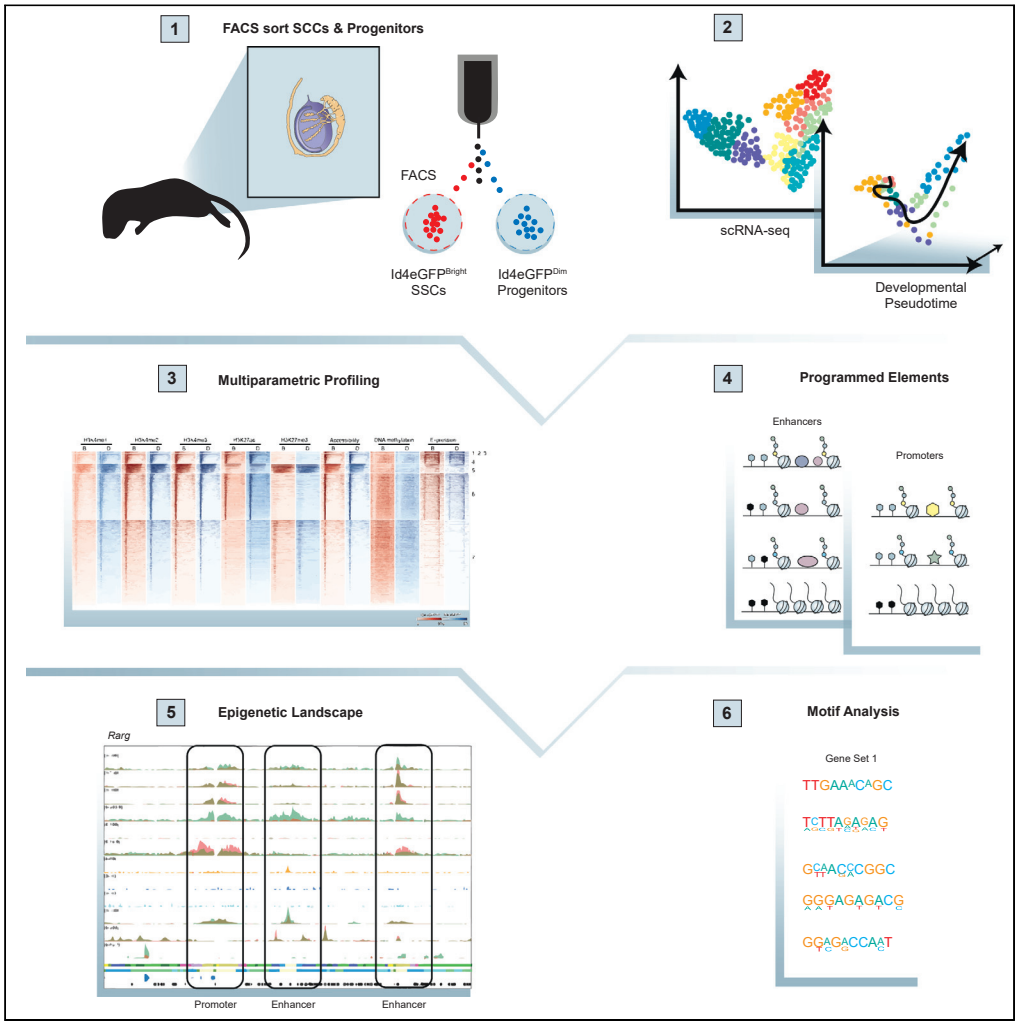
iScience

CellPress
OPEN ACCESS

Article

Unique Epigenetic Programming Distinguishes Regenerative Spermatogonial Stem Cells in the Developing Mouse Testis



Keren Cheng, I-Chung Chen, Ching-Hsun Eric Cheng, ..., Christopher B. Geyer, Jon M. Oatley, John R. McCarrey

john.mccarrey@utsa.edu

HIGHLIGHTS

Spermatogonial stem cells both self-renew and give rise to progenitor spermatogonia

Differential gene expression reflects stem versus differentiating cell fates

Differentially expressed genes show differential epigenetic programming

Motif analysis predicts regulators of cell type-specific epigenetic programming

Cheng et al., iScience 23, 101596
October 23, 2020
https://doi.org/10.1016/j.isci.2020.101596

iScience

CellPress
OPEN ACCESS

Article

Unique Epigenetic Programming Distinguishes Regenerative Spermatogonial Stem Cells in the Developing Mouse Testis

Keren Cheng,[1] I-Chung Chen,[1] Ching-Hsun Eric Cheng,[1] Kazadi Mutoji,[1] Benjamin J. Hale,[2] Brian P. Hermann,[1] Christopher B. Geyer,[2] Jon M. Oatley,[3] and John R. McCarrey[1,4,*]

SUMMARY

Spermatogonial stem cells (SSCs) both self-renew and give rise to progenitors that initiate spermatogenic differentiation in the mammalian testis. Questions remain regarding the extent to which the SSC and progenitor states are functionally distinct. Here we provide the first multiparametric integrative analysis of mammalian germ cell epigenomes comparable with that done for >100 somatic cell types by the ENCODE Project. Differentially expressed genes distinguishing SSC- and progenitor-enriched spermatogonia showed distinct histone modification patterns, particularly for H3K27ac and H3K27me3. Motif analysis predicted transcription factors that may regulate spermatogonial subtype-specific fate, and immunohistochemistry and gene-specific chromatin immunoprecipitation analyses confirmed subtype-specific differences in target gene binding of a subset of these factors. Taken together, these results show that SSCs and progenitors display distinct epigenetic profiling consistent with these spermatogonial subtypes being differentially programmed to either self-renew and maintain regenerative capacity as SSCs or lose regenerative capacity and initiate lineage commitment as progenitors.

INTRODUCTION

An average adult human male produces 85–100 million sperm per day, all of which emanate from the highly proliferative seminiferous epithelium in the testis (Johnson et al., 1980). Within this epithelium spermatogenesis is sustained by spermatogonial stem cells (SSCs), progeny of which either replenish the SSC pool or contribute to the spermatogenic differentiation pathway as transit-amplifying progenitors that give rise to differentiating spermatogonia (de Rooij, 2017). SSCs can be functionally distinguished in mice on the basis of their regenerative capacity measurable by a quantifiable transplantation assay (Brinster and Avarbock, 1994; Kubota and Brinster, 2018) analogous to that reliably used for decades to identify hematopoietic stem cells (Lewis and Trobaugh, 1964).

In the postnatal mouse testis, prospermatogonia give rise to spermatogonia, exhausting the prospermatogonia population by postnatal day 5 (P5) (Drumond et al., 2011; Kluin et al., 1982; Yoshida et al., 2006). By P6-8, distinct spermatogonial subtypes (SSC, progenitor, differentiating) can be distinguished on the basis of specific marker proteins (Buaas et al., 2004; Chan et al., 2014; Valli et al., 2014), lineage tracing (Hara et al., 2014; Sun et al., 2015), bulk (Helsel et al., 2017), and single-cell (sc) (Chen et al., 2018; Ernst et al., 2019; Green et al., 2018; Grive et al., 2019; Guo et al., 2018; Hermann et al., 2018; Jung et al., 2019; Law et al., 2019; Liao et al., 2019; Sohni et al., 2019; Velte et al., 2019) RNA-sequencing (RNA-seq), or transplantation analysis (McLaren, 2003; Shinohara et al., 2000). SSCs retain regenerative capacity and divide to either self-renew or generate progenitors that lose regenerative capacity as they become primed to initiate spermatogenic differentiation giving rise to differentiating spermatogonia (Drumond et al., 2011; Kluin et al., 1982; Yang et al., 2013).

In the developing testis of *Id4-eGfp* transgenic mice, eGFP marks undifferentiated spermatogonia (Chan et al., 2014), and selective FACS-based recovery of the brightest (ID4-eGFP^{Bright}) and dimmest (ID4-eGFP^{Dim}) portions of the ID4-eGFP+ population significantly enriches regenerative SSCs or non-regenerative progenitors,

[1]Department of Biology, University of Texas at San Antonio, San Antonio, TX, USA

[2]Department of Anatomy and Cell Biology at the Brody School of Medicine and East Carolina Diabetes and Obesity Institute, East Carolina University, Greenville, NC, USA

[3]Center for Reproductive Biology, School of Molecular Biosciences, College of Veterinary Medicine, Washington State University, Pullman, WA, USA

[4]Lead Contact

*Correspondence: john.mccarrey@utsa.edu

https://doi.org/10.1016/j.isci.2020.101596

respectively (Helsel et al., 2017). Collectively, ID4-eGFP^{Bright} spermatogonia from the P6-8 testis display characteristics of SSCs, including (1) a 5.5-fold higher colonization capacity following transplantation into a recipient testis (Helsel et al., 2017), (2) a distinct transcriptome that includes elevated expression of known SSC-expressed genes (*Gfra1, Id4, Bcl6b, Etv5*) as well as genes encoding factors favoring self-renewal and maintenance of a stem cell state including SSC marker transcripts such as *Sall4* and genes involved in cell cycle control and replication such as *Ccnb1, Cdc20, Cdk1, Mcm5*, and *Pcna* (Helsel et al., 2017; Hermann et al., 2018; Law et al., 2019; Mutoji et al., 2016), (3) elevated expression of known SSC marker proteins (GFRA1, CDH1, ZBTB16, ID4) (Chan et al., 2014; Niedenberger et al., 2015), (4) absence of differentiation markers (KIT, STRA8, RARG, SOHLH2) (Hermann et al., 2015), and (5) insensitivity to induction of differentiation by retinoic acid (RA) (Velte et al., 2019). In contrast, ID4-eGFP^{Dim} spermatogonia at P6-8 are enhanced for characteristics of progenitors and early differentiating spermatogonia including (1) a significant depletion of transplantable colonizing capacity (Helsel et al., 2017; Hermann et al., 2018; Law et al., 2019), (2) a transcriptome featuring elevated expression of genes associated with proliferation and commitment to spermatogenic differentiation (*Neurog3, Lin28a, Sohlh1, Rarg, Kit*) (Helsel et al., 2017; Hermann et al., 2018; Mutoji et al., 2016), (3) elevated expression of proteins unique to spermatogonia committed to differentiation (RARG, KIT, STRA8, SOHLH2, NDRG4) (Hermann et al., 2015; Lord et al., 2018), and (4) responsiveness to induction of differentiation by RA (Velte et al., 2019).

These characteristics indicate that progeny of SSCs adopt alternative fates manifest as retention of the SSC phenotype or adoption of the progenitor phenotype based on differential gene expression. The ENCODE and NIH Roadmap Epigenomics Mapping consortia reported distinct transcriptomes accompanied by up to 15 unique cell type-specific epigenetic programming profiles at promoters and enhancers for >100 different somatic cell types in mammals (ENCODE Project Consortium, 2012; Ernst et al., 2011; Gifford et al., 2013; Kellis et al., 2014; Ram et al., 2011; Roadmap Epigenomics Consortium et al., 2015; Shema et al., 2019; Zhu et al., 2013). However, none of these studies examined germ cells. We reasoned that similarly distinguishable epigenomic programming profiles should be associated with differentially expressed genes (DEGs) in SSCs and progenitors. We used FACS to selectively recover ID4-eGFP^{Bright} "SSC-enriched" spermatogonia and ID4-eGFP^{Dim} "progenitor-enriched" spermatogonia and performed multiparametric integrative analysis of genome-wide patterns of DNA methylation, six different histone modifications, and chromatin accessibility in conjunction with subtype-specific bulk and single-cell transcriptome analyses to identify unique epigenetic landscapes associated with DEGs in each spermatogonial subtype. We then performed *de novo* motif analysis followed by immunohistochemistry (IHC) and chromatin immunoprecipitation (ChIP) to identify candidate factors that may either directly establish or subsequently mediate the effects of differential epigenetic programming of spermatogonial-subtype specific genes. Our results provide unprecedented insight into the multiparametric epigenetic programming associated with DEG patterns that distinguish SSC- and progenitor-enriched spermatogonia and suggest that SSCs represent a unique spermatogonial subtype epigenetically programmed to retain SSC function, whereas progenitors have transitioned to a subtype associated with lineage commitment and spermatogenic differentiation.

RESULTS

Triplicate samples of regenerative SSC-enriched and non-regenerative progenitor-enriched spermatogonia were selectively recovered from testes of P6 *Id4-eGfp* transgenic mice by FACS sorting for relative eGFP epifluorescence as previously described (Helsel et al., 2017). Each epigenomic assay was run on identical aliquots of each replicate sample of ID4-eGFP^{Bright} and ID4-eGFP^{Dim} cells to assess genome-wide patterns of DNA methylation, six different histone modifications, chromatin accessibility, and gene expression, rendering results of each assay directly comparable with one another. Importantly, our previous scRNA-seq data demonstrated that FACS-based recovery of these subpopulations very effectively excluded contamination by somatic cells to <0.2% in the ID4-eGFP^{Bright} subpopulation and <1.3% in the ID4-eGFP^{Dim} subpopulation (Hermann et al., 2018).

Differential Gene Expression Distinguishes SSC- and Progenitor-Enriched Spermatogonial Subpopulations

Previous bulk and scRNA-seq analyses of ID4-eGFP^{Bright} SSC- and ID4-eGFP^{Dim} progenitor/differentiating spermatogonia-enriched (hereafter referred to as progenitor-enriched) spermatogonial subpopulations from the developing testis revealed distinct patterns of gene expression in each (Helsel et al., 2017; Hermann et al., 2018). Here, we first conducted bulk RNA-seq on aliquots of the same samples on which we also conducted bulk epigenomics analyses (Figure 1A). The bulk RNA-seq data revealed statistically

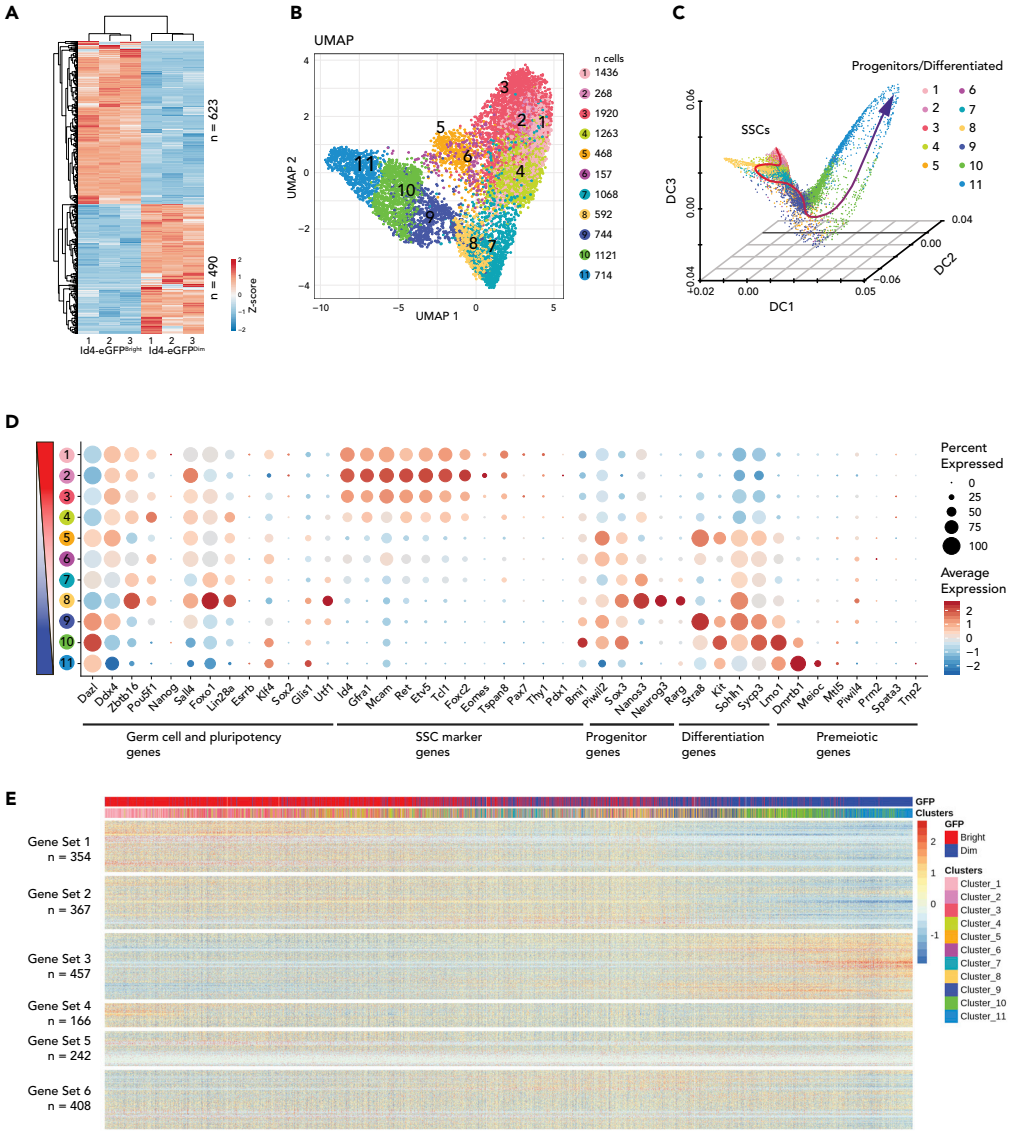


Figure 1. Differential Gene Expression in SSC-Enriched and Progenitor-Enriched Spermatogonial Subpopulations

(A) Differential gene expression profiling of ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia by bulk RNA-seq. Genes with >1.5 LFC (Log2 Fold Change difference >1.5, $p < 0.01$) were hierarchically clustered. Red–blue colors indicate high–low expression levels in *Z* score. Gene-specific results of bulk RNA-seq can be found in Table S1.

(B) Subpopulations among P6 ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia are revealed in this UMAP display of single-cell RNA-seq (scRNA-seq) data. Each dot represents a single cell. The cell number in each cluster is indicated via color coding.

(C) Developmental trajectory of neonatal spermatogonia. A dynamic diffusion map shows the developmental progression of ID4-eGFP^{Bright} to ID4-eGFP^{Dim} spermatogonia in pseudotime.

(D) Scaled average expression of selected marker genes and the percentage of cells within each cluster with detectable expression (dot radius) for each selected marker gene within each cell cluster. The cell cluster order was arranged by Diffusion Map from top to bottom.

(E) Scaled, normalized expression of the 2,000 most variably expressed genes in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia. Cells are ordered along their pseudotime developmental progression shown in Figure 1C. The top bar indicates the progression from ID4-eGFP^{Bright} (red) to ID4-eGFP^{Dim} (blue) cells based on intensity of ID4-eGFP expression. The lower bar represents the cell clusters identified according to the color code shown in Figure 1B. These 2,000 most variably expressed genes were hierarchically clustered into six gene sets based on expression patterns throughout the pseudotime progression. A detailed list of variably expressed genes is provided in Table S1.

significant ($p < 0.01$) differential gene expression patterns confirming significant enrichment of SSCs in the ID4-eGFP^{Bright} subpopulation and of progenitors in the ID4-eGFP^{Dim} subpopulation that was further corroborated by differential expression of known marker genes consistent with previous reports (Helsel et al., 2017; La et al., 2018). SSC marker transcripts including *Gfra1, Lhx1, Zbtb16, Etv5, Ret, T,* and *Bcl6b* were expressed at significantly higher levels in ID4-eGFP^{Bright} cells, whereas markers of progenitors including *Sox3, Nanos3, Neurog3, Rarg,* and *Kit* were expressed at higher levels in the ID4-eGFP^{Dim} cells (Figures 1D and S1A). Importantly, the absence of expression of genes recently reported to be markers of prospermatogonia, including *Etv4, Mmp9*, and *Ttc28* (Tan et al., 2020), supports the contention that, by P5, essentially all prospermatogonia have transitioned to undifferentiated spermatogonia (Drumond et al., 2011; Hilscher et al., 1974; Kluin and Rooij, 1981; Roosen-Runge and Leik, 1968; Wartenberg, 1976). A complete list of genes differentially expressed in the ID4-eGFP^{Bright} and ID4-eGFP^{Dim} subpopulations is provided in Table S1.

