/lu* mice are smaller than wild-type (WT), we found no significant difference in body weight between age-matched *lu/lu* and *Tg(Gdnf);lu/lu* mice (Fig. 1A). However, at four months of age, testis weight was significantly higher in *Tg(Gdnf);lu/lu* mice compared to *lu/lu* ($p<0.001$), although it was still lower than in *Tg(Gdnf)* ($p<0.0001$) animals (Fig. 1B). Periodic acid-Schiff staining of testes sections showed fewer agametic tubules (referred to as Sertoli cell only) in *Tg(Gdnf);lu/lu* mice compared to *lu/lu* at both 4 and 6 months of age (Fig. 1C and D). Increased testis weight in *Tg(GDNF);lu/lu* mice could therefore be due to an increase in the number of cells occupying individual tubules, reflected by a decrease in the number of tubules with a Sertoli cell only phenotype, and fewer tubules with loss of one or more cell populations at 6 months of age (Fig 1D).

To determine what germ cell populations were expanded in *Tg(Gdnf);lu/lu* testes, we immuno-stained both whole-mount tubules and sections for spermatogonia markers. Large clusters of GFRA1+ cells were observed in *Tg(Gdnf);lu/lu* tubules, suggesting an increased number of the earliest spermatogonia cell types, A_s and A_{pr} (Fig. S1A). Quantitation of LIN28A staining, which is detected in a subpopulation of A_s, and A_{pr}, A_{al} and A1 spermatogonia (*28*), on 12-week old testes sections (Fig. S1B) showed that 75% of *Tg(Gdnf);lu/lu* tubule cross-sections had >5 LIN28A+ cells per tubule compared to 30% in *lu/lu* (Fig. S1C, $p<0.001$). Only 8% of tubules lacked LIN28A-labeled cells compared to 30% in *lu/lu* testes ($p<0.01$) (Fig. S1C). To determine whether the A_1 to A_4 differentiated spermatogonia population increased in *Tg(Gdnf);lu/lu*

tubules, we immunostained for the marker SOHLH1, which is not expressed in PLZF+ A_s to A_{al} cells (Fig. S1D). Similar to LIN28A, quantification of sectioned tubules revealed a significant increase in the percent of tubules with >5 SOHLH1+ cells (Fig. S1E, F).

Using quantitative reverse transcriptase PCR (qRT-PCR), we compared mRNA levels of several known genes expressed in undifferentiated spermatogonia, using total RNA from 4 month-old *lu/lu* and *Tg(Gdnf);lu/lu* testes. Similar to our immunostaining results, we observed a significant increase in the undifferentiated spermatogonia markers *Gfra1*, *Lin28a* and *Sohlh1* in *Tg(Gdnf);lu/lu* compared to *lu/lu* (Fig. 1E). We also saw an increase in *Ret*, a gene that is part of the GDNF-GFRA1 complex. Thus, temporal ectopic expression of GDNF significantly increases the numbers of tubules with both undifferentiated and differentiated spermatogonia, leading to a partial rescue of testis weight in *lu/lu* mice.

Unbiased discovery of genes whose expression is altered in *Plzf* mutants

Seminiferous tubules from *Tg(Gdnf)* mice have an expanded GFRA1+ population and enhanced ability to repopulate the spermatogonial lineage in testes transplants compared to WT (*25*). Because *Tg(Gdnf)* and *Tg(Gdnf);lu/lu* testes have a comparable number of GFRA1+ spermatogonia (Fig. 1F), we reasoned that these mice could be used as a valuable tool for the discovery of genes involved in self-renewal. Using RNAseq, we compared the transcriptomes

of whole testes from *Tg(Gdnf)* and *Tg(Gdnf);lu/lu* mice and searched for genes with significant fold-changes in expression. Of 20,000 gene transcripts, ~1,500 were differentially expressed with an absolute $\log_2$ fold change >1.5 (FDR<0.05). Of this subset, 150 were upregulated two-fold or more, and 60 were down-regulated two- to seven-fold in *luxoid* mutants.

As expected, many previously known gene transcripts specific to undifferentiated spermatogonia had no significant change in expression between genotypes including: *Gfra1*, *Ret, Lin28a, Utf1, Nanos2, Nanos3, Sohlh2,* and *Sall4* (Fig. 2A), confirming that the undifferentiated spermatogonial populations were similar between the two genotypes. In the group of genes that were significantly down-regulated in *Tg(Gdnf);lu/lu* testes, we identified a number of genes that are involved in the maintenance of pluripotent stem cell fate or early embryonic development (Fig. 2B) including: *Pou5f1*, *Brachyury (T)* and *Eomes*. *Pou5f1*, essential for maintaining pluripotency of ES cells and for maintaining viability of the mammalian germline (*29, 30*), was down-regulated two-fold in *luxoid* testes overexpressing GDNF. *Brachyury (T)*, down-regulated seven-fold in *Tg(Gdnf);lu/lu* testes, is a T-box transcription factor important for germ cell development that when overexpressed results in testicular germ-cell tumors (*31, 32*). *Eomes*, a T-box transcription factor that was down-regulated 3-fold in *Tg(Gdnf);lu/lu* testes, is known to play a critical role in embryonic development (*33, 34*). Notably, the read counts per million (RCPM) for the two highly down-regulated genes shown, *Eomes* and *T*, were very low in either genotype and found at the distant end of the expression spectrum amongst 20,000 transcripts

(Fig. 2C, far right). *Eomes* had a value of only 1.4 RCPM in *Tg(Gdnf)* testis RNA compared to 5,509 RCPM for *Prm2*, a gene expressed in spermatids. This suggests that *Eomes* and *T* are targets of PLZF and are down-regulated in *lu/lu* mice, or that T and EOMES are expressed in a small subset of germ cells that are specifically reduced in the *lu/lu* mutant, possibly within a restricted population of A_s cells.

We next validated candidate genes at different developmental time points by RT-PCR. Using total testis RNA from *Tg(Gdnf)* and *Tg(Gdnf);lu/lu* mice, *Eomes* was reduced in *lu/lu* testes as early as 3 days after birth, and became barely detectable at 4 and 6 months of age (Fig. 2D). As mutations in *Plzf* lead to age-dependent germ cell loss, *Eomes* was a compelling candidate marking a functional SSC pool that is depleted in *luxoid* mutants.

EOMES is detected in a subpopulation of GFRA1+ cells in the testis

We first determined whether EOMES was expressed within the spermatogonia population of the testis. We immuno-stained whole mount WT seminiferous tubules for EOMES together with GFRA1, which marks A_s, A_{pr} and small chains of A_{al} spermatogonia (Fig. 2E, F, arrows). EOMES was expressed in 12% of the total GFRA1+ A_s population of cells (Fig. 2G). Fourteen percent of GFRA1+ A_{pr} clones also expressed EOMES+ in either one (5%) or both (9%) of the cells (Fig. 2F1, 2F2, and 2F3, arrows, and G). The clones counted as A_{pr} could be two lineage-independent, but adjacent, A_s cells, A_{pr} cells about to

184 complete cytokinesis to form two new A_s cells, or bona fide A_{pr} cells that will go
185 on to form a chain of A_{al} cells. In rare cases (4%) EOMES was detected in some
186 or all of the individual cells of GFRA1+ chains of A_{al} cells (Fig. 2F4, G and figure
187 legend). In all instances, EOMES was co-expressed with PLZF (Fig. 2E, arrows),
188 which marks all undifferentiated spermatogonia. EOMES was not detected in
189 LIN28A-expressing cells (Fig. 2E), which we detect in a subset of A_s cells and in
190 A_{pr} and A_{al} cells (*28*).

191 Because the EOMES+ population also expressed GFRA1, we
192 hypothesized that these cells might be under GDNF influence (*23*). We therefore
193 asked whether Sertoli-specific overexpression of GDNF affected this cell
194 population. In whole-mount *Tg(Gdnf)* tubules we detected EOMES+ cells in
195 clusters of PLZF+ and GFRA1+ cells (Fig. 2H, arrows). LIN28A stained only the
196 peripheries of these cell clusters, and was specifically excluded from the
197 EOMES+ cores (Fig. 2H, bottom row, asterisks). These data indicate that
198 EOMES primarily marks a subpopulation of cells that are GFRA1+, PLZF+ and
199 LIN28A-.

200 To determine if EOMES-expressing cells are totally absent in *lu/lu* testes,
201 we immuno-stained *lu/lu* tubules for EOMES and GFRA1. Despite the significant
202 decrease in *Eomes* RNA in whole testes, EOMES+ cells were detected in *lu/lu*
203 mutants, (Fig. 2I and I', arrows), suggesting that PLZF regulates the pool of
204 EOMES+ cells, but does not directly regulate EOMES expression.
205

206 **EOMES+ cells contribute to steady-state spermatogenesis**

One of the features of stem cells is their ability to maintain their population throughout adult life while giving rise to differentiating progeny. To study the stem cell behavior of EOMES expressing cells, we generated an inducible *iCreERt2/2A/tdTomato* knock-in allele (*Eomes* ^{iCreERt2}) at the *Eomes* locus to label and trace their progeny after tamoxifen treatment (Fig. 3A). Characterizing the mouse line for tdTOMATO expression, we observed single GFRA1+ tdTOMATO+ cells in whole mount seminiferous tubules from 4-week old males (Fig. 3B top row arrows). EOMES immunofluorescence marked the same cells as tdTOMATO (Fig. 3B', top row), and as with EOMES itself, tdTOMATO was detected in a subpopulation of GFRA1+ cells (Fig. 3B' bottom row, arrow). In testis sections, GFRA1+ tdTOMATO+ cells were detected within the seminiferous tubules (Fig. 3C - top) and along the basement membrane juxtaposed to peritubular myoid cells (Fig. 3C – bottom, arrowheads).

Next, we traced the progeny of *Eomes* ^{iCreERt2} cells during normal steady-state spermatogenesis after tamoxifen-induced recombination using the *R26R-confetti* reporter mice. Mice were fed a tamoxifen-enriched diet for 2 weeks and clone sizes were determined by the number of cells expressing a single fluorescent protein after recombination at the *R26R-confetti* locus and detected by fluorescence imaging of whole mount seminiferous tubules (Fig. 3D). One color clones indicate derivation from a single-labeled cell. We observed confetti-marked single cells and chains of 4, 8 and 16 by two weeks (Fig 3D, top row). The size of labeled clones increased with time (Fig. 3D, second and third row)

and large labeled clones of differentiated germ cells were observed at 4 and 7 months after *iCre* activation (Fig. 3D). Importantly, we observed single-labeled cells at all time points analyzed (Fig. 3D, arrows), suggesting the single-labeled cells were either maintained for the entire duration or had undergone self-renewal and migrated away from its sister cells during the time the clones were traced. The labeled clones gave rise to labeled sperm detected in the epididymis of these mice (Fig. 3E), and fluorescence positive pups were born to *Eomes-*^{iCreERt2}*: R26R-confetti* mice treated with tamoxifen when mated with WT C57BL6/J females (Fig. 3F). Labeled clones, or offspring, were not observed in the absence of TAM at any time points (Fig. S2). Together, these results demonstrate that EOMES-expressing cells contribute to long-term maintenance of A_{single} spermatogonia and to full-lineage differentiation to spermatozoa in mice.

EOMES+ cells are resistant to busulfan treatment

To determine if EOMES+ cells are capable of regeneration after chemical insult, we specifically killed proliferating spermatogonia with busulfan and observed how EOMES+ and EOMES− cells survived the insult. We chose a low dose (10mg/kg body weight) that specifically kills proliferating A_{pr} and A_{al} spermatogonia (*35*). Seminiferous tubules from treated mice were immunostained for EOMES and PLZF every 3 days between 3 and 15 days after busulfan injection (Fig. 4A). During the entire 15-day time period the EOMES+ population remained unchanged (Fig. 4B). In contrast, by 9 days after treatment,

the PLZF+ population had halved due to loss of A_{pr} and A_{al} cells. The population began to recover after 12 days (Fig. 4B). We also stained for EOMES and GFRA1. At day 0, ~20% of the total GFRA1-expressing cell population was EOMES+ (Fig. 4C). Comparison within the GFRA1+ population showed a selective increase of the EOMES+ fraction after busulfan treatment (Fig. 4C), suggesting that the EOMES− fraction was specifically lost.

