BOR Papers in Press. Published on January 7, 2015 as DOI:10.1095/biolreprod.114.125757

Transcriptional and translational heterogeneity among neonatal mouse spermatogonia[1]

Brian P. Hermann,[2,3] Kazadi N. Mutoji,[3] Ellen K. Velte,[4] Daijin Ko,[6] Jon M. Oatley,[7] Christopher B. Geyer,[4,5] and John R. McCarrey[3]
[3]Department of Biology, The University of Texas at San Antonio, San Antonio, Texas
[4]Department of Anatomy and Cell Biology at the Brody School of Medicine, East Carolina University, Greenville, North Carolina
[5]East Carolina Diabetes and Obesity Institute, East Carolina University, Greenville, North Carolina
[6]Department of Management Science and Statistics, The University of Texas at San Antonio, San Antonio, Texas
[7]Center for Reproductive Biology, School of Molecular Biosciences, Washington State University, Pullman, Washington

[1]Supported by NIH grants HD062687 (B.P.H.), HD061665 (J.M.O.), HD072552 (C.B.G.), and GM092334 (J.R.M.), NSF grant 1337513 (B.P.H.), the Max and Minnie Tomerlin Voelcker Fund, the Kerr Foundation, the Kleberg Foundation, and The University of Texas at San Antonio. A portion of these data were presented at the 2014 Gordon Research Conference on Mammalian Reproduction held August 10-15, 2014 at Colby-Sawyer College, New London, NH.
[2]Correspondence: Brian P. Hermann, The University of Texas at San Antonio, One UTSA Circle, San Antonio, TX 78249. E-mail: brian.hermann@utsa.edu

Running title: Heterogeneity among neonatal mouse spermatogonia

Summary Sentence: Spermatogonia in neonatal mouse testes exhibit substantial heterogeneity at the mRNA and protein levels indicative of multiple distinct sub-populations that may have different functional capacities.

Key Words: Spermatogonial stem cells, prospermatogonia, single-cell gene expression, first wave spermatogenesis, germline development.

ABSTRACT

Spermatogonial stem cells (SSCs) are a subset of undifferentiated spermatogonia responsible for ongoing spermatogenesis in mammalian testes. SSCs arise from morphologically homogeneous prospermatogonia, but growing evidence suggests that only a subset of prospermatogonia develops into the foundational SSC pool. This predicts that subtypes of undifferentiated spermatogonia with discrete mRNA and protein signatures should be distinguishable in neonatal testes. We used single-cell qRT-PCR to examine mRNA levels of 172 genes in individual spermatogonia from 6-day postnatal (P6) mouse testes. Cells enriched from P6 testes using the StaPut or THY1+ magnetic cell sorting methods exhibited considerable heterogeneity in the abundance of specific germ cell and stem cell mRNAs, segregating into one somatic and three distinct spermatogonial clusters. However, P6 *Id4*-eGFP+ transgenic spermatogonia, which are known to be enriched for SSCs, were more homogeneous in their mRNA levels, exhibiting uniform levels for the majority of genes examined (122/172). Interestingly, these cells displayed nonuniform (50/172) expression of a smaller cohort of these genes, suggesting there is

substantial heterogeneity even within the *Id4*-eGFP+ population. Further, while immunofluorescence staining largely demonstrated conformity between mRNA and protein levels, some proteins were observed in patterns that were disparate from those detected for the corresponding mRNAs in *Id4*-eGFP+ spermatogonia (e.g. *Kit, Sohlh2, Stra8*), suggesting additional heterogeneity is introduced at the post-transcriptional level. Taken together, these data demonstrate the existence of multiple spermatogonial subtypes in P6 mouse testes and raise the intriguing possibility that these subpopulations may correlate with the development of functionally distinct spermatogenic cell types.

INTRODUCTION

Spermatogonial stem cells (SSCs) are undifferentiated male germ cells that are essential for ongoing maintenance of spermatogenesis. In the adult testis, the SSC population balances self-renewal and differentiation to maintain the pool of SSCs and continually produce committed progenitor spermatogonia that will give rise to the remainder of the spermatogenic lineage. Undifferentiated spermatogonia comprise both SSCs and progenitor spermatogonia, but progenitors are distinct from stem cells in that they have a finite transient-amplifying replicative capacity. In the adult mouse testis, the prevailing view is that A_{single} (A_s) spermatogonia comprise the SSC population and these cells self-renew or initiate differentiation by giving rise to clones of progenitor A_{paired} (A_{pr}) and $A_{aligned}$ (A_{al}) spermatogonia which remain connected via intercellular cytoplasmic bridges to form chains and networks of spermatogonia, which, in turn, give rise to differentiating spermatogonia [1;2]. This *prima facie* model was built primarily on the fundamental principles of clonal amplification and observations of proliferation kinetics *in situ*. However, the number of A_s spermatogonia in a mouse testis (~35,000; [3]) is more than 10 times larger than the number of SSCs with regenerative capacity (~3,000) based on transplantation experiments [4], demonstrating that there is functional heterogeneity even among undifferentiated A_s spermatogonia which have similar morphological characteristics.

Visualization of spermatogonial clones in whole mount preparations of seminiferous tubules has enabled studies that have also reported phenotypic heterogeneity among undifferentiated A_s, A_{pr}, and A_{al} spermatogonia. Specifically, several proteins (e.g., BMI1, ID4, GFRA1, LIN28, NANOS2, NEUROG3, PAX7, ZBTB16/PLZF) show expression patterns that vary among undifferentiated spermatogonia with different clone lengths and between different spermatogonial clones of the same length [5-10]. In some cases, heterogeneous expression patterns have been reported within individual spermatogonial clones (e.g., GFRA1, NANOS2; [5;6]), suggesting undifferentiated spermatogonia exist as multiple dynamic sub-populations. Despite these data demonstrating phenotypic and functional heterogeneity among undifferentiated spermatogonia, there have been recent challenges to the A_s model. Results from live-imaging studies of spermatogonia expressing GFP under the control of the *Neurog3* and *Gfra1* promoters suggested that undifferentiated spermatogonia of any clone length (single, paired, aligned, and fragmented clones of various lengths) may contribute to maintenance of spermatogenesis in a steady-state [11;12], which is similar to the A_0/A_1 model that was originally advanced for rodents [13-16]. However, the A_s model is supported by studies that took into account stages of the seminiferous cycle and mapped the spermatogonia that remain after stage VIII (i.e., A_s and A_{pr}), when nearly all of the undifferentiated A_{al} spermatogonia transition to differentiating type A_1 spermatogonia [1;2]. Since the resurrected A_0/A_1 model based on the results of live imaging studies [11;12] does not account for seminiferous cycle stages, and it is not known if cells from fragmented clones persist after stage VIII, a requisite characteristic of

SSCs may not be fulfilled by this model. Identification of gene products that exhibit an expression pattern which is limited to SSCs might begin to reconcile these disparate observations, but to date, there have been no reports of strict SSC-specific markers.

Recently, the HLH transcriptional repressor ID4 was reported to be exclusively expressed by A_s spermatogonia in the testis from P6 into adulthood [8;17;18], and thus, has emerged as a candidate SSC-specific marker. Transplantation studies definitively demonstrated that SSCs were exclusively found within the *Id4* expressing fraction of cultured spermatogonia from mice bearing an *Id4*-eGFP reporter transgene [8]. Thus, since ID4 protein was only observed in A_s spermatogonia and SSCs were restricted to a population of *Id4*-expressing cells, these data support renewed confidence in the A_s model of SSC clonal amplification. However, it is not known whether ID4 is present in all A_s spermatogonia or in all SSCs *in vivo*. Indeed, there was only partial overlap between expression patterns of ID4 and ZBTB16 [17], a consensus marker of undifferentiated spermatogonia (including A_s), demonstrating molecular heterogeneity among A_s spermatogonia. Thus, despite the identification of ID4 as the most specific putative SSC marker to date, it does not resolve the discrepancy between the numbers of A_s spermatogonia and SSCs, respectively, per testis [3;4]. These data are, however, consistent with results of other studies demonstrating substantial phenotypic heterogeneity between mouse SSCs and other undifferentiated spermatogonia, and even among SSCs [12;19-21]. Moreover, the phenomenon of spermatogonial marker heterogeneity is not likely to be a mouse-specific phenomenon since there is recent evidence from rat testes of multiple A_s spermatogonial sub-populations on the basis of ERBB3 receptor tyrosine kinase expression [22], due at least in part to the stage of the cycle of the seminiferous epithelium.

The apparent phenotypic heterogeneity among SSCs and undifferentiated spermatogonia raises questions about the origin, functional significance, and full extent of heterogeneity among individual spermatogonia and sub-populations of spermatogonia. SSCs are descendants of XY primordial germ cells (PGCs), which, in mice, arrive at the developing testis at mid-gestation (~embryonic day (E) 10.5-12.5; [23;24]). Post-migratory PGCs become M-prospermatogonia at that time and undergo several rounds of mitotic cell division before becoming quiescent T1-prospermatogonia until birth. Between postnatal (P) days 0-3 (P0-P3), these cells become T2-prospermatogonia as they re-enter the cell cycle and migrate to the basement membrane of the seminiferous cords [23;24]. Between P3-P6, some T2-prospermatogonia transition directly into differentiating spermatogonia and produce the first spermatogenic wave, while others remain undifferentiated and become SSCs that sustain spermatogenesis, and still others undergo programmed cell death [25;26]. However, the mechanism(s) responsible for this functional divergence among undifferentiated spermatogonia that lead(s) some to become SSCs, but not others, is not known. One possibility is that fetal prospermatogonia diverge into separate sub-populations that are predisposed to form distinct sub-populations of undifferentiated spermatogonia [25;27]. This is consistent with analyses of spontaneous mutation frequencies suggesting that only a sub-population of prospermatogonia normally contribute to the ultimate SSC pool, and that these cells may be predetermined to this fate [28;29].

To better understand the extent of heterogeneity that exists among sub-populations of spermatogonia in the neonatal testis, we performed single-cell gene expression analyses on enriched populations of P6 spermatogonia and subsequently examined protein levels by immunofluorescence staining. The results of these studies demonstrate considerable differences between individual cells on both the mRNA and protein levels, including disparities between mRNA and protein expression consistent with regulation at the level of translation, all of which

support the concept that the P6 spermatogonial pool consists of multiple sub-populations with discrete gene expression signatures that potentially correlate with distinct cell fates.