We next mined our previously described scRNA-seq data (Hermann et al., 2018) to further delineate specific spermatogonial subpopulations among either the ID4-eGFP^{Bright} or ID4-eGFP^{Dim} populations (Figure 1B). We delineated 11 different clusters, 5 of which (clusters 1–5) were enriched in ID4-eGFP^{Bright} cells and the other 6 of which (6–11) were enriched in ID4-eGFP^{Dim} cells (Figure 1B). These clusters were then ordered by pseudotime analysis, which indicated that ID4-eGFP^{Bright} cells precede ID4-eGFP^{Dim} cells developmentally (Figure 1C). This developmental progression was further confirmed by differential expression of transcripts encoding markers known to initiate expression in prospermatogonia, SSCs, progenitors, differentiating spermatogonia, or premeiotic cells (Figure 1D). Finally, in addition to identifying clusters of spermatogonial subpopulations based on our scRNA-seq data, we also identified six distinct gene sets distinguished by their expression patterns among the different spermatogonial clusters (Figure 1E). Together the various approaches to assessing variation among spermatogonia in the developing testis confirm that this is a developmentally dynamic population of cells that is beginning to subdivide into distinct spermatogonial subtypes including SSCs and progenitors.

An analysis of the representation of transcripts characteristic of different phases of the cell cycle revealed some difference in representation of the M and M/G1 phases in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} cells, which could potentially reflect differences in cell division rates between SSCs and progenitors (Oatley and Brinster, 2008). However, we otherwise detected a very similar distribution of transcripts associated with the G1/S and S phases in these subpopulations (Figure S1E), consistent with previous assessments of cell cycle phase distributions in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia in the developing testis (Mutoji et al., 2016; Hermann et al., 2018). These observations were further corroborated by a cell cycle phase-specific DNA content analyses of live P6 ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia, which confirmed that both subpopulations contained cells in all phases of the cell cycle (Figure S1E). This analysis also detected a small but statistically significant difference in the proportion of cells in G1/G0 in each subpopulation ($p = 0.039$). Finally, GO analyses of the six differentially expressed gene sets revealed that sets 1 and 2 were enriched with genes related to stem cell maintenance, whereas set 3 was enriched for genes associated with spermatogenesis, oogenesis, and cell cycle (Figure S1E). Gene sets 4 and 5 were also enriched for cell cycle genes as well as genes for DNA replication and DNA repair, and gene set 6 was enriched for genes involved in epigenetic modifications. Taken together, these data reveal an ordered progression between SSC- and progenitor-enriched spermatogonial subpopulations that is consistently characterized by differential gene expression patterns related to a variety of cellular functions including stem cell maintenance, cell cycle, gametogenesis, and epigenetic reprogramming, but no one function appears to be disproportionately responsible for a majority of the differential gene expression.

Multiparametric Integrative Analysis Delineates Distinct Epigenetic Landscapes in Spermatogonial Subpopulations

We analyzed aliquots of the same samples of each spermatogonial subtype by (1) ChIP-seq to detect six different histone modifications—H3K4me1,2,3, H3K9me3, H3K27ac, and H3K27me3; (2) ATAC-seq to assess chromatin accessibility; and (3) MeDIP-seq to examine DNA methylation (Figure 2A). We chose to examine all three forms of H3K4 methylation because all tend to inhibit DNA methylation in the same region, but H3K4me1&2 tend to mark enhancers, whereas H3K4me3 tends to mark promoters. We examined H3K4me3 and H3K27me3 because the simultaneous presence of these marks is indicative of poised genes (Guo et al., 2017; Lesch et al., 2013, 2016). We assessed H3K27me3 and H3K27ac because these are repressive and active marks, respectively, that can function as a switch to promote expression or repression of

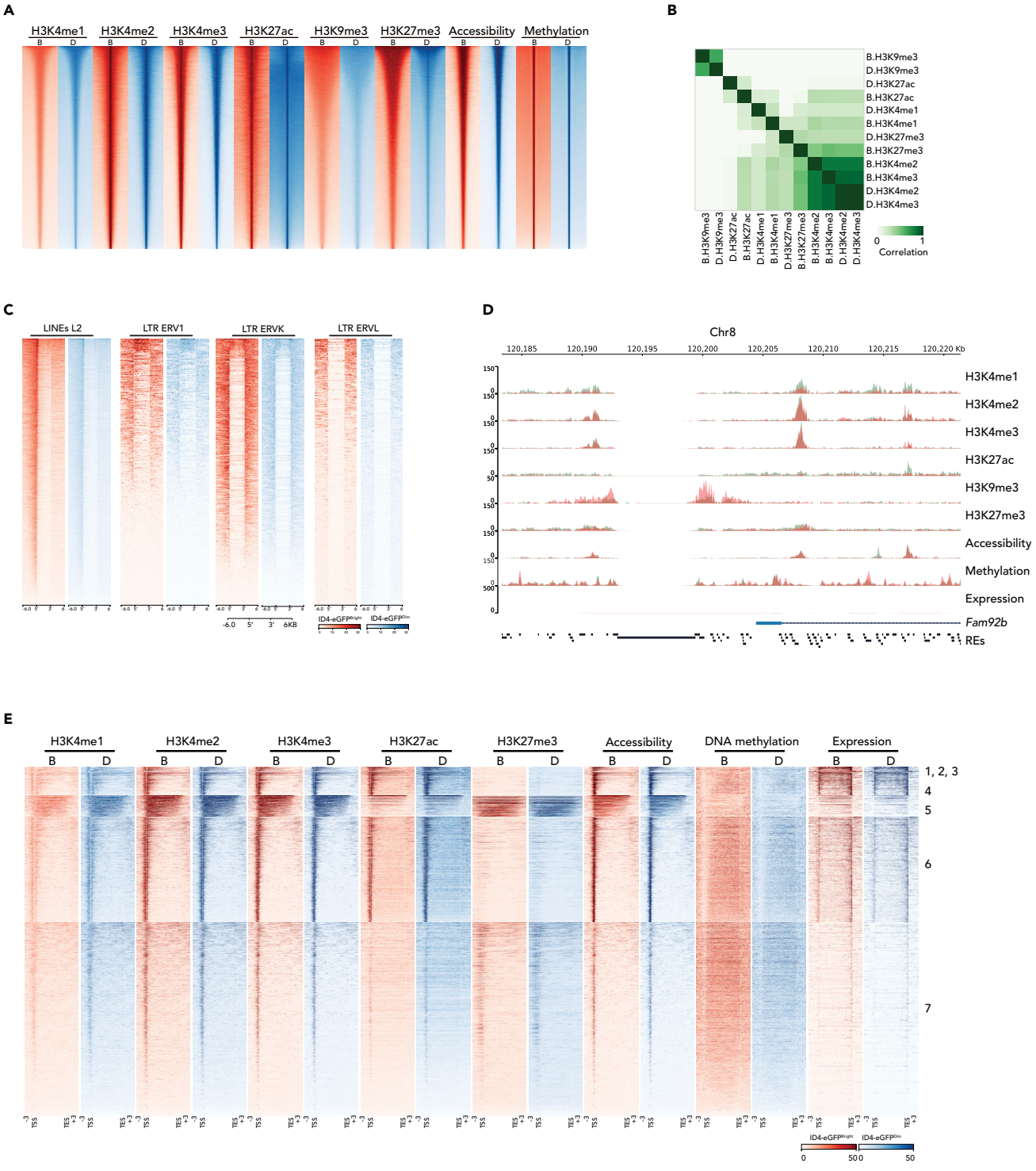


Figure 2. Epigenetic Profiling of Genic Regions in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} Spermatogonial Subpopulations

(A) A heatmap shows peaks of histone modifications (H3K4me1, H3K4me2, H3K4me3, H3K9me1, H3K27ac, and H3K27me3) deduced by ChIP-seq, chromatin accessibility deduced by ATAC-seq, and DNA methylation deduced by MeDIP-seq in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia. Red color indicates reads from ID4-eGFP^{Bright} B cells; blue color indicates reads from ID4-eGFP^{Dim} D cells.

(B) A heatmap depicts correlations among individual histone modification patterns in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia.

(C) Heatmaps of H3K9me3 deposition in four types of repeats—LINEs(*L2*) and LTRs (ERV1, ERVK, ERVL) in ID4-eGFP^{Bright} (red) and ID4-eGFP^{Dim} (blue) spermatogonia.

Figure 2. *Continued*

(D) A genome browser snapshot showing sequencing reads of six histone modifications, chromatin accessibility, DNA methylation, and transcript expression in ID4-eGFP^{Bright} (coral colored tracks) and ID4-eGFP^{Dim} spermatogonia (green colored tracks). The location of a gene that is expressed in both ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia (*Fam92b*) as well as that of repeat elements is shown at the bottom of this browser image.

(E) Epigenetic profiles within genic regions (TSS – TES + 3 kb upstream and 3 kb downstream) of all mouse Refseq annotated genes (n = 24,012) in ID4-eGFP^{Bright} (red tracks, B) and ID4-eGFP^{Dim} (blue tracks, D) spermatogonia together with corresponding transcript expression tracks. K-means clustering revealed seven distinct epigenetic profiles, including three that are too small to visually resolve (1–3). The order of genes represented in these tracks is shown in Table S2.

transcription (Katoh et al., 2018). Finally, we examined H3K9me3 because it tends to mark regions of the genome other than promoters or enhancers, especially repeat elements and heterochromatin (Becker et al., 2016). Assessment of the relationships among these parameters revealed that, as has been observed in somatic cell epigenomes (ENCODE Project Consortium, 2012; Zacher et al., 2017), patterns of H3K4me1,2,3 were highly correlated with one another in all spermatogonia, whereas H3K27me3 showed less correlation with H3K4me1,2,3 and H3K27ac showed higher correlation with H3K4me1,2,3, and high discordance with H3K27me3 (Figure 2B). Patterns of H3K9me3 peaks showed very little co-localization with any of the other histone modification peaks, presumably reflecting disparate functions of these modifications. Indeed, we found H3K9me3 peaks were enriched at repetitive elements such as LINEs, LTRs (ERV1, ERVK, ERVL) (Figures 2C, S2A, and S2C), and at 5′ AND-3′ UTRs of transposons, presumably contributing to repression of transposon mobility (Figure 2D). A majority of histone modifications were found in distal intergenic regions; however, peaks were also detected at or flanking gene promoters (Figure S2B). Within genic regions, peaks of chromatin accessibility and absence of DNA methylation correlated with peaks of H3K4me1,2,3 and H3K27ac, especially at gene promoters (Figures 2D and 2E). These data demonstrate that spermatogonia utilize complex combinations of epigenetic programming parameters to regulate gene expression in a manner similar to those reported to be used in numerous different somatic cell types by the ENCODE and NIH Roadmap Epigenomics Mapping consortia (ENCODE Project Consortium, 2012; Ernst et al., 2011; Gifford et al., 2013; Kellis et al., 2014; Ram et al., 2011; Roadmap Epigenomics Consortium et al., 2015; Shema et al., 2019; Zhu et al., 2013).

Genic Region Patterns of Chromatin Modifications Distinguish Genes Expressed in SSC- and Progenitor-Enriched Spermatogonial Subpopulations

K-means clustering of genic region data revealed seven different patterns of histone modifications (Figure 2E). Four modifications (H3K4me1,2,3 & H3K27ac) were enriched in genes expressed in one or both spermatogonial subpopulations, predominantly in promoter regions (Figure 2E, clusters 1,2,3,4,6), as has been observed in multiple somatic cell types (ENCODE Project Consortium, 2012; Zacher et al., 2017). Genes that were either not expressed or expressed at very low levels in one or both spermatogonial subpopulations (clusters 5,7) showed enrichment of the repressive H3K27me3 modification and depletion of the active H3K27ac modification, both within gene bodies (cluster 5) and at promoters (cluster 7). Repressed genes showed enrichment of H3K4me1,2,3 and H3K27me3 within gene bodies or downstream genic regions and lacked histone modification peaks at promoter regions. A GO analysis of gene promoters enriched for either H3K27ac or H3K27me3 was consistent with differential gene expression favoring enhanced maintenance of the stem cell state in SSCs and of lineage priming in progenitors (Figure S2D).

Exemplary sets of genes up-regulated in SSC-enriched spermatogonia (*Tspan8, Pax7, Lhx1, Egr2, Gfra1, Id4, Tcl1*) and genes up-regulated in progenitor-enriched spermatogonia (*Dmrtb1, Neurog3, Rarg, Kit, Lmo1, Crabp1*) illustrate differences in promoter programming associated with spermatogonial subtype-specific gene expression. Collectively, our assessments revealed that, in each respective spermatogonial subpopulation, promoters of up-regulated genes showed (1) enriched H3K4me1,2,3; (2) enriched H3K27ac; (3) depleted H3K27me3; (4) enhanced chromatin accessibility; and (5) hypomethylated DNA, whereas promoters of down-regulated genes showed (1) decreased H3K4me1,2,3; (2) depleted H3K27ac; (3) enriched H3K27me3; (4) decreased chromatin accessibility; and (5) hypomethylated DNA (Figures 3, S3A, and S3B). Thus, within genic regions, differential enrichment of H3K27ac or H3K27me3 in promoter regions correlated most closely with regulation of DEGs distinguishing ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia (Figures 3I and S3C–S3F). Notably, the correlations between differential patterns of H3K4me1/2/3 and H3K27me3 in promoters of genes that were up- or down-regulated in each spermatogonial subtype, respectively, were highly significant ($p < 2.2^{-16}$) (Figures S3A and S3B).

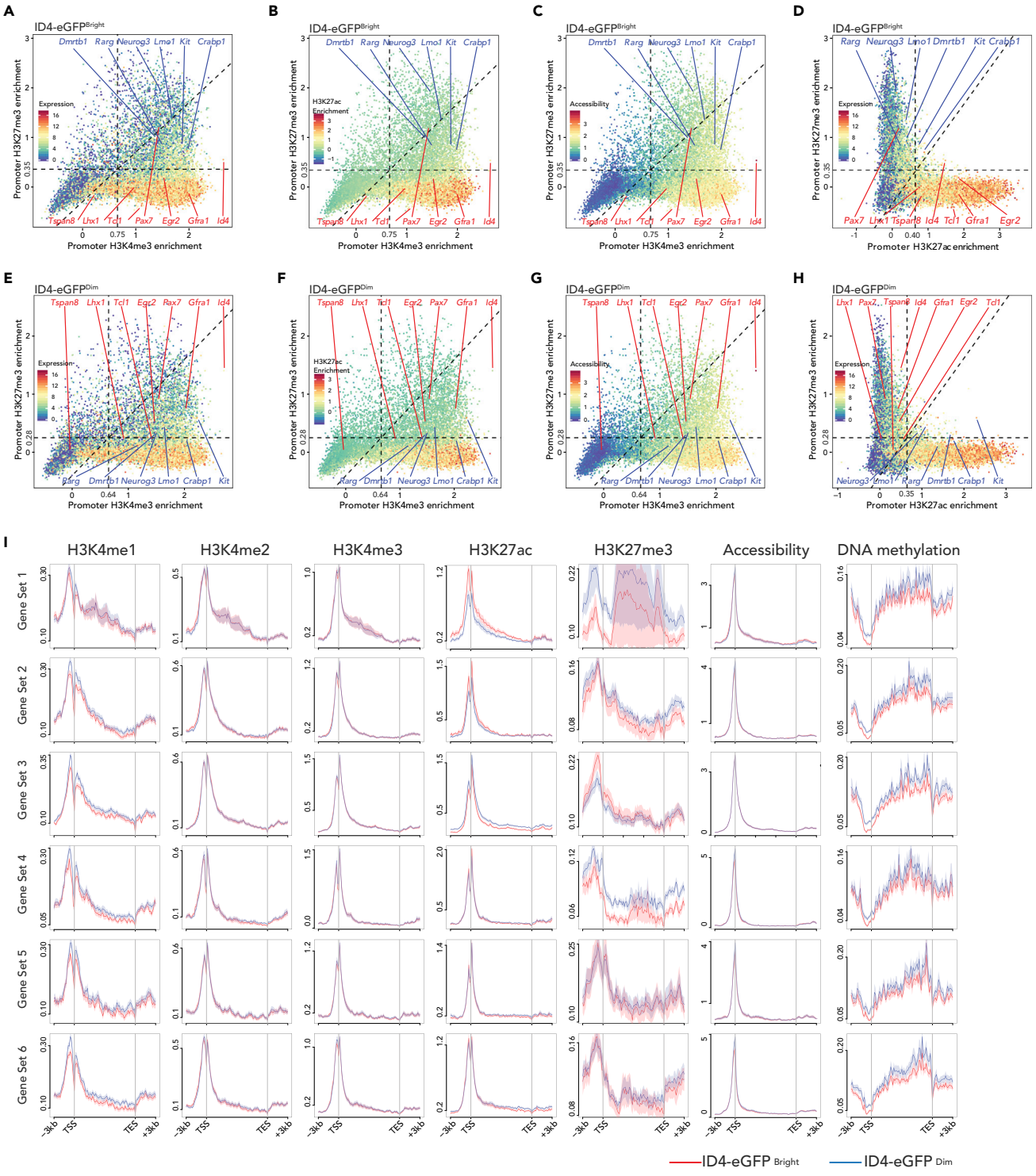


Figure 3. Epigenetic Profiling at Promoters in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} Spermatogonia

(A–H) Scatterplots correlating H3K4me3 (x axis) and H3K27me3 (y axis) enrichments (log2RPKM) at promoters (TSS ±500 bp) in ID4-eGFP^{Bright} (A–C) and ID4-eGFP^{Dim} (E–G) spermatogonia. Color codes indicate gene expression levels (A, E), H2K27ac enrichment (B and F), and chromatin accessibility (C and G) at promoters. H3K27ac/H3K27me3 double ChIP-seq signal enrichment correlated with gene expression levels at promoters in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia (D and H).

(I) Comparisons of histone modification, chromatin accessibility, and DNA methylation levels (RPKM, Reads Per Kilobase of sequence range per Million mapped reads) in the most differentially expressed gene sets in ID4-eGFP^{Bright} (red trace) and ID4-eGFP^{Dim} (blue trace) spermatogonia.

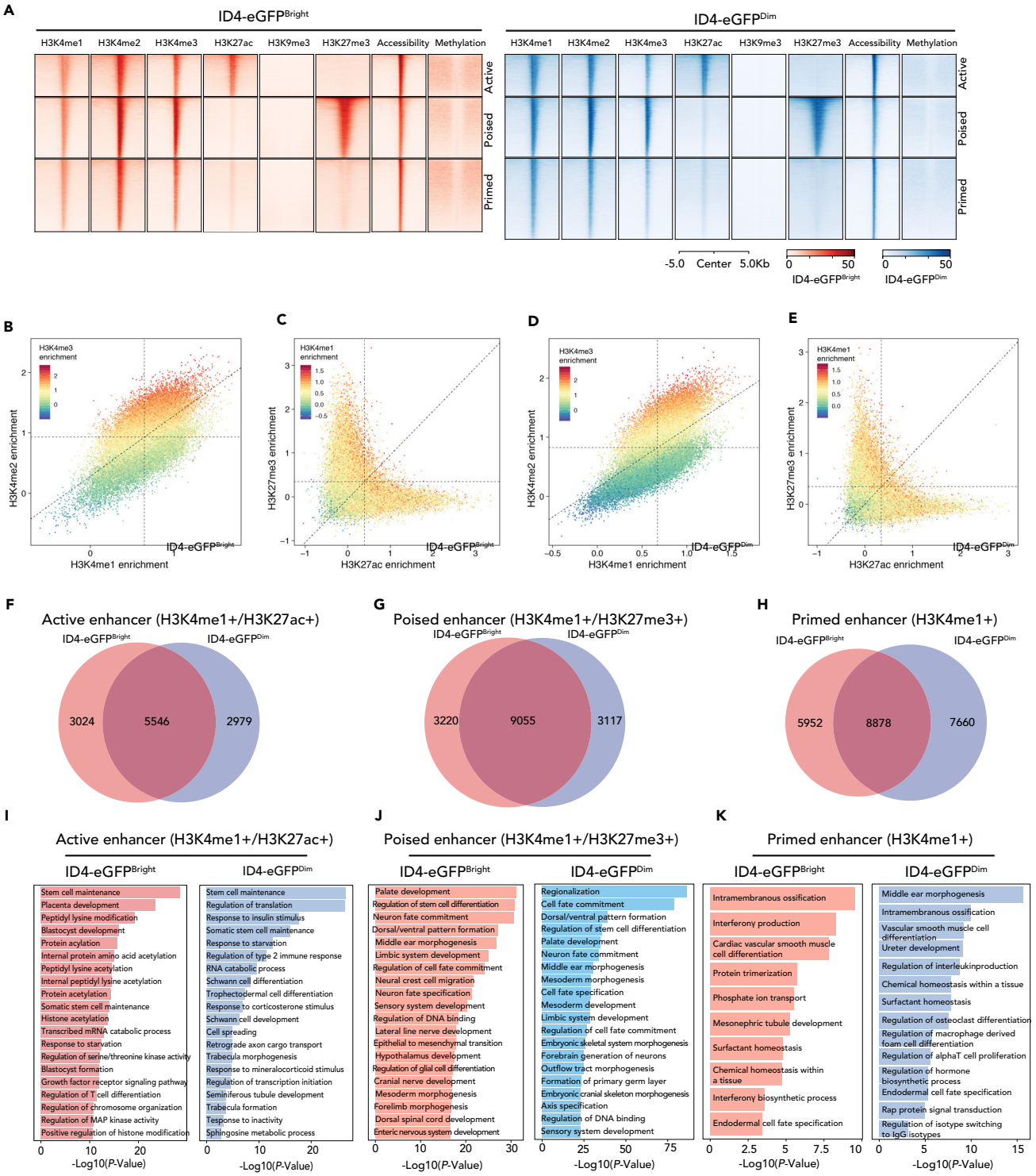


Figure 4. Epigenetic Profiling at Enhancers in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} Spermatogonia

(A) Heatmaps show profiles of each set of histone modification, chromatin accessibility, and DNA methylation peaks at sites of intergenic enhancers in ID4-eGFP^{Bright} (red) and ID4-eGFP^{Dim} (blue) spermatogonia. Peak profiles are shown for active, poised, and primed enhancers in each spermatogonial subtype. Genomic coordinates of enhancers shown in this figure are listed in Table S3.

