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

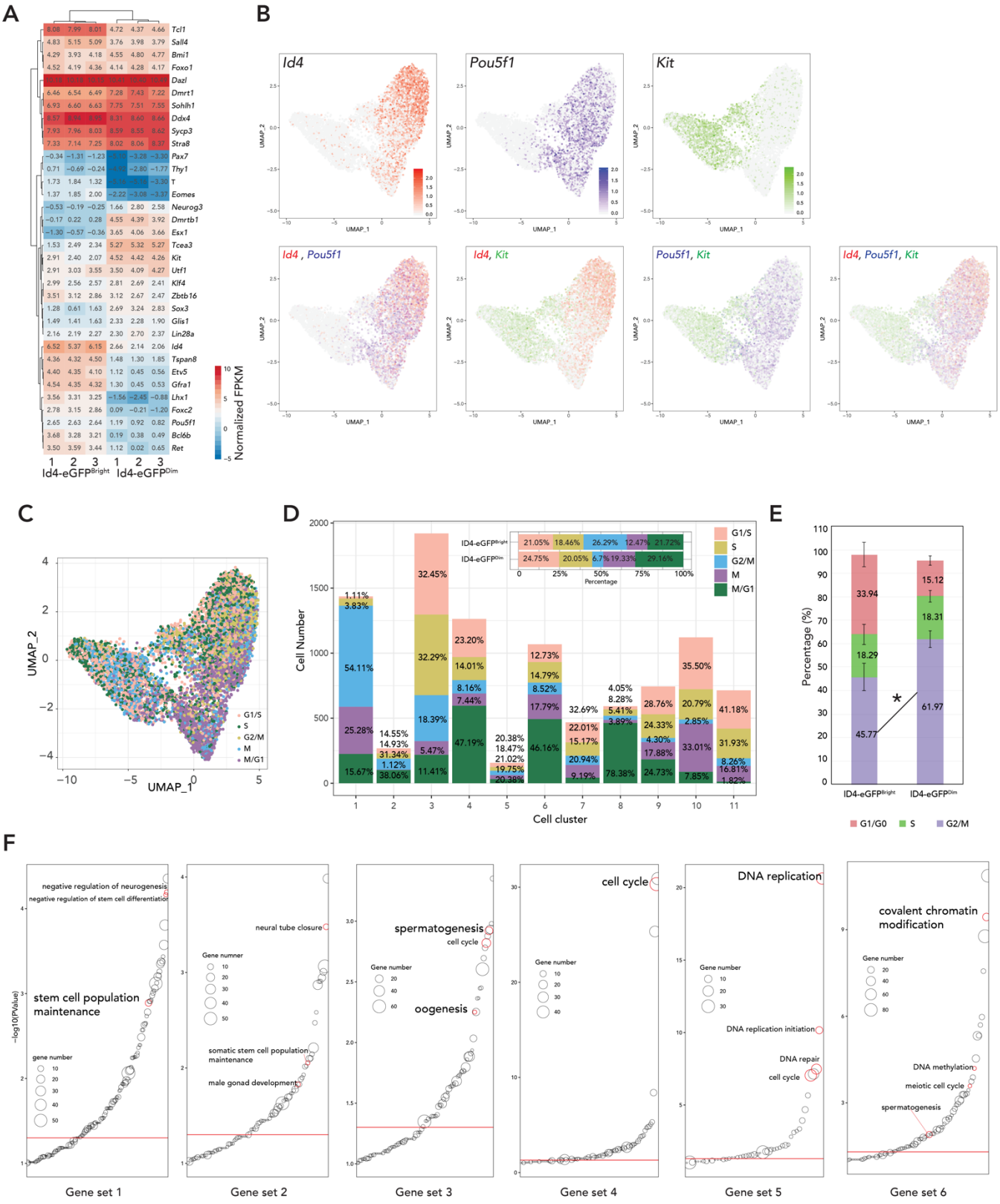

Figure S1. Differential gene expression in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia. Related to Figure 1 and Table S1.

(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.