MATERIALS AND METHODS

Animals

Male C57BL/6 mice from Jackson Laboratories and the *Id4*-eGFP (LT-11B6) transgenic reporter line [8] were maintained with ad libitum normal laboratory diet. All experiments utilizing animals were approved by the Institutional Animal Care and Use Committee of the University of Texas at San Antonio (Assurance A3592-01) and were performed in accordance with the NIH *Guide for the Care and Use of Laboratory Animals*.

Testes from 6-day postpartum (P6) mice were used to generate suspensions of cells following enzymatic digestion as described previously [30-32]. Testis cell suspensions were enriched for spermatogonia using StaPut gravity sedimentation, THY1+ MACS, or *Id4*-eGFP+ FACS and subsequently used for single-cell gene expression analyses (Supplemental Figure S1; Supplemental Data are available online at www.biolreprod.org). Testes from at least two mice were pooled for each cell isolation experiment.

StaPut gravity sedimentation

Populations of cells enriched for spermatogonia were prepared from P6 C57BL/6 mice using the StaPut method based on sedimentation velocity at unit gravity [33;34]. Briefly, testis cells (10^6-10^7) suspended in 2 ml of buffer plus 0.5% BSA were loaded onto a 50 ml gradient of 2-4% BSA [35] and allowed to sediment for 2.5 hr at 4°C. An aliquot of unfractionated cells (0.25 x10^6/ml) was also reserved as a control. Approximately one hundred 0.5 ml fractions were then collected in microcentrifuge tubes and analyzed for content of spermatogonia on the basis of morphology under phase contrast optics (which typically yields ≥85% purity). Fractions containing Sertoli cells or spermatogonia were pooled separately, concentrated (to 0.25 x10^6 cells/ml) and stored in buffer containing FBS on ice until use.

THY1+ spermatogonia isolation

Testis cell suspensions from P6 C57BL/6 males were generated as described above and enriched for spermatogonia by centrifugation through a cushion of 32-39% Percoll (Sigma-Aldrich) and subjected to selection for THY1 using antibodies conjugated to MACS microbeads (CD90.2, Miltenyi), as described previously [32]. THY1+, THY1-, and unselected cells (prior to Percoll cushion) were suspended in defined serum-free medium [32] and stored on ice until use.

Id4-*eGFP+ FACS*

Testis cell suspensions from P6 *Id4*-eGFP transgenic males were suspended (5-20 x 10^6 cells/ml) in ice-cold Dulbecco's phosphate-buffered saline (DPBS) containing 10% FBS (DPBS+S) and subjected to FACS sorting for eGFP epifluorescence using a FACS Aria (Supplemental Figure S2; BD, Franklin Lakes, NJ). Positive GFP epifluorescence was determined by comparison to testis cells from P6 *Id4*-eGFP-negative littermates. Propidium iodide (0.5μg/ml, BD Biosciences) was added for discrimination of dead cells. An aliquot of each sorted cell population was reanalyzed on the same instrument to determine purity following sorting, demonstrating an average purity of viable GFP+ cells equal to 88.1% (Supplemental Figure S2).

Single-cell qRT-PCR

Suspensions of StaPut-enriched spermatogonia, THY1+ spermatogonia and *Id4*-eGFP+ cells generated as noted above were used for specific target amplification (STA) qRT-PCR measurement of mRNA levels in individual cells using the C1 Single-Cell Autoprep System and BioMark HD instruments (Fluidigm) essentially as described [36]. Individual cells from each suspension were captured on a C1 STA integrated fluidic circuit (IFC) chip (10-17 μm cells) using the Fluidigm C1 instrument. Captured cells were subsequently stained with the LIVE/DEAD Cell Viability/Cytotoxicity Kit (Life Technologies), imaged on an AxioImager M1 to identify and exclude either dead cells (ethidium+) or captures of multiple cells prior to cell lysis and subsequent chemistry. Total time between addition of each cell suspension to an IFC and cell lysis was typically 90 min. Single-cell captures were replicated with three independent cell suspensions generated with each spermatogonia enrichment method (see Supplemental Table S1). Following imaging, pre-amplified cDNA was generated on-chip from each cell using the Single Cells-to-CT Kit (Life Technologies), pooled qPCR primers (Supplemental Table S2), Array Control RNA Spikes (#1, 4, and 7; Life Technologies), and Fluidigm STA reagents according to the manufacturer's recommendations. Off-chip "tube controls" were: 1) 250 cell aliquots (bulk) of each loaded cell suspension, 2) unselected cells, 3) negative cell populations (StaPut – Sertoli cells, THY1– cells, and *Id4*-eGFP–cells), and 4) no-template controls (no cells added). Each tube control was prepared using aliquots of the same lysis, reverse transcription and preamplification reagents used with the corresponding "on-chip" single-cell samples. A sample of each preamplified cDNA, including tube controls, was then used for high-throughput qPCR measurement of each gene of interest using the BioMark HD system as described [37] with modifications. The genes examined were chosen based on potential function in spermatogonia and other testicular cell types as reported in the literature and were categorized on the basis of putative functional roles or cell type marking (Supplemental Table S2). In all, expression of a total of 189 putative spermatogonial genes and three internal control RNA "spikes" was initially examined. A 2.25 μl aliquot of each cDNA was mixed with 2.5 μl of 2X SsoFast EvaGreen Supermix with Low ROX (Bio-Rad) and 0.25 μl of 20X DNA Binding Dye Sample Loading Reagent (Fluidigm), which was then pipetted into an individual sample inlet in a 96.96 Dynamic Array IFC chip (Fluidigm). Individual qPCR primer pairs (pool forward and reverse, 100μM each, Supplemental Table S2) were diluted 1:10 with TE (2.5 μl total volume), mixed with 2.5 μl Assay Loading Reagent (Fluidigm), and then individually pipetted into individual assay inlets in the same 96.96 Dynamic Array IFC chip. Samples and assays were loaded into the IFC chambers with an IFC Controller HX (Fluidigm) and qPCR was performed with the BioMark HD real-time PCR reader (Fluidigm) following the manufacturer's instructions using standard fast cycling conditions and melt-curve analysis, generating an amplification curve for each gene of interest in each sample (9,216 reactions per IFC). Quantitative PCR results were analyzed using Fluidigm's Real-time PCR Analysis software with the following parameters: 0.65 curve quality threshold, linear derivative baseline correction, automatic thresholding by assay (gene), and manual melt curve exclusion. Cycle threshold (Ct) values for each reaction from live single cells were exported and further analyzed using two R-script packages, SINGuLar Analysis Toolset 2.1 and 3.0 (Fluidigm), using a limit of detection of 24 and outlier exclusion, which generated the hierarchical clustering heatmap, PCA analyses, and violin plots of Log2-transformed Ct values for each gene of interest from the live, single cells. Additional statistical analyses for sample clustering (Gap test, partitioning around medoids) were

also performed in R studio. A thorough description of the single-cell data analysis methods employed is included in the Supplemental Information.

Immunostaining of* Id4-*eGFP testes

P6 *Id4*-eGFP+ testes were fixed with 4% PFA for 2 hr at 4°C, washed extensively with DPBS, soaked in 30% sucrose, embedded in OCT medium and frozen. Frozen sections (5μm) were cut and placed on positively-charged slides and stored at -80°C prior to use. Sections were blocked for 1 hr at room temperature in 1X PBS containing 3% BSA + 0.1% Triton X-100, stained for 1 hr in antibody diluted in blocking buffer (see list of antibodies and dilutions in Table 1), and washed with 1X PBS + 0.1% Triton X-100. Indirect immunofluorescence labeling was then performed with secondary antibodies (Table 1) plus phalloidin-635 (1:500, Life Technologies) for 1 hr at room temperature. Primary antibody was omitted as a negative control. After additional stringency washes, sections were mounted with Vectastain containing DAPI (Vector Laboratories), coverslipped, and images obtained using a Fluoview FV1000 confocal laser-scanning microscope (Olympus America). Each staining was performed in triplicate on testes from at least 2 different animals. Quantification of marker co-labeling was performed by counting the number of antibody-labeled and unlabeled GFP+ cells in 10 randomly selected testis cords. Counting was performed three times per sample and the average reported. GFP epifluorescence intensity (bright vs dim) was determined using an intensity thresholding approach, where cells with intensity above 50% of maximum pixel exposure were considered bright, while those with expression below 50% were counted as dim.

RESULTS

We hypothesized that sub-populations of spermatogonia with discrete mRNA signatures could be identified in neonatal mouse testes. To test this hypothesis, we performed single-cell qRT-PCR on enriched populations of spermatogonia from P6 testes that were isolated using methods that differ in their efficiency of selecting SSCs. Specifically, StaPut gravity sedimentation and THY1+ MACS, which have both been used extensively for preparation of enriched spermatogonial populations, were used to capture the full complement of P6 spermatogonia, which we expected would include undifferentiated spermatogonia (including nascent SSCs) and early differentiating spermatogonia (Supplemental Figure S1). In addition, FACS for eGFP-expressing cells from mice bearing an *Id4*-eGFP reporter transgene [8] was used to isolate spermatogonia that are more highly enriched for SSCs (Supplemental Figures S1 and S2). Suspensions of cells prepared using each method were subjected to single-cell qRT-PCR to measure the mRNA levels of 189 genes. Data from a total of 584 single viable cells from 9 independent cell preparations (three each using each cell isolation method) were included in subsequent analyses (Supplemental Tables S1 and S3). Results from mRNA measurements of 172 genes (of the 189 genes measured) were included in subsequent analyses, while the remaining 17 genes were eliminated from further analysis because of assay failure in one or more of the sample groups (Supplemental Tables S2 and S3). Pairwise global comparisons of the gene expression measurements (Log2ex) from each of the nine samples (three replicates each of cells isolated from P6 testes using StaPut, THY1+ MACS, and *Id4*¬eGFP+ FACS) demonstrated robust reproducibility between individual sample replicates for *Id4*-eGFP+ cells ($r = 0.925$ to 0.989), and somewhat lower correlation among replicate StaPut isolations ($r = 0.869$-0.961) and THY1+ isolations ($r = 0.686$-0.893; Supplemental Figure S3). A complete

description of the single-cell gene expression analysis methods employed is included in the Supplemental Information.