(B–E) Scatterplots showing positive or negative combinatorial enrichment of different histone modifications at enhancers. Dashed lines indicate the minimum threshold value indicative of enrichment within each bimodal distribution. Each dot is 1 enhancer. (B) Enrichment of the H3K4me3 modification

Figure 4. *Continued*

correlates positively with enrichment of the H3K4me2 and H3K4me1 modifications. (C) Enrichment of the H3K4me1 modification correlates positively with enrichment of either the H3K27me3 or the H3K27ac modifications, but enrichment of the H3K27me3 modification correlates negatively with enrichment of the H3K27ac modification, and vice versa. (D) Enrichment of the H3K4me3 modification correlates positively with enrichment of either the H3K27me3 or the H3K27ac modifications, but enrichment of the H3K27me3 modification correlates negatively with enrichment of the H3K27ac modification, and vice versa.

(F–H) Venn diagrams show proportions of active (F), poised (G), and primed (H) enhancers unique to ID4-eGFP^{Bright} spermatogonia, common to both ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia, or unique to ID4-eGFP^{Dim} spermatogonia.

(I–K) GREAT GO analysis of functions encoded by genes associated with active (I), poised (J), or primed (K) enhancers in ID4-eGFP^{Bright} or ID4-eGFP^{Dim} spermatogonia.

Intergenic Enhancers Are Differentially Programmed in SSC- and Progenitor-Enriched Spermatogonial Subpopulations

ATAC-seq peaks identify open chromatin regions indicative of promoters and enhancers (Buenrostro et al., 2013). Nearly 50% of accessible chromatin regions resided in distal intergenic regions in both ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia (Figure S2B), and nearly all intergenic ATAC-seq peaks co-localized with peaks of H3K4me1,2,3 enrichment and hypomethylated DNA (Figure 2D), indicative of enhancers (Calo and Wysocka, 2013). We integrated histone ChIP-seq, MeDIP-seq, and ATAC-seq data to identify three distinct types of enhancers: (1) active enhancers marked by H3K27ac; (2) primed enhancers marked by H3K4me1/2; and (3) poised enhancers marked by H3K4me1 and H3K27me3 (Figure 4A). As at promoters, H3K4me1,2,3 marks co-localized within enhancers (Figures 4B and 4D), whereas H3K27me3 or H3K27ac marks appeared to be mutually exclusive in both ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia (Figures 4C and 4E). For each type of enhancer, we observed significant overlap in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia, but we also observed active, poised, and primed enhancers unique to either ID4-eGFP^{Bright} or ID4-eGFP^{Dim} cells (Figures 4F–4H). We next utilized GREAT for Gene Ontology (GREAT-GO) analysis (McLean et al., 2010) to identify categories of biological functions of genes predicted to be associated with these different types of enhancers (Figures 4I–4K). Active enhancers in ID4-eGFP^{Bright} spermatogonia were associated with genes that were enriched for stem cell maintenance functions (e.g., *Dappa2, Eomes, Lin28a, Pou5f1, Nodal*, and *Tcl1*) and other early developmental functions. In ID4-eGFP^{Dim} spermatogonia we observed enrichment of active enhancers associated with differentiation functions plus some enrichment of active enhancers associated with stem cell maintenance. As expected, genes associated with active enhancers within each spermatogonial subtype were up-regulated in that subtype (Figure S4). These results confirm that, as in somatic cells, individual genes in spermatogonia are typically regulated by a single promoter plus multiple enhancers and that epigenetic programming of enhancers is more variable than that of promoters (Shen et al., 2012). These results also indicate that differential programming of enhancers is a primary distinction between SSC-enriched ID4-eGFP^{Bright} and progenitor-enriched ID4-eGFP^{Dim} spermatogonia.

Genes Differentially Expressed in SSCs and Progenitor Spermatogonia Show Similarly Reduced Promoter Region DNA Methylation in Both Spermatogonial Subtypes

Our MeDIP-seq analysis of genic region DNA methylation patterns revealed constitutively hypomethylated promoters of DEGs in SSC- and progenitor-enriched spermatogonia, respectively (Figures 3A and 3J). We observed >300,000 MeDIP-seq peaks genome wide in ID4-eGFP^{Bright} and/or ID4-eGFP^{Dim} spermatogonia, of which ~62% were common to both spermatogonial subtypes, ~25% were unique to ID4-eGFP^{Bright} spermatogonia and the remaining ~13% were unique to ID4-eGFP^{Dim} spermatogonia (Figure 5A). The latter two categories reflect the 1,486 differentially methylated regions (DMRs) we observed between ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia (Figure 5D). In genic regions, DNA methylation occurred within gene bodies, in both exons and introns (Figure 5B). Genome wide, a majority of the MeDIP-seq peaks occurred in distal intergenic regions or introns (Figure 5C). GREAT-GO analysis of DMRs showed that many were predicted to be associated with genes enriched in cell fate commitment and stem cell population maintenance (Figure 5E).

De Novo Motif Analysis Reveals Potential Regulators of Differential Epigenetic Programming Associated with Distinct Spermatogonial Subtypes

We performed *de novo* motif analysis of promoter and enhancer regions associated with genes expressed differentially or constitutively in each spermatogonial subpopulation (Figure 6). We performed *de novo*

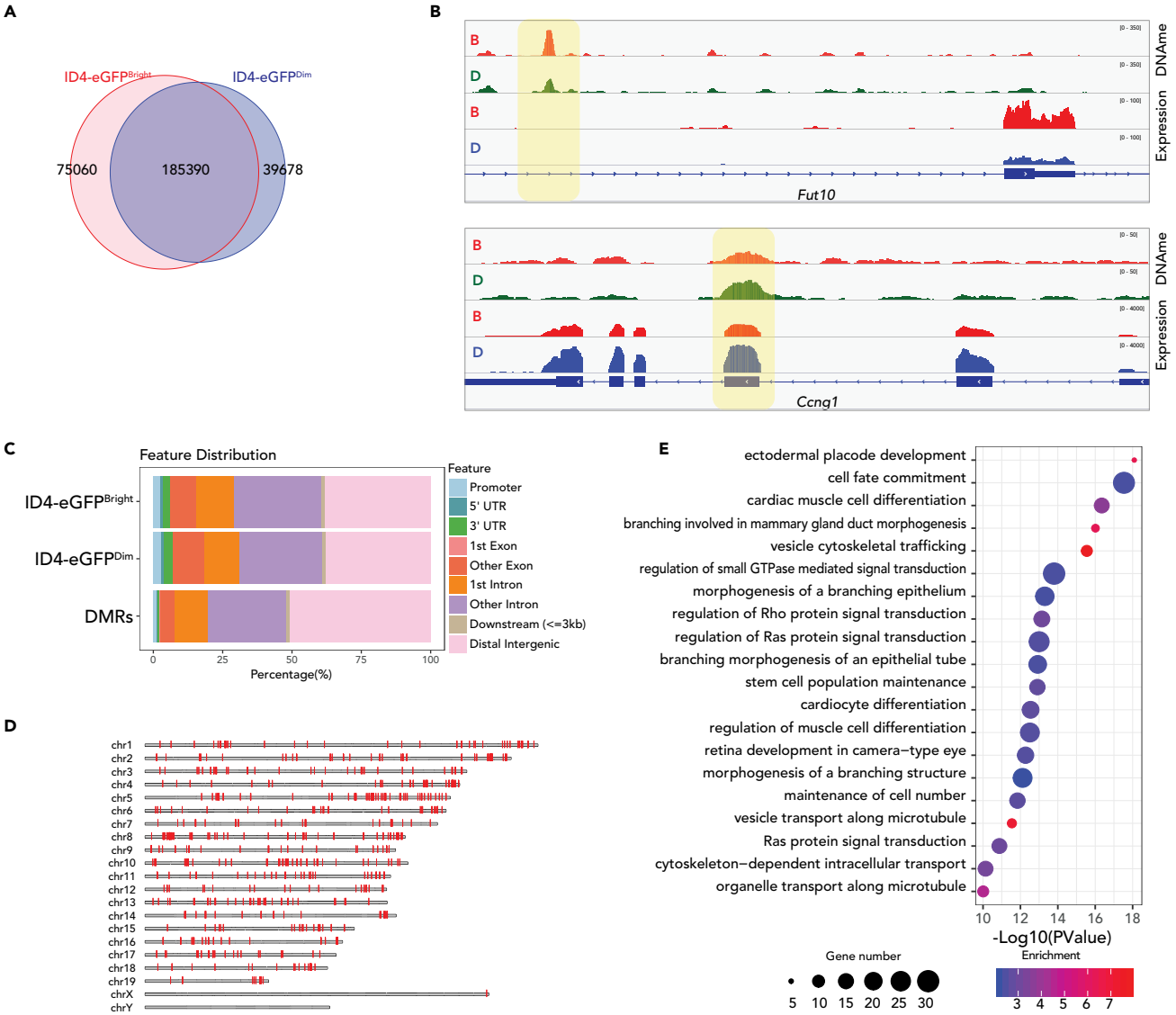


Figure 5. DNA Methylation Profiles in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} Spermatogonia

(A) Venn diagrams show proportions of DNA methylation peaks unique to ID4-eGFP^{Bright} spermatogonia, common to both ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia, or unique to ID4-eGFP^{Dim} spermatogonia.

(B) Genome browser snapshots showing correlations or lack thereof between DNA methylation and transcript peaks in ID4-eGFP^{Bright} (B, orange tracks) and ID4-eGFP^{Dim} (D, green or blue tracks) spermatogonia in regions encompassing a gene that is up-regulated in ID4-eGFP^{Bright} spermatogonia (*Fut10*) and a gene that is up-regulated in ID4-eGFP^{Dim} spermatogonia (*Ccng1*). Differentially methylated regions within gene bodies, but not at promoters, appear to be positively correlated with gene expression levels.

(C) The distribution of DNA methylation peaks in each spermatogonial subtype and DMRs between the two subtypes is shown. DMRs were rare at promoters but abundant in intra- and intergenic regions.

(D) DMRs were found throughout the autosomes, but, with one exception, not on the sex chromosomes. Genomic coordinates of DMRs are shown in Table S4.

(E) GREAT-GO analysis of genes associated with DMRs. The color code indicates the enrichment levels of DMR-associated genes (blue–red = low–high). Dot diameters indicate the number of DMR-associated genes within each GO term.

motif discovery for putative promoter-binding factors enriched in genes in each of the six different sets described in Figure 1E, for which consensus binding sequences are shown in Figure 6A. Examples of TF binding sites over-represented in promoters of each gene set include those for CEBPB, BACH2, BCL6B, and NF-Y in genes differentially up-regulated in SSC-enriched spermatogonia; MXI1, ZBTB3, EBF1/2, SPI1, and HLTF in genes up-regulated in progenitor-enriched spermatogonia; and NF-Y, IRF5, ZFP711,

A

de novo motifs in Gene Set 1	P-value % of Targets	Matches
	1e-11 5.52%	CHR, CEBPB BACH2
	1e-10 19.63%	BCL6, BCL6B, STAT5
	1e-9 15.03%	NR2F2, YY1, ELF1
	1e-9 19.94%	KLF5, SMAD2, KLF14
	1e-9 8.59%	NFYC, NFYA, NFYB

de novo motifs in Gene Set 2	P-value % of Targets	Matches
	1e-11 8.85%	GLIS2, PLAGL1, NR2F2
	1e-10 12.68%	NFYB, NFYA, MSX2
	1e-10 7.96%	YY2, HINFP, ZFP128
	1e-10 5.31%	SOX3, ZFP7, SOX15
	1e-9 6.49%	SOX17, SOX6, NR5A1

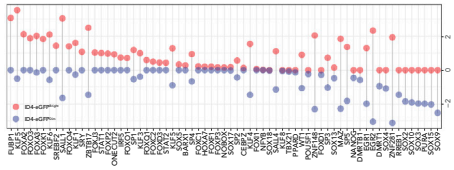
de novo motifs in Gene Set 3	P-value % of Targets	Matches
	1e-12 39.07%	MYC, MXI1, MAX
	1e-9 7.21%	ZBTB3, ZFP691 ZFP161
	1e-9 6.74%	EBF1/2 TFAP2A, TFAP2C
	1e-9 5.12%	SPI1, IRF1, IRF8
	1e-9 3.95%	HLTF, MYB, NR4A1

de novo motifs in Gene Set 4	P-value % of Targets	Matches
	1e-16 24.05%	CHR, LIN54, DUX
	1e-11 33.54%	NFYA, NFYB, NFYC
	1e-9 6.96%	IRF5, HNF1B, HMBOX1
	1e-9 13.29%	ZFP161, NR5A1, ZFP711
	1e-9 7.59%	SIX1, SIX4, HOXA5

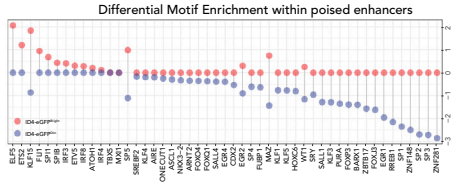
de novo motifs in Gene Set 5	P-value % of Targets	Matches
	1e-23 17.62%	E2F family
	1e-10 6.19%	RELA, MEF2A, MEF2C
	1e-9 7.62%	ZFP281, GCM1, NFYA
	1e-9 8.10%	HAND1, TWIST1, ZBTB18
	1e-9 9.05%	SIX1, ZEB2, SIX2

de novo motifs in Gene Set 6	P-value % of Targets	Matches
	1e-11 10.46%	HAND1, SMAD3, SMAD4
	1e-9 1.79%	YY1, YY2
	1e-9 6.70%	SPIC, HINFP, MZF1
	1e-9 7.33%	GLI3, GLI2, KLF4
	1e-8 0.15%	POU3F1, POU3F3 POU1F1

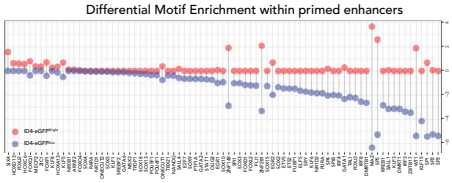
B



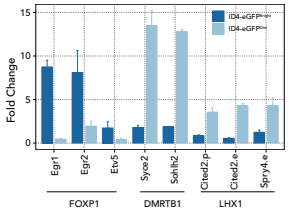
C



D



E



F

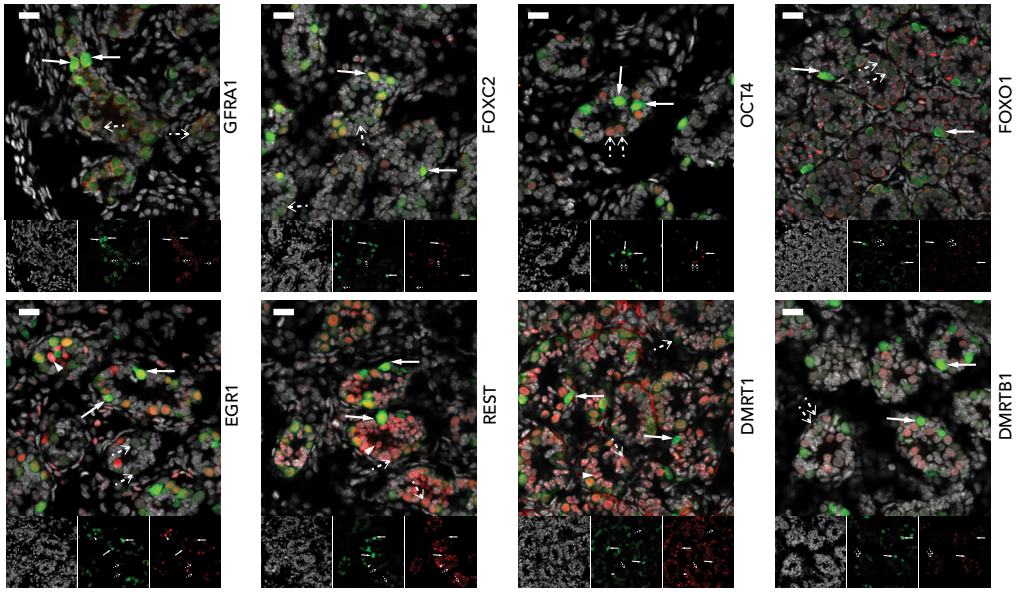


Figure 6. Enrichment of Transcription Factor Binding Sites Predicts Cell Type and Gene-Specific Patterns of Expression and Binding of Specific Transcription Factors in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} Spermatogonia

(A) *De novo* motif discovery within promoters of genes most differentially expressed in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia.

(B–D) Differential enrichment of transcription factor binding motifs within active, poised, or primed enhancers in ID4-eGFP^{Bright} (red) and ID4-eGFP^{Dim} (blue) spermatogonia, including active enhancers (B), poised enhancers (C), and primed enhancers (D).

(E) ChIP-qPCR confirms predicted binding patterns of specific transcription factors in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia. FOXP1 preferentially binds to promoters of *Egr1*, *Egr2*, and *Etv5* in ID4-eGFP^{Bright} spermatogonia. DMRTB1 preferentially binds to promoters of *Syce2* and *Sohlh2* in ID4-eGFP^{Dim} spermatogonia. LHX1 preferentially binds to the *Cited2* gene promoter (*Cited2.p*), the *Cited2* proximal enhancer (*Cited2.e*), and the *Spry4* enhancer (*Spry4.e*).

(F) Immunochemistry staining (IHC) of marker proteins in seminiferous cords in whole-mount sections of P6 mouse testes. GFRA1 and FOXC2 localized specifically to ID4-eGFP^{Bright} SSCs, OCT4 appeared in both ID4-eGFP^{Bright} SSCs and progenitor-enriched ID4-eGFP^{Dim} spermatogonia but was more prominent in the latter, FOXO1 and DMRTB1 localized specifically in progenitor-enriched ID4-eGFP^{Dim} spermatogonia, DMRT1 appeared in ID4-eGFP^{Dim} spermatogonia and some Sertoli cells, but not in ID4-eGFP^{Bright} SSCs, and EGR1 and REST co-localized with both ID4-eGFP^{Bright} SSCs and ID4-eGFP^{Dim} progenitor/early differentiating spermatogonia as well as in some somatic Sertoli cells. ⟵, ID4-eGFP^{Bright} spermatogonium; ⟵ ▪, ID4-eGFP^{Dim} spermatogonium; ◀, Sertoli cell. Scale bars, 20 μm.

SIX4, the E2F family, RELA, and TWIST1 in genes up-regulated in both SSC- and progenitor-enriched spermatogonia (Figure 6A). Expression patterns of transcripts encoding each of these factors and others in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia are shown in Figure S5A. Interestingly, in addition to the presence of binding motifs for factors expressed in either SSC- or progenitor-enriched spermatogonia, or both, in promoters of genes expressed in these subtypes, we also observed binding motifs for several factors that are not expressed in spermatogonia (Figure S5A). We were surprised to see evidence of expression of the *Mef2a* gene in spermatogonia but note that this evidence refers only to transcript-level expression and does not comment on expression at the protein level.