To assess if EOMES+ cells contribute to regeneration after chemical injury, 3wk old *Rosa26confetti: Eomes-*^{iCreERt2} mice were placed on a TM diet for 12 days to activate CreERt2, treated with a low dose of busulfan (10mg/kg body weight) on day 12, and then analyzed for labeled clones on day 45 (Fig. 4D). Whole mount seminiferous tubules showed labeled clones at varying stages of differentiation, supporting the conclusion that EOMES+ GFRA1+ PLZF+ spermatogonia are resistant to busulfan insult compared to their EOMES− counterparts, and that the EOMES+ population reconstitutes the lineage after chemical injury.

EOMES+ A_s cells are slow-cycling spermatogonia

Proliferating cells are more susceptible to busulfan insult because the damaging alkylating activity crosslinks DNA during replication. We hypothesized that the resistance of the EOMES+ population to busulfan might be due to a difference in cell cycle rate compared to other spermatogonial populations. To test this, we used 5-ethynyl-2'-deoxyuridine (EdU) labeling to identify cells in S-

phase, coupled with GFRA1 or EOMES immunostaining 24 hours after EdU injection (Fig. 5A). In WT tubules, only 10% of EOMES+ cells were labeled with EdU compared to 24% of the entire GFRA1+ population (Fig. 5B), suggesting that the EOMES+ subpopulation of GFRA1+ cells has a lower proliferation index.

In *luxoid* testes, we observed that EOMES+ cell numbers were greatly reduced. We speculated that progressive germ cell loss in the absence of *Plzf* could be due to an alteration in SSC cycling rates resulting in SSC exhaustion. To test this hypothesis, we injected EdU into *luxoid* mice and probed for cells in S-phase in conjunction with EOMES and GFRA1 immunostaining. Although fewer EOMES+ cells were present in *lu/lu* testes compared to WT, a greater fraction of the EOMES+ cells had incorporated EdU compared to WT (36% vs. 10%, Fig. 5A and B). Additionally, EdU labeling was significantly increased for the entire GFRA1+ population, but to a lesser degree (35% vs. 24%, Fig. 5B). These results indicate that EOMES+ cells have a higher than normal proliferative index in the absence of PLZF, leading to an overall increase in the number of total cycling GFRA1+ A_S spermatogonia.

To determine if the fraction of cycling EOMES+ cells changes during regeneration, mice were injected with EdU at 9 or 12 days post busulfan injection and assayed by immunofluorescence in whole mount seminiferous tubules. At both time points the EdU labeling index was increased in the EOMES+ and GFRA1+ populations (Fig. 5C) suggesting that busulfan is either directly inducing proliferation or the cells are responding to a decrease in SSC density.

Single cell RNA sequencing of EOMES $^{tdTOMATO+}$ cells

299

300 The expression or function of a growing collection of genes have been
301 linked to either SSC self-renewal or progenitor cell expansion (Table 1). To
302 address if these SSC identity genes are expressed in EOMES+ cells, we
303 performed single cell RNA sequencing (scRNAseq) on 897 fluorescently
304 activated cell sorted (FACS) EOMES $^{tdTOMATO+}$ cells from 4wk old *Eomes-*
305 *iCreERt2-tdTomato ; Plzf* $^{+/+}$ and 352 EOMES $^{tdTOMATO+}$ cells from *Eomes-*
306 *iCreERt2-tdTomato ; Plzf* $^{lu/lu}$ mice (Fig. 6A). The mean number of expressed
307 genes per cell in *Plzf* $^{+/+}$ and *Plzf* $^{lu/lu}$ was 3104 and 3230, while the mean number
308 of transcripts with unique molecular identifiers was 7254 and 8134, respectively
309 (Fig. S3).
310 An expression heat-map of the 11 identity genes indicates that the majority of the
311 cells express *Gfra1*, *Bmi1*, *Id4* and *Plzf*, while fewer cells express *T*, *Eomes*,
312 *Nanos*, *Neuorg3* and *Pax7 (*Fig. 6B*)*. The most striking finding is the considerable
313 degree of heterogeneity in expression across the entire scRNAseq data set.
314 None of the cells expressed all of the genes, while 90% (807/897) of cells
315 expressed at least one of the genes and 73% (655/897) expressed two or more
316 of the genes. *Gfra1* was detected in 566/897 cells (63%), while PAX7 was
317 detected in only 10/897 cells (1.11%). *Neurog3* (*36*), which is expressed in early
318 progenitor cells, was detected at relatively low levels in 27/897 cells (3.01%),
319 suggesting that *Neurog3* is expressed in some SSCs, or that the sorted cells
320 include early progenitor cells. *Eomes* transcripts were detected in 85/897 (9.48%)
321 EOMES $^{tdTOMATO+}$ cells. The low number of cells could be due to the inherent low

322 expression of *Eomes* relative to other expressed genes, early progenitor cells
323 that are tdTOMATO positive but are no longer expressing *Eomes* transcript, or
324 the cell cycle stage in which the cells were collected. Together, the
325 heterogeneity in transcript detection in the scRNAseq data mirrors previously
326 reported heterogeneity in protein expression as detected by immunofluorescence
327 (*8*).

328 To determine the relationship between the cells expressing the identify
329 genes, cluster analysis was performed on the 100 most differentially expressed
330 genes, plus 10/11 identity genes, from *Plzf* $^{+/+}$ sorted cells*. Pax7* was excluded
331 from the analysis because its normalized mean expression was too low to be
332 meaningful. The bulk of the cells clustered together (cluster 2) as a group of 740
333 cells (Fig. 6C). The remaining cells were defined by a separate group of 5
334 different smaller clusters. Cells assigned to cluster 0 could not be confidently
335 assigned to any cluster. The expression of *Dazl,* a marker of germ cells (*37*), at
336 moderate to high levels in all clusters, confirms that each cluster represents germ
337 cells (Fig. 6C). The expression of *Ldhc*, which is highly restricted to germ cells
338 and whose protein is first detected in early pachytene spermatocytes (*38*), is
339 expressed at highest levels in clusters 3-5, which are represented by a few cells,
340 suggesting that *Ldhc* mRNA is expressed at low levels in SSCs or early
341 progenitor cells. Although *Gfra1* was detected in all clusters, its highest mean
342 expression was in cluster 1, as was *Id4* and *Bmi* (Fig. 6D). Although cluster 1
343 contained only 13 cells, *Eomes*, *T* and *Nanos2* were expressed in a higher
344 proportion of cells (e.g. 23%-38% in cluster 1 compared to 9%-15% in cluster 2)

than in any other cluster. Nonetheless, the presence of multiple clusters expressing different identity genes suggests that EOMES $^{tdTOMATO+}$ sorted cells represent a collection of cells with different transcriptional profiles and supports the conclusion that SSCs are transcriptionally heterogeneous.

To determine the impact of loss of *Plzf* on the expression of the identity genes, we computed the mean expression level of each gene across all cells in *Plzf* $^{+/+}$ and *Plzf* $^{lu/lu}$ mice. In *Plzf* $^{lu/lu}$ mice, the mean expression of the identity genes *Bmi1*, *Id4, Pax7, Nanos2* and *Neurog3* were unchanged compared to *Plzf* $^{+/+}$ mice, while the mean expression of *Gfra1*, *Eomes* and T were significantly decreased (Fig. 6B and Table 2). In *Plzf* $^{lu/lu}$ mutants, the fraction of cells expressing *Gfra1* was reduced from 63.10% in *Plzf* $^{+/+}$ cells to 38.07% in *Plzf* $^{lu/lu}$ mutants. *Eomes* and *T* were only expressed in (1.14%) of *Plzf* $^{lu/lu}$ mutant cells, compared to 9.48% and 14.94%, in *Plzf* $^{+/+}$ cells, respectively, suggesting that mutation of *Plzf* differentially affects a subpopulation of SSCs expressing *Gfra1*, *T* and *Eomes*.

Given that ID4-EGFP has previously been reported to be restricted to a subpopulation of A_s cells (*39*), that it has been proposed to mark the ultimate pool of SSCs (*39*), and that we did not detect significant changes in *Id4* transcript in EOMES $^{tdTOMATO+}$ sorted cells in *Plzf* $^{lu/lu}$ mutants, we investigated the expression of ID4 by immunofluorescence. Using a rabbit polyclonal antibody, which detects a protein of the correct molecular weight by western blotting in extracts prepared from WT testes but not in extracts from *Id4* $^{-/-}$ testes (Fig. 7A), we detected ID4 in PLZF+ A_s, A_{pr}, and A_{al} cells (Fig. 7C). Because this staining

pattern differed from that previously reported for an *Id4-EGFP* transgene, we performed double-staining for ID4 and LIN28A. We again found that we could detect ID4 in chains of A_{al} cells that were also positive for LIN28A (Fig. 7D). ID4 immunostaining was not detected in seminiferous tubules of two different *Id4* $^{-/-}$ mutants (Fig. 7E). We assayed for co-expression of tdTOMATO and ID4 and found that 40% of the tdTOMATO+ A_s cells were also positive for ID4 (Fig 7F, upper row). We again detected ID4 in chains of A_{al} cells (Fig. 7F, bottom row), none of which were positive for tdTOMATO. In aggregate, we determined that only 20% of ID4+ cells were positive for tdTOMATO. To determine if *Id4* was required for expression of EOMES, we performed immunofluorescence on testis from *Id4*$^{-/-}$ mutants. EOMES was detected in GFRA1+ cells (Fig. 7G), consistent with the detection of *Eomes* RNA by rtPCR in *Id4*$^{-/-}$ mutants (Fig. 7B). We conclude that ID4 is expressed in a subpopulation of EOMES+ cells, but the majority (80%) of the ID4+ population is EOMES-.

DISCUSSION

Exploiting a transgenic overexpression model for GDNF in *Plzf* $^{+/+}$ and *Plzf* $^{lu/lu}$ mice, we discovered a rare population of EOMES-expressing spermatogonia. Lineage tracing studies during steady state spermatogenesis and during regeneration demonstrated that EOMES+ cells are SSCs. EOMES+ cells cycle more slowly than the bulk of GFRA1+ cells, and in *Plzf* $^{lu/lu}$ mice, both populations cycle more frequently. The significant loss of *Eomes*- and *T*- expressing cells

391 between *Plzf* $^{+/+}$ and *Plzf* $^{lu/lu}$ suggest that EOMES and T represent a
392 subpopulation of GFRA1+ cells that are uniquely depleted in *Plzf* $^{lu/lu}$ mice.
393 The concept of a slow-cycling or long-term SSC in rodents, referred to as
394 the A_0 "reserve stem cell", was first proposed 50 years ago (*40*). In retrospect,
395 the authors were most likely distinguishing between rapidly proliferating A1-A4
396 spermatogonia and undifferentiated A_s-A_{al} spermatogonia. The A_0 cell was never
397 discovered, and the field turned towards the 'A_s' model, which stipulated that A_s
398 cells were the functional SSC population that both self-renewed and proliferated
399 irreversibly into differentiated cell subtypes (*2, 41*). Early 3[H]-thymidine
400 incorporation studies hinted at the existence of both rapid-cycling and long-
401 cycling A_s SSCs (*42*). However, similar studies in Chinese hamsters did not
402 support the earlier studies in rats (*43*).
403 With our discovery of slow-cycling EOMES+ cells under PLZF-mediated
404 proliferative control, we now suggest that the A_s population is heterogeneous in
405 its cycling status. In our working model, there are two subtypes of SSCs under
406 steady-state conditions: a slow-cycling ($A_{s\text{-}SC}$) long-term stem cell that expresses
407 EOMES and a more rapid-cycling ($A_{s\text{-}RC}$) short-term stem cell that is EOMES
408 negative. Both populations express GFRA1 and PLZF. This functional hierarchy
409 ensures long-term tissue homeostasis of the stem cell compartment and
410 continued spermatogenesis. We propose that during steady state conditions, the
411 $A_{s\text{-}RC}$ cells divide regularly to both self-renew and differentiate. Although these
412 cells populate the spermatogonia lineage and support the normal cycle of the
413 epithelium, we suggest their capacity is limited. Upon depletion or reduction of $A_{s\text{-}}$

414 $_{RC}$ cells in a given region of the epithelium, $A_{s\text{-}SC}$ cells are recruited and convert to
415 EOMES- $A_{s\text{-}RS}$ cells. Key to this model is the limitation of cycling in $A_{s\text{-}SC}$ cells by
416 PLZF, which prevents stem cell exhaustion and the progressive germ cell loss
417 observed in *luxoid* mutants. By extension, our data suggest that the failure of
418 *luxoid* mutant germ cells to successfully colonize after transplantation is a
419 consequence of their severely depleted population of EOMES+ $A_{s\text{-}SC}$ cells and
420 the altered proliferative index of the entire population of GFRA1+ cells. This
421 paradigm could extend to the regulation of germline stem cells in humans, as
422 PLZF is expressed in human SSCs (*44*), which include a proposed quiescent A_s
423 subpopulation.