Using these data, we performed an unsupervised hierarchical analysis to group the individual cell samples based on Euclidean distance (Figure 1) and, as expected, many divisions evident in the dendrogram formed sample (cell) clusters. Indeed, statistical analyses of these data supported existence of 8-10 distinct clusters of cells among P6 testis cells (see Supplemental Information, Supplemental Table S4, and Supplemental Figure S4). At the first division in the dendrogram, one major group containing 183 cells (Sample Cluster 1) exhibited gene expression profiles consistent with somatic cells, including low or absent values for germ cell genes and presence of mRNAs for genes expressed specifically by Sertoli cells, Leydig cells and/or peritubular myoid cells (Figure 1, Supplemental Tables S3 and S5). Only nine cells in this somatic cell group were derived from *Id4*-eGFP+ samples (out of 229 *Id4*-eGFP+ cells), demonstrating the relative efficiency of isolating enriched spermatogonial populations using this FACS-based method (Figure 1, Supplemental Table S5). The second major group (Sample Cluster 2) consisted of 401 cells that contained abundant germ cell messages. This group could be further subdivided into at least five major clusters (2-1 through 2-5), four of which contained cells from all three isolation methods (Figure 1, Supplemental Table S5). Specifically, *Id4*-eGFP+ cells were absent from one of these germ cell clusters (2-1) and overall, were largely restricted to the other four germ cell clusters (2-2 through 2-5). Thus, testis cells, and more specifically, spermatogonia isolated from P6 mouse testes by StaPut or THY1+ MACS, exhibited the greatest amount of heterogeneity in mRNA abundance, while spermatogonia expressing the *Id4*-eGFP transgene were more homogeneous in their steady-state mRNA levels as a population. Nevertheless, *Id4*-eGFP+ spermatogonia (non-somatic cells) still segregated into four discrete germ cell groups, demonstrating that even *Id4*-eGFP+ cells represent a heterogeneous population characterized by multiple gene expression signatures (Figure 1, Supplemental Table S5).

In addition to hierarchical sample clustering shown on the X-axis in Figure 1, the mRNA abundance data were used to cluster genes on the basis of their Pearson correlations as shown on the Y-axis in Figure 1. Genes that clustered in these analyses tended to be measured at similar levels in the same cells. The two major divisions in the gene dendrogram largely split between genes expressed in somatic cells or genes which were poorly expressed or absent from all samples (Gene Cluster 1), and germ cell genes (Gene Cluster 2; Figure 1 and Supplemental Table S6). Gene Cluster 2 was further subdivided into six major groups, and included genes associated with SSC self-renewal (e.g., *Bcl6b, Id4*; 2-1), a group of poorly expressed stem cell genes (e.g., *Nanog*; 2-2), and genes associated with germ cells (e.g., *Gpr125*; 2-3), genes restricted to undifferentiated spermatogonia (e.g., *Zbtb16* and *Sall4*; 2-5), genes associated with differentiating spermatogonia (e.g., *Kit* and *Stra8*; 2-4) and genes highly expressed in all spermatogonia (e.g., *Ddx4* and *Tex14*; 2-6) (Figure 1 and Supplemental Table S6). Genes with the most tightly correlated mRNA levels were found within the undifferentiated spermatogonia cluster (2-5) and included *Zbtb16*, *Sall4*, and *Igf2bp1* (r=0.9526) and *Dazl*, *Sohlh2*, and *Sycp3* (r= 0.9533) (Supplemental Table S6).

Principal component analysis (PCA) was used to simplify the sample clustering by reducing the data dimensionality while still taking into account the majority of heterogeneity among P6 testis cells (Figure 2, see Supplemental Information). The biological significance of this analysis became evident as the gene expression signatures of cell clusters were analyzed (see next paragraph). Presumed somatic cells clustered to a distinct region of the two-dimensional PCA

plot (Figure 2A) that was further separable into three groups in the third dimension (Figure 2B-J; Supplemental Movie S1). By definition, the first principal component, which is an algebraic description of the majority of the variance in the dataset, was the major driving force separating presumed somatic cells from presumed spermatogonia (Figure 2A, Supplemental Figure S5). P6 spermatogonia isolated by StaPut or THY1+ MACS, which fell outside the somatic cell cluster, were heterogeneous on the basis of abundance of specific mRNAs and fell into three distinct clusters representing potentially distinct sub-populations of spermatogonia (Spermatogonial Signatures 1, 2, and 3; Figure 2A). Although the three spermatogonial clusters could not be obviously subdivided into additional groups using the third principal component (Figure 2B-J), previous PAM statistical tests suggested germ cells might be divisible into five discrete groups (Supplemental Figure S4C). Spermatogonia recovered from *Id4*-eGFP transgenic mice fell only into two of the putative spermatogonia clusters (Spermatogonial Signatures 1 and 2), and thus, were generally more homogeneous as a population based on levels of the specific mRNAs analyzed (Figure 2A). Statistical analyses of the *Id4*-eGFP+ cells using the PAM algorithm suggested that at least two and perhaps three clusters of cells could be discriminated using the mRNA levels from the 172 genes included in this study (Supplemental Figure S6). Thus, these data demonstrate that the considerable gene expression heterogeneity observed among P6 spermatogonia can be used to define cells with discrete gene expression signatures.

Based on this PCA, covariance among genes demonstrated similar gene clustering to the previous unsupervised hierarchical clustering (Supplemental Figure S7; Supplemental Table S6). We extracted the gene expression data from cells that fell within the 'Somatic Cell' and three 'Spermatogonial Signatures' to explore the gene expression profiles of each cluster (Supplemental Figure S8A, Supplemental Table S5). As expected, cells in the somatic cluster exhibited high levels of mRNAs characteristic of Sertoli cells (e.g., *Nr5a1*, *Sox9*, *Wt1*; Supplemental Figure S8B) and generally undetectable mRNA levels of markers of undifferentiated spermatogonia (e.g., *Boll*, *Dazl*, *Sall4*, *Zbtb16*; Supplemental Figure S8B). Likewise, cells in the three spermatogonial clusters exhibited high mRNA levels for undifferentiated spermatogonia markers (e.g., *Ddx4*, *Lin28a*, *Sall4*, *Zbtb16*; Supplemental Figure S8B). Distinctions among the three 'Spermatogonial Signatures' shown in Figure 2A appeared to predominantly reflect quantitative rather than qualitative differences in mRNA levels for most genes assessed (Supplemental Figure S8B). However, cells in the 'Spermatogonial Signature 3' cluster exhibited some surprising differences in mRNA abundance compared with cells in the 'Spermatogonial Signatures 1 and 2' clusters (Supplemental Table S5) including high levels of mRNAs characteristic of Sertoli cells (*Nr5a1*, *Sox9*, *Wt1*; Supplemental Figure S8B). Thus, in addition to expressing spermatogonial genes (e.g., *Dazl, Sall4, Zbtb16*), cells in this cluster appeared to exert less rigid control over somatic cell gene expression, similar to a previous report of *Wt1* expression in mouse PGCs [38]. By contrast, cells in the clusters labeled Spermatogonial Signatures 1 and 2 appeared to display greater fidelity of germ-cell specific gene expression, with very few, if any, showing expression of somatic-cell specific genes.

In order to identify genes with expression profiles that differ substantially among undifferentiated P6 spermatogonia (and might define sub-populations), we examined the expression levels (log2ex values) of individual genes among the individual cells (Figure 3). This analysis was focused on cells isolated from *Id4*-eGFP mice since these cell preparations contained the fewest contaminating somatic cells (Figures 1-2). As expected, the abundance of *Id4* mRNA was largely uniform among *Id4*-eGFP+ cells (Figure 3Ai; first three violin histograms – red, green, blue) and was detectable in nearly all GFP+ cells. GFP

immunofluorescence staining in P6 *Id4*-eGFP testes (Figure 3Aii & iv; red) showed 100% colocalization with eGFP epifluorescence (Figure 3Aiii-iv, green), which served as a baseline for all further colocalization studies. The majority of genes examined (122/172; Supplemental Table S6 and Supplemental Figure S9) exhibited uniform mRNA abundance among *Id4*-eGFP+ spermatogonia, as demonstrated by limited normal distribution around a mean expression value (Figure 3B; row i, red, green and blue violin histograms in each plot). However, because of the greater heterogeneity in cell types isolated using THY1+ MACS and StaPut, these mRNA levels were typically more heterogeneous in populations isolated by these methods [Figure 3B, second group of violins (tan, purple, pink) and third group of violins (dark green, violet, maroon), respectively]. Five exemplary gene products in this "uniform mRNA abundance" category (Figure 3B, ai-*Dazl,* bi-*Ddx4,* ci-*Foxo1,* di-*Sohlh2,* and ei-*Zbtb16*), among *Id4*-eGFP+ cells, were examined on the protein level by immunostaining P6 *Id4*-eGFP+ testes (Figure 3B). Among these, DAZL (3Ba ii-iv), DDX4 (3Bb ii-iv), FOXO1 (3Bc ii-iv), and ZBTB16 (3Be ii-iv) were found to be present uniformly among GFP+ cells. Although all DDX4+ spermatogonia were GFP+ (Figure 3Ba i), we found that 24% of DDX4+ spermatogonia exhibited bright GFP+ epifluorescence and the remainder (76%) were GFP-dim (Figure 3Ba ii-iv, arrowheads), consistent with previous observations [8]. SOHLH2 protein, however, did not appear to be present at uniform levels among GFP+ cells (Figure 3Bd ii-iv). Specifically, cells with dim GFP epifluorescence exhibited high apparent levels of SOHLH2 staining (Figure 3Bd ii-iv, arrows), while cells with high GFP epifluorescence had lower SOHLH2 staining (Figure 3Bd ii-iv, arrowheads), which is consistent with the proposed role of SOHLH2 in spermatogonial differentiation [39;40]. Similar results were observed for SOHLH1 (data not shown). Thus, while uniform mRNA levels typically predicted uniformity at the protein level, there was a disparity between mRNA and protein levels for certain genes such as *Sohlh1* and *Sohlh2*, which encode known markers of spermatogonial differentiation [39;40].