Similarly, *de novo* motif analysis revealed over-representation of TF-binding sites present in active enhancers including those for FUBP1, KLF5, FOXO3, KLF6, ZBTB17, and FOXC2 in ID4-eGFP^{Bright} spermatogonia and SOX15, SOX3, SOX10, RREB1, DMRTB1, and EGR1 in ID4-eGFP^{Dim} spermatogonia (Figure 6B). TF-binding sites over-represented in poised enhancers included those for ELF5, ETS2, KLF15, and ETV5 in ID4-eGFP^{Bright} spermatogonia and ZNF281, SP3, SP2, EGR1, FOXPs, and EGR2 in ID4-eGFP^{Dim} spermatogonia (Figure 6C), and those in primed enhancers included sites for SIX4, HOXD13, and ZIC1 in ID4-eGFP^{Bright} spermatogonia and SP3, SP2, DMRT1, KLF153, ZBTB17, and DMRTB1 in ID4-eGFP^{Dim} spermatogonia (Figure 6D). Expression patterns of transcripts encoding each of these factors and others in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia are shown in Figures S5B and S5C.

We next used immunohistochemistry to validate the predicted presence of several of these factors in SSC-enriched ID4-eGFP^{Bright} or ID4-eGFP^{Dim} spermatogonia *in situ* (Figure 6F), including detection of GFRA1 and FOXC2 in SSC-enriched ID4-eGFP^{Bright} spermatogonia, FOXO1 and DMRTB1 in progenitor-enriched ID4-eGFP^{Dim} spermatogonia, OCT4 in both SSC-enriched ID4-eGFP^{Bright} spermatogonia and progenitor-enriched ID4-eGFP^{Dim} spermatogonia (though more prominently in the latter), DMRT1 in progenitor-enriched ID4-eGFP^{Dim} spermatogonia and some Sertoli cells but not in ID4-eGFP^{Bright} cells, and EGR1 and REST in both ID4-eGFP^{Bright} SSCs and ID4-eGFP^{Dim} progenitor/early differentiating spermatogonia as well as in some somatic Sertoli cells. We then validated a representative sample of spermatogonial-subtype specific differential binding patterns of specific factors to specific enhancers on the basis of factor/locus-specific ChIP-qPCR, which demonstrated elevated binding of FOXP1 to enhancers of three genes, *Egr1, Egr2,* and *Etv5*, specifically in SSC-enriched ID4-eGFP^{Bright} spermatogonia, and elevated binding of DMRTB1 to enhancers of two genes, *Syce2* and *Sohlh2,* and of LHX1 to enhancers of two genes, *Cited2* and *Spry4*, in progenitor-enriched ID4-eGFP^{Dim} spermatogonia (Figure 6E). Interestingly, transcripts encoding these three transcription factors are also differentially expressed in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia, respectively, with those encoding FOXP1 and LHX1 being expressed at higher levels in ID4-eGFP^{Bright} spermatogonia and those encoding DMRTB1 being expressed at higher levels in D4-eGFP^{Dim} spermatogonia. Taken together, these data demonstrate the power of multiparametric integrative analysis of epigenetic parameters as a means to reveal differential epigenetic programming of elements potentially regulating differential gene expression in SSC-enriched ID4-eGFP^{Bright} or progenitor-enriched ID4-eGFP^{Dim} spermatogonia, respectively. Confirmation of spermatogonial subtype-specific binding of specific factors for which motifs are overrepresented in these elements affords a means to identify candidate regulators that may either establish and/or mediate this differential epigenetic programming.

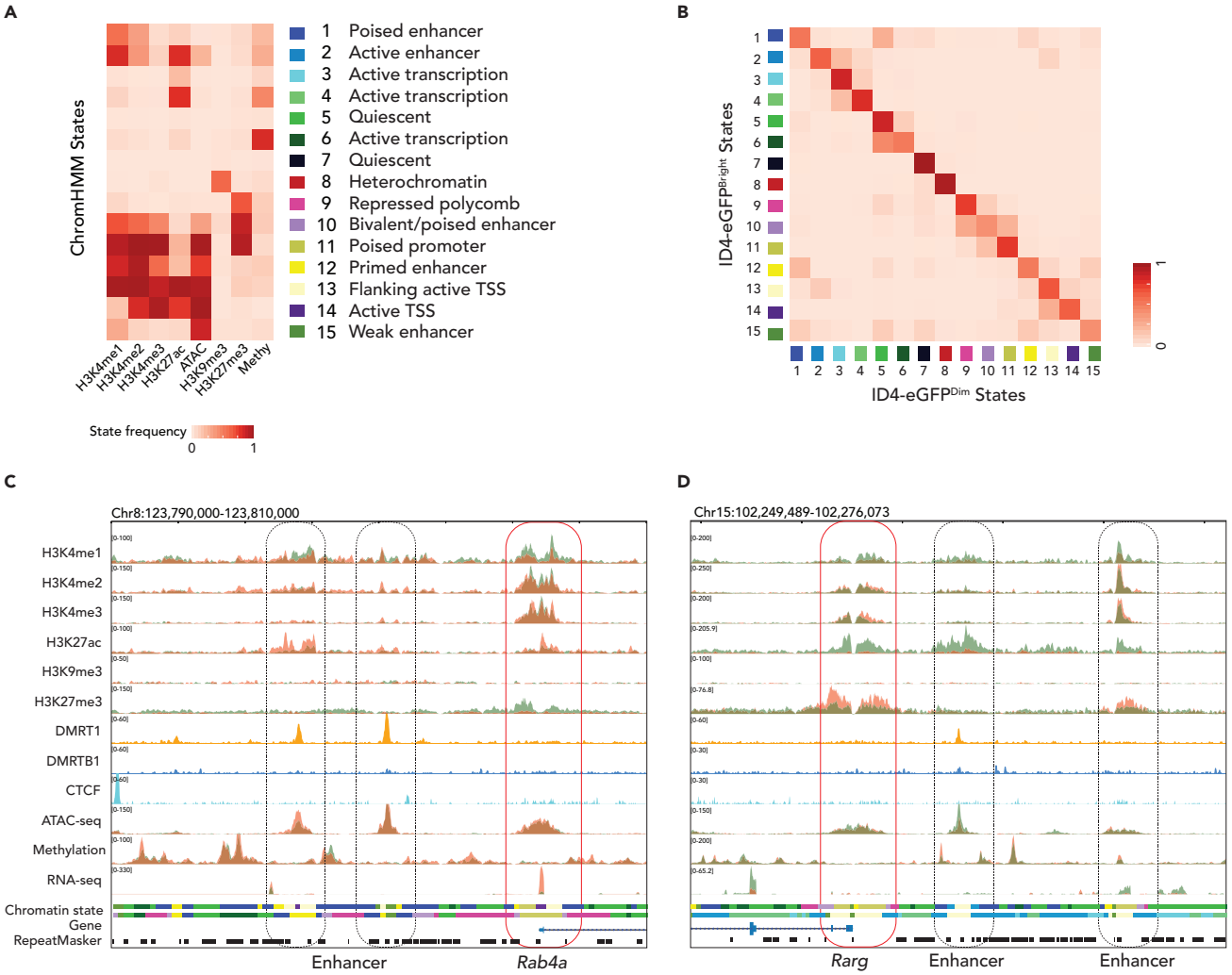


Figure 7. Coordinated Epigenetic Programming Correlates with Differential Gene Expression in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} Spermatogonia

(A) A heatmap shows that 15 distinct chromatin states are predicted by ChromHMM analysis of unique combinational patterns of eight different epigenetic parameters.(B) A heatmap reveals spermatogonial-subtype specific transitions among the 15 chromatin states color-coded to match those shown in Figure 7A.

(C and D) Genome browser views from ID4-eGFP^{Bright} spermatogonia (coral colored tracks) and ID4-eGFP^{Dim} spermatogonia (green colored tracks) of epigenetic parameters associated with a gene (Rab4a) up-regulated in ID4-eGFP^{Bright} spermatogonia (C), and a gene (Rarg) up-regulated in ID4-eGFP^{Dim} spermatogonia (D). Note the elevated H3K27ac levels associated with both the enhancer and promoter regions of each gene in the spermatogonial subtype in which the gene is up-regulated and the elevated H3K27me3 levels associated with both the enhancer and promoter regions of each gene in the spermatogonial subtype in which the gene is down-regulated.

Differential Fates of ID4-eGFP^{Bright} and ID4-eGFP^{Dim} Spermatogonia Are Associated with Coordinated, Multiparametric Programming of Differentially Expressed Genes

Ultimately, it is specific combinations of chromatin parameters defined by epigenetic signatures and specific TF interactions that drive differential gene expression, which, in turn, establishes distinct fates of different cell types or subtypes (Ernst and Kellis, 2017). Thus, we integrated our ChIP-seq, MeDIP-seq, ATAC-seq, and RNA-seq data using multivariate Hidden Markov Model building by ChromHMM to identify and characterize 15 different chromatin states within the spermatogonial genome in agreement with ENCODE project methods (Bailey et al., 2013; Landt et al., 2012) (Figure 7A, S6A, and S6B). These 15 different states were assessed in SSC- and progenitor-enriched spermatogonia to compare the extent to which transitions among each were associated with unique fates of spermatogonial subtypes (Figure 7B). For instance, the pattern of quiescent enhancers in state 7 showed little or no difference between

ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia, likely representing enhancers involved with gene expression in non-spermatogonial cell types. In contrast, a portion of enhancers displaying either an inactive (state 1) or active (state 6) status in SSC-enriched spermatogonia resolved to a quiescent status (state 5) in progenitor-enriched spermatogonia (Figure 7B).

Parallel visualization of tracks indicative of read intensities derived from each epigenomic analysis facilitated the most direct, comprehensive, locus-specific comparisons of coordinated epigenetic programming in SSC- and progenitor-enriched spermatogonia, respectively (Figures 7C and 7D). This revealed unique epigenetic signatures associated with promoters or enhancers of DEGs in these two spermatogonial subtypes. Thus, as noted above, programming patterns associated with promoters of DEGs up-regulated in one or the other spermatogonial subpopulation included enriched H3K4me1,2,3 and H3K27ac, depleted H3K27me3, hypomethylated DNA, and elevated chromatin accessibility (Figures 7C and 7D). Interestingly, promoters of many genes showed enriched H3K4me1,2,3, hypomethylated DNA and elevated chromatin accessibility in both spermatogonial subpopulations, despite the fact that these genes were differentially expressed. However, enrichment of H3K27ac or H3K27me3 varied directly with up- or down-regulation of these genes in each subpopulation, respectively. This suggests differential enrichment of these two promoter region modifications contributes directly to differential regulation of gene expression in SSCs versus progenitors. Enhancers of DEGs showed enrichment of H3K27ac and depletion of H3K4me1 for up-regulated genes and enrichment of H3K27me3 and H3K4me1 for down-regulated genes (Figures 7C and 7D).

We augmented these data with those from published reports of genome-wide binding patterns of DMRT1 (Murphy et al., 2015), DMRTB1 (Zhang et al., 2014), and CTCF (Yue et al., 2014) in adult testis tissue. In a study by Murphy et al. in which no distinction was made between SSC- and progenitor-enriched spermatogonia, or even between spermatogenic and somatic cells, it is noteworthy that peaks of DMRT1 and DMRTB1 binding were detected at enhancers of many of the spermatogonial subtype-specific genes identified in our study (Figures 7C and 7D) (Murphy et al., 2015). In addition, in a study by Rivero-Hinojosa et al. binding motifs for CTCF were observed predominantly in distal flanking intergenic regions throughout the genome consistent with reports that CTCF plays an important role in 3-dimensional organization of the genome to mediate long-range enhancer-promoter interactions contributing to cell fate determination (Ren et al., 2017; Rivero-Hinojosa et al., 2017). Finally, using the code shown in Figure 7A, we were able to predict the arrangement of chromatin states in a linear context in each spermatogonial subtype (Figures 7C, 7D, and S6C), as well as the extent to which these states varied either between genes differentially expressed in each spermatogonial subtype or within the same genes in each spermatogonial subtype (Figures 7C and 7D). Additional data regarding the genome-wide distribution of potential regulatory elements and chromatin states identified by our ChromHMM analysis are shown in Figure S6. These results further demonstrate the correlation between spermatogonial subtype-specific differential epigenetic programming and spermatogonial subtype-specific differential gene expression.

DISCUSSION

The ENCODE (ENCODE Project Consortium, 2012), NIH Roadmap Epigenomics Mapping Consortium (Zacher et al., 2017) and related (Ernst et al., 2011) studies characterized the variable epigenetic states of key regulatory elements throughout the genomes of >100 different somatic cell types on the basis of multiparametric integrative methodology (Ernst and Kellis, 2017) but did not examine any germ cell types. One previous study (Hajkova et al., 2008) provided an initial characterization of histone modifications in fetal mouse prospermatogonia, but did not examine postnatal germ cells, and therefore did not characterize epigenetic programming distinguishing SSCs from progenitors. Fetal and postnatal spermatogenic cell types have been assessed for poised genes, but those studies were restricted to a limited set of histone modifications (Lesch et al., 2013). Several other studies assessed one or a few different epigenetic parameters in various spermatogenic cell types, including SSCs (Gaysinskaya et al., 2018; Hammoud et al., 2015, 2014; Kubo et al., 2015; Lambrot et al., 2019; Maezawa et al., 2018a, 2018b), but these were too limited in scope to facilitate integrative profiling analogous to that developed by the ENCODE and NIH Roadmap Epigenomics Mapping consortia (ENCODE Project Consortium, 2012; Ernst et al., 2011; Gifford et al., 2013; Kellis et al., 2014; Ram et al., 2011; Roadmap Epigenomics Consortium et al., 2015; Shema et al., 2019; Zhu et al., 2013). Our simultaneous analysis of nine different parameters in aliquots of the same samples of SSC- or progenitor-enriched spermatogonial subpopulations yielded combinatorial data sufficient to distinguish the same 15 different chromatin states that were described for somatic cell types by the ENCODE study (ENCODE Project Consortium, 2012).

Results from transplantation studies have shown that regenerative SSC capacity resides in a small subpopulation of undifferentiated spermatogonia (Kubota and Brinster, 2018); however, the advent of the *Id4-eGfp* transgenic mouse has facilitated selective recovery of spermatogonial subpopulations highly enriched for, or significantly depleted of, this capacity (Chan et al., 2014; Helsel et al., 2017). Importantly, the fact that these different spermatogonial subpopulations are recovered from the intact testis in a similar manner but then display consistent, significant differences in regenerative capacity suggests that this reflects an inherent distinction between SSCs and progenitors rather than any sort of artifactual effect imposed by the isolation, dissociation, or transplantation of these cells. Indeed, use of the *Id4-eGfp* transgenic mouse and the transplantation assay to subdivide and validate regenerative capacity among different spermatogonial subpopulations or subtypes affords a robust, objective means by which to monitor regenerative capacity characteristic of SSCs and typically absent from progenitors.

Our bulk and single-cell RNA-seq data corroborate multiple recent studies confirming consistent differences in gene expression patterns in SSC- and progenitor-enriched subpopulations (Green et al., 2018; Guo et al., 2018; Helsel et al., 2017; Hermann et al., 2018; Law et al., 2019; Mutoji et al., 2016), indicating these spermatogonial subtypes represent the emergence of distinct cell fates driven by distinct transcriptomes. Here, we have extended these analyses by conducting the first comprehensive analysis of epigenetic programming associated with spermatogonial-subtype specific DEGs. This revealed distinct epigenetic landscapes associated with DEGs in SSCs and progenitors, which we further mined to identify binding motifs for specific factors that may either direct establishment of this differential epigenetic programming or mediate its effects to coordinate subsequent differential expression of genes required to either maintain the SSC phenotype or transition to the progenitor phenotype.

Among the parameters we measured, differential enrichment of H3K27me3 and H3K27ac at gene promoters and related intergenic enhancers correlated most closely with differential spermatogonial subtype-specific gene expression patterns. Specifically, elevated representation of H3K27me3 was associated with down-regulated DEGs in each spermatogonial subtype, whereas that of H3K27ac was associated with up-regulated DEGs in each case. Interestingly, several other chromatin parameters, including patterns of H3K4me1,2,3, chromatin accessibility and DNA methylation at promoters, and those of H3K4me2,3 and chromatin accessibility at enhancers, showed no significant variation among DEGs regardless of whether the gene was expressed in both or only one spermatogonial subpopulation. Collectively, this is consistent with the concept that development of distinct cell fates from a common lineage involves an ordered series of changes in epigenetic programming to first initiate and subsequently stabilize differential gene expression. Similar results regarding DNA methylation and chromatin accessibility patterns have been reported for undifferentiated and differentiating human spermatogonia (Guo et al., 2017, 2020).

Previous reports have described epigenetic poising of genes (promoters simultaneously marked with H3K4me3 and H3K27me3) in the spermatogenic lineage that appears to predispose the capacity of the paternal genome to rapidly transition to an embryonic transcriptome following fertilization (Guo et al., 2017; Lesch et al., 2013, 2016). We found that many non-poised genes expressed in SSC-enriched spermatogonia become poised and repressed in progenitor-enriched spermatogonia. This raises the intriguing possibility that, in addition to marking initiation of commitment to the spermatogenic differentiation pathway, the SSC-progenitor transition may also demarcate initiation of a final phase of epigenetic programming to prepare the paternal genome for post-fertilization functions.

Enhancers are rich in transcription factor binding sites (Stadler et al., 2011), and it has been shown that transcription factors act as key drivers of differential states of activity or inactivity at enhancers (Zentner and of Chemistry, 2012). Our *de novo* motif analysis revealed many binding sites common to regulatory regions in both spermatogonial subtypes. However, we also identified differential enrichment of certain motifs in promoters or enhancers regulating DEGs in ID4-eGFP^{bright} and ID4-eGFP^{dim} spermatogonia, and these formed the basis for testable predictions of differential expression and/or binding of specific transcription factors in these spermatogonial subpopulations. We confirmed the predicted differential expression of eight different TFs (FOXC2, OCT4, FOXO1, EGR1, REST, DMRT1, DMRTB1, EZH2) and the predicted differential, locus-specific binding of three TFs (FOXP1, DMRTB1, LHX1). Thus, we validated spermatogonial subpopulation-specific differences in TF expression at the RNA level, prevalence/intracellular location at the protein level, and binding to enhancers of target DEGs at the genomic level. Interestingly, we observed enrichment of binding motifs for six factors that were also identified by

de novo motif analysis of regulatory regions in human SSCs (Guo et al., 2017) including CTCF, DMRT1, CTCFL, NFY, FOXP1, and SOX3.

Our results reveal that SSC- and progenitor-enriched spermatogonia manifest distinct patterns of epigenetic programming associated with consistent differential patterns of gene expression distinguishing these spermatogonial subtypes. We suggest this differential epigenetic programming drives cell fate divergence between SSCs and progenitors, thereby directing a significant developmental switch between retention of SSC fate and initiation of spermatogenic differentiation, respectively. We have identified specific epigenetic parameters that correlate with these states that can now be monitored within the normal developmental context of spermatogenesis.

Limitations of the Study

Our multiparametric integrative analysis showed that differential gene expression is accompanied by differential epigenetic programming in SSCs and progenitors and that SSCs maintain expression of stemness genes while progenitors initiate expression of spermatogenic differentiation genes. However, the extent to which progenitors derived from SSCs retain the potential to reverse these differences in epigenetic programming and gene expression and transition back into SSCs remains unclear, as does the identity of regulators of the differential epigenetic programming that distinguishes SSCs and progenitors. Also unknown is the extent to which differential epigenetic programming regulating changes in gene expression required to promote developmental progression from one cell fate to the next becomes primed earlier in the lineage, prior to transcriptional activation of cell fate-specific genes.

Resource Availability

Lead Contact

Further information and requests for resources and questions should be directed to and will be fulfilled by the lead contact John R McCarrey (John.McCarrey@UTSA.edu).

Materials Availability

All material used are listed in Transparent Methods section and Key Resources Table in Supplemental Information, and any further information and requests for resources and questions should be directed to and will be fulfilled by the Lead Contact.