424 Overexpression of GDNF led to clusters of tightly packed nests of
425 undifferentiated spermatogonia. Cells in the center of the nests tended to be
426 EOMES+ LIN28A- while cells at the periphery were EOMES- LIN28A+. We
427 speculate that the significance of the geometry of the nests could be a
428 consequence of increased paracrine signaling, perhaps driven by intercellular
429 coupling of GDNF with its co-receptor GFRA1/RET, or the physiological state in
430 the center of the nests, that favors SSC self-renewal.

431 In addition to expression in A_s cells, EOMES was occasionally detected in
432 one of two adjacent GFRA1+ cells. The clones, which we counted as A_{pr}, could
433 have also been two lineage-independent, but adjacent, A_s cells. Alternatively, the
434 two cells could be daughter cells, suggesting that EOMES expression can be
435 asymmetrically inherited. Interestingly, in rare cases EOMES was expressed
436 heterogeneously in chains of GFRA1+ A_{al} cells, again suggesting that the fate of

437 cells with clones of A_{al} is not uniform and supporting the possibility that cells
438 within chains are subject to reversion to SSCs following fragmentation (*6*).
439
440 **Relation of EOMES cells to other SSC markers**
441
442 Single-cell RNA sequencing of EOMES $^{tdTOMATO+}$ cells from *Plzf* $^{lu/lu}$
443 mutants suggests that EOMES-expressing A_s cells constitute a unique population
444 of A_s cells. In recent studies, other GFRA1+ A_s subpopulations have also been
445 identified. *Pax7* lineage-marked cells are capable of repopulating germ cell-free
446 testes in transplants and are resistant to busulfan insult. (*14*). However, unlike
447 EOMES+ cells, PAX7 cells are rapid cycling, and *Pax7* transcript was detected in
448 only 10/897 EOMES $^{tdTOMATO+}$ cells, suggesting the possibility that they
449 correspond to rapid-cycling EOMES- GFRA1+ A_s SSCs. A *Bmi1* ^{CreERT} allele has
450 also been shown to mark SSCs (*16*). We detected *Bmi1* transcript in EOMES
451 $^{tdTOMATO+}$ cells, and long-term labeling of SSCs has been observed in *Bmi1*
452 lineage tracing experiments, however, the mean expression of *Bmi1* was similar
453 between *Plzf* $^{+/+}$ and *Plzf* $^{lu/lu}$ cells, suggesting that *Bmi1* expression is not
454 restricted to EOMES+ cells.
455 An *Id4-gfp* allele also marks a subset of A_s spermatogonia (*12, 13, 15, 45*)
456 and recent studies have shown that transplantable SSC activity resides in the
457 ID4-GFP-High population (*39*). Because ~40% of EOMES+ cells are ID4+, and
458 *Id4* transcripts were co-expressed in 49 of the 85 *Eomes*+ cells, it is likely that
459 there is substantial overlap in the population of cells marked by EOMES, ID4-

GFP-High, and T, which has recently been shown to also be expressed in a subpopulation of A_s cells with SSC activity (*17*). However, our data suggest that the *Id4-gfp* transgene may not reflect the full expression pattern of the ID4 protein itself. Although we detected ID4 in about 40% of the EOMES+ cells, the majority of the ID4+ cells (80%) were EOMES- and included A_{pr} and A_{al} cells. Differences in half-lives of the ID4 and EGFP proteins, or differences in translational control elements in their respective mRNAs, may be responsible for the discordance in their expression.

At face value our results are inconsistent with the fragmentation model for SSC self-renewal (*6, 7*). However, it is possible that the *Ngn3* and *GFRa1* lineage tracing studies preferentially labeled the equivalent of our EOMES-GFRA1+ rapid-cycling SSCs, which we propose function as short-term SSCs. It also remains a possibility that EOMES- A_s, A_{pr} and A_{al} cells can de-differentiate to EOMES+ A_s slow-cycling SSCs following fragmentation.

Function of PLZF

To date, it is unclear how PLZF mechanistically regulates SSC self-renewal, as it is expressed in the entire undifferentiated spermatogonia population. Several studies have provided *in vitro* evidence for a role of PLZF in cell cycle regulation. PLZF associates with CDC2 kinase activity *in vitro* (*46*) and when overexpressed in myeloid progenitors, PLZF decreases proliferation by reducing entry into S-phase (*47*). Furthermore, PLZF suppresses growth of

483 32Dcl3 cells *in vitro* by repressing transcription of the cyclin A2 gene and other
484 cyclin-dependent complexes involved in the G_1/S transition (*48, 49*). Our studies
485 show that PLZF is critical for maintaining SSCs in a low proliferative state and
486 suggest that it regulates the cycling status of SSCs. Analogous to the
487 hematopoietic system, which contains both slow-cycling long-term (LT)-HSCs,
488 and rapid-cycling short-term HSCs, (*50-52*), and where over-cycling of LT-HSCs
489 can lead to stem cell exhaustion and consequent loss of differentiated cell types
490 (*53*), we propose that loss of PLZF results in proliferative exhaustion of SSCs.
491

492 **Acknowledgements**

493 We are extremely grateful to Dr. Matthew Havrda for providing us with *Id4*$^{-/-}$ mice,
494 to Dr. Dayana Krawchuk for her help in manuscript preparation and to The Single
495 Cell Biology Lab at The Jackson Lab for Genomic Medicine.
496

497 **Author contributions**

498 M.S. designed, executed and wrote the manuscript. A.S. analyzed the RNAseq
499 data. H.E.F and D.B. contributed to experimental design and execution. W.F.F.
500 performed the scRNAseq analysis. R.E.B contributed to experimental design,
501 interpretation of the data and writing of the manuscript.
502

503 **Funding**

504 This work was supported by a grant from NICHD/NIH (HD042454UW, to R.E.B.)
505 and from NCI (CA34196, to The Jackson Laboratory) in support of The Jackson

506 Laboratory's shared services.

507

508 **References**

509 1. E. F. Oakberg, A new concept of spermatogonial stem-cell renewal in the
510 mouse and its relationship to genetic effects. *Mutat. Res.* **11**, 1-7 (1971).
511 2. E. F. Oakberg, Spermatogonial stem-cell renewal in the mouse. *Anat. Rec.* **169**,
512 515-531 (1971).
513 3. C. Huckins, Cell cycle properties of differentiating spermatogonia in adult
514 Sprague-Dawley rats. *Cell Tissue Kinet* **4**, 139-154 (1971).
515 4. H. Chiarini-Garcia, L. D. Russell, High-resolution light microscopic
516 characterization of mouse spermatogonia. *Biol. Reprod.* **65**, 1170-1178
517 (2001).
518 5. H. Chiarini-Garcia, L. D. Russell, Characterization of mouse spermatogonia by
519 transmission electron microscopy. *Reproduction* **123**, 567-577 (2002).
520 6. T. Nakagawa, M. Sharma, Y. Nabeshima, R. E. Braun, S. Yoshida, Functional
521 hierarchy and reversibility within the murine spermatogenic stem cell
522 compartment. *Science* **328**, 62-67 (2010).
523 7. K. Hara *et al.*, Mouse spermatogenic stem cells continually interconvert
524 between equipotent singly isolated and syncytial states. *Cell Stem Cell* **14**,
525 658-672 (2014).
526 8. H. Suzuki, A. Sada, S. Yoshida, Y. Saga, The heterogeneity of spermatogonia is
527 revealed by their topology and expression of marker proteins including the

germ cell-specific proteins Nanos2 and Nanos3. *Dev. Biol.* **336**, 222-231 (2009).

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

10. V. Barroca *et al.*, Mouse differentiating spermatogonia can generate germinal stem cells in vivo. *Nat Cell Biol* **11**, 190-196 (2009).

11. A. Sada, A. Suzuki, H. Suzuki, Y. Saga, The RNA-binding protein NANOS2 is required to maintain murine spermatogonial stem cells. *Science* **325**, 1394-1398 (2009).

12. F. Sablitzky *et al.*, Stage- and subcellular-specific expression of Id proteins in male germ and Sertoli cells implicates distinctive regulatory roles for Id proteins during meiosis, spermatogenesis, and Sertoli cell function. *Cell Growth Differ.* **9**, 1015-1024 (1998).

13. F. Chan *et al.*, Functional and molecular features of the Id4+ germline stem cell population in mouse testes. *Genes Dev.* **28**, 1351-1362 (2014).

14. G. M. Aloisio *et al.*, PAX7 expression defines germline stem cells in the adult testis. *J. Clin. Invest.* **124**, 3929-3944 (2014).

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

16. Y. Komai *et al.*, Bmi1 expression in long-term germ stem cells. *Sci Rep* **4**, 6175 (2014).

17. M. Tokue *et al.*, SHISA6 Confers Resistance to Differentiation-Promoting Wnt/beta-Catenin Signaling in Mouse Spermatogenic Stem Cells. *Stem Cell Reports* **8**, 561-575 (2017).

18. A. Buageaw *et al.*, GDNF family receptor alpha1 phenotype of spermatogonial stem cells in immature mouse testes. *Biol. Reprod.* **73**, 1011-1016 (2005).

19. C. K. Naughton, S. Jain, A. M. Strickland, A. Gupta, J. Milbrandt, Glial cell-line derived neurotrophic factor-mediated RET signaling regulates spermatogonial stem cell fate. *Biol. Reprod.* **74**, 314-321 (2006).

20. D. S. Johnston, E. Olivas, P. DiCandeloro, W. W. Wright, Stage-specific changes in GDNF expression by rat Sertoli cells: a possible regulator of the replication and differentiation of stem spermatogonia. *Biol. Reprod.* **85**, 763-769 (2011).

21. T. Sato *et al.*, In vitro production of functional sperm in cultured neonatal mouse testes. *Nature* **471**, 504-507 (2011).

22. M. Grasso *et al.*, Distribution of GFRA1-expressing spermatogonia in adult mouse testis. *Reproduction* **143**, 325-332 (2012).

23. X. Meng *et al.*, Regulation of cell fate decision of undifferentiated spermatogonia by GDNF. *Science* **287**, 1489-1493 (2000).

24. L. Y. Chen, W. D. Willis, E. M. Eddy, Targeting the Gdnf Gene in peritubular myoid cells disrupts undifferentiated spermatogonial cell development. *Proc Natl Acad Sci U S A* **113**, 1829-1834 (2016).

572 25. M. Sharma, R. E. Braun, Cyclical expression of GDNF is required for
573 spermatogonial stem cell homeostasis. *Development*, (2018).
574 26. F. W. Buaas *et al.*, Plzf is required in adult male germ cells for stem cell self-
575 renewal. *Nat. Genet.* **36**, 647-652 (2004).
576 27. J. A. Costoya *et al.*, Essential role of Plzf in maintenance of spermatogonial
577 stem cells. *Nat. Genet.* **36**, 653-659 (2004).
578 28. P. Chakraborty *et al.*, LIN28A marks the spermatogonial progenitor
579 population and regulates its cyclic expansion. *Stem Cells* **32**, 860-873 (2014).
580 29. J. Nichols *et al.*, Formation of pluripotent stem cells in the mammalian
581 embryo depends on the POU transcription factor OCT4. *Cell* **95**, 379-391
582 (1998).
583 30. J. Kehler *et al.*, Oct4 is required for primordial germ cell survival. *EMBO Rep* **5**,
584 1078-1083 (2004).
585 31. A. R. Sangoi *et al.*, Specificity of brachyury in the distinction of chordoma
586 from clear cell renal cell carcinoma and germ cell tumors: a study of 305
587 cases. *Mod. Pathol.* **24**, 425-429 (2011).
588 32. R. Tirabosco *et al.*, Brachyury expression in extra-axial skeletal and soft
589 tissue chordomas: a marker that distinguishes chordoma from mixed
590 tumor/myoepithelioma/parachordoma in soft tissue. *Am. J. Surg. Pathol.* **32**,
591 572-580 (2008).
592 33. I. Costello *et al.*, The T-box transcription factor Eomesodermin acts upstream
593 of Mesp1 to specify cardiac mesoderm during mouse gastrulation. *Nat Cell*
594 *Biol* **13**, 1084-1091 (2011).