A second group of genes (50/172; Supplemental Table S6 and Supplemental Figure S9) exhibited mRNA levels that were not uniform among *Id4*-eGFP+ cells. For these nonuniform genes, there were often wide ranges in message levels among cells which appeared to delineate two or more discrete sub-populations (Figure 3Ca,b i). In some cases, these nonuniform genes appeared to be repressed within one sub-population of cells in which mRNA was undetectable (off) and expressed in a second sub-population in which mRNA was detectable (on). Interestingly, immunofluorescence staining for two exemplary genes in this category (*Kit* and *Stra8*) revealed a discrepancy between mRNA abundance and apparent protein abundance (Figure 3Ca,b ii-iv). That is, while most *Id4*-eGFP+ cells contained detectable mRNAs for both *Kit* and *Stra8*, only 18% and 35% of dim GFP+ cells co-labeled for KIT or STRA8 protein, respectively (Figure 3Ca,b ii-iv; arrows). Neither KIT nor STRA8 immunoflourescence staining was observed in bright GFP+ cells (Figure 3Ca,b ii-iv; arrowheads). Taken together, these data indicate that multiple sub-populations among *Id4*-eGFP+ cells in the P6 testis are distinguishable on the basis of both mRNA and protein abundance. This suggests that both transcriptional and translational regulatory mechanisms contribute to this heterogeneity and, potentially, to the development of discrete functional sub-populations among undifferentiated spermatogonia.

DISCUSSION

It is becoming increasingly evident that populations of genetically-identical cells of the same type maintained under identical conditions exhibit heterogeneous transcriptomes, and thus, may have previously unappreciated heterogeneous functions and/or fates [41]. This cellular

heterogeneity appears to be common among varied processes dictating cell fate, including cell cycle/mitosis [42], apoptosis [43], and developmental potential of stem cells [44-46]. The foundational population of the adult male germ line – spermatogonial stem cells – emerges early in postnatal development from a subset of undifferentiated spermatogonia. Mounting evidence suggests that SSCs are remarkably heterogeneous on the phenotypic level [5-10;12;17;19-21]. Yet, the origin and extent of phenotypic heterogeneity among undifferentiated spermatogonia has remained poorly understood. The advent of high-throughput single-cell gene expression technology has allowed us to begin to address this knowledge gap by performing a series of gene expression studies using single-cell qRT-PCR to examine mRNA abundance for a panel of genes in individual P6 testis cells and to identify substantial heterogeneity among undifferentiated spermatogonia.

The distinct gene expression signatures we identified among individual P6 spermatogonia may be explained in several ways. It is possible that gene expression heterogeneity is at least partially the result of biological noise induced by subtle differences between cells in their extrinsic microenvironment or by the intrinsic stochastic nature of biochemical reactions. For instance, transcript levels may vary according to the phase of the cell cycle [47-49]. While we cannot exclude the possibility that cell cycle state contributes to the gene expression heterogeneity we observed in P6 spermatogonia, empirical dissection of mRNA levels among cells by cell-cycle phase indicates less than 20% of the heterogeneity in mRNA levels (and as low as 5%) is the result of cell cycle [48;49]. On the other hand, gene expression heterogeneity among cells in a population may be a regulated component of higher-order population stability and function [50;51]. Such regulated gene expression heterogeneity may reflect either the existence of multiple discrete cell populations [52-55], or alternatively, the presence of metastable variants of the same cell population [56-58]. We favor the interpretation that the gene expression heterogeneity revealed here by single-cell studies in P6 spermatogonia supports the existence of multiple subtypes of undifferentiated spermatogonia at this stage.

Certainly, different cell types in any tissue or cell lineage may be distinguished on the basis of divergent biological functions, but ultimately, any differences in biological functions among genetically identical (or nearly identical) cells are rooted in differences at the level of gene expression. The intrinsic and extrinsic circumstances leading to these gene expression differences among cells (at the transcriptional or post-transcriptional level) are likely to be many and form the foundation by which cellular differentiation arises and is maintained. In the present study, we report the existence of heterogeneity among postnatal day 6 (P6) spermatogonia at the mRNA and protein levels. In should be noted, however, that the threshold of gene expression differences among cells that is sufficient to constitute distinct cell types is not clear. Despite this uncertainty, our findings raise the intriguing possibility that the distinct gene expression signatures we have described for each unique sub-population correlate with distinct functional characteristics (Figure 4), including differentiating spermatogonia, SSCs, and progenitor spermatogonia. While it is tempting to equate the three spermatogonial sub-types proposed in Fig 4 to be present in the P6 testis with the three distinct spermatogonial signatures we identified in Fig 2A on the basis of mRNA abundance groupings, such correlations must await functional evidence which will be the subject of future studies.

After the present study was initiated, a new candidate SSC marker, PAX7, was reported [9]. Like ID4, PAX7 was reported to be exclusively expressed by A_s spermatogonia in the neonatal and adult testis [8;9;17;18]. Lineage-tracing demonstrated that progeny of *Pax7*+ cells can produce complete spermatogenesis and have regenerative capacity after cytotoxic insult [9],

supporting the notion that cells with a PAX7+ phenotype may have stem cell capacity, although this gene is also known to be dispensable for spermatogenesis [9]. Surprisingly, though, whole-transcriptome evaluation of mRNA levels in *Id4*-eGFP+ and *Id4*-eGFP- cultured spermatogonia failed to identify any *Pax7* transcripts in either population [8], raising the possibility that ID4 and PAX7 mark completely separate cell populations. Although we did not investigate mRNA levels of *Pax7* in the present study, the potential that PAX7 and ID4 are expressed by different spermatogonial populations is consistent with our results, and would further support the notion of substantial phenotypic heterogeneity among neonatal undifferentiated spermatogonia. Further, although we were unable to utilize the samples generated in this study to interrogate the relationship between *Pax7* and *Id4* mRNA levels in single spermatogonia, future studies utilizing unbiased mRNA detection modes may begin to address the question of whether these two genes are expressed in distinct germ cell sets, *in vivo*.

Our single-cell gene expression data from P6 testes beg the question of whether there is heterogeneity among germ cells at earlier developmental stages, and whether such heterogeneity would predispose formation of the distinct gene expression signatures we detect among undifferentiated spermatogonia in the P6 testis. Indeed, results of spontaneous mutation frequency analyses support a predetermination theory in which a distinct sub-population of prospermatogonia gives rise to the foundational SSC pool. Specifically, the frequency of point mutations in mice bearing a *lacI* mutation-reporter transgene was higher in prospermatogonia at E15.5 and in either StaPut-enriched or THY1+ spermatogonia at P6 than in primary spermatocytes at P18 [28;29]. While this decline in mutation frequency between P6 and P18 correlates with a known wave of apoptosis, it is likely that the prospermatogonia which give rise to the surviving germ cells at P18 maintain the integrity of their genomes more stringently than those prospermatogonia that give rise to differentiating spermatogonia that are subsequently eliminated by apoptosis. It follows that SSC specification, either through a mechanism of predetermination or selection, may occur as a result of molecular divergence that first emerges among prospermatogonia during an earlier developmental window, and subsequently establishes sub-populations of undifferentiated spermatogonia that are more or less likely to produce foundational SSCs (Figure 4). Future studies to determine the point during male germline development when heterogeneity at the mRNA and/or protein levels first appears will further elucidate the mechanisms and timing of SSC specification in mice.

Previous studies of mouse and human PGCs demonstrated considerable heterogeneity among individual cells. Sorted *Pou5f1/Oct4*-eGFP+ PGCs from E9.25, E9.5, E10.25, E10.5 and E11.5 male mouse embryos all demonstrated heterogeneity in the mRNA levels of a limited gene repertoire (combinations of the following genes: *Dazl, Dnd1, Dppa3, Gapdh, Prdm1, Prdm14, Tet1, Tet2,* and *Tet3*; [59;60]). Similarly, single-cell evaluation of post-migratory human male PGCs isolated from 9.5 and 16-week fetal testes revealed considerable heterogeneity in the mRNA abundance for 7 key genes (*DAZL*, *DDX4, GAPDH, NANOS2, NANOS3, POU5F1, and PRDM1*; [61]). Thus, heterogeneity at the transcript level may be a fundamental feature of the germ line that persists from the earliest stages of male germline specification in the epiblast through to the stage at which foundational SSCs seed the seminiferous epithelium to initiate spermatogenesis. The full extent to which this heterogeneity in mRNA abundance extends to the rest of the transcriptome and its functional relevance to formation and maintenance of the male germ line are yet to be determined. PGC-like cells generated from human and mouse pluripotent stem cells and human KIT+ PGCs have been used to derive population-level transcriptomes for PGCs, and these can serve as a baseline for the identification of gene

expression differences among individual cells [59;61]. Recently, single-cell RNA-seq analyses were performed on E6.5 and E7.5 mouse PGCs [62], but since only two individual PGCs were sequenced from each stage, the degree of message heterogeneity within the PGC population as a whole could not be assessed.

A surprising finding in this study was the identification of a subpopulation of P6 *Id4*-eGFP-negative spermatogonia, 'Spermatogonial Signature 3,' which paradoxically contained mRNAs characteristic of both spermatogonia and Sertoli cells. Cells in this group were only observed in the THY1+ and StaPut preparations (i.e., cells with this phenotype were not observed in the *Id4*-eGFP+ sorted populations), but were observed in all three replicates of cells prepared by each of these two methods, suggesting this cluster of cells is neither a technical artifact due to preparation method nor an artifact of the particular animal batch. Moreover, the detection of somatic cell mRNAs was not observed in 'Spermatogonial Signatures 1 and 2,' demonstrating that detection of these mRNAs in 'Spermatogonial Signature 3' was not a technical artifact of the qRT-PCR methodology. It is intriguing to consider the possibility that the cells in this group, which comprise ~17% of the presumed spermatogonia evaluated in this study and arise from an *Id4*-eGFP- phenotype, maintain less stringent control over cell-type specific transcription than other spermatogonia. On the flipside, the presumed spermatogonia in 'Spermatogonial Signatures 1 and 2' appear to exhibit tighter control over cell-type gene expression, which is similar to previous observations of enhanced maintenance of genetic integrity in a distinct sub-population of prospermatogonia/spermatogonia [28;29]. Since WT1 protein and mRNA have previously been detected uniformly among mid-gestation prospermatogonia [38], another potential explanation for these data is that cells in the 'Spermatogonial Signature 3' cluster are more similar to fetal prospermatogonia. In either case, we are not aware of any reports in which Sertoli cell markers (e.g., NR5A1, SOX9, WT1) have been detected among spermatogonia (or likewise, in which spermatogonial marker expression has been detected among somatic cells), and thus, utilization of these mRNAs at translation, spuriously generated or not, must be tightly controlled. Future studies could examine whether the overlap between spermatogonial and somatic cell mRNAs in Spermatogonial Signature 3 extends to the protein level.