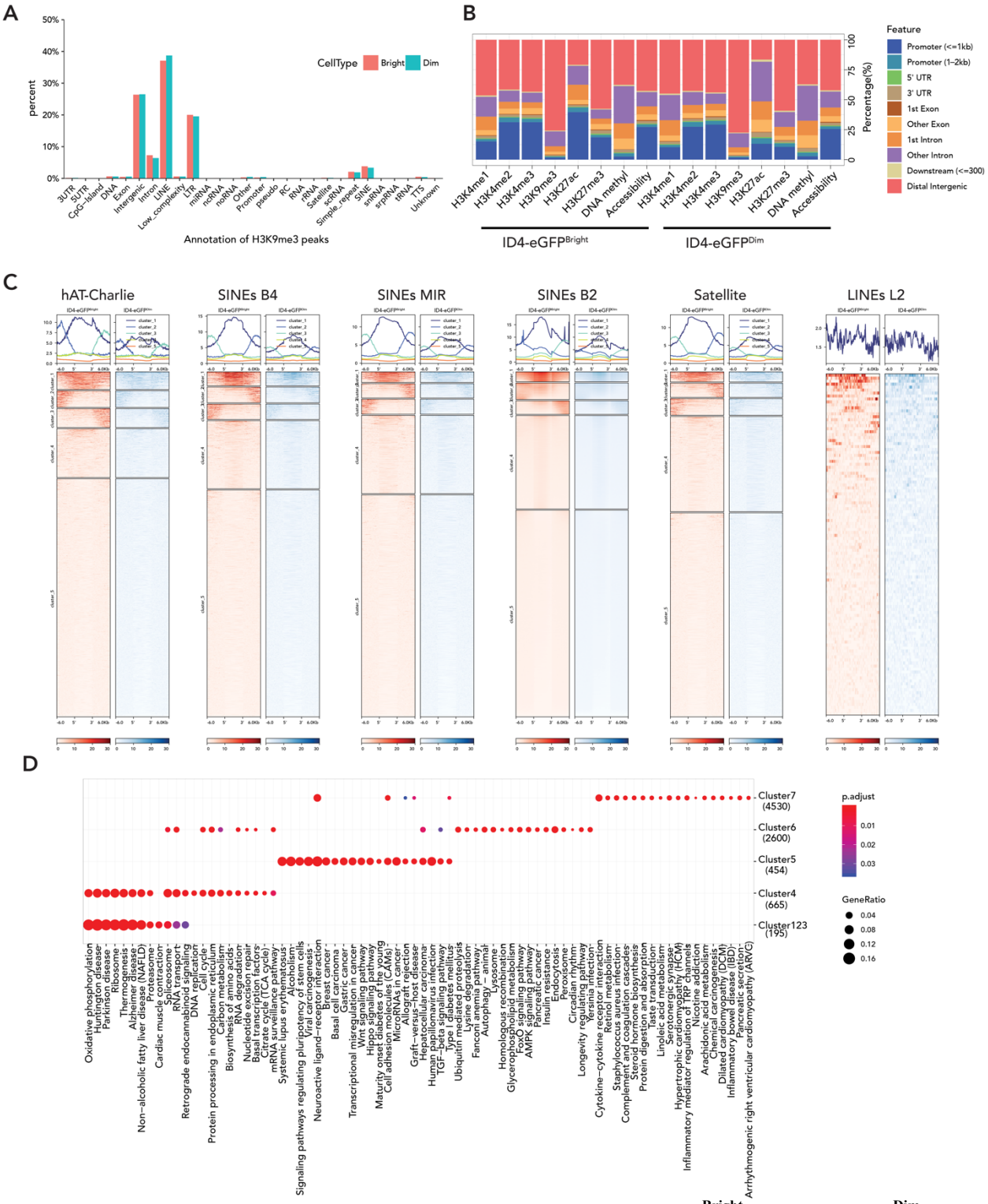

Figure S2. Genomic distributions of histone modifications in ID4-eGFP^{Bright} and ID4-eGFP^{Dim} spermatogonia. Related to Figure 2.