595 34. S. J. Arnold, U. K. Hofmann, E. K. Bikoff, E. J. Robertson, Pivotal roles for
596 eomesodermin during axis formation, epithelium-to-mesenchyme transition
597 and endoderm specification in the mouse. *Development* **135**, 501-511 (2008).
598 35. C. J. van Keulen, D. G. de Rooij, Spermatogonial stem cell renewal in the
599 mouse. II. After cell loss. *Cell Tissue Kinet* **6**, 337-345 (1973).
600 36. S. Yoshida *et al.*, Neurogenin3 delineates the earliest stages of
601 spermatogenesis in the mouse testis. *Dev. Biol.* **269**, 447-458 (2004).
602 37. H. J. Cooke, M. Lee, S. Kerr, M. Ruggiu, A murine homologue of the human DAZ
603 gene is autosomal and expressed only in male and female gonads. *Hum. Mol.*
604 *Genet.* **5**, 513-516 (1996).
605 38. E. Goldberg, E. M. Eddy, C. Duan, F. Odet, LDHC: the ultimate testis-specific
606 gene. *J. Androl.* **31**, 86-94 (2010).
607 39. A. R. Helsel *et al.*, ID4 levels dictate the stem cell state in mouse
608 spermatogonia. *Development* **144**, 624-634 (2017).
609 40. Y. Clermont, E. Bustos-Obregon, Re-examination of spermatogonial renewal
610 in the rat by means of seminiferous tubules mounted "in toto". *Am J Anat* **122**,
611 237-247 (1968).
612 41. C. Huckins, The spermatogonial stem cell population in adult rats. I. Their
613 morphology, proliferation and maturation. *Anat. Rec.* **169**, 533-557 (1971).
614 42. C. Huckins, The spermatogonial stem cell population in adult rats. 3. Evidence
615 for a long-cycling population. *Cell Tissue Kinet* **4**, 335-349 (1971).

616 43. D. Lok, M. T. Jansen, D. G. de Rooij, Spermatogonial multiplication in the
617 Chinese hamster. IV. Search for long cycling stem cells. *Cell Tissue Kinet* **17**,
618 135-143 (1984).
619 44. B. P. Hermann, M. Sukhwani, M. C. Hansel, K. E. Orwig, Spermatogonial stem
620 cells in higher primates: are there differences from those in rodents?
621 *Reproduction* **139**, 479-493 (2010).
622 45. F. Sun, Q. Xu, D. Zhao, C. Degui Chen, Id4 Marks Spermatogonial Stem Cells in
623 the Mouse Testis. *Sci Rep* **5**, 17594 (2015).
624 46. H. J. Ball, A. Melnick, R. Shaknovich, R. A. Kohanski, J. D. Licht, The
625 promyelocytic leukemia zinc finger (PLZF) protein binds DNA in a high
626 molecular weight complex associated with cdc2 kinase. *Nucleic Acids Res.* **27**,
627 4106-4113 (1999).
628 47. S. Doulatov *et al.*, PLZF is a regulator of homeostatic and cytokine-induced
629 myeloid development. *Genes Dev.* **23**, 2076-2087 (2009).
630 48. R. Shaknovich *et al.*, The promyelocytic leukemia zinc finger protein affects
631 myeloid cell growth, differentiation, and apoptosis. *Mol. Cell. Biol.* **18**, 5533-
632 5545 (1998).
633 49. P. L. Yeyati *et al.*, Leukemia translocation protein PLZF inhibits cell growth
634 and expression of cyclin A. *Oncogene* **18**, 925-934 (1999).
635 50. S. H. Cheshier, S. J. Morrison, X. Liao, I. L. Weissman, In vivo proliferation and
636 cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proc*
637 *Natl Acad Sci U S A* **96**, 3120-3125 (1999).

638 51. S. J. Morrison, I. L. Weissman, The long-term repopulating subset of
639 hematopoietic stem cells is deterministic and isolatable by phenotype.
640 *Immunity* **1**, 661-673 (1994).
641 52. G. J. Spangrude, S. Heimfeld, I. L. Weissman, Purification and characterization
642 of mouse hematopoietic stem cells. *Science* **241**, 58-62 (1988).
643 53. F. Arai, T. Suda, Maintenance of quiescent hematopoietic stem cells in the
644 osteoblastic niche. *Ann. N. Y. Acad. Sci.* **1106**, 41-53 (2007).
645 54. L. Madisen *et al.*, A robust and high-throughput Cre reporting and
646 characterization system for the whole mouse brain. *Nat Neurosci* **13**, 133-
647 140 (2010).
648 55. K. Yun, A. Mantani, S. Garel, J. Rubenstein, M. A. Israel, Id4 regulates neural
649 progenitor proliferation and differentiation in vivo. *Development* **131**, 5441-
650 5448 (2004).
651 56. R. K. Patel, M. Jain, NGS QC Toolkit: a toolkit for quality control of next
652 generation sequencing data. *PLoS One* **7**, e30619 (2012).
653 57. D. Kim *et al.*, TopHat2: accurate alignment of transcriptomes in the presence
654 of insertions, deletions and gene fusions. *Genome Biol* **14**, R36 (2013).
655 58. D. W. Barnett, E. K. Garrison, A. R. Quinlan, M. P. Stromberg, G. T. Marth,
656 BamTools: a C++ API and toolkit for analyzing and managing BAM files.
657 *Bioinformatics* **27**, 1691-1692 (2011).
658 59. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package
659 for differential expression analysis of digital gene expression data.
660 *Bioinformatics* **26**, 139-140 (2010).

661

662

663 **Table 1: Genes reported to be expressed in SSCs or progenitor cells**

Gene	Expression		Functional Assay	References
Gfra1	A_s, A_{pr}, A_{al} (4-8)	*Gfra1-CreERT2*	Lineage tracing Transplant	(*7*)
Bmi1	A_s	*Bmi1CreERT*	Lineage tracing	(*16*)
Id4	A_s	*Id4-eGFP*	Transplant	(*15*)
Plzf	A_s, A_{pr}, A_{al} (4-16)	Antibody staining	Transplant	(*26*)
Lin28a	A_s, A_{pr}, and A_{al} (4-16)	Antibody staining	Transplant	(*28*)
T	A_s	*T(nEGFP-CreERT2)*	Lineage tracing	(*17*)
Eomes	A_s	*Eomes-iCre ERt2 tdTomato*	Lineage tracing	This manuscript
Nanos2	A_s, A_{pr}	Antibody staining	Knockout	(*11*)
Nanos3	A_s, A_{pr}, A_{al}	Antibody staining	Knockout	(*8*)
Neurog3	A_{al} (4-16)	*Ngn3 CreERT*	Lineage tracing	(*6*)
Pax7	A_s	*Pax7-CreERT2*	Lineage tracing	(*14*)

664
665

666

Table 2: Mean expression* of SSC identity transcripts in *Plzf* $^{+/+}$ and *Plzf* $^{lu/lu}$ mutant cells

Gene	*Plzf* $^{+/+}$ Mean	*Plzf* $^{lu/lu}$ Mean		P value	Log2 FC
Gfra1	**0.7**	**0.35**		**1.69E-06**	**0.7943**
Eomes	**0.06**	**0.004**		**6.68E-04**	**3.0602**
Plzf	**0.26**	**0.06**		**3.02E-11**	**1.9324**
Id4	**0.32**	**0.28**		**9.61E-01**	**0.0245**
Bmi1	**0.38**	**0.34**		**9.17E-01**	**0.0414**
Pax7	**0.007**	**0.012**		**6.70E-01**	**-0.8002**
T	**0.12**	**0.006**		**1.93E-07**	**4.5234**
Lin28a	**0.18**	**0.18**		**9.18E-01**	**-0.0648**
Nanos2	**0.14**	**0.14**		**8.92E-01**	**-0.1503**
Neurog3	**0.02**	**0.01**		**3.83E-01**	**1.2448**

* Mean expression = log(normalized expression + 1)

Figure Legends

Figure 1. Sertoli-cell overexpression of GDNF partially reverses the loss of undifferentiated spermatogonia in *luxoid* mice. (A) No difference was observed in body weight between *lu*/*lu* and *Tg*(*Gdnf)*;*lu*/*lu* male mice at 4 months of age. ns, not significant; ***, p <0.005; number in bar = n animals. (B) Significant differences were observed in testis weight between *lu*/*lu* and *Tg(Gdnf);lu/lu* animals. **, p<0.005; ***, p<0.0001; number in bar = n testes. (C) Representative images of periodic-acid-Schiff stained cross-sections of 4- and 6-month old testes showing agametic tubules (asterisks) in *lu*/*lu*. (D) Fewer tubules with germ cell loss were present at 6 months in *Tg(Gdnf):lu/lu*. A total of 200-300 tubules were counted for each genotype (n=3). P < 0.005. (E) qRT-PCR on testis RNA from 4-month old mice shows a significant increase in SSC gene expression in *lu/lu* mice overexpressing GDNF in Sertoli cells. Relative mRNA levels are normalized to a *β-actin* internal control. *, p<0.01. (n = 3) (F) Average number of GFRA1+ cells per 500μm length of tubules in *Tg(Gdnf)* and *Tg(Gdnf);lu/lu* in 10wk old mice (n=3).

Figure 2. EOMES marks a subpopulation of SSCs. (A) Relative gene expression (read counts per million, RCPM) of selected transcripts from RNAseq performed on testis RNA from 4-month old *Tg(Gdnf)* and *Tg(Gdnf);lu/lu* mice. No significant change is detected in known spermatogonial marker genes. (B) Fold change in RNAseq transcripts for a subset of pluripotency and developmental genes in *Tg(Gdnf);lu/lu* compared to *Tg*(*Gdnf)* testes. ***, p<0.0001. (C)

693 Relatively low RCPM of the highest-fold change transcripts shown in (B),
694 compared to the spermatid-expressed protamine 2 gene *Prm2*, and
695 differentiating spermatogonia marker *Stra8*. X-axis represents 20,000 genes
696 numbered in descending order by RCPM. (D) Semi-quantitative RT-PCR of total
697 testis RNA showing decreasing *Eomes* expression over time in *lu*/*lu* testes. (E)
698 Immunostaining for EOMES and undifferentiated spermatogonial markers in WT
699 whole-mount testes. EOMES marks some (arrows), but not all (asterisks),
700 GFRA1+ PLZF+ A_s cells. Weak EOMES staining can occasionally be detected in
701 chains of GFRA1+ A_{al} cells (arrowhead). EOMES+ A_s cells do not express
702 LIN28A (bottom row, arrows). (F) Examples of EOMES+ subset of GFRA1+ cells.
703 F1 shows an example of two adjacent cells where one is EOMES+ (arrow) and
704 one is EOMES- (asterisk). F2 and F2 illustrate two examples of two juxtaposed
705 cells that are both EOMES+. F4 illustrates an example of a chain of 6 A_{al} cells
706 with two EOMES+ (arrow) and four EOMES– (asterisks) cells. (G) Graph
707 showing the percent of EOMES+ cell in the GFRA1+ population clone types in
708 WT adult mice (631 GFRA1+ cells from 3 mice). n = number of clone types
709 counted (e.g. A_{pr} = 114 pairs of cells = 228 GFRA1+ total cells). Of the total
710 GFRA1+ A_{pr} population, in 9% of cases both cells were positive for EOMES,
711 while 5% had only one cell positive for EOMES. Within the A_{al} fraction, 4% of the
712 clones express EOMES in a heterogeneous pattern (two A_{al} chains where 3 of 3
713 were EOMES+; two A_{al} chains of 4 where all were EOMES+; one A_{al} chain of 4
714 where only 1 was EOMES+; one A_{al} chain of 6 where only 2 of 6 cells were
715 EOMES+). (H) EOMES is expressed in tightly packed GFRA1+ PLZF+ clusters

(arrows) found in *Tg*(*Gdnf)* whole-mount tubules. EOMES is not expressed in the LIN28A+ cells in the cortical region of the clusters (bottom row, asterisks). (I) Detection of EOMES+ GFRA1+ cells (arrows) in *Plzf* $^{lu/u}$ mutants. Asterisks mark an EOMES- GFRA1+ cell.