In addition and consequently, results from the current study reinforce the concept that while regulation of gene expression at the mRNA level (i.e., transcription and RNA stability) is an important contributor to heterogeneity among spermatogonia, regulation of message utilization (i.e., translational control) can also contribute to this heterogeneity, and both of these mechanisms appear to be operating in developing male germ cells [63;64]. Thus, while mRNA was detected for the differentiation markers, *Stra8* and *Kit,* in all and nearly all *Id4*-eGFP+ spermatogonia, respectively, the corresponding proteins were detected in only a small fraction of *Id4*-eGFP+ cells by immunofluorescence. Indeed, KIT and STRA8 protein were only localized to spermatogonia with dim eGFP epifluorescence, but since the half-life of eGFP is estimated to be >26 hours [65;66], it is possible that eGFP+ cells which are also KIT+ or STRA8+ had transitioned out of an SSC state to become progenitor spermatogonia. While lack of available technology to interrogate protein expression in single cells has limited complementary protein studies to the extent to which mRNA levels can now be interrogated [36], new technologies such as mass cytometry are emerging that may allow parallel quantitative examination of >10 individual proteins within individual cells [67]. It is also possible that differences in mircoRNA (miRNA) expression may contribute to the disparities we observed between mRNA and protein levels for certain gene products, and future studies could utilize existing strategies to catalog

mature miRNA heterogeneity among individual spermatogonia [68]. Thus, a comprehensive understanding of the molecular differences that distinguish sub-populations of male germ cells clearly requires consideration of the transcriptome (mRNA and miRNA) and proteome in each case.

In conclusion, this study provides the first examination of molecular heterogeneity among individual mouse spermatogonia and defines the extent of differences in abundance of specific mRNAs, and in some cases proteins as well, among individual cells. From among the population of male germ cells collectively termed undifferentiated spermatogonia in the neonatal testis, we have identified sub-populations of cells with discrete mRNA abundance signatures, and suggest that these correlate with specific spermatogonial subtypes that differ in their functional capacities. This study sets a new precedent for the extent to which heterogeneity among otherwise potentially homogeneous cell populations can and should be investigated during male germline development, and provides a resource that can be employed in future investigations to identify the timing and mechanisms of specification of SSCs. Indeed, the novel methodologies employed here could be used in future studies to begin to resolve differing opinions regarding formation and maintenance of the SSC pool, and may begin to address long-standing questions regarding the underlying molecular mechanisms responsible for specification and renewal of SSCs.

ACKNOWLEDGEMENTS

The authors thank Dr. Bernard Arulanandam for use of LSRII flow cytometer and the staff of the UTSA LARC for animal care. The authors also gratefully acknowledge Dr. Thomas Forsthuber and Julie Tudyk for their assistance with cell sorting studies. Other cell sorting studies were performed by the University of Texas Health Science Center at San Antonio (UTHSCSA) Flow Cytometry Shared Resource Facility supported by UTHSCSA, P30CA54174 and UL1RR025767.

REFERENCES

[1] Huckins C. The spermatogonial stem cell population in adult rats. I. Their morphology, proliferation and maturation. Anat Rec 1971; 169: 533-557.
[2] Oakberg EF. Spermatogonial stem-cell renewal in the mouse. Anat Rec 1971; 169: 515-531.
[3] Tegelenbosch RAJ, de Rooij DG. A quantitative study of spermatogonial multiplication and stem cell renewal in the C3H/101 F1 hybrid mouse. Mutat Res 1993; 290: 193-200.
[4] Nagano MC. Homing efficiency and proliferation kinetics of male germ line stem cells following transplantation in mice. Biol Reprod 2003; 69: 701-707.
[5] Grisanti L, Falciatori I, Grasso M, Dovere L, Fera S, Muciaccia B, Fuso A, Berno V, Boitani C, Stefanini M, Vicini E. Identification of Spermatogonial Stem Cell Subsets by Morphological Analysis and Prospective Isolation. Stem Cells 2009; 27: 3043-3052.
[6] Suzuki H, Sada A, Yoshida S, Saga Y. The heterogeneity of spermatogonia is revealed by their topology and expression of marker proteins including the germ cell-specific proteins Nanos2 and Nanos3. Dev Biol 2009; 336: 222-231.
[7] Zheng K, Wu X, Kaestner KH, Wang PJ. The pluripotency factor LIN28 marks undifferentiated spermatogonia in mouse. BMC Dev Biol 2009; 9: 38.
[8] Chan F, Oatley MJ, Kaucher AV, Yang QE, Bieberich CJ, Shashikant CS, Oatley JM. Functional and molecular features of the Id4+ germline stem cell population in mouse testes. Genes Dev 2014; 28: 1351-1362.

[9] Aloisio GM, Nakada Y, Saatcioglu HD, Pena CG, Baker MD, Tarnawa ED, Mukherjee J, Manjunath H, Bugde A, Sengupta AL, Amatruda JF, Cuevas I, Hamra FK, Castrillon DH. PAX7 expression defines germline stem cells in the adult testis. J Clin Invest 2014; 124: 3929-3944.
[10] Komai Y, Tanaka T, Tokuyama Y, Yanai H, Ohe S, Omachi T, Atsumi N, Yoshida N, Kumano K, Hisha H, Matsuda T, Ueno H. Bmi1 expression in long-term germ stem cells. Sci Rep 2014; 4: 6175.
[11] Hara K, Nakagawa T, Enomoto H, Suzuki M, Yamamoto M, Simons BD, Yoshida S. Mouse spermatogenic stem cells continually interconvert between equipotent singly isolated and syncytial states. Cell Stem Cell 2014; 14: 658-672.
[12] Nakagawa T, Sharma M, Nabeshima Y, Braun RE, Yoshida S. Functional hierarchy and reversibility within the murine spermatogenic stem cell compartment. Science 2010; 328: 62-67.
[13] Clermont Y, Bustos-Obregon E. Re-examination of spermatogonial renewal in the rat by means of seminiferous tubules mounted "in toto". Am J Anat 1968; 122: 237-247.
[14] Dym M, Clermont Y. Role of spermatogonia in the repair of the seminiferous epithelium following x-irradiation of the rat testis. Am J Anat 1970; 128: 265-282.
[15] Clermont Y, Hermo L. Spermatogonial stem cells in the albino rat. Am J Anat 1975; 142: 159-175.
[16] Bartmanska J, Clermont Y. Renewal of type A spermatogonia in adult rats. Cell Tissue Kinet 1983; 16: 135-143.
[17] Oatley MJ, Kaucher AV, Racicot KE, Oatley JM. 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 2011; 85: 347-356.
[18] Hermann BP, Phillips BT, Orwig KE. The elusive spermatogonial stem cell marker? Biol Reprod 2011; 85: 221-223.
[19] Yoshida S, Nabeshima Y, Nakagawa T. Stem cell heterogeneity: actual and potential stem cell compartments in mouse spermatogenesis. Ann N Y Acad Sci 2007; 1120: 47-58.
[20] Morimoto H, Kanatsu-Shinohara M, Takashima S, Chuma S, Nakatsuji N, Takehashi M, Shinohara T. Phenotypic plasticity of mouse spermatogonial stem cells. PLoS ONE 2009; 4: e7909.
[21] Shinohara T, Ishii K, Kanatsu-Shinohara M. Unstable side population phenotype of mouse spermatogonial stem cells in vitro. J Reprod Dev 2011; 57: 288-295.
[22] Abid SN, Richardson TE, Powell HM, Jaichander P, Chaudhary J, Chapman KM, Hamra FK. A-single spermatogonia heterogeneity and cell cycles synchronize with rat seminiferous epithelium stages VIII-IX. Biol Reprod 2014; 90: 32, 1-15.
[23] Hilscher B, Hilscher W, Bulthoff-Ohnolz B, Kramer U, Birke A, Pelzer H, Gauss G. Kinetics of gametogenesis. I. Comparative histological and autoradiographic studies of oocytes and transitional prospermatogonia during oogenesis and prespermatogenesis. Cell Tissue Res 1974; 154: 443-470.
[24] McCarrey JR. Toward a more precise and informative nomenclature describing fetal and neonatal male germ cells in rodents. Biol Reprod 2013; 89: 47.
[25] Kluin PM, de Rooij DG. A comparison between the morphology and cell kinetics of gonocytes and adult type undifferentiated spermatogonia in the mouse. Int J Androl 1981; 4: 475-493.
[26] Yoshida S, Sukeno M, Nakagawa T, Ohbo K, Nagamatsu G, Suda T, Nabeshima Y. The first round of mouse spermatogenesis is a distinctive program that lacks the self-renewing spermatogonia stage. Development 2006; 133: 1495-1505.