Data and Code Availability

All datasets generated from this study have been archived in the NCBI GEO database, with the accession number GSE131657 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE131657).

P6 ID4-eGFP + single cell RNA-seq data were obtained from GSE109049 (Hermann et al., 2018). Adult mouse testis ChIP-seq data for CTCF, DMRT1, and DMRTB1 was obtained from GSM918711 (Yue et al., 2014), GSE64892 (Murphy et al., 2015), and GSM1480189 (Zhang et al., 2014), respectively.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101596.

ACKNOWLEDGMENTS

The authors thank Mr. Sean Vargas for assistance with procedures conducted in the UTSA Genomics Core. Funding for this research was provided by NIH grants HD078679 and HD098593 (to J.R.M.), HD090083 (to C.B.G.), HD061665 (to J.M.O.), HD090007 (to B.P.H.); as well as from the Robert J. Kleberg and Helen C. Kleberg Foundation, and the Nancy Smith Hurd Foundation. Results were generated in part with assistance from the UTSA Genomics Core supported by NIH grant G12-MD007591, NSF grant DBI-1337513, and UTSA. This work received computational support from UTSA's HPC cluster SHAMU, operated by University Technology Solutions.

AUTHOR CONTRIBUTIONS

K.C. and J.R.M. conceived and designed the experiments. K.C. performed all RNA-seq, ChIP-seq, ATAC-seq, and MeDIP-seq experiments and analyzed all of the data from each of those experiments and prepared all of the figures. K.C. also performed the *de novo* motif analyses and gene-specific ChIP experiments as well as the IHC with assistance from C.-H.E.C. I.-C.C. assisted with cell sorting. J.M.O. provided the *Id4-egfp* transgenic mouse line and related spermatogonial transplantation results. B.J.H. and C.B.G. assisted with interpretation of the IHC results. K.C. mined single-cell transcriptome data from datasets previously prepared by K.C., B.P.H., and J.R.M. B.P.H. assisted with interpretations of the scRNA-seq data and directs the UTSA Genomics Core in which K.C. prepared libraries for sequencing. All authors contributed to interpretation of data. J.R.M. wrote the manuscript with very significant input from K.C. plus additional input from all other authors. All authors read and accepted the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.

Received: June 9, 2020
Revised: August 17, 2020
Accepted: September 17, 2020
Published: October 23, 2020

REFERENCES

Bailey, T., Krajewski, P., Ladunga, I., Lefebvre, C., Li, Q., Liu, T., Madrigal, P., Taslim, C., and Zhang, J. (2013). Practical guidelines for the comprehensive analysis of ChIP-seq data. PLoS Comput. Biol. *9*, e1003326.

Becker, J.S., Nicetto, D., and Zaret, K.S. (2016). H3K9me3-Dependent heterochromatin: barrier to cell fate changes. Trends Genet. *32*, 29–41.

Brinster, R.L., and Avarbock, M.R. (1994). Germline transmission of donor haplotype following spermatogonial transplantation. Proc. Natl. Acad. Sci. U S A *91*, 11303–11307.

Buaas, F.W., Kirsh, A.L., Sharma, M., McLean, D.J., Morris, J.L., Griswold, M.D., De Rooij, D.G., and Braun, R.E. (2004). Plzf is required in adult male germ cells for stem cell self-renewal. Nat. Genet. *36*, 647–652.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol. Cell *49*, 825–837.

Chan, F., Oatley, M.J., Kaucher, A.V., Yang, Q.E., Bieberich, C.J., Shashikant, C.S., and Oatley, J.M. (2014). Functional and molecular features of the Id4+ germline stem cell population in mouse testes. Genes Dev. *28*, 1351–1362.

Chen, Y., Zheng, Y., Gao, Y., Lin, Z., Yang, S., Wang, T., Wang, Q., Xie, N., Hua, R., Liu, M., et al. (2018). Single-cell RNA-seq uncovers dynamic processes and critical regulators in mouse spermatogenesis. Cell Res. *28*, 879–896.

Drumond, A.L., Meistrich, M.L., and Chiarini-Garcia, H. (2011). Spermatogonial morphology and kinetics during testis development in mice: a high-resolution light microscopy approach. Reproduction *142*, 145–155.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Ernst, J., and Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. Nat. Protoc. *12*, 2478–2492.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43–49.

Ernst, C., Eling, N., Martinez-Jimenez, C.P., Marioni, J.C., and Odom, D.T. (2019). Staged developmental mapping and X chromosome transcriptional dynamics during mouse spermatogenesis. Nat. Commun. *10*, 1251.

Gaysinskaya, V., Miller, B.F., De Luca, C., van der Heijden, G.W., Hansen, K.D., and Bortvin, A. (2018). Transient reduction of DNA methylation at the onset of meiosis in male mice. Epigenetics Chromatin *11*, 15.

Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R., et al. (2013). Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell *153*, 1149–1163.

Green, C.D., Ma, Q., Manske, G.L., Shami, A.N., Zheng, X., Marini, S., Moritz, L., Sultan, C., Gurczynski, S.J., Moore, B.B., et al. (2018). A comprehensive Roadmap of murine spermatogenesis defined by single-cell RNA-seq. Dev. Cell *46*, 651–667.e10.

Grive, K.J., Hu, Y., Shu, E., Grimson, A., Elemento, O., Grenier, J.K., and Cohen, P.E. (2019). Dynamic transcriptome profiles within spermatogonial and spermatocyte populations during postnatal testis maturation revealed by single-cell sequencing. Plos Genet. *15*, e1007810.

Guo, J., Grow, E.J., Yi, C., Mlcochova, H., Maher, G.J., Lindskog, C., Murphy, P.J., Wike, C.L., Carrell, D.T., Goriely, A., et al. (2017). Chromatin and single-cell RNA-seq profiling reveal dynamic signaling and metabolic transitions during human spermatogonial stem cell development. Cell Stem Cell *21*, 533–546.e6.

Guo, J., Grow, E.J., Mlcochova, H., Maher, G.J., Lindskog, C., Nie, X., Guo, Y., Takei, Y., Yun, J., Cai, L., et al. (2018). The adult human testis transcriptional cell atlas. Cell Res. *28*, 1141–1157.

Guo, J., Nie, X., Giebler, M., Mlcochova, H., Wang, Y., Grow, E.J., Kim, R., Tharmalingam, M., Matilionyte, G., Lindskog, C., et al. (2020). The dynamic transcriptional cell atlas of testis development during human puberty. Cell Stem Cell *26*, 262–276.e4.

Hajkova, P., Ancelin, K., Waldmann, T., Lacoste, N., Lange, U.C., Cesari, F., Lee, C., Almouzni, G., Schneider, R., and Surani, M.A. (2008). Chromatin dynamics during epigenetic reprogramming in the mouse germ line. Nature *452*, 877–881.

Hammoud, S.S.S., Low, D.H.P.H.P., Yi, C., Carrell, D.T.T., Guccione, E., and Cairns, B.R.R. (2014). Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis. Cell Stem Cell *15*, 239–253.

Hammoud, S.S., Low, D.H.P., Yi, C., Lee, C.L., Oatley, J.M., Payne, C.J., Carrell, D.T., Guccione, E., and Cairns, B.R. (2015). Transcription and imprinting dynamics in developing postnatal male germline stem cells. Genes Dev. *29*, 2312–2324.

Hara, K., Nakagawa, T., Enomoto, H., Suzuki, M., Yamamoto, M., Simons, B.D.D., and Yoshida, S. (2014). Mouse spermatogenic stem cells continually interconvert between equipotent singly isolated and syncytial states. Cell Stem Cell *14*, 658–672.

Helsel, A.R., Yang, Q.-E., Oatley, M.J., Lord, T., Sablitzky, F., and Oatley, J.M. (2017). ID4 levels dictate the stem cell state in mouse spermatogonia. Development *144*, dev.146928–634.

Hermann, B.P., Mutoji, K.N., Velte, E.K., Ko, D., Oatley, J.M., Geyer, C.B., and McCarrey, J.R. (2015). Transcriptional and translational heterogeneity among neonatal mouse Spermatogonia1. Biol. Reprod. *92*, 54.

Hermann, B.P., Cheng, K., Singh, A., Roa-De La Cruz, L., Mutoji, K.N., Chen, I.C., Gildersleeve, H., Lehle, J.D., Mayo, M., Westernströer, B., et al. (2018). The mammalian spermatogenesis single-cell transcriptome, from spermatogonial stem cells to spermatids. Cell Rep. *25*, 1650–1667.e8.

Hilscher, B., Hilscher, W., Bülthoff-Ohnolz, B., Krämer, U., Birke, A., Pelzer, H., and Gauss, G. (1974). Kinetics of gametogenesis: I. Comparative histological and autoradiographic studies of oocytes and transitional prospermatogonia during oogenesis and prespermatogenesis. Cell Tissue Res. *154*, 443–470.

Johnson, L., Petty, C.S., and Neaves, W.B. (1980). A comparative study of daily sperm production and testicular composition in humans and rats. Biol. Reprod. *22*, 1233–1243.

Jung, M., Wells, D., Rusch, J., Ahmad, S., Marchini, J., Myers, S.R., and Conrad, D.F. (2019). Unified single-cell analysis of testis gene regulation and pathology in five mouse strains. ELife *8*, e43966.

Katoh, N., Kuroda, K., Tomikawa, J., Ogata-Kawata, H., Ozaki, R., Ochiai, A., Kitade, M., Takeda, S., Nakabayashi, K., and Hata, K. (2018). Reciprocal changes of H3K27ac and H3K27me3 at the promoter regions of the critical genes for endometrial decidualization. Epigenomics *10*, 1243–1257.

Kellis, M., Wold, B., Snyder, M.P., Bernstein, B.E., Kundaje, A., Marinov, G.K., Ward, L.D., Birney, E., Crawford, G.E., Dekker, J., et al. (2014). Defining functional DNA elements in the human genome. Proc. Natl. Acad. Sci. U S A *111*, 6131–6138.

Kluin, Ph.M., and Rooij, D.G. (1981). A comparison between the morphology and cell kinetics of gonocytes and adult type undifferentiated spermatogonia in the mouse. Int. J. Androl. *4*, 475–493.

Kluin, P.M., Kramer, M.F., and Derooij, D.G. (1982). Spermatogenesis in the immature mouse proceeds faster than in the adult. Int. J. Androl. *5*, 282–294.

Kubo, N., Toh, H., Shirane, K., Shirakawa, T., Kobayashi, H., Sato, T., Sone, H., Sato, Y., Tomizawa, S., Tsurusaki, Y., et al. (2015). DNA methylation and gene expression dynamics during spermatogonial stem cell differentiation in the early postnatal mouse testis. BMC Genomics *16*, 617–624.

Kubota, H., and Brinster, R.L. (2018). Spermatogonial stem cells. Biol. Reprod. *99*, 52–74.

La, H.M., Mäkelä, J.-A., Chan, A.-L., Rossello, F.J., Nefzger, C.M., Legrand, J.M.D., Seram, M., Polo, J.M., and Hobbs, R.M. (2018). Identification of dynamic undifferentiated cell states within the male germline. Nat. Commun. *9*, 2819.

Lambrot, R., Siklenka, K., Lafleur, C., and Kimmins, S. (2019). The genomic distribution of histone H3K4me2 in spermatogonia is highly conserved in sperm. Biol. Reprod. *100*, 1661–1672.

Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. *22*, 1813–1831.

Law, N.C., Oatley, M.J., and Oatley, J.M. (2019). Developmental kinetics and transcriptome dynamics of stem cell specification in the spermatogenic lineage. Nat. Commun. *10*, 2787.

Lesch, B.J., Dokshin, G.A., Young, R.A., McCarrey, J.R., and Page, D.C. (2013). A set of genes critical to deèelopment is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis. Proc. Natl. Acad. Sci. U S A *110*, 16061–16066.

Lesch, B.J., Silber, S.J., McCarrey, J.R., and Page, D.C. (2016). Parallel evolution of male germline epigenetic poising and somatic development in animals. Nat. Genet. *48*, 888–894.

Lewis, J.P., and Trobaugh, F.E. (1964). Hæmatopoietic stem cells. Nature *204*, 589–590.

Liao, J.Y., Ng, S.H., Luk, A.C., Suen, H.C., Qian, Y., Lee, A.W.T., Tu, J.J., Fung, J.C.L., Tang, N.L.S., Feng, B., et al. (2019). Revealing cellular and molecular transitions in neonatal germ cell differentiation using single cell RNA sequencing. Development *146*, dev174953.

Lord, T., Oatley, M.J., and Oatley, J.M. (2018). Testicular architecture is critical for mediation of retinoic acid responsiveness by undifferentiated spermatogonial subtypes in the mouse. Stem Cell Rep. *10*, 538–552.

Maezawa, S., Yukawa, M., Alavattam, K.G., Barski, A., and Namekawa, S.H. (2018a). Dynamic reorganization of open chromatin underlies diverse transcriptomes during spermatogenesis. Nucleic Acids Res. *46*, 593–608.

Maezawa, S., Hasegawa, K., Alavattam, K.G., Funakoshi, M., Sato, T., Barski, A., and Namekawa, S.H. (2018b). SCML2 promotes heterochromatin organization in late spermatogenesis. J. Cell Sci. *131*, jcs217125.

McLaren, A. (2003). Primordial germ cells in the mouse. Dev. Biol. *262*, 1–15.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of \textlessi\textgreatercis\textless/i\textgreater-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Murphy, M.W., Lee, J.K., Rojo, S., Gearhart, M.D., Kurahashi, K., Banerjee, S., Loeuille, G.A., Bashamboo, A., McElreavey, K., Zarkower, D., et al. (2015). An ancient protein-DNA interaction underlying metazoan sex determination. Nat. Struct. Mol. Biol. *22*, 442–U26.

Mutoji, K., Singh, A., Nguyen, T., Gildersleeve, H., Kaucher, A.V., Oatley, M.J., Oatley, J.M., Velte, E.K., Geyer, C.B., Cheng, K., et al. (2016). TSPAN8 expression distinguishes spermatogonial stem cells in the prepubertal mouse testis. Biol. Reprod. *95*, 117.

Niedenberger, B.A., Busada, J.T., and Geyer, C.B. (2015). Marker expression reveals heterogeneity of spermatogonia in the neonatal mouse testis. Reproduction *149*, 329–338.

Oatley, J.M., and Brinster, R.L. (2008). Regulation of spermatogonial stem cell self-renewal in mammals. Annu. Rev. Cell Dev. Biol. *24*, 263–286.

Ram, O., Goren, A., Amit, I., Shoresh, N., Yosef, N., Ernst, J., Kellis, M., Gymrek, M., Issner, R., Coyne, M., et al. (2011). Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. Cell *147*, 1628–1639.

Ren, G., Jin, W., Cui, K., Rodrigez, J., Hu, G., Zhang, Z., Larson, D.R., and Zhao, K. (2017). CTCF-mediated enhancer-promoter interaction is a critical regulator of cell-to-cell variation of gene expression. Mol. Cell *67*, 1049–1058.e6.

Rivero-Hinojosa, S., Kang, S., Lobanenkov, V.V., and Zentner, G.E. (2017). Corrigendum: testis-specific transcriptional regulators selectively occupy BORIS-bound CTCF target regions in mouse male germ cells. Sci. Rep. *7*, 46891.

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

de Rooij, D.G. (2017). The nature and dynamics of spermatogonial stem cells. Development *144*, 3022–3030.

Roosen-Runge, E.C., and Leik, J. (1968). Gonocyte degeneration in the postnatal male rat. Am. J. Anat. *122*, 275–299.

Shema, E., Bernstein, B.E., and Buenrostro, J.D. (2019). Single-cell and single-molecule epigenomics to uncover genome regulation at unprecedented resolution. Nat. Genet. *51*, 19–25.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., et al. (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116–120.

Shinohara, T., Orwig, K.E., Avarbock, M.R., and Brinster, R.L. (2000). Spermatogonial stem cell enrichment by multiparameter selection of mouse testis cells. Proc. Natl. Acad. Sci. U S A *97*, 8346–8351.

Sohni, A., Tan, K., Song, H.W., Burow, D., de Rooij, D.G., Laurent, L., Hsieh, T.C., Rabah, R., Hammoud, S.S., Vicini, E., et al. (2019). The neonatal and adult human testis defined at the single-cell level. Cell Rep. *26*, 1501–1517.e4.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., Tiwari, V.K., et al. (2011). DNA-

binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490–495.

Sun, F., Xu, Q., Zhao, D., and Chen, C. (2015). Id4 marks spermatogonial stem cells in the mouse testis. Sci. Rep. *5*, 17594.

Tan, K., Song, H.-W., and Wilkinson, M.F. (2020). Single-cell RNAseq analysis of testicular germ and somatic cell development during the perinatal period. Development *147*, dev183251.

Valli, H., Sukhwani, M., Dovey, S.L., Peters, K.A., Donohue, J., Castro, C.A., Chu, T., Marshall, G.R., and Orwig, K.E. (2014). Fluorescence- and magnetic-activated cell sorting strategies to isolate and enrich human spermatogonial stem cells. Fertil. Steril. *102*, 566–580.e7.

Velte, E.K., Niedenberger, B.A., Serra, N.D., Singh, A., Roa-DeLaCruz, L., Hermann, B.P., and Geyer, C.B. (2019). Differential RA responsiveness directs formation of functionally-distinct spermatogonial populations at the initiation of spermatogenesis in the mouse. Development *146*, dev173088.

Wartenberg, H. (1976). Comparative cytomorphologic aspects of the male germ cells, especially of the "gonia"*. Andrologia *8*, 117–130.

Yang, Q.-E., Gwost, I., Oatley, M.J., and Oatley, J.M. (2013). Retinoblastoma protein (RB1) controls fate determination in stem cells and progenitors of the mouse male germline. Biol. Reprod. *89*, 113.

Yoshida, S., Sukeno, M., Nakagawa, T., Ohbo, K., Nagamatsu, G., Suda, T., and Nabeshima, Y. (2006). The first round of mouse spermatogenesis is a distinctive program that lacks the self-renewing spermatogonia stage. Development *133*, 1495–1505.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*, 355–364.

Zacher, B., Michel, M., Schwalb, B., Cramer, P., Tresch, A., and Gagneur, J. (2017). Accurate promoter and enhancer identification in 127 ENCODE and Roadmap epigenomics cell types and tissues by GenoSTAN. PLoS One *12*, e0169249.

Zentner, G.E., and of Chemistry, S.-P.C. (2012). The chromatin fingerprint of gene enhancer elements. J. Biol. Chem. *287*, 30888–30896.

Zhang, T., Murphy, M.W., Gearhart, M.D., Bardwell, V.J., and Zarkower, D. (2014). The mammalian Doublesex homolog DMRT6 coordinates the transition between mitotic and meiotic developmental programs during spermatogenesis. Development *141*, 3662–3671.