(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.

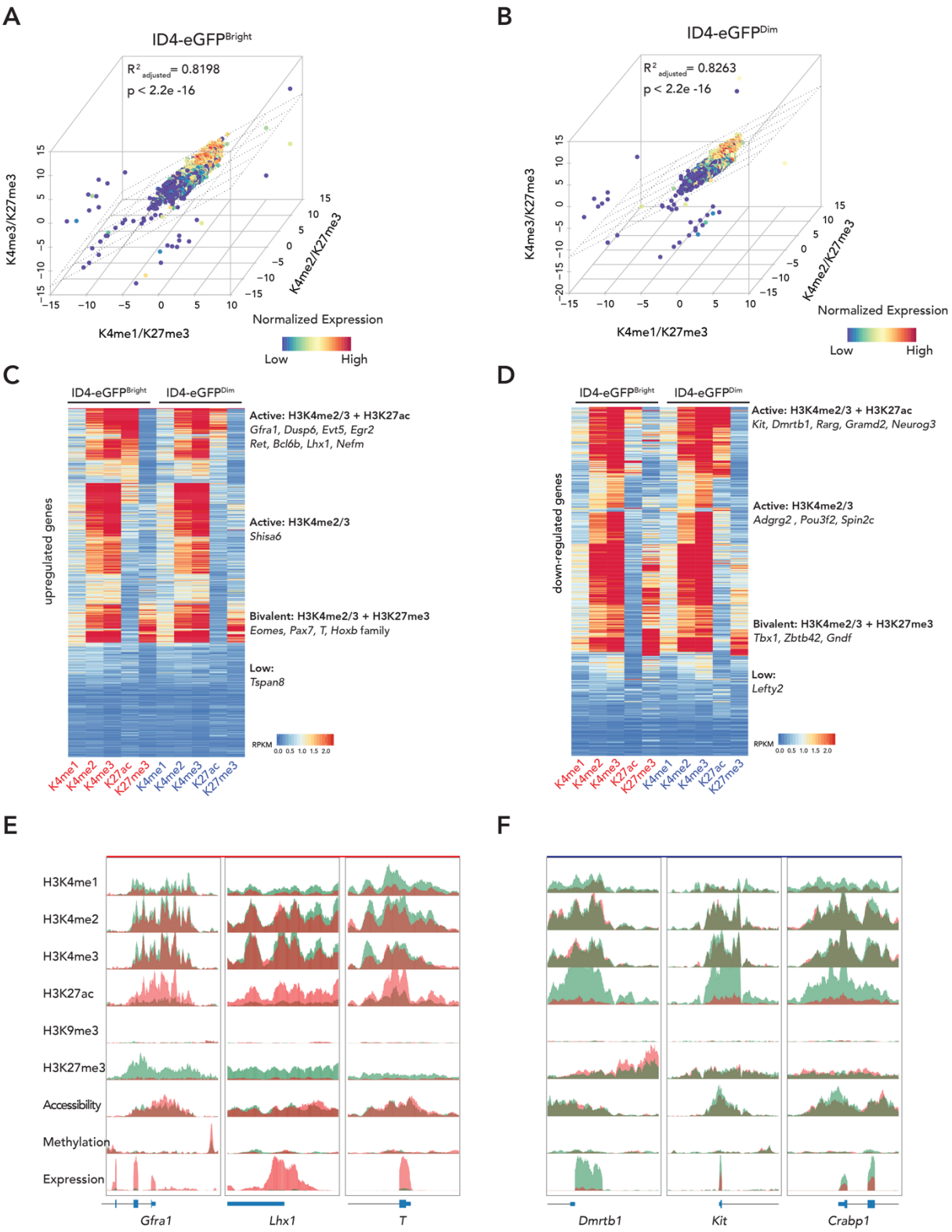

Figure S3. Related to Figure 3.

(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.