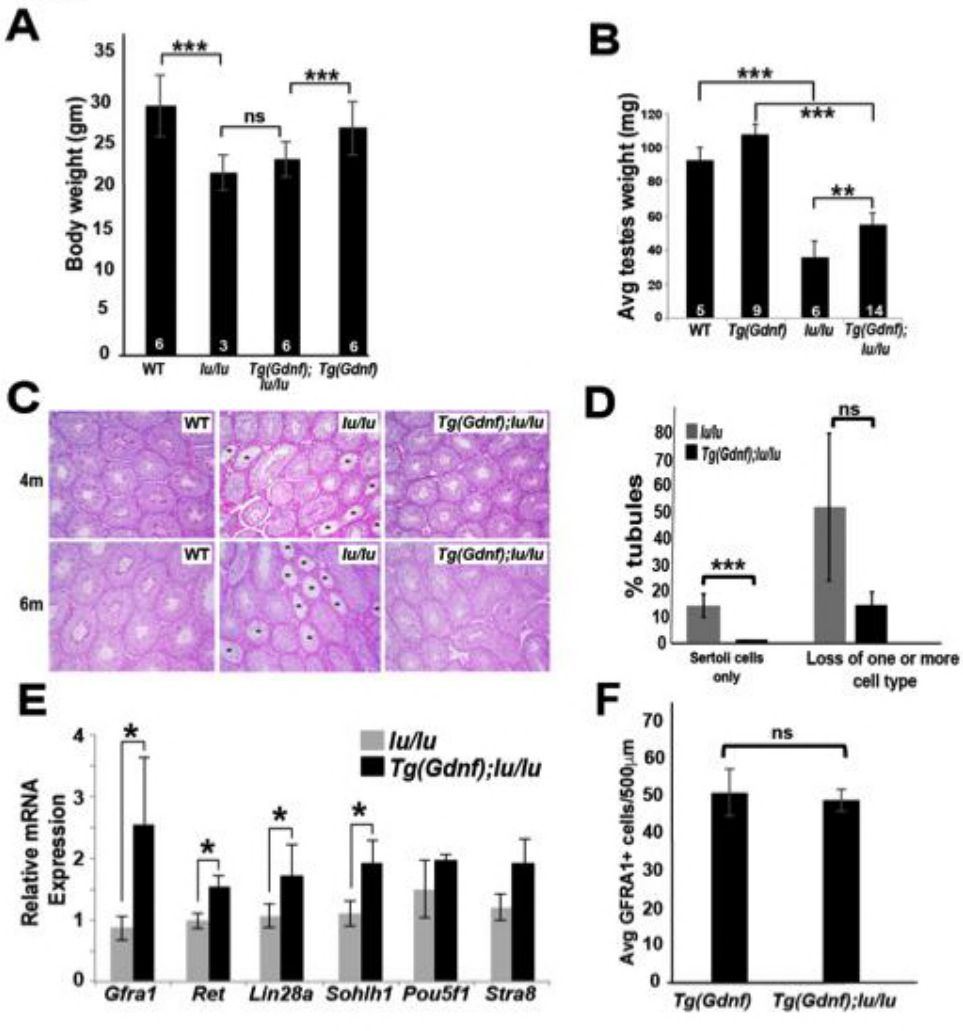
Figure 3. EOMES+ cells are long-term SSCs. (A) Schematic diagram of the *Eomes* locus on chromosome 9 and insertion cassette used to generate an *Eomes/2AiCreERT2/2A/tdTomato* knock-in allele. (B) GFRA1+ tdTomato+ A_s cell in testis of 4wk old mouse (arrow). GFRA1+ tdTOMATO- A_s cell (asterick). (B') tdTOMATO is co-expressed with EOMES (top row, arrow). tdTOMATO is expressed in GFRA1+ A_s cells but not in chains of GFRA1+ A_{al} cells (bottom row). (C) tdTOMATO cells (Red), positive for GFRA1 (Green), are present within the seminiferous tubules. Top - An optical section of confocal image showing the localization of EOMES+ cell along the basement membrane. Bottom – Magnified view of GFRA1+ EOMES $^{tdTOMATO+}$ cells shown above. Arrowheads mark myoid cell (M) nucleus, which are negative for EOMES ^{tdTOMATO}. * Background fluorescence in the interstitial space associated with GFRA1 staining. (D) Experimental flow chart showing the induction times and clonal analysis of *Eomes* ^{iCreERt2} targeted cells labeled in mice crossed to Rosa26 Confetti reporter and fed a tamoxifen-enriched diet. Fluorescence images of whole-mount seminiferous tubules showing the different size of labeled clones. Representative clones at 2 (n=5, animals), 3 (n=4), and 4wks (n=3), and 4 (n=3) and 7 (n=3) months after tamoxifen induction. (E) Images of labeled sperm from the

739 epididymis of mice 7months after tamoxifen induced recombination. (F)
740 Fluorescence positive offspring born to WT female mated with ROSA26confetti:
741 *Eomes-iCreERt2* 4months after tamoxifen-induced recombination at Confetti
742 locus by *Eomes-iCreERt2.*
743
744 **Figure 4. EOMES+ cells are resistant to busulfan and contribute in**
745 **regeneration.** (A) Whole-mount tubule immunostaining for EOMES and PLZF
746 after low-dosage busulfan treatment (10mg/mL). (B) Cell counts from (A) show
747 that EOMES+ cell numbers remain constant while PLZF+ cell numbers drop after
748 busulfan treatment. (n = 3 per time point, 8 weeks old) (C) The EOMES+ fraction
749 of GFRA1+ cells increased following busulfan treatment. **, p<0.005; ***,
750 p<0.0001. (D) Schedule to assay the regeneration properties of EOMES+ cells.
751 Three-week old *R26R-confetti*: *Eomes* ^{iCreERt2} mice were put on a tamoxifen-
752 enriched diet for 1 week, treated with busulfan, and analyzed on day 45. Images
753 of whole mount seminiferous tubules (n = 5 mice) show labeled clones at
754 different developmental stages during regeneration.
755
756 **Figure 5. EOMES+ cells have a low proliferative index.** (A) Whole-mount
757 seminiferous tubules from 8-week old WT and *lu/lu* mice injected with EdU and
758 immunostained 24 hours after injection. EdU+ EOMES+ or EdU+ GFRA1+ cells
759 (arrows); EdU- EOMES+ or EdU- GFRA1+ cells (asterisks). (B) Quantification of
760 the EdU labeling index in WT (n = 4 mice) and *lu/lu* (n = 5 mice) EOMES+ and
761 GFRA1+ cells. GFRA1+ cells have a higher labeling index than EOMES+ cells in

762 WT (p = 0.001). EOMES+ cells have a higher labeling index in *lu/lu* than in WT (p
763 = 0.001). GFRA1+ cells have a higher labeling index in lu/lu than in WT (p =
764 0.005). (C) Increase in proliferation rate of EOMES+ and GFRA1+ cells during
765 the recovery phase after busulfan treatment. EdU incorporation was assayed 24
766 hours post EdU injection (n = 3 mice per timepoint; *** p = 0.001; *p < 0.05).
767
768 **Figure 6. Single cell RNA sequencing of EOMES ^{tdTOMATO} cells.** (A) FACscan
769 profile of EOMES $^{tdTOMATO+}$ cells sorted from *Plzf* $^{+/+}$ and Plzf $^{lu/lu}$ mutant testis.
770 The boxed region in the upper right indicated cells with background
771 autofluorescene that were excluded from the analysis. (B) Heat map and
772 dendogram of SSC identity and progenitor genes expressed in EOMES ^{tdTOMATO}
773 cells from *Plzf* $^{+/+}$ and Plzf $^{lu/lu}$ mutant testis. (C) Clusters generated from WT
774 cells. Cellular expression profiles of 110 genes were embedded into a
775 2dimensional latent space using UMAP (see supplemental information for
776 details). Cells in cluster 0 could not be confidently placed in any of the other 6
777 clusters. (D) Violin plots showing the mean expression of each SSC identity gene
778 per cluster in cells sorted from *Plzf* $^{+/+}$ testes.
779
780 **FIGURE 7. ID4 is expressed in A_s, A_{pr} and A_{al} spermatogonia.** A**)** Western
781 blot of total testis protein lysate from 2 WT and 2 *Id4*$^{-/-}$ mutant mice probed with
782 ID4 antibody. For loading control, western blot was probed with α-tubulin
783 antibody. (B) RT-PCR using total RNA from two *Id4*$^{-/-}$ mutant and 2 WT mice for
784 *Eomes, Lin28a, Pou5f1 and β-actin*. (C) Detection of PLZF and ID4 by

785 immunofluorescence in whole mount seminiferous tubules. Both ID4 in PLZF+
786 were detected in $A_{s\ (arrow)}$, A_{pr}, and A_{al} spermatogonia in WT adult mouse testis.
787 (D) ID4 protein was co-localized in chains of LIN28A+ A_{al} spermatogonia (arrow).
788 (E) ID4+ cells are not detected in whole mount tubules from prepared from *Id4*$^{-/-}$
789 mutant mice. (F) Top row, examples of tdTOMATO/ID4 double-positive cells
790 (arrowhead), tdTOMATO+ ID4- (**), and two adjacent cells where one is
791 tdTOMATO+ and the other is tdTOMATO/ID4 double-positive (magenta arrow).
792 Bottom row, examples of tdTOMATO- ID4+ A_s and A_{al} chains of 4 and 8 cells.
793 (G) EOMES+ GFRA1+ cells are present in *Id4*$^{-/-}$ mutants (arrow).
794
795
796
797