[27] Lei L, Spradling AC. Mouse primordial germ cells produce cysts that partially fragment prior to meiosis. Development 2013; 140: 2075-2081.
[28] Walter CA, Intano GW, McCarrey JR, McMahan CA, Walter RB. Mutation frequency declines during spermatogenesis in young mice but increases in old mice. Proc Natl Acad Sci U S A 1998; 95: 10015-10019.
[29] Murphey P, McLean DJ, McMahan CA, Walter CA, McCarrey JR. Enhanced genetic integrity in mouse germ cells. Biol Reprod 2013; 88: 6.
[30] Ogawa T, Aréchaga JM, Avarbock MR, Brinster RL. Transplantation of testis germinal cells into mouse seminiferous tubules. Int J Dev Biol 1997; 41: 111-122.
[31] Nagano M, Ryu BY, Brinster CJ, Avarbock MR, Brinster RL. Maintenance of mouse male germ line stem cells in vitro. Biol Reprod 2003; 68: 2207-2214.
[32] Oatley JM, Brinster RL. Spermatogonial stem cells. Methods Enzymol 2006; 419: 259-282.
[33] Romrell LJ, Bellve AR, Fawcett DW. Separation of mouse spermatogenic cells by sedimentation velocity. A morphological characterization. Dev Biol 1976; 49: 119-131.
[34] Bellve AR, Cavicchia JC, Millette CF, O'Brien DA, Bhatnagar YM, Dym M. Spermatogenic cells of the prepuberal mouse. Isolation and morphological characterization. J Cell Biol 1977; 74: 68-85.
[35] McCarrey JR, Berg WM, Paragioudakis SJ, Zhang PL, Dilworth DD, Arnold BL, Rossi JJ. Differential transcription of Pgk genes during spermatogenesis in the mouse. Dev Biol 1992; 154: 160-168.
[36] Wu AR, Neff NF, Kalisky T, Dalerba P, Treutlein B, Rothenberg ME, Mburu FM, Mantalas GL, Sim S, Clarke MF, Quake SR. Quantitative assessment of single-cell RNA-sequencing methods. Nat Methods 2014; 11: 41-46.
[37] Dalerba P, Kalisky T, Sahoo D, Rajendran PS, Rothenberg ME, Leyrat AA, Sim S, Okamoto J, Johnston DM, Qian D, Zabala M, Bueno J, Neff NF, Wang J, Shelton AA, Visser B, Hisamori S, Shimono Y, van de Wetering M, Clevers H, Clarke MF, Quake SR. Single-cell dissection of transcriptional heterogeneity in human colon tumors. Nat Biotechnol 2011; 29: 1120-1127.
[38] Natoli TA, Alberta JA, Bortvin A, Taglienti ME, Menke DB, Loring J, Jaenisch R, Page DC, Housman DE, Kreidberg JA. Wt1 functions in the development of germ cells in addition to somatic cell lineages of the testis. Dev Biol 2004; 268: 429-440.
[39] Ballow D, Meistrich ML, Matzuk M, Rajkovic A. Sohlh1 is essential for spermatogonial differentiation. Dev Biol 2006; 294: 161-167.
[40] Suzuki H, Ahn HW, Chu T, Bowden W, Gassei K, Orwig K, Rajkovic A. SOHLH1 and SOHLH2 coordinate spermatogonial differentiation. Dev Biol 2012; 361: 301-312.
[41] Waks Z, Silver PA. Nuclear origins of cell-to-cell variability. Cold Spring Harb Symp Quant Biol 2010; 75: 87-94.
[42] Willardsen MI, Link BA. Cell biological regulation of division fate in vertebrate neuroepithelial cells. Dev Dyn 2011; 240: 1865-1879.
[43] Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature 2009; 459: 428-432.
[44] Takahasi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 2007; 131: 861-872.
[45] Graf T, Stadtfeld M. Heterogeneity of embryonic and adult stem cells. Cell Stem Cell 2008; 3: 480-483.

[46] Chan EM, Ratanasirintrawoot S, Park IH, Manos PD, Loh YH, Huo H, Miller JD, Hartung O, Rho J, Ince TA, Daley GQ, Schlaeger TM. Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. Nat Biotechnol 2009; 27: 1033-1037.
[47] Chen WC, Wu PH, Phillip JM, Khatau SB, Choi JM, Dallas MR, Konstantopoulos K, Sun SX, Lee JS, Hodzic D, Wirtz D. Functional interplay between the cell cycle and cell phenotypes. Integr Biol (Camb) 2013; 5: 523-534.
[48] McDavid A, Dennis L, Danaher P, Finak G, Krouse M, Wang A, Webster P, Beechem J, Gottardo R. Modeling bi-modality improves characterization of cell cycle on gene expression in single cells. PLoS Comput Biol 2014; 10: e1003696.
[49] Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, Cahill DP, Nahed BV, Curry WT, Martuza RL, Louis DN, Rozenblatt-Rosen O, Suva ML, Regev A, Bernstein BE. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science 2014; 344: 1396-1401.
[50] Snijder B, Pelkmans L. Origins of regulated cell-to-cell variability. Nat Rev Mol Cell Biol 2011; 12: 119-125.
[51] Paszek P, Ryan S, Ashall L, Sillitoe K, Harper CV, Spiller DG, Rand DA, White MR. Population robustness arising from cellular heterogeneity. Proc Natl Acad Sci U S A 2010; 107: 11644-11649.
[52] Kawaguchi A, Ikawa T, Kasukawa T, Ueda HR, Kurimoto K, Saitou M, Matsuzaki F. Single-cell gene profiling defines differential progenitor subclasses in mammalian neurogenesis. Development 2008; 135: 3113-3124.
[53] Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, Schwartz S, Yosef N, Malboeuf C, Lu D, Trombetta JJ, Gennert D, Gnirke A, Goren A, Hacohen N, Levin JZ, Park H, Regev A. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature 2013; 498: 236-240.
[54] Pollen AA, Nowakowski TJ, Shuga J, Wang X, Leyrat AA, Lui JH, Li N, Szpankowski L, Fowler B, Chen P, Ramalingam N, Sun G, Thu M, Norris M, Lebofsky R, Toppani D, Kemp DW, Wong M, Clerkson B, Jones BN, Wu S, Knutsson L, Alvarado B, Wang J, Weaver LS, May AP, Jones RC, Unger MA, Kriegstein AR, West JA. Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. Nat Biotechnol 2014; 32: 1053-1058.
[55] Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH, Desai TJ, Krasnow MA, Quake SR. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature 2014; 509: 371-375.
[56] Hanna J, Markoulaki S, Mitalipova M, Cheng AW, Cassady JP, Staerk J, Carey BW, Lengner CJ, Foreman R, Love J, Gao Q, Kim J, Jaenisch R. Metastable pluripotent states in NOD-mouse-derived ESCs. Cell Stem Cell 2009; 4: 513-524.
[57] Dubois A, Deuve JL, Navarro P, Merzouk S, Pichard S, Commere PH, Louise A, Arnaud D, Avner P, Morey C. Spontaneous reactivation of clusters of X-linked genes is associated with the plasticity of X-inactivation in mouse trophoblast stem cells. Stem Cells 2014; 32: 377-390.
[58] Hough SR, Thornton M, Mason E, Mar JC, Wells CA, Pera MF. Single-cell gene expression profiles define self-renewing, pluripotent, and lineage primed states of human pluripotent stem cells. Stem Cell Reports 2014; 2: 881-895.
[59] Vincent JJ, Li Z, Lee SA, Liu X, Etter MO, az-Perez SV, Taylor SK, Gkountela S, Lindgren AG, Clark AT. Single cell analysis facilitates staging of Blimp1-dependent primordial germ cells derived from mouse embryonic stem cells. PLoS ONE 2011; 6: e28960.

[60] Vincent JJ, Huang Y, Chen PY, Feng S, Calvopina JH, Nee K, Lee SA, Le T, Yoon AJ, Faull K, Fan G, Rao A, Jacobsen SE, Pellegrini M, Clark AT. Stage-specific roles for tet1 and tet2 in DNA demethylation in primordial germ cells. Cell Stem Cell 2013; 12: 470-478.
[61] Gkountela S, Li Z, Vincent JJ, Zhang KX, Chen A, Pellegrini M, Clark AT. The ontogeny of cKIT+ human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and in vitro differentiation. Nat Cell Biol 2013; 15: 113-122.
[62] Magnusdottir E, Dietmann S, Murakami K, Gunesdogan U, Tang F, Bao S, Diamanti E, Lao K, Gottgens B, Azim SM. A tripartite transcription factor network regulates primordial germ cell specification in mice. Nat Cell Biol 2013; 15: 905-915.
[63] Chappell VA, Busada JT, Keiper BD, Geyer CB. Translational activation of developmental messenger RNAs during neonatal mouse testis development. Biol Reprod 2013; 89: 61.
[64] Busada JT, Kaye EP, Renegar RH, Geyer CB. Retinoic acid induces multiple hallmarks of the prospermatogonia-to-spermatogonia transition in the neonatal mouse. Biol Reprod 2014; 90: 64.
[65] Li X, Zhao X, Fang Y, Jiang X, Duong T, Fan C, Huang CC, Kain SR. Generation of destabilized green fluorescent protein as a transcription reporter. J Biol Chem 1998; 273: 34970-34975.
[66] Corish P, Tyler-Smith C. Attenuation of green fluorescent protein half-life in mammalian cells. Protein Eng 1999; 12: 1035-1040.
[67] Lepore M, Kalinichenko A, Colone A, Paleja B, Singhal A, Tschumi A, Lee B, Poidinger M, Zolezzi F, Quagliata L, Sander P, Newell E, Bertoletti A, Terracciano L, De LG, Mori L. Parallel T-cell cloning and deep sequencing of human MAIT cells reveal stable oligoclonal TCRbeta repertoire. Nat Commun 2014; 5: 3866.
[68] Petriv OI, Kuchenbauer F, Delaney AD, Lecault V, White A, Kent D, Marmolejo L, Heuser M, Berg T, Copley M, Ruschmann J, Sekulovic S, Benz C, Kuroda E, Ho V, Antignano F, Halim T, Giambra V, Krystal G, Takei CJ, Weng AP, Piret J, Eaves C, Marra MA, Humphries RK, Hansen CL. Comprehensive microRNA expression profiling of the hematopoietic hierarchy. Proc Natl Acad Sci U S A 2010; 107: 15443-15448.

Figure Legends

Figure 1. Unsupervised hierarchical clustering of normalized mRNA levels among individual P6 testis cells identifies cell sub-populations and co-expressed genes. Gene expression data (single-cell mRNA levels for 172 genes) were used for unsupervised hierarchical clustering to group data and generate a dendrogram based on similarities between samples (cells) and mRNA levels for each gene interrogated. Log2-transformed gene expression data from the entire dataset were converted to a global Z-score to allow comparison between all samples and gene assay and were displayed in the heat-map based on relative mRNA levels where blue is the lowest value and red is highest. The legend at the top left indicates sample group identifiers. Genes (vertical) were clustered together using the Pearson method (0-1, with 1 being the highest correlation), and samples (horizontal) were clustered together using the Euclidean method (0-0.5, with 0 being the closest). In each case, the complete linkage method was then used to find similar clusters. Sample cluster and gene cluster identifiers are indicated to the bottom and right of the heat map, respectively. Samples falling into each noted cluster (Sample Cluster 1 – somatic; Sample clusters 2-1 through 2-5, spermatogonia) are shown in Supplemental Table S5. Likewise, genes

falling into each cluster (Gene Cluster 1 and Gene Clusters 2-1 through 2-6) are listed in Supplemental Table S6.