Zhu, J., Adli, M., Zou, J.Y.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L., et al. (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell *152*, 642–654.

iScience, Volume *23*

Supplemental Information

Unique Epigenetic Programming Distinguishes Regenerative Spermatogonial Stem Cells in the Developing Mouse Testis

Keren Cheng, I-Chung Chen, Ching-Hsun Eric Cheng, Kazadi Mutoji, Benjamin J. Hale, Brian P. Hermann, Christopher B. Geyer, Jon M. Oatley, and John R. McCarrey

Supplementary Figures and Legends

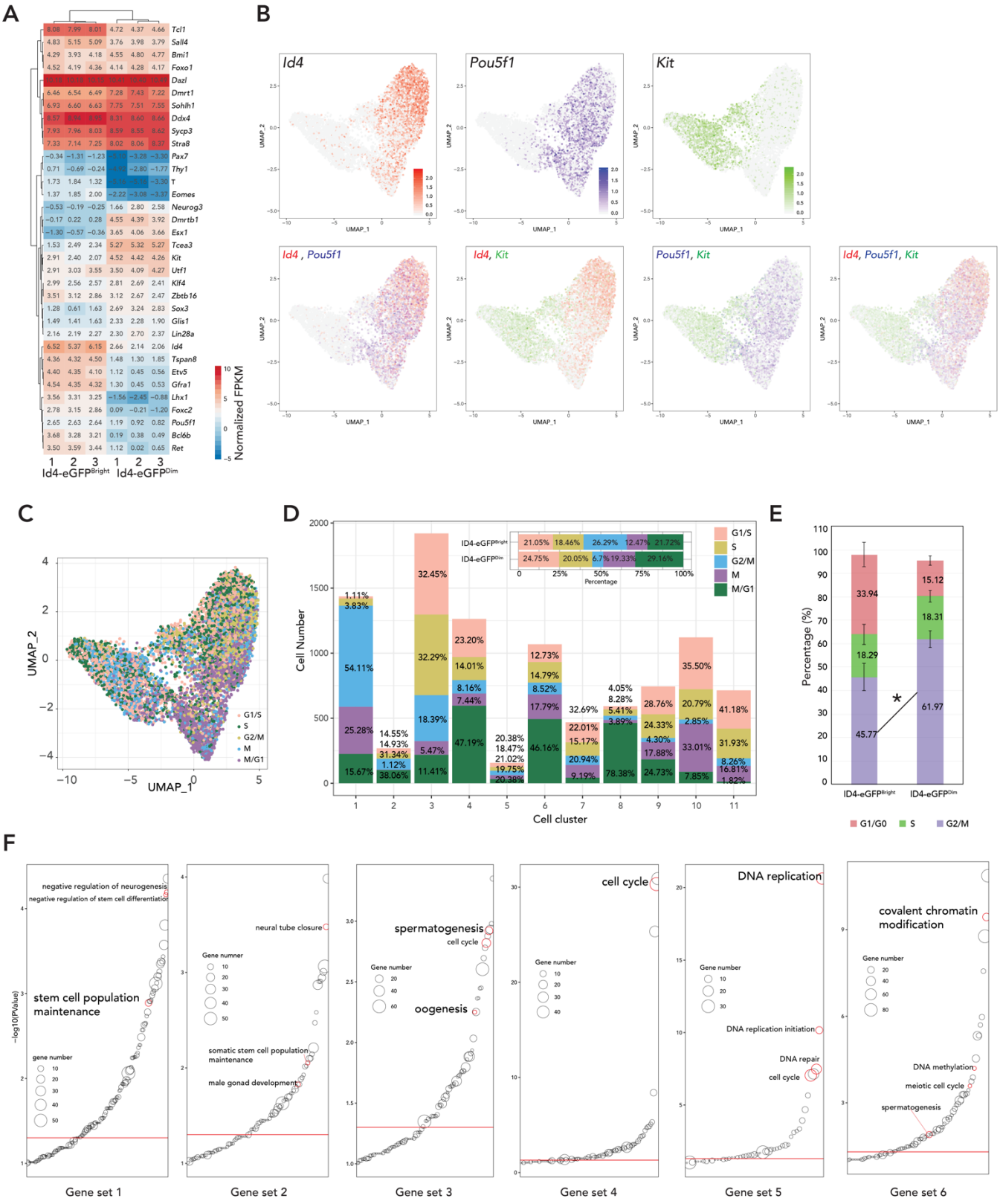


Figure S1. Differential gene expression in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia. Related to Figure 1 and Table S1.

(A) A heatmap showing expression of selected marker genes in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia. Color indicates library-size normalized FPKM (fragments per kilobase/million) (red – blue = high – low).

(B) *Id4, Pou5f1* and *Kit* expression patterns revealed by scRNA-seq of P6 ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia.

(C, D) Cell cycle analysis based on scRNA-seq. The percentage of cells in each different cell cycle phase is detected by UMAP analysis (C) and quantitatively summarized in a bar graph (D).

(E) Cell cycle analysis based on DNA content. Stacked bars show percentages of cells in different cell cycle phases. *Significant difference ($p<0.05$).

(F) GO analyses of the six differentially expressed gene sets described in Figure1e**.** Particularly relevant GO terms are highlighted in red, and detailed GO results can be found in Table S1.

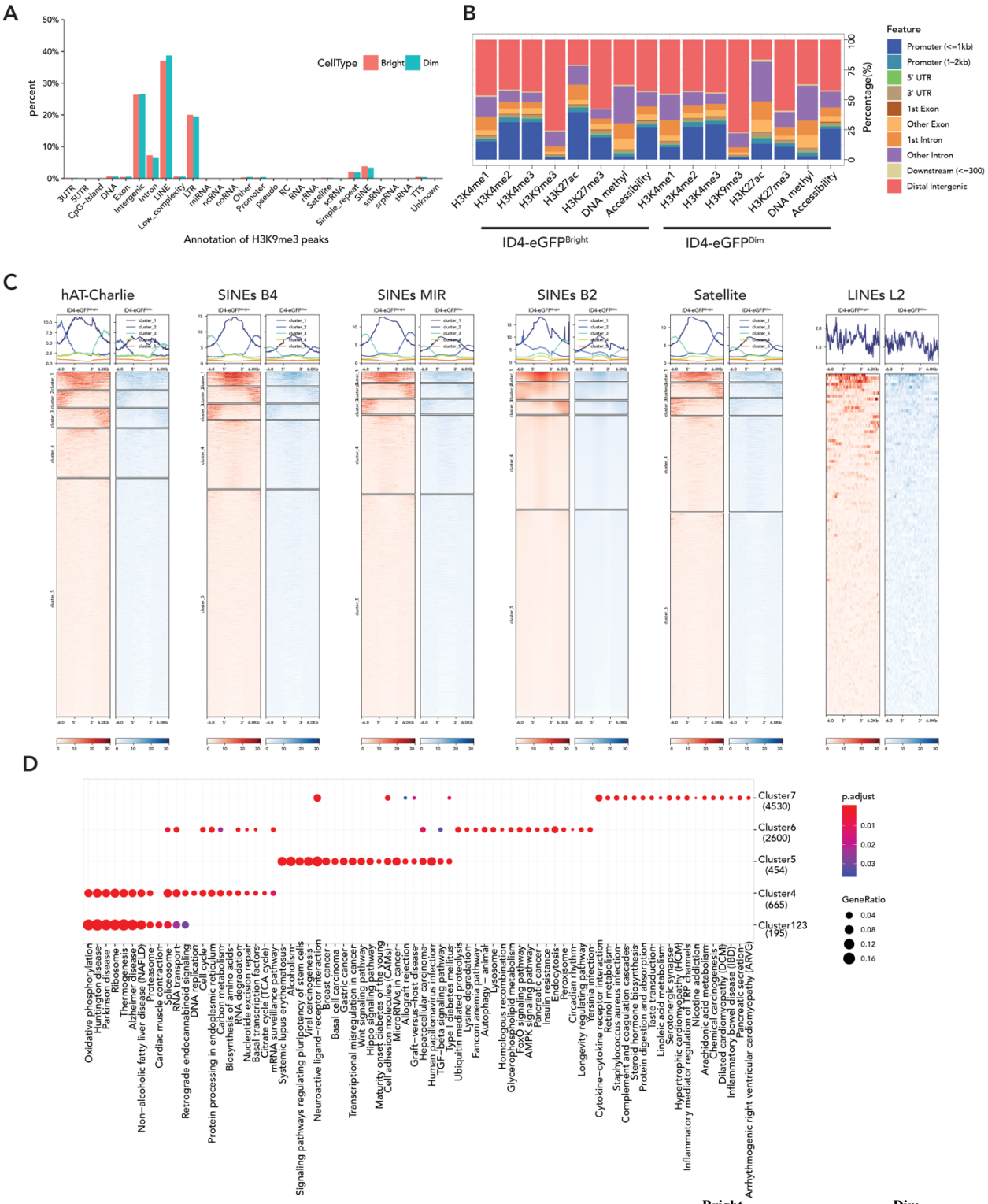


Figure S2. Genomic distributions of histone modifications in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia. Related to Figure 2.

(A) Genomic annotation of H3K9me3 peaks. H3K9me3 peaks were enriched at LINEs, LTRs, intergenic regions, introns, SINEs, and simple repeats.

(B) Proportional genomic distribution of histone modification, chromatin accessibility and DNA methylation peaks.

(C) Heatmaps showing enrichment of H3K9me3 on repetitive elements. Red color indicates ID4-eGFP^{Bright} spermatogonia, and blue color indicates ID4-eGFP^{Dim} spermatogonia.
(D) Comparison of KEGG analyses showing the function of genes with different epigenetic patterns shown in Figure 2C.

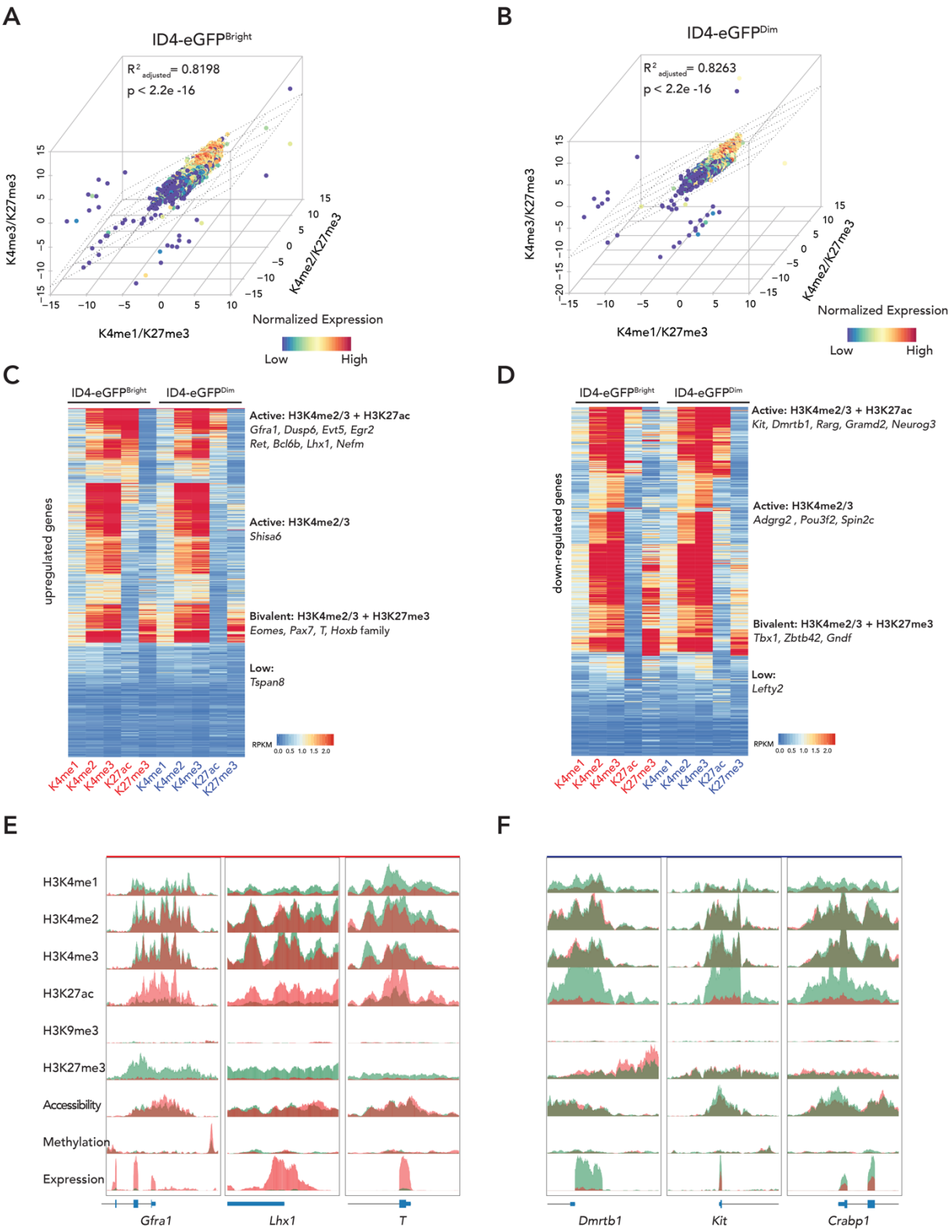


Figure S3. Related to Figure 3.

(A, B) – 3-Dimensional correlations of histone modification and transcript levels (RPKM) at promoters of up-regulated genes in ID4-eGFP^{Bright} (A) and ID4-eGFP^{Dim} (B) spermatogonia. In each spermatogonial subtype, up-regulated genes show elevated H3K4me1,2,3 levels relative to

H3K27me3 levels, while down-regulated genes show elevated H3K27me3 levels relative to H3K4me1,2,3 levels.
(C, D) Heatmaps show relative abundancies of specific histone modifications associated with enhancers of genes up- (C) or down- (D) regulated in each spermatogonial subtype.
(E, F) Genome browser screenshots for promoters of up-regulated and down-regulated genes, showing the epigenetic status.

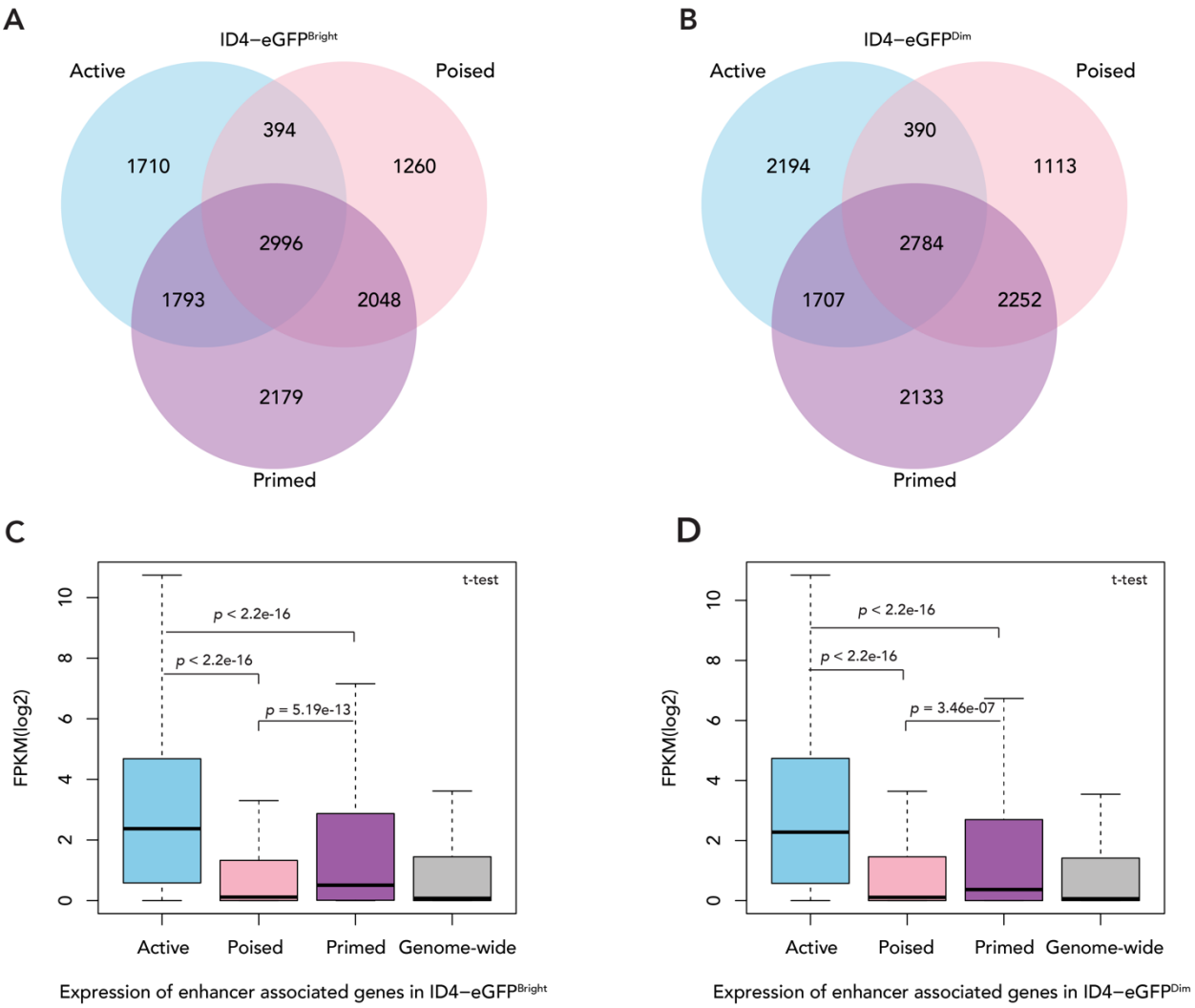


Figure S4: Differential expression of enhancer-associated genes. Related to Figure 4.
(A, B) Enhancer-associated genes in ID4-eGFP^{Bright} (A) and ID4-eGFP^{Dim} (B) spermatogonia. Many genes are associated with more than one enhancer, and those enhancers can be in different states. Venn diagrams show the distribution of different types of enhancers associated with individual genes, including the number of enhancers of each type uniquely associated with individual genes and the number of enhancers of different types jointly associated with individual genes.
(C, D) Genes associated with active enhancers show higher expression than those associated with poised or primed enhancers. Overall, genes associated with poised enhancers display very low expression levels similar to those of randomly selected genes.

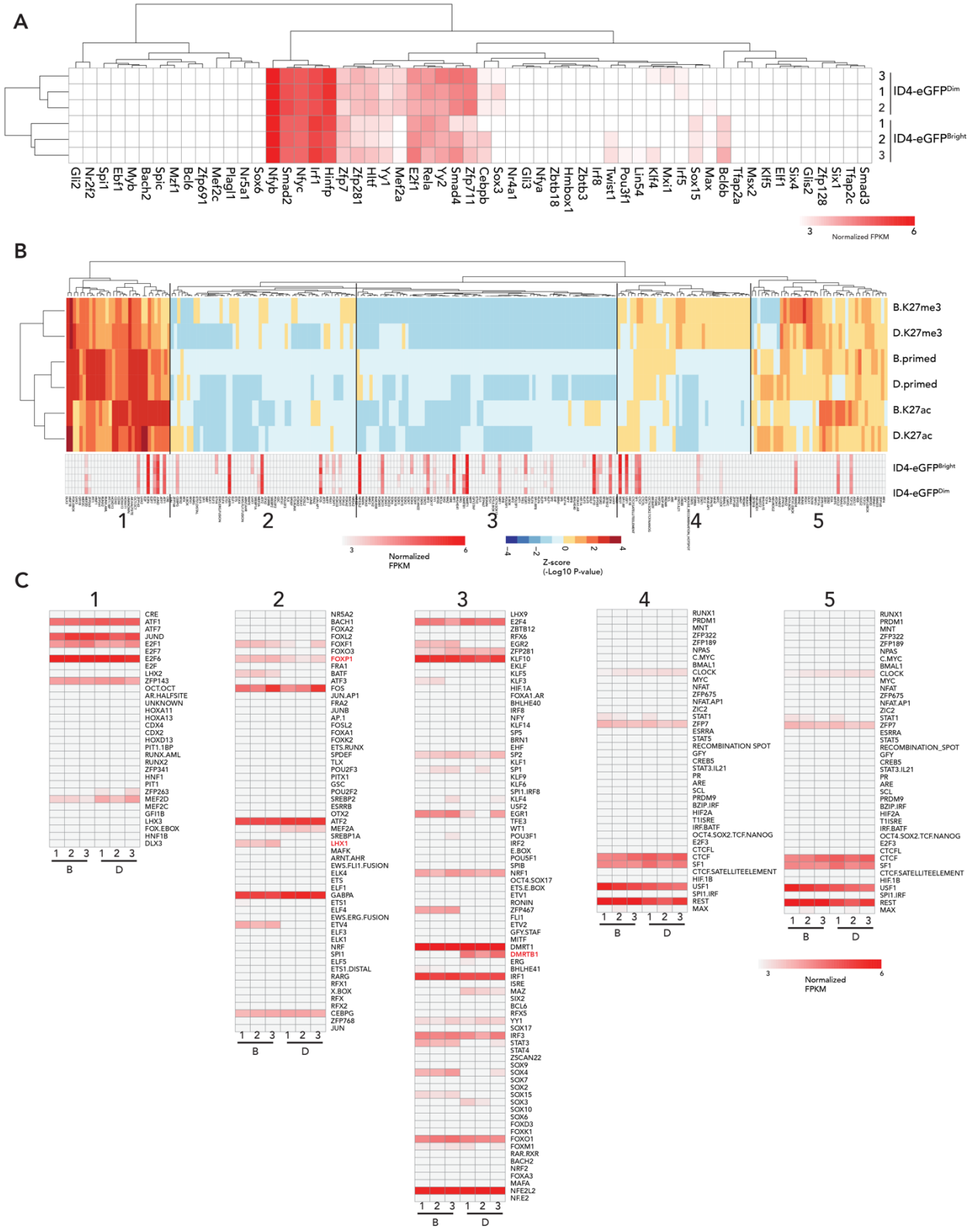


Figure S5. Related to Figure 6.