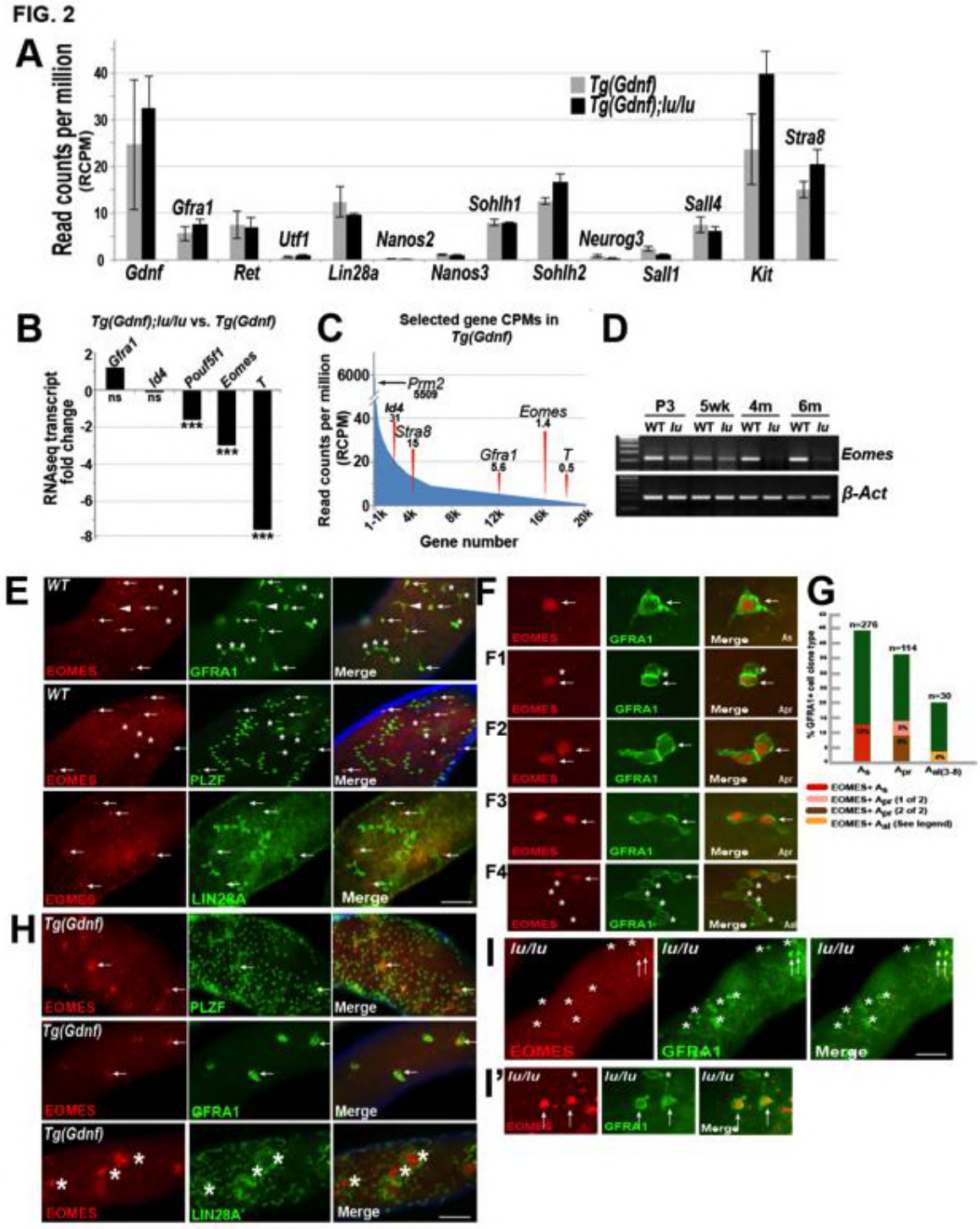
FIG. 1



798

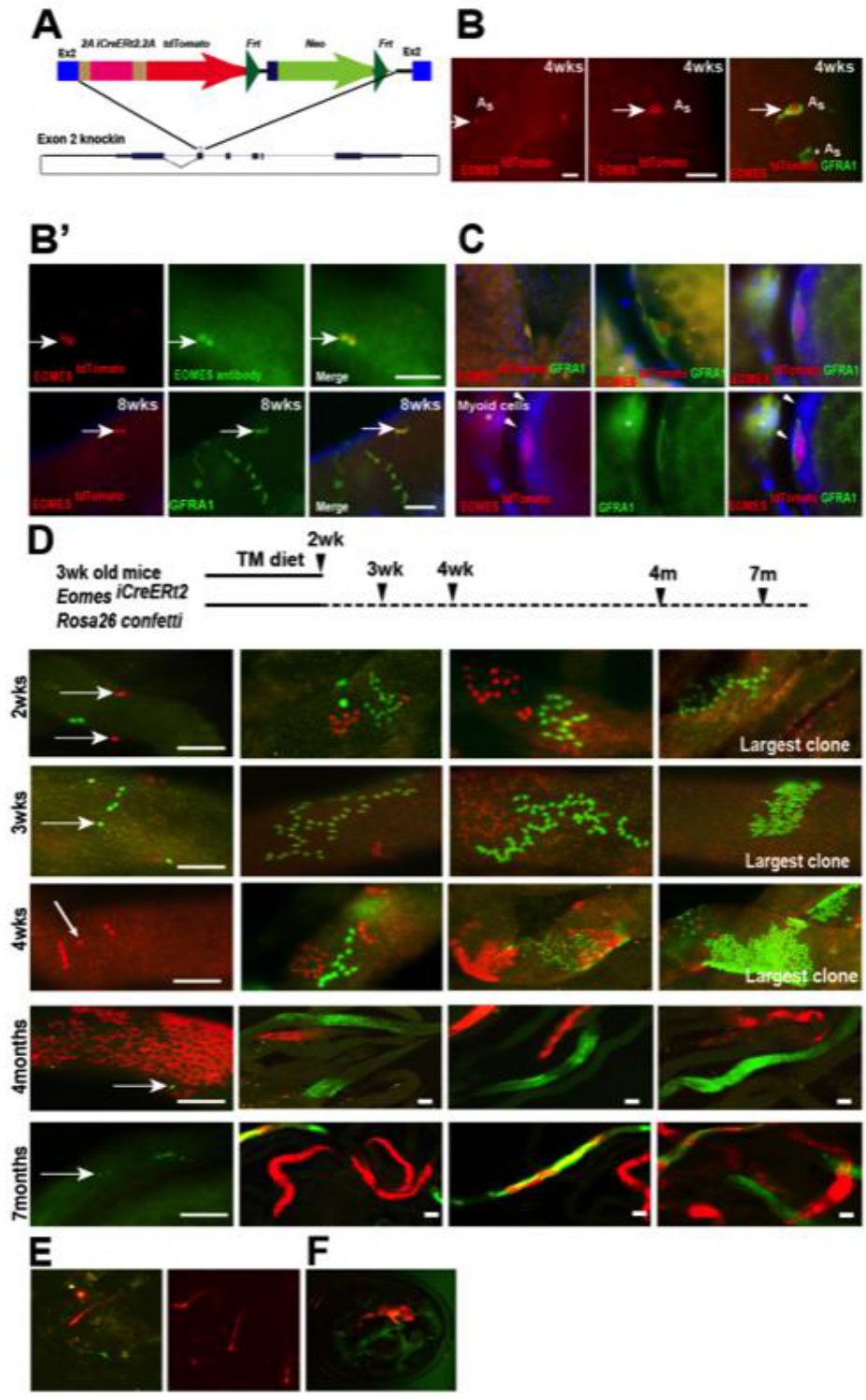
799
800
801
802
803

FIG. 2



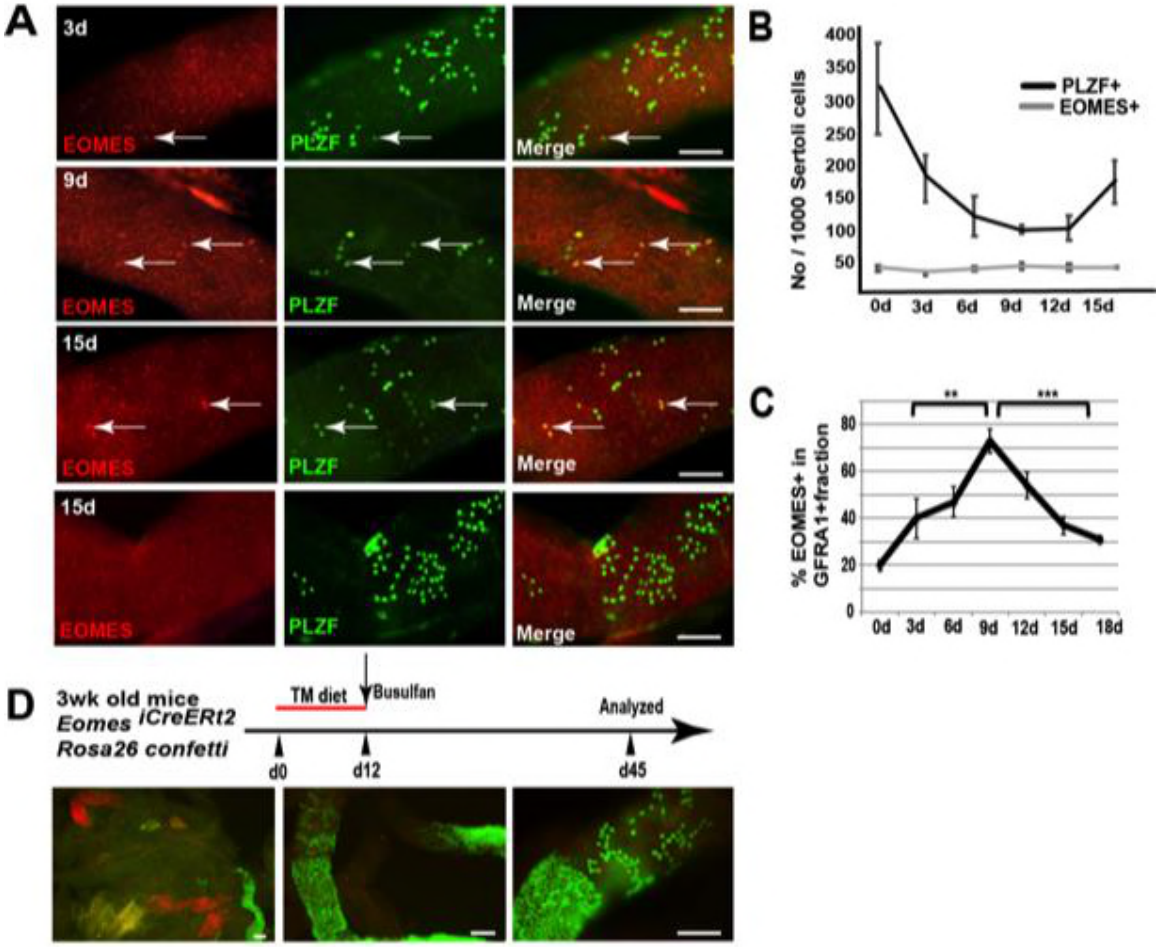
804
805

FIG. 3



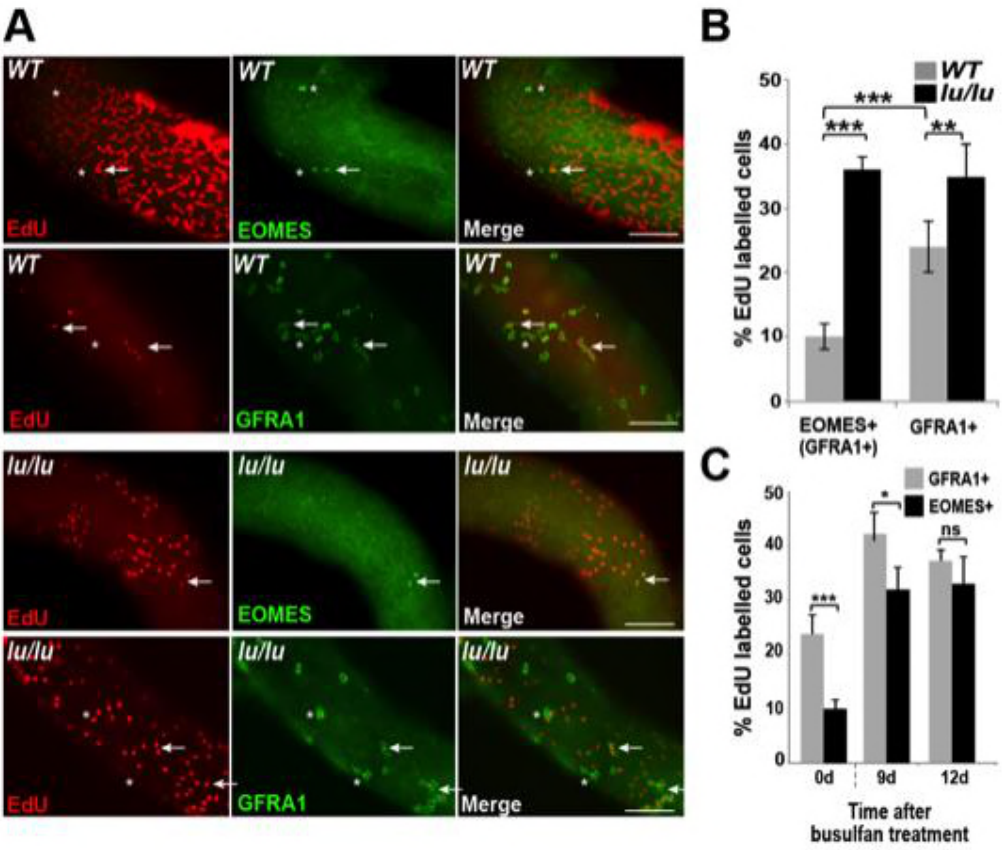
806

FIG. 4



807

FIG. 5



808
809
810
811

812
813

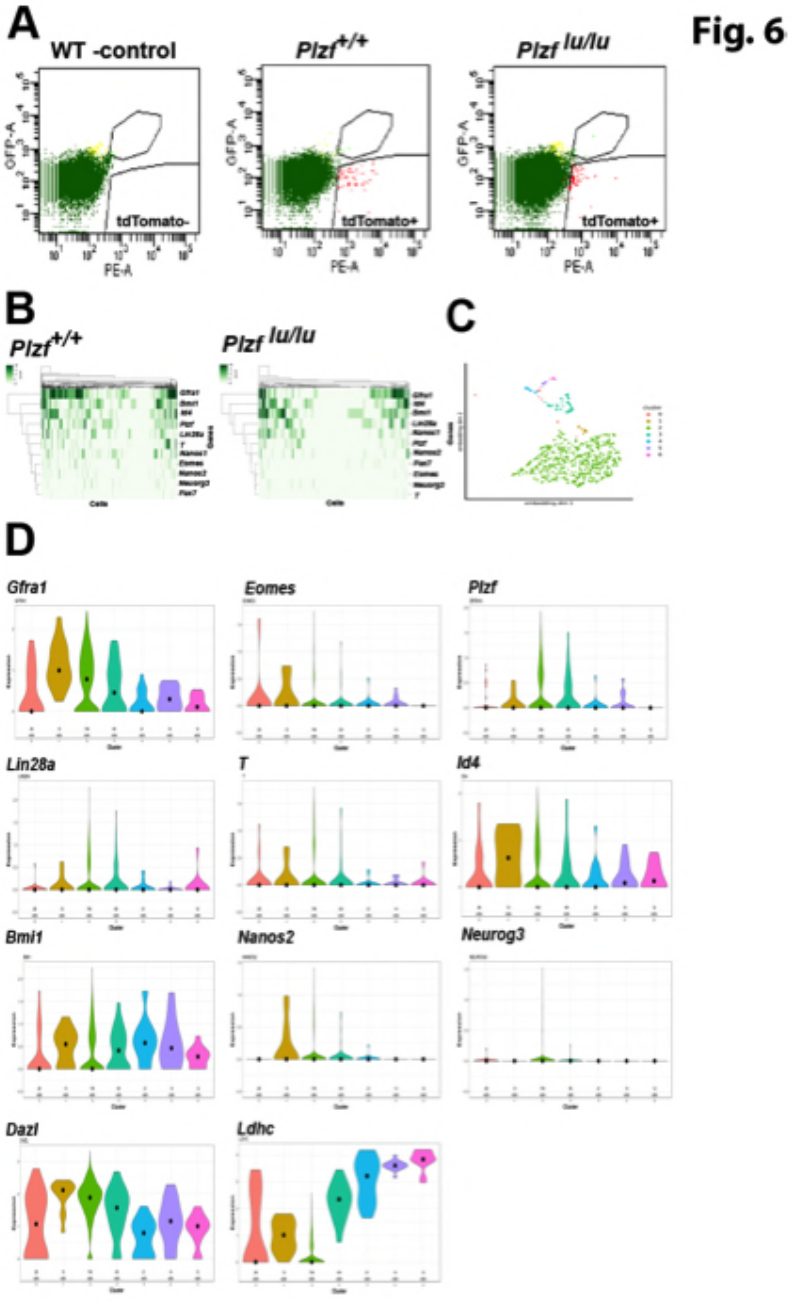


Fig. 6

814

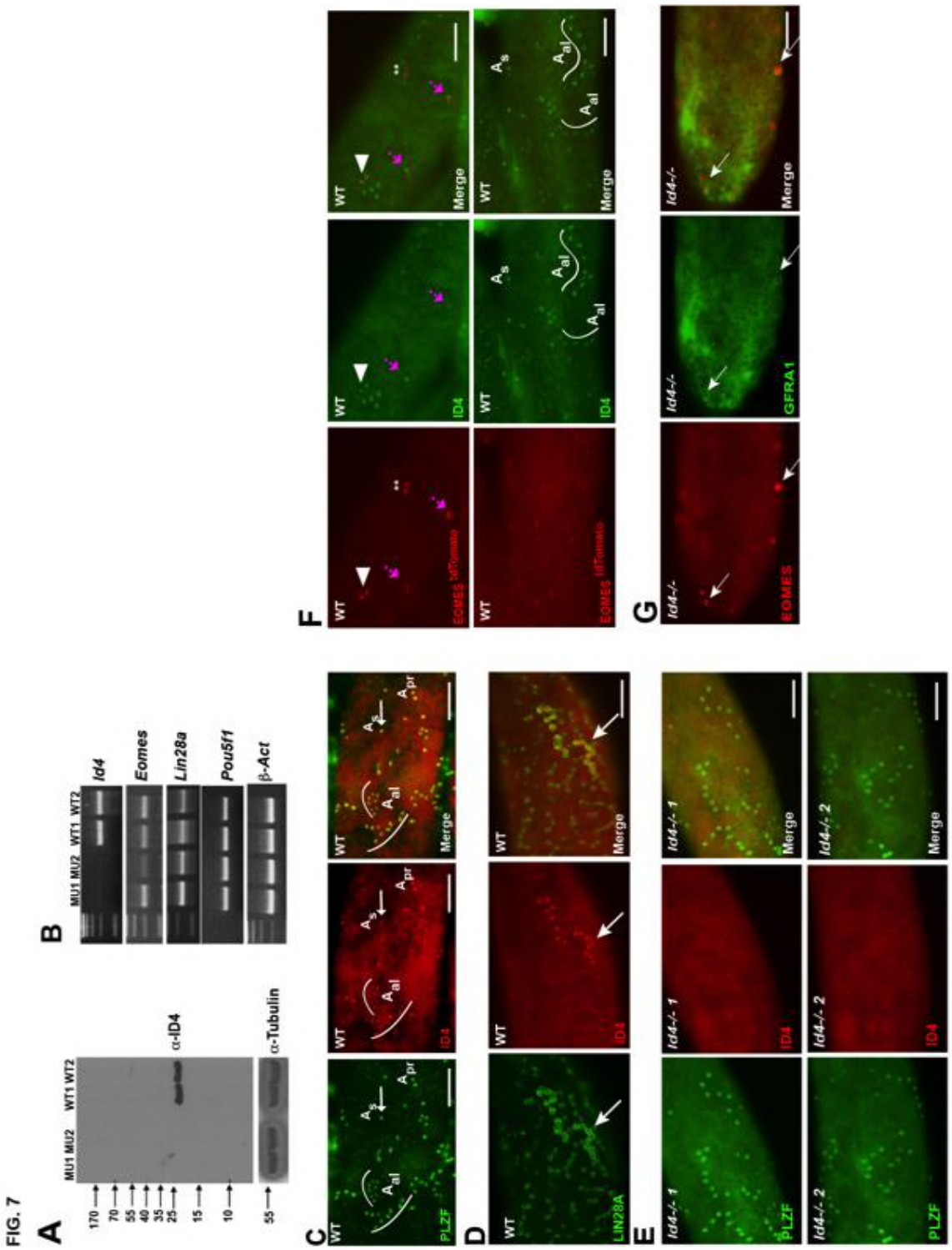
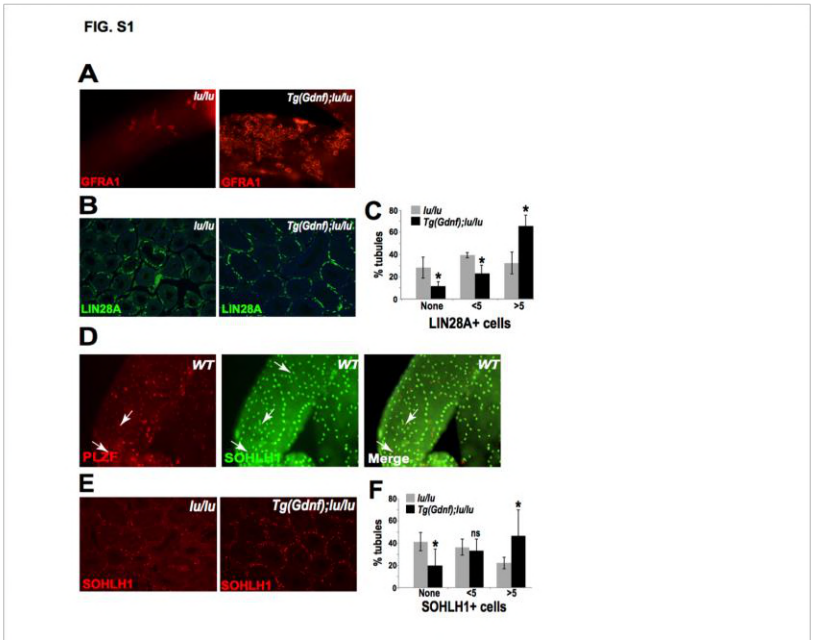


FIG. 7

Supplementary Information

Figure S1. (Related to Fig. 1) Sertoli-cell overexpression of GDNF increases the SSC population in *Plzf-mutant luxoid mice*. (A) Immunostained whole-mount tubules from 12-week old *Tg(Gdnf);lu/lu* mice show large clusters of GFRA1+ cells compared to *lu/lu*. (B) Representative LIN28A immunostaining of *lu/lu* and *Tg(Gdnf);lu/lu* tubule paraffin sections. (C) Quantification of immunostained sections as in (B) show 70% more tubules with 5 or more LIN28A+ cells in *Tg(Gdnf);lu/lu* compared to *lu/lu* (n=4). *, p<0.05. (D) In WT SOHOLH1 expression is detected in A_{al} cells transitioning to A1 and differentiating cells A2-A4, absent from A_s, A_{pr}, and A_{al} cells. (E) Representative