Figure 2. Distinct P6 spermatogonial subtypes emerge based on principal component analysis of mRNA abundance in single cells. Principal component analysis was used to compare the similarity of mRNA levels in individual cells (without respect to sample source) based on the first three principal components (see Supplemental Table S6) where each individual point is an individual cell. **A**) The first two principal components, which describe the vast majority of variability between all cells (see Supplemental Figure S5), clearly separated somatic cells (red ellipse) from spermatogonia. Among spermatogonia, three distinct populations were evident, labeled Spermatogonial Signature 1, 2, and 3 (green, violet and blue ellipses, respectively). *Id4*-eGFP+ cells were almost exclusively found in the green and violet populations. The legend shows sample group identifiers. A three-dimensional PCA plot was used to further examine sample differences by including the third principal component (see Supplemental Movie S1). **B**, **E**, **H**) Two dimensional projections of the 3D plot are shown. In each plot, one or more stars noted with red curved arrows indicate rotation axes and directions. **B**) The plot is rotated by grabbing the four-point star and dragging in the direction of the red arrow (c) to produce plot **C**. **C)** The plot is subsequently rotated in the direction of red arrow (d) to produce plot **D**. **E**) The plot is pivoted by grabbing the four-point star and rotating in the direction of red arrow (f) to produce plot **F** or is turned by grabbing the five-point star and rotating the plot in the direction of red arrow (g) to produce plot **G**. **H**) Likewise, the plot is turned by grabbing the four-point star and rotating in the direction of red arrow (i) to produce plot **I** or is pivoted by grabbing the five-point star and rotating the plot in the direction of red arrow (j) to produce plot **J**. These rotations allow visualization of the PCA plot in the third dimension and mimic the rotations indicated in Supplemental Movie S1.

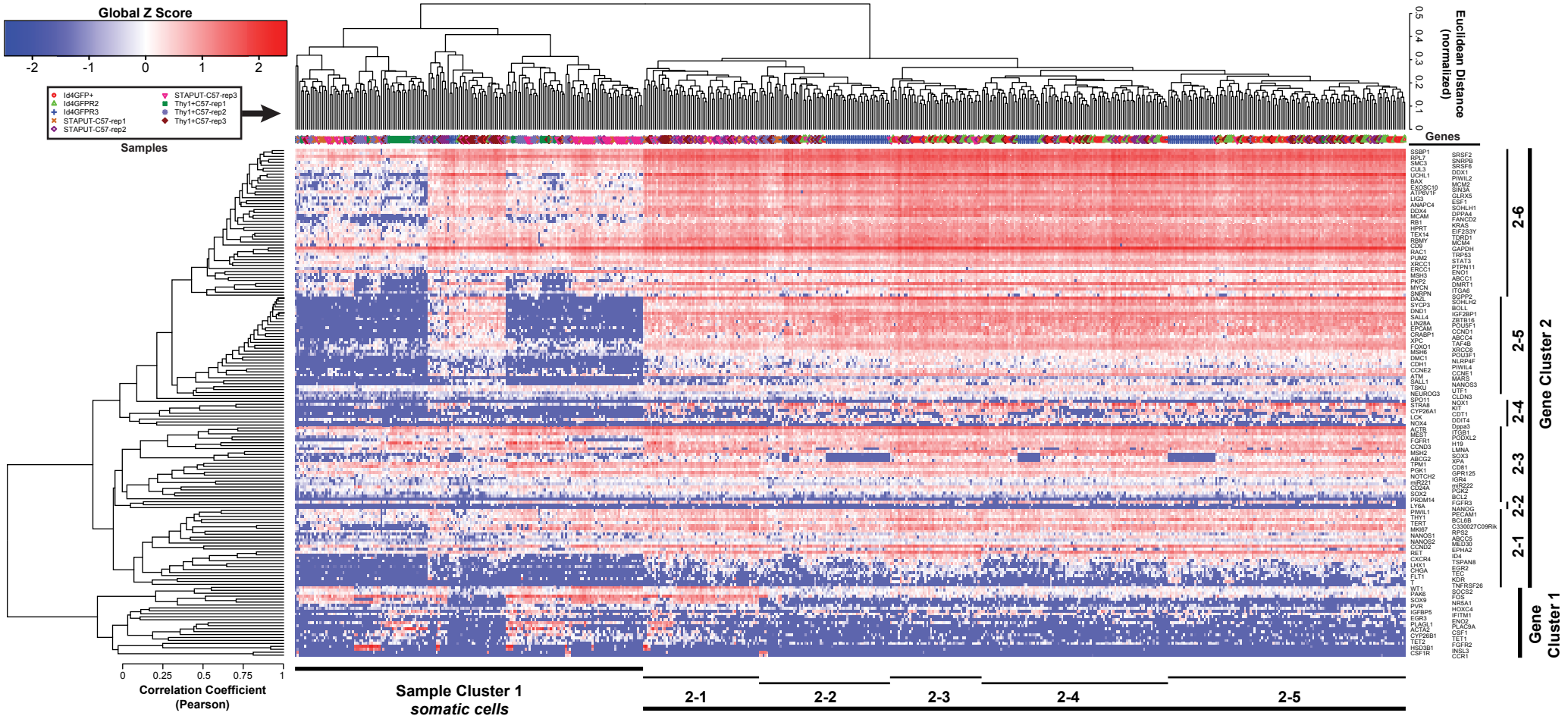
Figure 3. Distinct mRNA and protein abundance patterns characterize P6 *Id4*-eGFP+ undifferentiated spermatogonia. Different patterns in mRNA abundance were evident from the single-cell analyses and these mRNA patterns were compared to protein levels using immunofluorescence staining in P6 *Id4*-eGFP testes. As a visual representation of the mRNA levels among individual cells, this figure presents violin plots for each gene in which individual violin histograms are shown for each of the three replicate samples for each preparation method (nine violins total, according to the sample legend). The first three violins in each plot are from *Id4*-eGFP+ samples, the second three are from StaPut isolations, and the last three from THY1+ isolated cells. The width of the violin histograph at any vertical position reflects the relative proportion of samples (cells) with that degree of mRNA abundance (Log2ex values). Thus, higher vertical position indicates relatively higher mRNA abundance. See Supplemental Figure S9 for violin plots for all 172 genes. As a baseline, **(Ai)** a violin plot of *Id4* mRNA abundance is shown as a reference and demonstrates that *Id4* mRNA is present in most of the *Id4*-eGFP+ cells, as expected. Among *Id4*-eGFP+ samples (first three violins in each gene plot), different patterns of mRNA abundance were evident: uniform or unimodal abundance **(B)**, for example we show *Dazl* **(Bai)**, *Ddx4* **(Bbi)**, *Foxo1* **(Bci)**, *Sohlh2* **(Bdi)**, and *Zbtb16* **(Bei)**; and nonuniform abundance **(C)**, including *Kit* (**Cai**) and *Stra8* (**Cbi**). To correlate mRNA levels depicted in row i with protein, (**A**, **B**, **C** – row **ii**) Immunofluorescence staining using antibodies against the noted proteins (red) was compared with **(A, B**, **C** – row **iii**) eGFP epifluorescence (green) in P6 testes from *Id4*-eGFP LT-11 testes. **A**, **B**, **C** – row **iv**) Overlay of protein immunoflourescence, *Id4*-

eGFP epifluorescence, and F-actin staining (Phalloidin) is shown. Arrowheads denote example bright GFP+ cells. Arrows in **Bd, ii-iv** note examples of dim GFP+ with bright SOHLH2 stainng. Arrows in **Ca, ii-iv and Cb, ii-iv** identify examples of KIT+ and STRA8+ spermatogonia, respectively. Bar = 50 μm.