(A) The expression in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia of TFs predicted by de novo motif discovery as described in Figure6A. The color code indicates the library size normalized FPKM for each transcript in each replicate sample of each spermatogonial subpopulation. The color code is indicative of

relative levels of transcript detection. White = undetectable or negligible/background transcript levels, light – dark red = increasingly higher transcript levels above background.
(B) Overall de novo motif analysis of different kinds of enhancers. Rows (Enhancers) and columns (Motifs) are hierarchically clustered. Bottom bars indicate the normalized expression of corresponding matched TFs. The color code indicates the library size normalized FPKM.
(C) An enlarged version of the gene expression data from the clustered heatmap shown in part b grouped per the hierarchical clusters shown in part b. B = ID4-eGFP^{Bright} cells and D = ID4-eGFP^{Dim} cells.

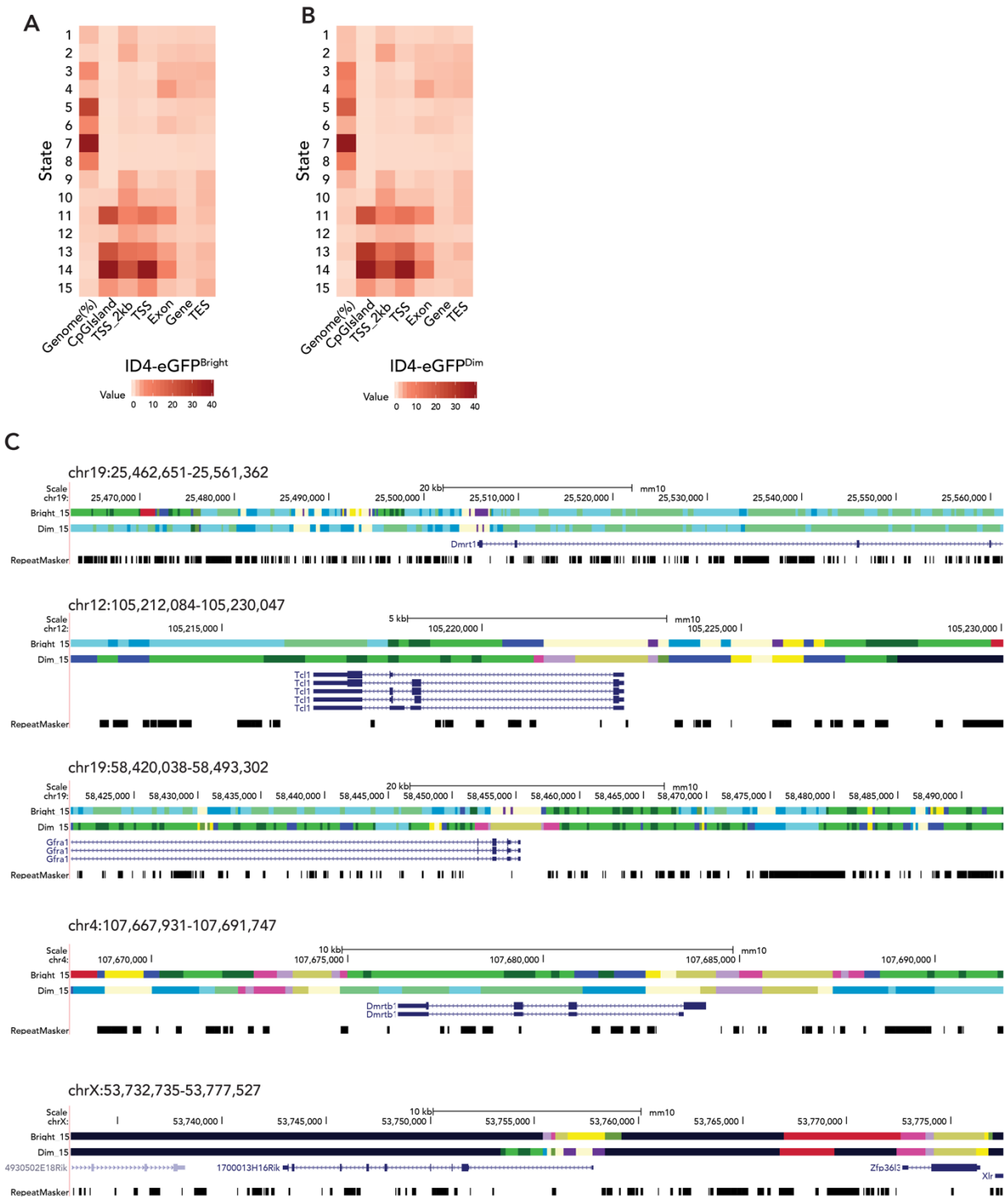


Figure S6. Differential chromatin states are associated with spermatogonial subtype-specific differential gene expression. Related to Figure 7.

(A, B) The relative genome-wide distribution of each of the 15 chromatin states shown in Figure 7a as a function of specific genomic features in ID4-eGFP^{Bright} (A) and ID4-eGFP^{Dim} (B) spermatogonial subtypes.

(C) Browser views of ChromHMM genome annotations representative of the 15 different chromatin states described in Figure 7A and in parts Fig.S6A and Fig.S6B above. Data is shown for one gene (*Dmrt1*) expressed at similar levels in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia, two genes (*Tcl1* &

Gfra1) up-regulated in ID4-eGFP^{Bright} spermatogonia, and three genes (*Dmrtb1, 4930502E18Rik* & 1700013H16Rik) up-regulated in ID4-eGFP^{Dim} spermatogonia. Color coding of chromatin states is as shown in Figure 7A.

Transparent Methods

Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-Histone H3 (mono methyl K4)	Abcam	Cat#ab8895, RRID:AB_306847
Rabbit anti-Histone H3 (di methyl K4)	Abcam	Cat#ab7766, RRID:AB_2560996
Rabbit anti-Histone H3 (tri methyl K4)	Abcam	Cat#ab8580, RRID:AB_306649
Rabbit anti-Histone H3 (tri methyl K9)	Abcam	Cat#ab8898, RRID:AB_306848
Rabbit anti-Histone H3K27ac	Active Motif	Cat#39133, RRID:AB_2561016
Rabbit anti-Histone H3 (tri methyl K27)	Abcam	Cat#ab6002, RRID:AB_305237
Mouse anti-5-methylcytosine (5-mC) clone 33D3	Diagenode	Cat#C15200081-100 RRID:AB_2572207
Rabbit anti-FOXP1	Abcam	Cat#ab16645, RRID:AB_732428
Rabbit anti-DMRTB1 1:200	Abcam	Cat#ab241275
Mouse anti-LHX1	Santa Cruz Biotechnology	Cat#sc-515631
Mouse IgG	Abcam	Cat#ab18413, RRID:AB_2631983
Rabbit IgG	Abcam	Cat#ab171870, RRID:AB_2687657
Goat anti-GFRA1 1:250	R&D Systems	Cat#AF560, RRID:AB_2110307
Goat anti-FOXC2 1:100	R&D Systems	Cat#AF6989, RRID:AB_10973139
Rabbit anti-OCT4 1:200	Abcam	Cat#ab19857, RRID:AB_445175
Mouse anti-DMRT1 1:100	Santa Cruz Biotechnology	Cat#sc-377167
Rabbit anti-EGR1	Novus Biologicals	Cat#NBP2-56100
Rabbit anti-FOXO1	Abcam	Cat#ab39670, RRID:AB_732421
Rabbit anti-REST	Abcam	Cat#ab202962
Donkey anti-Goat IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor Plus 594	Invitrogen	Cat#A32758, RRID:AB_2762828
Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor Plus 647	Invitrogen	Cat#A32795, RRID:AB_2762835
Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 546	Invitrogen	Cat#A10036, RRID:AB_2534012
Chemicals, Peptides, and Recombinant Proteins		
DPBS, no calcium, no magnesium	Gibco	Cat#14190250
Hank's Balanced Salt Solution	Gibco	Cat#14170112
HEPES (1 M)	Gibco	Cat#15630080
Fetal Bovine Serum	Gibco	Cat#10082147
Trypsin-EDTA (0.25%), phenol red	Gibco	Cat#25200056
Deoxyribonuclease I	Sigma-Aldrich	Cat#DN25-10MG
Propidium Iodide	Invitrogen	Cat#P1304MP

Dihydrochloride	Invitrogen	Cat#D1306
Micrococcal Nuclease	New England Biolabs	Cat#M0247S
UltraPure™ 1M Tris-HCl, pH 8.0	Invitrogen	Cat#15568025
UltraPure™ 0.5M EDTA, pH 8.0	Invitrogen	Cat#15575020
cOmplete™ Protease Inhibitor Cocktail	Roche	Cat#11697498001
NP-40	Thermo Scientific	Cat#28324
Sodium dodecyl sulfate	Sigma-Aldrich	Cat#71725
Bovine Serum Albumin	Sigma-Aldrich	Cat#A2058
Triton™ X-100	Sigma-Aldrich	Cat#T8787
Sodium deoxycholate	Sigma-Aldrich	Cat#30970
Sodium chloride	Sigma-Aldrich	Cat#S3014
Sodium bicarbonate	Sigma-Aldrich	Cat#S5761
Digitonin	Sigma-Aldrich	Cat#D141
TWEEN® 20	Sigma-Aldrich	Cat#P9416
UltraPure™ Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v)	Invitrogen	Cat#15593031
T4 DNA Polymerase	New England Biolabs	Cat#M0203S
DNA Polymerase I, Large (Klenow) Fragment	New England Biolabs	Cat#M0210S
Klenow Fragment (3'→5' exo-)	New England Biolabs	Cat#M0212S
NEBNext® Multiplex Oligos for Illumina® (Dual Index Primers Set 1)	New England Biolabs	Cat#E7600S
AMPure XP	Beckman Coulter	Cat#A63881
Proteinase K	VIOGENE	Cat#PK0100
NEB buffer 2	New England Biolabs	Cat#B7002S
Lambda DNA	New England Biolabs	Cat#N3011S
Sodium phosphate dibasic	Sigma-Aldrich	Cat#S0876
Sodium phosphate monobasic	Sigma-Aldrich	Cat#S3139
NEBNext Ultra II Q5 Master Mix	New England Biolabs	Cat#M0544
Luna® Universal qPCR Master Mix	New England Biolabs	Cat# M3003L
Critical Commercial Assays		
Nextera DNA Sample Preparation Kit	Illumina	Cat#FC-121-1031
Nextera XT DNA Library Preparation Kit	Illumina	Cat#FC-131-1002
Nuclei Isolation Kit	Sigma-Aldrich	Cat#NUC101-1KT
Quick Ligation™ Kit	New England Biolabs	Cat#M2200L
SMART-Seq™ v4 Ultra Low Input RNA Kit	Clontech Laboratories	Cat#634888
Zymo DNA Clean and Concentrator-5 Kit	Zymo research	Cat#D4014
Qubit dsDNA HS Assay Kit	Invitrogen	Cat#Q32854
Dynabeads™ Protein A Immunoprecipitation Kit	Invitrogen	Cat#10006D
Dynabeads™ Protein G Immunoprecipitation Kit	Invitrogen	Cat#10007D
Deposited Data		
Raw and analyzed data	This paper	GEO:GSE131657
RNA-seq	This paper	GEO:GSE131653
ChIP-seq	This paper	GEO:GSE131657
ATAC-seq	This paper	GEO:GSE131655
MeDIP-seq	This paper	GEO:GSE131654
Single Cell RNA-seq	(Hermann et al., 2018)	GEO:GSE109049
CTCF ChIP-seq	(Yue et al., 2014)	GEO:GSM918711
DMRT1 ChIP-seq	(Murphy et al., 2015)	GEO:GSE64892

DMRTB1 RNA-seq	(Zhang et al., 2014)	GEO:GSM1480189
Experimental Models: Organisms/Strains		
Mus Musculus: Strain Id4-eGfp (LT-11B6)	(Chan et al., 2014)	N/A
Mus Musculus: Strain C57Bl6/JJ	The Jackson Laboratory	Cat#000664
Mus Musculus: Strain *Rosa26-lacZ*	The Jackson Laboratory	Cat#002073
Oligonucleotides		
Primer: Negative Control primer1 Forward: GCTCTGAAGAAATGCCCAGC	This paper	N/A
Primer: Negative Control primer1 Reverse: AGCCAGCAACGTTTCACCTA	This paper	N/A
Primer: Negative Control primer1 Forward: GGTTCAGTTCTCAGCACCCA	This paper	N/A
Primer: Negative Control primer1 Reverse: GGATTCCCCTCTCAGCTGTC	This paper	N/A
Primer: FOXP1 binds to Egr1 Forward: TCAGCCTGAGTCCTTACCCA	This paper	N/A
Primer: FOXP1 binds to Egr1 Reverse: CAGCCCGAAGCAGAACAAAC	This paper	N/A
Primer: FOXP1 binds to Egr2 Forward: CGGAATGGCTCCCAAACAAG	This paper	N/A
Primer: FOXP1 binds to Egr2 Reverse: GGAGGAATTCCGGTTCTCCG	This paper	N/A
Primer: FOXP1 binds to Etv5 Forward: TTAAGAGTCGCGGAGCGTTT	This paper	N/A
Primer: FOXP1 binds to Etv5 Reverse: TACAGAAGCGGGGTGCAAG	This paper	N/A
Primer: LHX1 binds to Spry4 enhancer Forward: CTGGATGTAGAGATTTGGGGTGA	This paper	N/A
Primer: LHX1 binds to Spry4 enhancer Reverse: ACAACATCCTGTTCTTTTGTGAGAC	This paper	N/A
Primer: LHX1 binds to Cited2 promoter Forward: AGAAATCGCAAAGACGGAAGGT	This paper	N/A
Primer: LHX1 binds to Cited2 promoter Reverse: GCACATCCTGTTGTTATTCCCC	This paper	N/A
Primer: LHX1 binds to Cited2 enhancer Forward: ATGTAACTATCAGCGGTCACCC	This paper	N/A
Primer: LHX1 binds to Cited2 enhancer Reverse: CCTTGCTAAGTTGTTGGGCTTT	This paper	N/A
Primer: DMRTB1 binds to Syce2 Forward: CGAGTTCGCCGCCCCC	This paper	N/A
Primer: DMRTB1 binds to Syce2 Reverse: CTCACTCCGTGGCGCTC	This paper	N/A
Primer: DMRTB1 binds to Sohlh2 Forward: GTGTTCAAGTGAGCTGCG	This paper	N/A
Primer: DMRTB1 binds to Sohlh2 Reverse: TATTCAGCCCTGGTTCAG	This paper	N/A
Primer: Lambda DNA unmethylated primer Forward: GGCTAGAACTGACCAGACAGAC	This paper	N/A
Primer: Lambda DNA unmethylated primer Reverse: ATCTGTAGCCAATCCTAGAGCA	This paper	N//A
Primer: Lambda DNA methylated primer Forward: CATGGCCCACAAAGTAATAAAA	This paper	N/A
Primer: Lambda DNA methylated primer Reverse: AACGACTTACAACGAGCTCAAA	This paper	N/A

Software and Algorithms		
ImageJ	(Schneider et al., 2012)	https://imagej.nih.gov/ij/
FastQC v0.11.6	Babraham Institute	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Samtools v1.19	(Li et al., 2009)	http://samtools.sourceforge.net/
Picard	Broad Institute	https://broadinstitute.github.io/picard/
MACS2	(Zhang et al., 2008)	https://github.com/hbctraining/Intro-to-ChIPseq
ChIPseeker v1.24	(Yu et al., 2015)	https://www.bioconductor.org/packages/release/bioc/html/ChIPseeker.html
DAVID v6.8	(Huang et al., 2009)	https://david.ncifcrf.gov/
GREAT v3.00	(McLean et al., 2010)	http://great.stanford.edu/public/html/
clusterProfiler v3.16	(Yu, 2018)	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
HOMER v4.11	(Heinz et al., 2010)	http://homer.ucsd.edu/homer/index.html
MEME Suite v5.1.1		http://meme-suite.org/
ChromHMM v1.20	(Ernst and Kellis, 2017)	http://compbio.mit.edu/ChromHMM/
Primer-BLAST		https://www.ncbi.nlm.nih.gov/tools/primer-blast/
Rsubread v2.2.2	(Liao et al., 2019)	RRID:SCR_016945; http://bioconductor.org/packages/Rsubread/
QuasR v1.28.0	(Gaidatzis et al., 2015)	http://bioconductor.org/packages/release/bioc/html/QuasR.html
edgeR v3.29	(Robinson et al., 2010)	https://www.bioconductor.org/packages/release/bioc/html/edgeR.html
MEDIPS v1.40	(Lienhard et al., 2013)	https://www.bioconductor.org/packages/release/bioc/html/MEDIPS.html
Mclust v5.4.6	(Scrucca et al., 2016)	https://cran.r-project.org/web/packages/mclust/index.html

deepTools v2.0	(Ramirez et al., 2016)	https://deeptools.readthedocs.io/en/develop/
ngs.plot v2.61	(Shen et al., 2014)	https://github.com/shenlab-sinai/ngsplot
pyGenomeTracks v3.2	(Ramirez et al., 2018)	https://github.com/deeptools/pyGenomeTracks
pheatmap v1.0.12	(Kolde, 2015)	https://cran.r-project.org/web/packages/pheatmap/index.html
ggplot2 v3.1.0	(Wickham, 2011)	RRID:SCR_014601; https://cran.r-project.org/web/packages/ggplot2/index.html
Seurat v3.0	(Butler et al., 2018)	https://cran.r-project.org/web/packages/Seurat/index.html
Destiny v3.2.0	(Angerer et al., 2016)	https://bioconductor.org/packages/release/bioc/html/destiny.html
R Project for Statistical Computing v3.5.0	(Team, 2013)	http://www.r-project.org/

Mice and cells

All experiments utilizing animals were preapproved 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 National Institutes of Health (NIH) *Guide for the Care and Use of Laboratory Animals*. Testes were recovered from 6-day old (P6) F1 male offspring of a cross between *Id4-eGfp* (LT-11B6) and either C57Bl6/JJ or *Rosa26-lacZ* (The Jackson Laboratory #000664, #002073) mice and used to generate suspensions of cells by enzymatic digestion as described[28]. Briefly, ID4-eGFP+ testes were distinguished by fluorescence microscopy and then subjected to dissociation and FACS sorting. After removing the tunica albuginea, testes were digested with DNAaseI + trypsin (0.05% trypsin, 10 µg/ml Dnase I in HBSS with 0.5 mM EDTA) to generate a single cell suspension. The resulting dissociated cells were washed and resuspended in Dulbecco's phosphate-buffered saline (DPBS) + 10% FBS, filtered through a 40 micron strainer to remove Sertoli cells and cell clumps prior to being subjected to fluorescence-activated cell sorting (FACS) as previously described[28]. Cells at a concentration of approximately 15×10^6 cells/ml DPBS + 10% FBS were subjected to flow cytometry using BD FACS Aria. Propidium iodide (PI) was added to discriminate dead cells at $5\mu l/10^6$ cells. Positive ID4-eGFP epifluorescence was determined by comparison to testis cells from testes of wild-type mice lacking the P6 *Id4-eGfp* transgene. The gating area of eGFP positive was subdivided into thirds to define the ID4-eGFP+ subsets as being Dim (lower third) or Bright (upper third) by fluorescent intensity as described(Helsel et al., 2017, Hermann et al., 2018).

Bulk RNA-seq

Aliquots of cells from each of three replicate samples of each spermatogonial subpopulation were used for separate bulk RNA-seq analyses to catalogue gene expression in each spermatogonial subtype. Populations of at least ≥ 1000 ID4-eGFP^{Bright} or ID4-eGFP^{Dim} cells recovered by FACS sorting were counted, pelleted, and subjected to direct cDNA synthesis using the SMART-Seq v4 ultra Low Input RNA Kit for Sequencing (Clontech Laboratories #634888). Approximately 250pg of cDNA was used for preparation of dual-indexed libraries using the Nextera XT DNA Library Preparation Kit (Illumina # FC-131-1002) following the manufacturer's procedures.