SOHLH1 immunostaining of *lu/lu* and *Tg(Gdnf);lu/lu* tubule paraffin sections. (F) Quantification of immunostained sections as in (E) show a significant increase in the number of tubules with 5 or more SOHLH1+ cells in *Tg(Gdnf);lu/lu* compared to *lu/lu*. (n=4). *, p<0.05, ns, not significant.



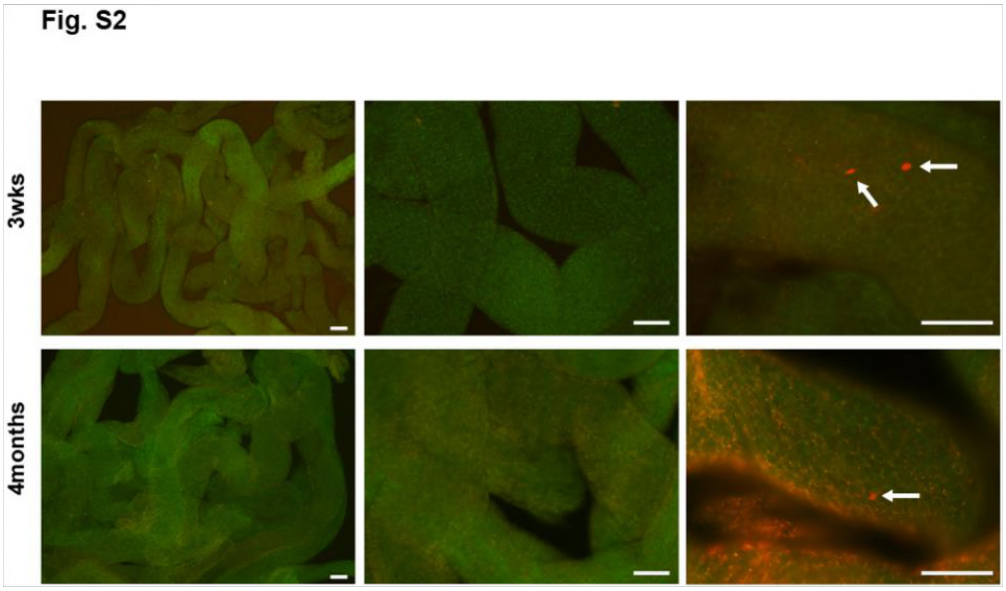
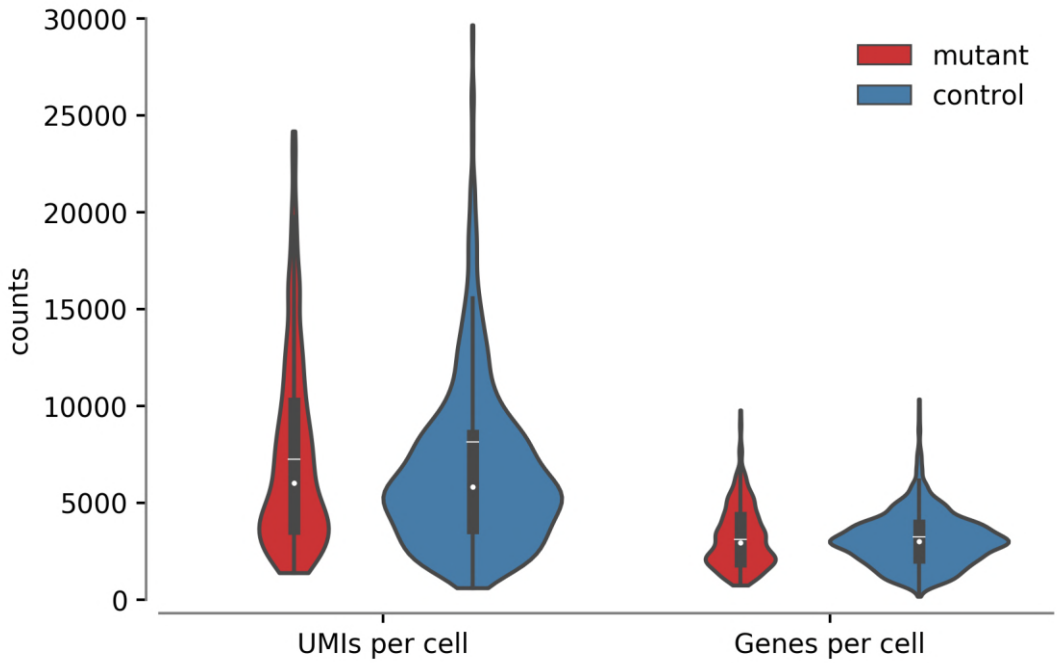


Figure S2. (Related to Fig. 3.) No TAM controls for lineage tracing of *Eomes*^{iCreERt2} targeted cells in mice crossed to Rosa26 Confetti reporter on a normal (no tamoxifen) diet. Representative images at 3wks and 4mos post completion of TM diet fed to experimental cohorts shown in Fig. 3. Arrows indicate tdTOMATO+ cells. Scale bar = 100μm



847

848 **Fig. S3 (Related to Fig. 6) Distribution of unique molecular identifiers (UMIs)**

849 **and genes per cell.** Median (white dot), mean (white line), and interquartile

850 range (dark box) are shown for each of the four distributions, with control (*Plzf* $^{+/+}$ *;*

851 *Eomes-iCreERt-tdTomato*+) in blue and mutant (*Plzf* $^{lu/lu}$ *; Eomes-iCreERt-*

852 *tdTomato*+) in red.