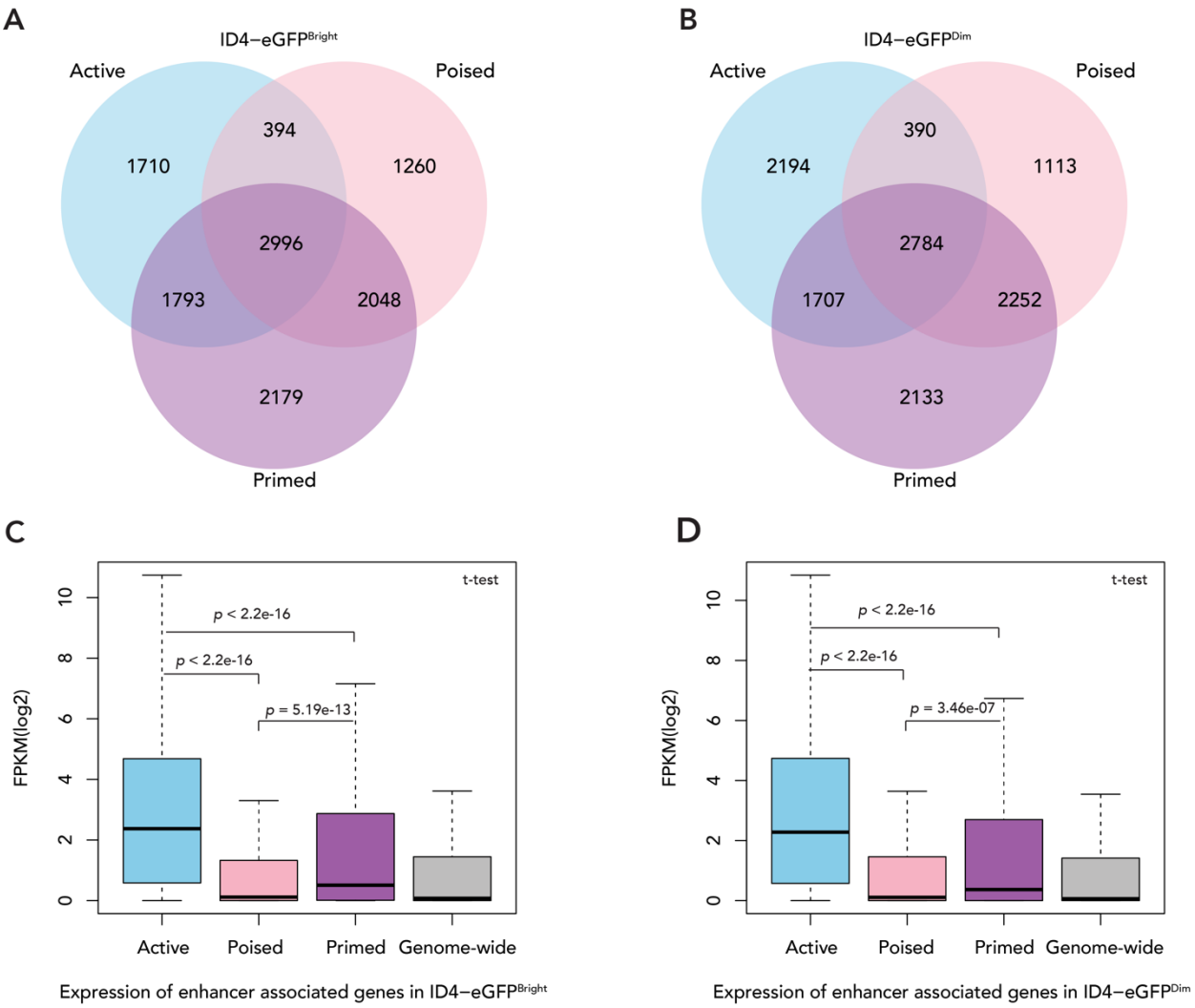

Figure S4: Differential expression of enhancer-associated genes. Related to Figure 4.
(A, B) Enhancer-associated genes in ID4-eGFP^{Bright} (A) and ID4-eGFP^{Dim} (B) spermatogonia. Many genes are associated with more than one enhancer, and those enhancers can be in different states. Venn diagrams show 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.