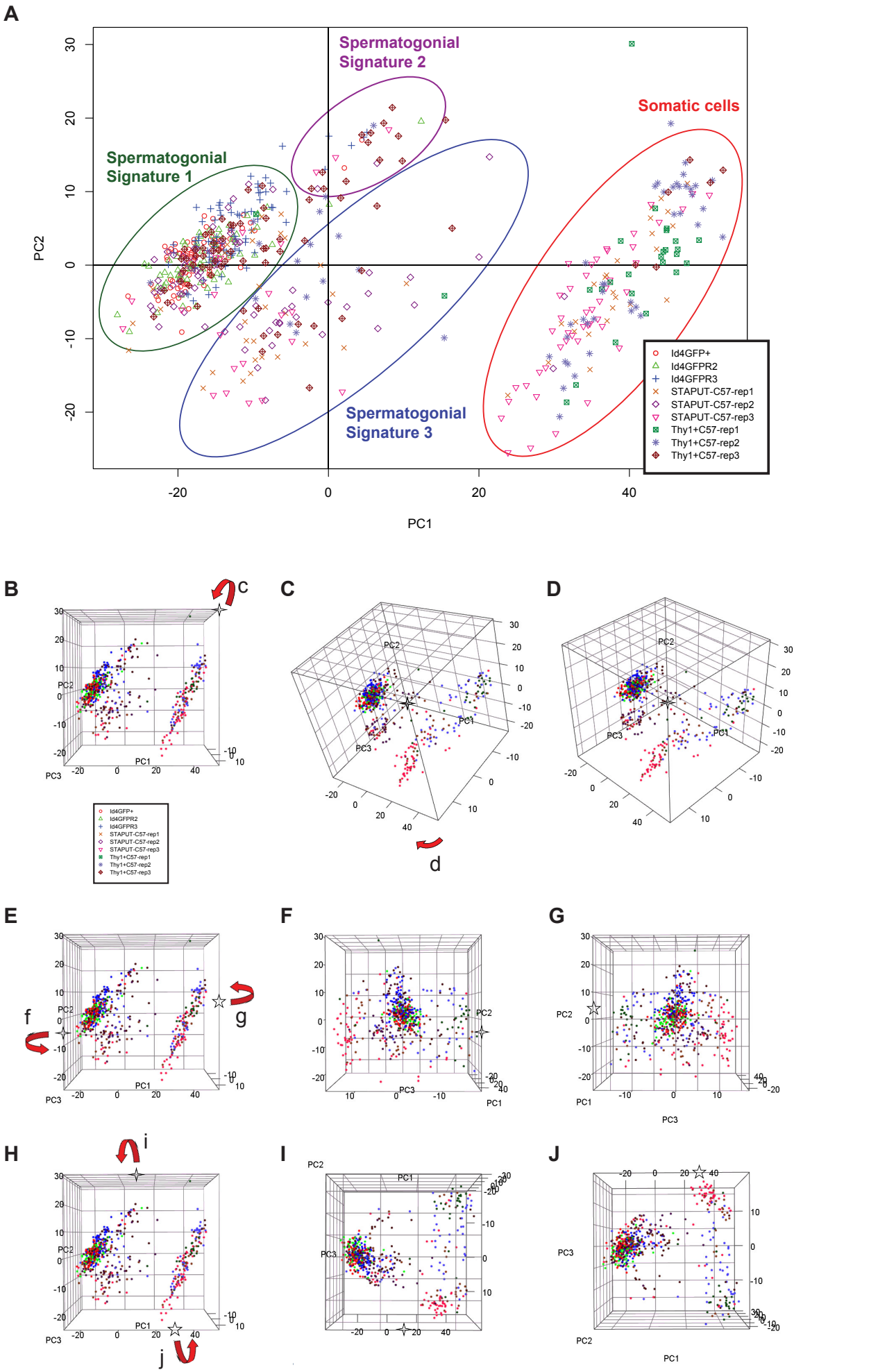
Figure 4. Neonatal spermatogonial heterogeneity. The single-cell gene expression results of this study demonstrate existence of multiple populations of spermatogonia at P6. A portion of this population appears to correspond to differentiating spermatogonia (blue), at least some of which may be destined to later undergo apoptosis, and the remainder contributing to the first wave of spermatogenesis. Multiple populations of undifferentiated spermatogonia were also evident in the P6 testis which may correspond to SSCs (green) and progenitor spermatogonia (violet); future studies will assign functional activities to these subpopulations. The heterogeneity in the germ cell populations that produce the spermatogonial and SSC pool may have earlier origins in germ line development, possibly tracing back to fetal development. Future investigations will need to clarify both the temporal origin of the gene expression heterogeneity observed among neonatal spermatogonia as well as the full extent of this phenomenon (at the whole transcriptome level). Such knowledge may reveal the timing of SSC specification and the mechanisms driving formation of the foundational SSC pool. Diff Sg = differentiating spermatogonia; Sct = spermatocyte; Std = spermatid; Spz = spermatozoa; Apoptotic Sg = apoptotic spermatogonia; and proSg = prospermatogonia.

TABLE 1. Primary and Secondary antibodies.

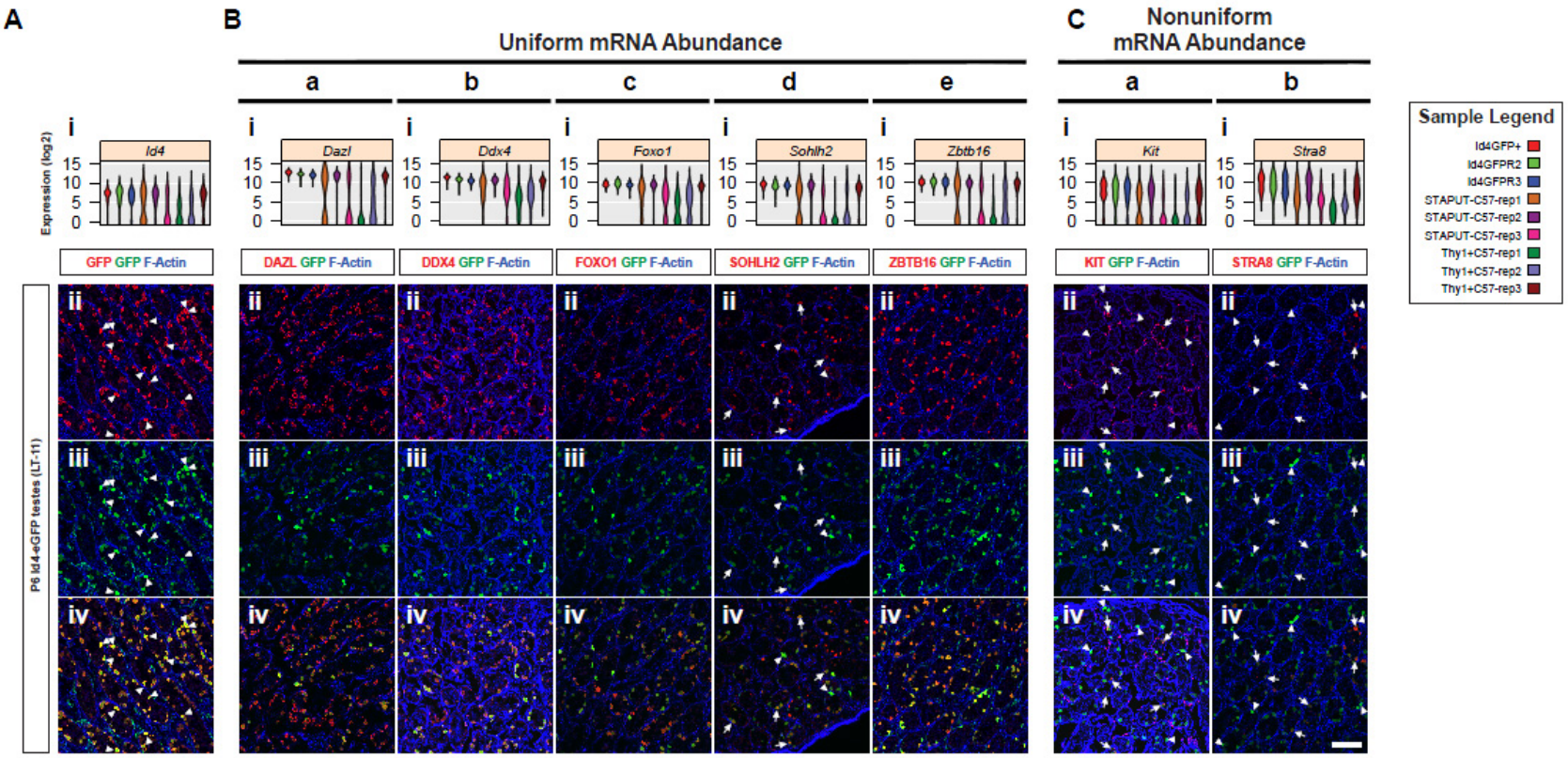
Category	Primary antibody	Secondary antibody (Life Technologies; 1:500)
Controls	GFP (Biovision, 1:200)	Donkey anti-rabbit Alexa Fluor 555
Uniform mRNA abundance	DAZL (Abcam; 1:100)	Donkey anti-rabbit Alexa Fluor 555
	DDX4 (R&D Systems; 1:800)	Donkey anti-goat Alexa Fluor 594
	SOHLH1 (A. Rajkovic; 1:200)	Donkey anti-rabbit Alexa Fluor 555
	SOHLH2 (A. Rajkovic; 1:200)	Donkey anti-guinea pig Alexa Fluor 555
	FOXO1 (Cell Signaling; 1:100)	Donkey anti-rabbit Alexa Fluor 555
	ZBTB16 (Santa Cruz Biotech; 1:500)	Donkey anti-rabbit Alexa Fluor 555
Nonuniform mRNA abundance	KIT (Cell Signaling; 1:1000)	Donkey anti-rabbit Alexa Fluor 555
	STRA8 (Abcam; 1:3000)	Donkey anti-rabbit Alexa Fluor 555



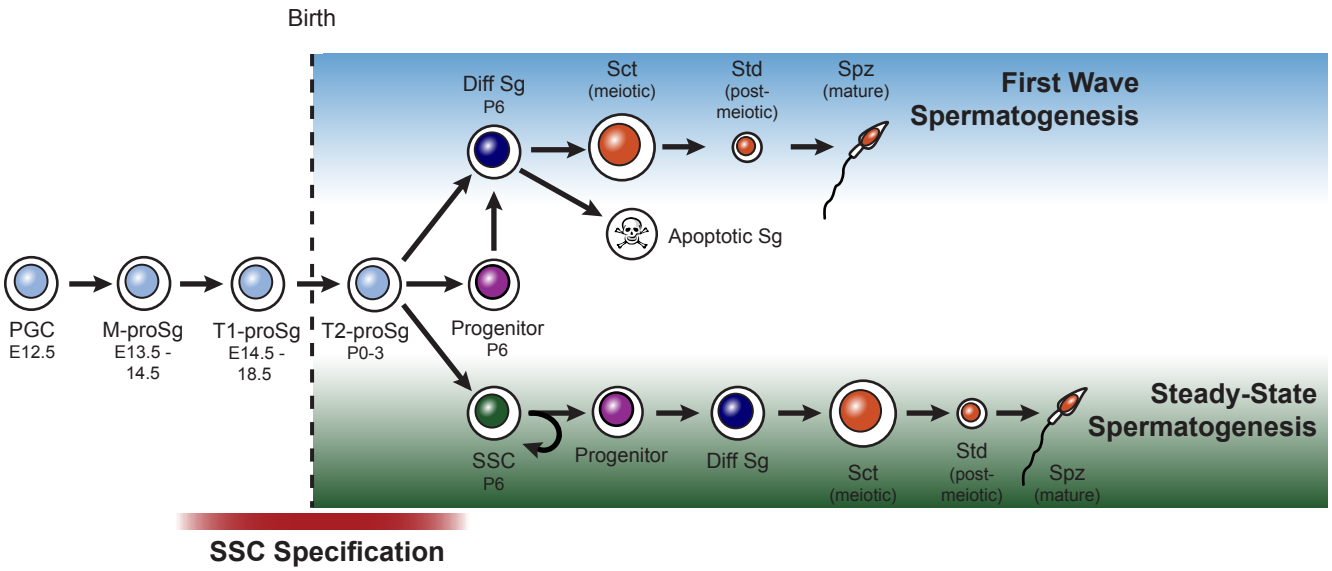
Hermann et al., Figure 1



Hermann et al., Figure 2



Hermann et al., Figure 3



Hermann et al., Figure 4