ChIP-seq

Aliquots of cells from each of three replicate samples of each spermatogonial subpopulation were used for separate ChIP-seq analyses to detect genome-wide enrichment patterns of six different histone modifications – H3K4me1, H3K4me2, H3K4me3, H3K9me3, H3K27me3 and H3K27ac. Approximately 1×10^6 cells were used for each immunoprecipitation (IP). ULI-NChIP-seq was performed as previously described(Brind'Amour et al., 2015). Briefly, FACS-sorted cells were pelleted and re-suspended in nuclear isolation buffer (Sigma #NUC101-1KT). Depending on input size, chromatin was fragmented for 5-7.5 min using MNase, and diluted in NChIP immunoprecipitation buffer (20mM Tris-HCl pH 8.0, 2mM EDTA, 15 mM NaCl, 0.1% Triton X-100, 1×EDTA-free protease inhibitor cocktail and 1mM phenylmethanesulfonyl fluoride). 10% of each sample was reserved as input control. Chromatin was pre-cleared with 5 or 10 µl of 1:1 protein A:G Dynabeads (Life Technologies #10015D) and immunoprecipitated with H3K4me1 (Abcam #ab8895), H3K4me2 (Abcam #ab7766), H3K4me3 (Abcam #ab8580), H3K9me3 (Abcam #ab8898), H3K27ac (Active Motif #39133) and H3K27me3 (Abcam #ab6002) antibody-bead complexes overnight at 4°C. IPed complexes were washed twice with 400 µl of ChIP wash buffer I (20 mM Tris-HCl, pH 8.0, 0.1% SDS, 1% Triton X-100, 0.1% deoxycholate, 2 mM EDTA and 150mM NaCl) and twice with 400 µl of ChIP wash buffer II (20mM Tris-HCl (pH 8.0), 0.1% SDS, 1% Triton X-100, 0.1% deoxycholate, 2mM EDTA and 500mM NaCl). Protein-DNA complexes were eluted in 30 µl of ChIP elution buffer (100mM $NaHCO_3$ and 1% SDS) for 2h at 68°C. IPed material was purified by Phenol:Chloroform:Isoamyl Alcohol (PCI), 25:24:1 V/V (Invitrogen #15593031), ethanol-precipitated and raw ChIP material was resuspended in 10 mM Tris-HCl pH 8.0. Fragment length was confirmed using a Bioanalyzer (Aglient Technology), and DNA concentration was determined by the Qubit dsDNA HS Assay Kit (Invitrogen #Q32854). Illumina libraries were constructed using a modified custom paired-end protocol. In brief, samples were end-repaired in 1×T4 DNA ligase buffer, 0.4 mM dNTP mix, 2.25U T4 DNA polymerase, 0.75U Klenow DNA polymerase and 7.5U T4 polynucleotide kinase for 30 min at 21-25°C, then A-tailed in 1×NEB buffer 2, 0.4 mM dNTPs and 3.75U of Klenow(exo-) for 30 min at 37°C and then ligated in 1×rapid DNA ligation buffer plus 1mM Illumina PE adapters and 1,600U DNA ligase for 1-8h at 21-25°C. Ligated fragments were amplified using dual-indexed primers for Illumina (NEB #E7600S) for 8-10 PCR cycles. DNA was purified with 1.8×volume Ampure XP DNA purification beads (Beckman Coulter #A63881) between each step. Fragment length was again checked by Bioanalyzer (Aglient Technology), and DNA concentration was determined using the Qubit dsDNA HS Assay Kit (Invitrogen #Q32854).

ATAC-seq

After FACS sorting, each aliqout of fresh cells (~50,000 cells/aliquot) was pelleted and re-suspended in transposition mix (25µl 2×TD buffer, 2.5 µl Tn5 transposase (100 nM final), 16.5µl

PBS, 0.5 µl 1% Digitonin, 0.5µl 10% Tween-20, 5µl H_2O) and incubated at 37°C for 30 min in a thermomixer. The mix was then cleanuped with a Zymo DNA Clean and Concentrator-5 Kit (Zymo research #D4014). Transposed fragments were amplified for 5 cycles using 25 µl NEB Q5 master mix, 2.5 µl Illumina i5 index primer and 2.5µl Illumina i7 index primer. Cycling conditions was 72°C for 5 min, 98°C for 10 sec, then 5 cycles of 98°C for 10 sec, 63°C for 30 sec, 72°C for 1 min. qPCR of amplified products was determined to add appropriate additional cycles.

MeDIP-seq

MeDIP-seq libraries were constructed as previously described(Taiwo et al., 2012). After FACS sorting, each aliqout (~50,000) of fresh cells was pelleted and re-suspended in lysis buffer (50mM Tris pH 8.0, 10 mM EDTA, 100 mM NaCl, 1% SDS, 0.5 mg/ml proteinase K) and incubated at 55°C for 5h. Genomic DNA was isolated using PCI, and sheared using a Bioruptor (Diagenode UCD-200). 10% raw sheared DNA was retained to serve as input control. Samples were end-repaired in 1×T4 DNA ligase buffer, 0.4mM dNTP mix, 2.25U T4 DNA polymerase, 0.75U Klenow DNA polymerase and 7.5U T4 polynucleotide kinase for 30 min at 21-25°C, then A-tailed in 1×NEB buffer 2, 0.4mM dNTPs and 3.75U of Klenow(exo-) for 30 min at 37°C, and then ligated in 1×rapid DNA ligation buffer, 1mM Illumina PE adapters and 1,600U DNA ligase for 1-8 h at 21-25°C. Samples were denatured at 95°C for 10 min, then transfered immediately to ice to prevent re-annealing. 0.2pM λ-DNA fragments (50% methylated) were used as a spike-in control. MeDIP on purified adapter-ligated DNA with spike-in was performed in 0.1 M Na_2HPO_4/NaH_2PO_4, 35µl of 2 M NaCl, 2.5 µl of 10% Triton X-100 and 1µl of anti-methylcytidine antibody (1 mg/ml Diagenode #MAb-081-100) overnight. DNA-IgG complexes were captured by protein A/G agarose beads. DNA was extracted by PCI. Recovery (%) of MeDIP was calculated as 2^amplification efficiency$^{(\text{Adjusted InputCt - MeDIPCt})}$ ×100%. Specificity of MeDIP is calculated as: Specificity = 1-(unmeth recovery/meth recovery). Only libraries with specificity ≥95% and unmethylated recovery of < 1% were used for further analysis.

Next Generation DNA Sequencing

Libraries were quantified by PCR using the NEBNext Library Quant Kit from Illumina (NEB #E7630L). After quantification, libraries were pooled in equal molar concentrations. RNA-seq libraries were sequenced on an Illumina HiSeq 2000 (PE100) at the University of Texas Southwestern Medical Center Sequencing Core. ChIP-seq, ATAC-seq and MeDIP-seq libraries were sequenced on an Illumina Hiseq 3000 sequencer (PE100) at the University of Texas Health Science Center at San Antonio Sequencing Core according to standard Illumina protocols.

Bioinformatics Analyses

Sequencing and Alignments: All raw fastq files were mapped to the UCSC mm10 genome reference using Rsubread(Liao et al., 2019) or QuasR(Gaidatzis et al., 2015).

RNA-seq analysis: Count matrices assigned to genes were obtained using featureCounts(Liao et al., 2014). Differential expression was inferred using edgeR(Robinson et al., 2010). Genes with $p < 0.01$ and LFC >1.5 were considered significantly differentially expressed.

Single cell RNA-seq analysis: Raw count table of P6 ID4-eGFP Bright/Dim spermatogonia were downloaded from GEO:GSE109049. Seurat and Diffusion Map were used for analyzing scRNA-seq data.

ChIP-seq analysis: RPKM of histone H3 modifications including H3K4me1, H3K4me2, H3K4me3, H3K9me3, H3K27ac and H3K27me3 on promoters (TSS ± 500bp) were determined,

log-transformed and defined as positive if their enrichment value was ⩾ a threshold established by fitting a two-component Gaussian mixture model using Mclust(Scrucca et al., 2016). Read coverage, K-means clustering and heatmap visualization were performed by deepTools(Ramirez et al., 2016) and ngs.plot.r(Shen et al., 2014). Repeat element annotations (RepeatMasker) were downloaded from the UCSC table browser (mm10)(Haeussler et al., 2019). Repeat consensus sequences were downloaded from Dfam3.1(Hubley et al., 2016).

ATAC-seq analysis: To identify potential enhancer loci, sequence within +/– 1kb from each ATAC-seq peak was examined. All ATAC-peaks not overlapping with promoters, known gene bodies, or extended transcription end sites were examined. The histone enrichment in these regions was determined by fitting a two-component Gaussian mixture model using Mclust(Scrucca et al., 2016).

MeDIP-seq analysis: Genome-wide differential coverage analysis of MeDIP-seq data was conducted using MEDIPS(Lienhard et al., 2013). Differentially methylated regions were annotated by ChIPseeker and interpreted by GREAT(McLean et al., 2010).

Peak calling: Duplicated reads were removed by Picard (http://broadinstitute.github.io/picard/). Regions enriched for H3K9me3 or H3K27me3 were determined using MACS2 peak callers on non-duplicated, uniquely aligned reads(Zhang et al., 2008). Broad peaks (H3K9me3, H3K37me3) were identified using MACS2 broadpeaks ($p < 1\times10^{-6}$, FDR < 0.01) and narrow peaks (H3K4me1, H3K4me2, H3K4me3, H3K27ac, ATAC-seq and MeDIP-seq) were identified with MACS2 ($p < 1\times10^{-6}$, FDR < 0.01). Peaks closer than 2 kb apart were merged and peaks larger than 0.5 kb were included in our analysis. Peaks were compared and annotated using ChIPseeker(Yu et al., 2015).

Gene Ontology analysis: GO analysis were determined using DAVID(Huang et al., 2009) or clusterProfiler(Yu, 2018). Functional interpretation of enhancer-like regions was performed using GREAT using default parameters(McLean et al., 2010).

Motif analysis: Promoter *de novo* motif discovery was performed by using HOMER(Heinz et al., 2010). De novo motif analysis within enhancer regions was analyzed with MEME-suit with default parameters(Bailey et al., 2009). All known motifs used in our study were defined by HOMER and HOCOMOCO mouse full V11(Kulakovskiy et al., 2018).

Integrating chromatin states: Chromatin states, were assigned after the mouse genome was discretized into 200bp bins and subjected to a 15-state Hidden Markov modeling analyses using the ChromHMM method with default parameters(Ernst and Kellis, 2017). CTCF(Rivero-Hinojosa et al., 2017), DMRT1(Murphy et al., 2015) and DMRTB1(Zhang et al., 2014) ChIP-seq coverage from published studies of adult mouse testes and data from our analyses of P6 mouse testes were integrated and visualized by pyGenomeTracks and UCSC genome browser(Kent et al., 2002).

Cell cycle analyses. These analyses were done as described previously (Mutoji et al., 2016). Briefly, cells from P6 *Id4-eGfp*+ mouse testes were suspended in DPBS+S and treated with 50μM verapamil (Sigma-Aldrich) for 5 min at 37°C, labeled with 5-10μM Vybrant® DyeCycle™ Violet Stain (ThermoFisher Scientific) for an additional 30 minutes at 37 ˚C, and then cooled on ice for 5 minutes. Evaluation of cell staining was performed utilizing an LSRII cytometer and cell cycle state was determined from these data using FlowJo v.10.0.7 with the Cell Cycle Univariate analysis (Watson et al., 1987). Results were from four independent labeling experiments.

Factor/Gene-Specific ChIP and Real-time PCR

FACS-sorted populations of P6 ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia were fixed in freshly prepared cross-linking buffer (0.1M NaCl, 1mM EDTA, 0.5mM EGTA, 50mM HEPES (pH 8.0), 11% Formaldehyde). Cells were lysed in buffer L1 (140mM NaCl, 1mM EDTA, 50mM

HEPES, 10% Glycerol, 0.5% NP-40, 0.25% Triton X-100) and nuclei were isolated using buffer L2 (200mM NaCl, 1mM EDTA, 0.5mM EGTA, 10mM Tris). Chromatin was sheared to average size of 500bp using a Bioruptor (Diagenode UCD-200). FOXP1 (Abcam #ab16645), DMRTB1 (Abcam #ab241275), LHX1 (Santa Cruz Biotechnology #sc-515631) and IgG (Abcam #ab37355 & ab171870) antibodies were coupled to DynBeads in DPBS (5 mg/ml BSA) by incubating overnight on a rotating platform at 4˚C. Chromatin was precipitated by antibody-bead complexes in IP buffer (1% TritonX-100, 0.1% deoxycholate sodium salt, 1× Complete protease inhibitor, 10mM Tris-Cl (pH 8.0), 1mM EDTA) overnight on a rotating platform at 4˚C. DNA-antibody-bead complexes were washed 10 times using freshly prepared RIPA buffer (50mM HEPES, 1 mM EDTA, 1% NP-40, 0.7% deoxycholate sodium salt, 0.5M LiCl, 1× complete protease inhibitor). DNA was eluted in elution buffer (10 mM Tris (pH 8.0), 1mM EDTA, 1% SDS) and cross-linkages were reversed overnight at 65˚C. After proteinase K digestion, DNA was purified with PCI. ChIP-qPCR was performed on QuantStudio 3 instrument (Applied Biosystems) using Luna® Universal qPCR Master Mix (NEB #M3003S) following instructions in the reagent manual. ChIP DNA and control DNA were used as templates. Primers flanking potential factor-binding sites were designed by Primer-BLAST[90]. Fold Enrichment was calculated by $2^{-\Delta\Delta Ct}$.

Immunohistochemistry

Immunolabeling was done as previously described(Serra et al., 2017). Briefly, testes were immersion-fixed in fresh 4% paraformaldehyde, washed in PBS, incubated overnight in 30% sucrose at 4°C, and frozen in O.C.T. Five micrometer sections were incubated in blocking reagent (PBS containing 3% BSA and 0.1% Triton X-100) for 30 min at room temperature. Primary antibodies were used against GFRA1 (R&D Systems #AF560, 1:250), FOXC2 (R&D Systems #AF6989, 1:100), POU5F1 (Abcam #ab19857, 1:200), DMRT1 (Santa Cruz Biotechnology #sc-377167, 1:200), DMRTB1 (Abcam #ab241275, 1:200), EGR1 (Novus Biologicals #NBP2-56100, 1:100), FOXO1 (Abcam #ab39670, 1:500), REST (Abcam #ab202962, 1:200). Primary antibodies were diluted with blocking reagent and incubated on tissue sections for 1 h at room temperature. Primary antibody was omitted as a negative control. Following stringency washes, sections were incubated in secondary antibody (1:500, Donkey anti-Goat IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor Plus 594, Invitrogen #A32758; Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor Plus 647, Invitrogen #A32795) with phalloidin-405 (at 1:500, Invitrogen) for 1h at room temperature. Blocking and antibody incubations were done in PBS containing 3% BSA and 0.1% TritonX-100, and stringency washes were done with PBS and 0.1% TritonX-100. Cover slips were mounted with Vectastain containing DAPI (Vector Laboratories #H-1200), and images obtained using a confocal laser-scanning microscope (ZEISS 710), processing with ImageJ(Schneider et al., 2012).

Supplemental References

Angerer, P., Haghverdi, L., Buttner, M., Theis, F.J., Marr, C., and Buettner, F. (2016). destiny: diffusion maps for large-scale single-cell data in R. Bioinformatics *32*, 1241–1243.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME Suite: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202–W208.

Brind'Amour, J., Liu, S., Hudson, M., Chen, C., Karimi, M.M., and Lorincz, M.C. (2015). An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. Nat. Commun. *6*, 6033.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Chan, F., Oatley, M.J., Kaucher, A.V., Yang, Q.E., Bieberich, C.J., Shashikant, C.S., and Oatley, J.M. (2014). Functional and molecular features of the Id4+ germline stem cell population in mouse testes. Genes Dev. *28*, 1351–1362.

Ernst, J., and Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. Nat Protoc *12*.

Gaidatzis, D., Lerch, A., Hahne, F., and Stadler, M.B. (2015). QuasR: quantification and annotation of short reads in R. Bioinforma. Oxf. Engl. *31*, 1130–1132.

Haeussler, M., Zweig, A.S., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J., Lee, C.M., Lee, B.T., Hinrichs, A.S., Gonzalez, J.N., et al. (2019). The UCSC Genome Browser database: 2019 update. Nucleic Acids Res. *47*, D853–D858.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell *38*, 576–589.

Helsel, A.R., Yang, Q.-E., Oatley, M.J., Lord, T., Sablitzky, F., and Oatley, J.M. (2017). ID4 levels dictate the stem cell state in mouse spermatogonia. Development *144*, dev.146928–634.

Hermann, B.P., Cheng, K., Singh, A., Roa-De La Cruz, L., Mutoji, K.N., Chen, I.C., Gildersleeve, H., Lehle, J.D., Mayo, M., Westernströer, B., et al. (2018). The Mammalian Spermatogenesis Single-Cell Transcriptome, from Spermatogonial Stem Cells to Spermatids. Cell Rep. *25*, 1650–1667.e8.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Hubley, R., Finn, R.D., Clements, J., Eddy, S.R., Jones, T.A., Bao, W., Smit, A.F.A., and Wheeler, T.J. (2016). The Dfam database of repetitive DNA families. Nucleic Acids Res. *44*, D81-89.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, and D. (2002). The Human Genome Browser at UCSC. Genome Res. *12*, 996–1006.

Kolde, R. (2015). pheatmap: Pretty heatmaps [Software].

Kulakovskiy, I.V., Vorontsov, I.E., Yevshin, I.S., Sharipov, R.N., Fedorova, A.D., Rumynskiy, E.I., Medvedeva, Y.A., Magana-Mora, A., Bajic, V.B., Papatsenko, D.A., et al. (2018). HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. Nucleic Acids Res. *46*, D252–D259.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Liao, Y., Smyth, G.K., and Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res.

Lienhard, M., Grimm, C., Morkel, M., Herwig, R., and Chavez, L. (2013). MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. Bioinforma. Oxf. Engl. *30*, 284–286.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of \textlessi\textgreatercis\textless/i\textgreater-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Murphy, M.W., Lee, J.K., Rojo, S., Gearhart, M.D., Kurahashi, K., Banerjee, S., Loeuille, G.A., Bashamboo, A., McElreavey, K., Zarkower, D., et al. (2015). An ancient protein-DNA interaction underlying metazoan sex determination. Nat. Struct. Mol. Biol. *22*, 442–U26.

Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res *44*, W160–5.

Ramirez, F., Bhardwaj, V., Arrigoni, L., Lam, K.C., Grüning, B.A., Villaveces, J., Habermann, B., Akhtar, A., and Manke, T. (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat. Commun. *9*, 189.

Rivero-Hinojosa, S., Kang, S., Lobanenkov, V.V., and Zentner, G.E. (2017). Testis-specific transcriptional regulators selectively occupy BORIS-bound CTCF target regions in mouse male germ cells. Sci. Rep. *7*, 1–13.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Scrucca, L., Fop, M., Murphy, T.B., and Raftery, A.E. (2016). mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. R J *8*, 289–317.

Serra, N.D., Velte, E.K., Niedenberger, B.A., Kirsanov, O., and Geyer, C.B. (2017). Cell-autonomous requirement for mammalian target of rapamycin (Mtor) in spermatogonial proliferation and differentiation in the mouse†. Biol. Reprod. *96*, 816–828.

Shen, L., Shao, N., Liu, X., and Nestler, E. (2014). ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics *15*, 284.

Taiwo, O., Wilson, G.A., Morris, T., Seisenberger, S., Reik, W., Pearce, D., Beck, S., and Butcher, L.M. (2012). Methylome analysis using MeDIP-seq with low DNA concentrations. Nat Protoc *7*, 617–636.

Team, R.C. (2013). R: A language and environment for statistical computing.

Watson, J.V., Chambers, S.H., and Smith, P.J. (1987). A pragmatic approach to the analysis of DNA histograms with a definable G1 peak. Cytometry *8*, 1–8.

Wickham, H. (2011). ggplot2: ggplot2. Wiley Interdiscip. Rev. Comput. Stat. *3*, 180–185.

Yu, G. (2018). clusterProfiler: An universal enrichment tool for functional and comparative study. BioRxiv 256784.

Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics *31*, 2382–2383.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*.

Zhang, T., Murphy, M.W., Gearhart, M.D., Bardwell, V.J., and Zarkower, D. (2014). The mammalian Doublesex homolog DMRT6 coordinates the transition between mitotic and meiotic developmental programs during spermatogenesis. Development *141*, 3662–3671.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.