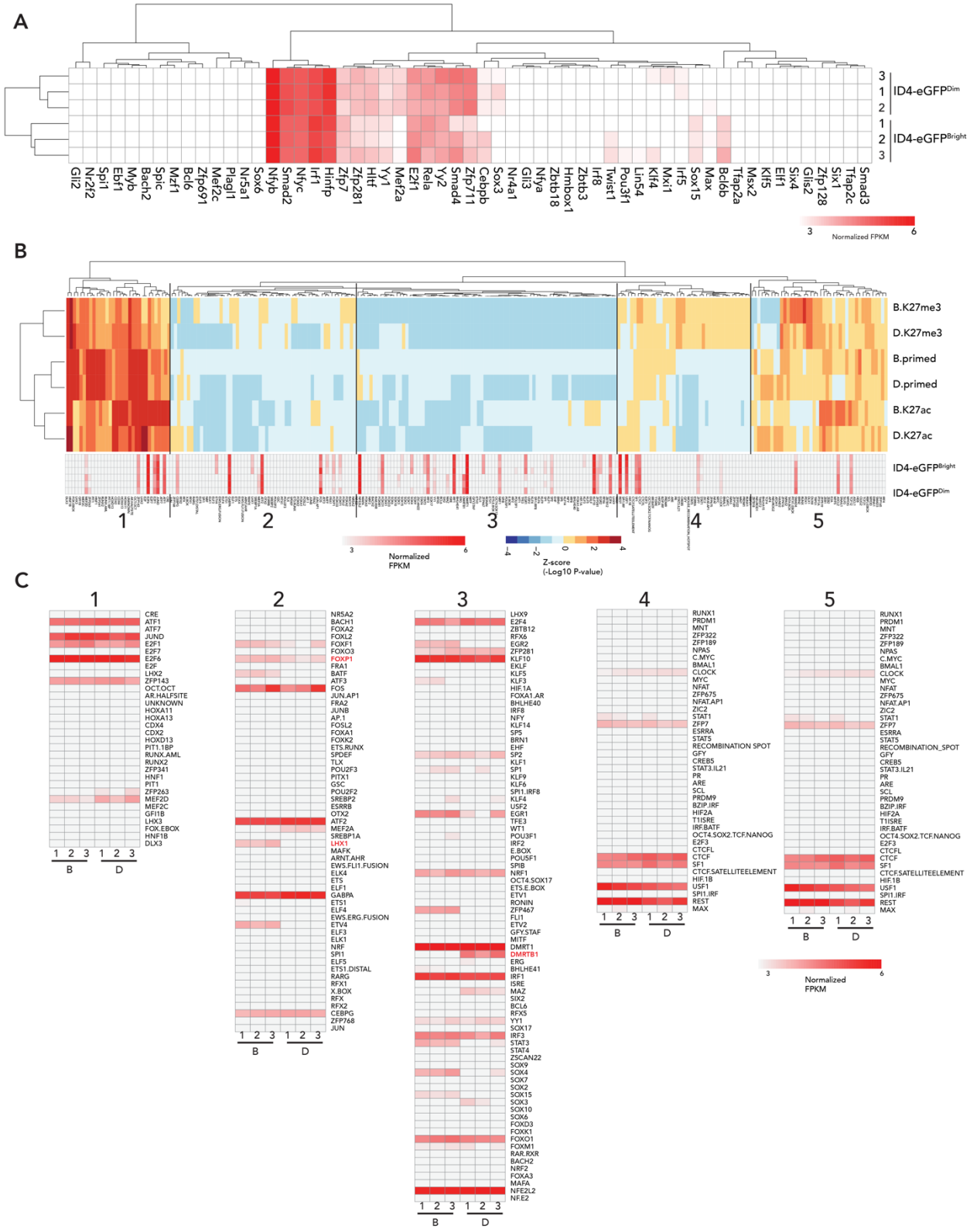

Figure S5. Related to Figure 6.

(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.