853

854

MATERIALS AND METHODS

Animals

For *Plzf* $^{lu/lu}$, we used the Green's luxoid or 'luxoid' (as referred to here) *Zbtb16*^{lu}/J allele from The Jackson Laboratory. B6.Cg-*Gt(ROSA)26Sor*$^{tm6(CAG-ZsGreen1)Hze}$/J (CAG(ZsGreen) mice (*54*) were obtained from the The Jackson Laboratory (JR # 007906). *Id4*$^{-/-}$ mice (*55*) were provided by Mathew Havrda (Dartmouth University). All animals were housed in a barrier facility at The Jackson Laboratory. All experimental procedures were approved by the Jackson Laboratory Institution Animal Care and Use Committee (ACUC) and were in accordance with accepted institutional and government policies.

Generation of Tamoxifen inducible *iCre ERt2-tdTomato* knock-in into the *Eomes* locus

The BAC clone RP23-235G22 was used to clone the *2A:CreERt2-FRT-PGK-Neor-FRT* cassette into exon 2 of mouse *Eomes* gene. Linearized vector was electroporated into B6N-JM8 ES cells from B6N mouse strain. Chimera were generated using B6(Cg)-$^{Tyr\ c\text{-}2J}$/J the albino mice (are C57BL/6J mice carrying a mutation in the tyrosinase gene). The chimeras were crossed to B6N parental strain to generated F1 progeny. The reporter mice used to assay for Cre recombination were *B6.129P2-Gt(ROSA) 26* $^{Sortm1(CAG-Brainbow2.1)Cle}$*/J* also known as R26R-confetti conditional mice stock number # 017492 from Jackson laboratory. For activation of *iCre* the mice were fed TD.130859 (TAM diet). The food was formulated for 400mg tamoxifen citrate per kg diet that would provide

878 ~40mg tamoxifen per kg body weight per day.

879 **Testes weight and histology**

880 Mice were genotyped with specific primer sets and assessed for body weight and
881 testes weight at each time point analyzed. For histological analysis, testes were
882 fixed overnight in Bouin's and paraffin-embedded sections were stained with PAS
883 after dewaxing.

884

885 **Antibodies**

886 Antibodies were purchased from the following companies. Mouse anti-PLZF,
887 antibody (Santa Cruz Biotech Inc., Santa Cruz, CA); Rat anti-GFRA1 (R&D
888 Systems, Minneapolis, MN); Rat anti-LIN28A (gift from Dr. Eric G. Moss, Dept. of
889 Mol. Biology, UMDNJ, New Jersey); Rabbit anti-SOHLH1 and anti-EOMES from
890 (Abcam, Cambridge, MA), and Rabbit monoclonal anti-mouse /human ID4
891 antibody (# BCH-9/#82-12) from BioCheck, Inc. Foster City, CA.

892

893 **Whole mount immunostaining**

894 Seminiferous tubules were dispersed in PBS by removing the tunica from the
895 testes. They were rinsed with PBS and fixed in 4% paraformaldehyde (Electron
896 Microscopy Sciences, Fort Washington, PA) for 4-6h at 4°C. For PLZF, LIN28A,
897 and SOHLH1 antibodies, the tubules were permeabilized with 0.25% NP-40 in
898 PBS-T (0.05% Tween in PBS) for 25 min at RT before blocking in 5% normal
899 goat serum. The tubules were then incubated with primary antibodies and
900 incubated overnight (GFRA1 and EOMES) at RT or 4°C for PLZF, LIN28, and

SOHLH1. The next day the tubules were washed three times with PBS-Tween20 for 5 min, and incubated with secondary antibodies conjugated to Alexa Fluors (Molecular Probes) for 1 h at room temperature. After washing in PBS-Tween20, the tubules were mounted in VectaShield® with DAPI (Vector Laboratories, Burlingame, CA) and imaged using a Nikon Eclipse E600 epifluorescence microscope equipped with an EXi Aqua Camera from Q-Imaging (Surrey, BC, Canada).

Proliferation assay

To assay the proliferative index of spermatogonial stem cells, mice were injected intra-peritoneally with 50mg/kg body weight of EdU. EdU staining was performed using the Click-iT^{Tm} EdU imaging Kit (Invitrogen, Carlsbad, CA) according to manufacturer's protocol. Briefly, seminiferous tubules were dissected 24 hrs after EdU injection and first probed with GFRA1 or EOMES antibodies before staining for EdU. Tubules were incubated with Click-iT ^{TM} reaction cocktail containing buffer, $CuSO_4$, Alex Fluor 594 Azide, and reaction additive buffer for 30 min at room temperature. Tubules were then washed in PBS-Tween20 buffer for more than 2 hours and then mounted in Vectashield mounting media (Vector Laboratories Inc., Burlingame, CA) with DAPI and imaged on a Nikon Eclipse E600 epifluorescence microscope equipped with an EXi Aqua Camera from Q-Imaging (Surrey, BC, Canada).

Quantitative Real Time (qRT)-PCR

924 Total RNA was prepared from testes using the PureLink™ Kit (Ambion). qRT-
925 PCR was performed using SYBR® Green Master Mix with gene-specific primer
926 sets in a One-step RT-PCR reaction (Applied Biosystems) on an equal amount of
927 total RNA from testes of each genotype and analyzed using ABI 7500 system
928 sequence detection software (Applied Biosystems, Carlsbad, California).
929 Arcturus™ PicoPure™ RNA kit from Applied Biosystems (Waltham, MA) was
930 used to isolated RNA from the FACsorted cells.
931
932 **mRNA sequencing and analysis**
933 An equal amount (4µg) of total RNA from testes of 4 month-old mice was used
934 for sequencing using the TruSeq RNA Sample Prep Kit v2. The libraries were
935 amplified by 15 PCR cycles and quantitated using a Kapa Biosystems
936 Quantification Kit. The quality of each library was assessed on a Bioanalyzer
937 Agilent DNA 1000 Chip. The libraries were then normalized to 2nM and then
938 pooled into 1 sample. The pooled samples were loaded onto 3 lanes of a V3
939 Flow Cell and then sequenced on the Illumina HiSeq 2000 (100bp paired-end
940 run). RNAseq data for all samples were subjected to quality-control check by
941 NGSQCtoolkit v 2 (*56*), and samples with base qualities ≥ 30 over 70 nucleotides
942 (100 bp reads) were used in the analysis. Orphaned reads after the quality-
943 control step were used as single-ended reads in the downstream analysis. Read
944 mapping was carried out using *TopHat v 2.0.7* (*57*) with supplied annotations at
945 parameters (--read-mismatches 2, --read-gap-length 3 and --read-edit-dist 3)
946 against the mouse genome (build-mm9). Bamtools v 1.0.2 (*58*) was used to

947 calculate the mapping statistics. *HTSeq-count* script was used at default
948 parameters (http://www-huber.embl.de/users/anders/HTSeq/doc/count.html) to
949 count the reads mapped to known genes for both orphaned and paired-end
950 reads, and final counts were merged prior to differential expression analysis.
951 Expression differences between *Tg(Gdnf)* and *Tg(Gdnf);lu/lu* mice were
952 determined by using *edgeR v 2.6.10* (*59*) (*exactTest* function was used to
953 determine the differences in expression). Genes with FDR ≤ 0.05 and Absolute
954 $\log_2$ fold change ≥ 1 were considered differentially expressed.
955

956 **Regeneration of SSCs after busulfan treatment**

957 A single dose of 10mg/kg body weight of busulfan was injected intraperitoneally
958 in WT C57BL/6 adult mice. Three mice were analyzed at each time point. Their
959 testes were processed for whole-mount immunofluorescence. For a percentage
960 of GFRA1 population, 300-500 cells were counts for each mouse. The EOMES+
961 and PLZF+ cells were counted over equal length of seminferous tubules, for each
962 mouse. The data are presented as the number of cells ± SE per 1,000 Sertoli
963 cells.
964

965 **Flow Cytometry and single cell sequencing**

966 Single cell suspensions from *Plzf* $^{+/+}$ and *Plzf* $^{lu/lu}$ testes carrying the *Eomes-*
967 ^{tdTomato} knockin allele were prepared using a two-step collagenase and trypsin
968 enzymatic digestion procedure. Cells were filtered first with an 40um nylon
969 membrane followed by a 25um filter. DNase I was added to the cell suspension

to prevent clumping. Cells were washed once in PBS and finally suspended at a concentration of 1-2x10^6 cells/ml, analyzed and sorted on Becton Dickinson FACSVantage cell sorter. Cells from littermate mice that were negative for *Eomes*-^{tdTomato} were used to gate the positive staining cell population and DAPI was used to gate out the dead cells. Single cells were flow sorted into individual wells of a Bio-Rad hard shell 384 well plate. The plate was immediately transferred and stored in a -80C freezer. Custom designed Drop-Seq barcodes (see below) from Integrated DNA Technologies (IDT) were delivered into each well of each 384 well plate. All primers in one well shared the same unique cell barcode and billions of different unique molecular identifiers (UMIs). An Echo 525 liquid handler was used to dispense 1ul of primers and reaction reagents into each well in the plate for the cell lysis and cDNA synthesis. Following cDNA synthesis, the contents of each well were collected and pooled into one tube using a Caliper SciClone Liquid Handler. After treatment with exonuclease I to remove unextended primers, the cDNA was PCR amplified for 13 cycles. The cDNA was fragmented and amplified for sequencing with the Nextera XT DNA sample prep kit (Illumina) using custom primers (see below) that enabled the specific amplification of only the 3′ ends. The libraries were purified, quantified, and sequenced on an Illumina NextSeq 500.

Barcode oligo primer: 5'-

AAGCAGTGGTATCAACGCAGAGTACJJJJJJJJJJJJNNNNNNNN

TTTTTTTTTTTTTTTTTTTTTTTTTTTTVN-3'

Custom primer sequence: 5'-
AATGATACGGCGACCACCGAGATCTACACGCCTGTCCGCGGAAGCAGTGG
TATCAACGCAGAGT*A*C-3'

Custom read 1 sequence: 5'-CGGAAGCAGTGGTATCAACGCAGAGTAC-3'

Single cell sequencing analysis

Cells derived from the mutant and control samples were clustered separately. Genes with fewer than 2 counts in fewer than 2 cells were removed from the digital expression matrix and the expression was normalized by relative library size (computed as total counts per cell divided by median counts per cell) and log transformed. 100 genes with the largest variance / mean ratio and 10 genes used for cell type verification (*Gfra1*, *Eomes*, *Id4*, *T*, *Lin28a*, *Plzf*, *Bmi1*, *Pax7*, *Nanos2* and *Nanos3*) were used to subset the expression matrix for dimensionality reduction and clustering. Cellular expression profiles at these 110 genes were embedded into a 2-dimensional latent space using UMAP (McInnes L and Healy J. Uniform manifold approximation and projection for dimension reduction. ArXiv e-prints 2018, version 0.2.1), and clusters of cells were identified using hierarchical density-based spatial clustering (HDBSCAN, Campello Ricardo JGB et al., Hierarchical density estimates for data clustering, visualization, and outlier detection. ACM transactions on Knowledge Discovery from Data (TKDD) 2015; McInnes L, Healy J, and Astels S. Hierarchical density

based clustering. Journal of open Sourse Software (JOSS) 2017, version 0.8.12). Cells assigned to cluster 0 are those which could not be confidently assigned to any cluster.