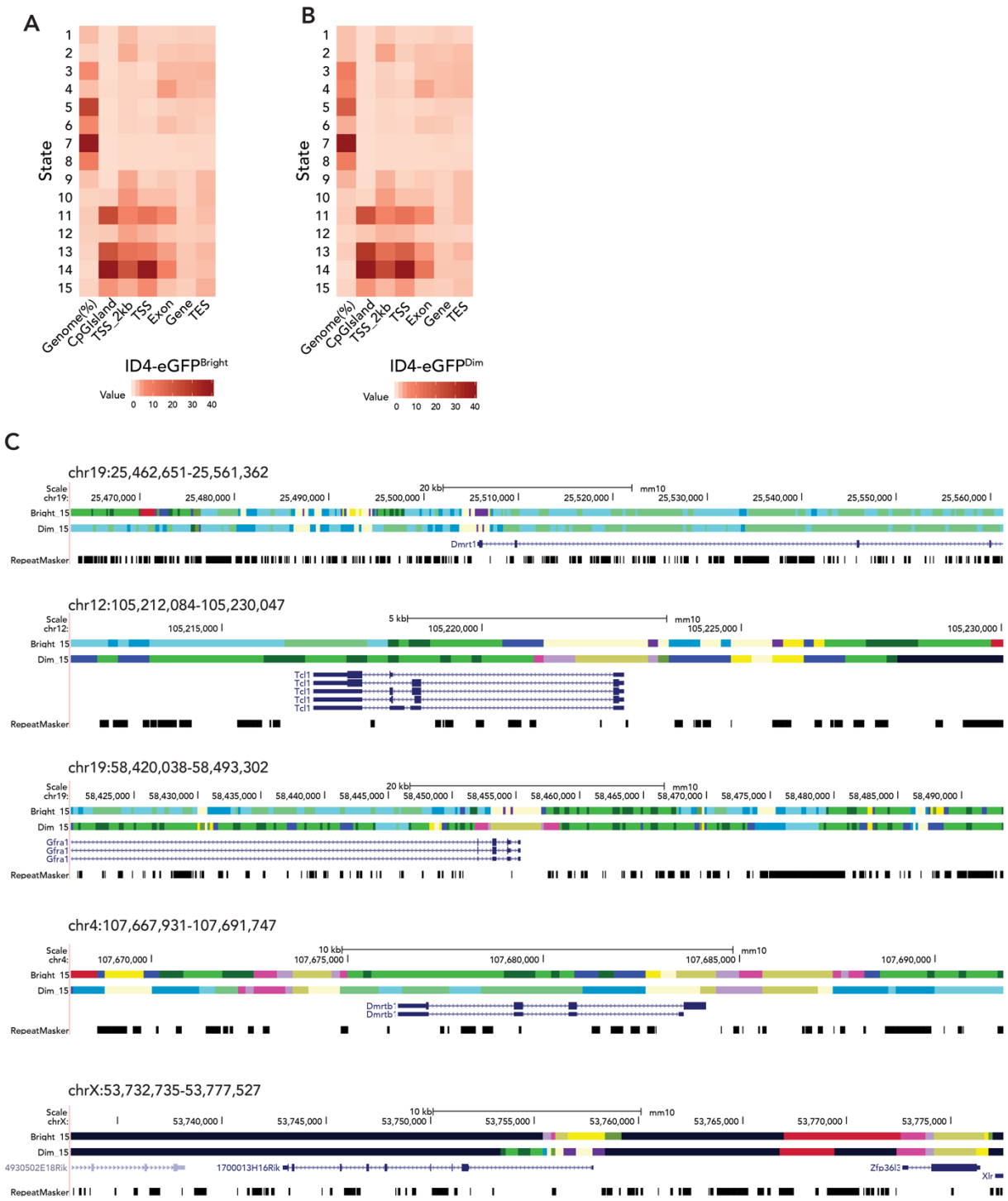

Figure S6. Differential chromatin states are associated with spermatogonial subtype-specific differential gene expression. Related to Figure 7.

(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.