The Dnmt3L ADD Domain Controls Cytosine Methylation Establishment during Spermatogenesis

Georgios Vlachogiannis, Chad E. Niederhuth, Salih Tuna, Athanasia Stathopoulou, Keijo Viiri, Dirk G. de Rooij, Richard G. Jenner, Robert J. Schmitz, and Steen K.T. Ooi
Figure S1

A

5' Probe

B

8.94kb

C

6.83kb

D

Dnmt3L

E

CG

F

% Methylation

G

ES PSG

H

LINE1-Tf

I

ES 1dpp testes

Dnmt3L

Tubulin

Dnmt3A

Dnmt3B

Cdc2
A Genes up-regulated in Dnm3L<sup>AA</sup> spermatogonia

| Ref Seq ID | Gene Name | Ref Seq ID | Gene Name | Ref Seq ID | Gene Name |
|-----------|-----------|-----------|-----------|-----------|-----------|
| NM_001381 | Ercc2     | NM_013596 | Kif13a    | NM_01003450 | Wnda     |
| NM_019974 | Aprh1     | NM_01042805 | Cd74    | NM_007247 | Nprl1   |
| NM_008225 | Hcls1     | NM_011955 | Ifil2q    | NM_007609 | Nmda    |
| NM_010658 | MafB      | NM_032633 | Tgfbr3    | NM_022018 | Fam129a |
| NM_011691 | Vav2      | NM_010268 | Prkcs     | NM_019971 | Pdcdg   |
| NM_010096 | Pck1c     | NM_010288 | Gla1      | NM_009592 | Ccc6ex88 |
| NM_008590 | Ccdd       | NM_007699 | Lpl        | NM_010186 | Lgr1    |
| NM_010186 | Lgr1      | NM_010142 | Lpl        | NM_011691 | Vav2    |
| NM_019971 | Pdcdg     | NM_010288 | Gla1      | NM_009592 | Ccc6ex88 |
| NM_008590 | Ccdd       | NM_007699 | Lpl        | NM_010186 | Lgr1    |
| NM_010186 | Lgr1      | NM_010142 | Lpl        | NM_011691 | Vav2    |

Endopeptidase inhibitor activity

| Ref Seq ID | Gene Name | Ref Seq ID | Gene Name | Ref Seq ID | Gene Name |
|-----------|-----------|-----------|-----------|-----------|-----------|
| NM_010825 | Gm8000    | NM_009977 | Gata3     | NM_001366 | Hoxa1    |
| NM_01003450 | Wnda   | NM_010574 | Gata3     | NM_001366 | Hoxa1    |
| NM_010289 | Bcl2a1d   | NM_010728 | Bcl2a1d   | NM_010289 | Bcl2a1d   |
| NM_010713 | Bcl2a1d   | NM_010289 | Bcl2a1d   | NM_010713 | Bcl2a1d   |
| NM_010713 | Bcl2a1d   | NM_010289 | Bcl2a1d   | NM_010713 | Bcl2a1d   |

B Genes down-regulated in Dnm3L<sup>AA</sup> spermatogonia

| Ref Seq ID | Gene Name | Ref Seq ID | Gene Name | Ref Seq ID | Gene Name |
|-----------|-----------|-----------|-----------|-----------|-----------|
| NM_008091 | Gata3     | NM_017820 | sla3      | NM_008527 | Tgfbr3    |
| NM_016889 | Insm1     | NM_009862 | Ifil1q3   | NM_008882 | Cpr617367 |
| NM_008592 | Gsm1376    | NM_014956 | Ccdd6     | NM_17259  | Sap1     |
| NM_009503 | Bcl2a1d   | NM_021459 | Sft       | NM_008495 | Lats2    |
| NM_008997 | Scul2     | NM_008760 | Eya5h15a  | NM_008471 | Knt19    |
| NM_014657 | Gata2     | NM_008899 | Sou32q2   | NM_017638 | Cbr3     |
| NM_008295 | Lcol10045166 | NM_001197 | Six1      | NM_008287 | Hhex      |
| NM_001023 | Fzd5       | NM_001044 | Hoxa1     | NM_003327 | Zfp243   |
| NM_001273 | Tll13      | NM_002478 | Wnta3     | NM_054033 | Prkrs     |
| NM_00117152 | Dccn    | NM_007865 | Dll1      | NM_007131 | Lhx8    |

Cell fate commitment

| Ref Seq ID | Gene Name | Ref Seq ID | Gene Name | Ref Seq ID | Gene Name |
|-----------|-----------|-----------|-----------|-----------|-----------|
| NM_008259 | Foxa1     | NM_008259 | Lcol10047556 | NM_008891 | Pak11q3 |
| NM_008091 | Gata3     | NM_001054 | Dll2     | NM_021459 | Sft       |
| NM_001097 | Dkk1      | NM_001027 | Ser1     | NM_008935 | Foxn4    |
| NM_001053 | Dkk1      | NM_007865 | Dll1      | NM_017013 | Lhx8    |

Cellular developmental process

| Ref Seq ID | Gene Name | Ref Seq ID | Gene Name | Ref Seq ID | Gene Name |
|-----------|-----------|-----------|-----------|-----------|-----------|
| NM_008091 | Gata3     | NM_010254 | Dlx2      | NM_009849 | Lhx5     |
| NM_00159407 | Dlg6     | NM_008849 | Hoxa1     | NM_001057 | Mx3      |
| NM_178068 | Gata3     | NM_010728 | Bcl2a1d   | NM_01003450 | Wnda   |
| NM_00159407 | Dlg6     | NM_008849 | Hoxa1     | NM_001057 | Mx3      |
| NM_0159407 | Dlg6     | NM_008849 | Hoxa1     | NM_001057 | Mx3      |

Multicellular organismal development

| Ref Seq ID | Gene Name | Ref Seq ID | Gene Name | Ref Seq ID | Gene Name |
|-----------|-----------|-----------|-----------|-----------|-----------|
| NM_008091 | Gata3     | NM_006529 | Tgfb1    | NM_006544 | Hoxa1     |
| NM_001053 | Gata3     | NM_006529 | Tgfb1    | NM_006544 | Hoxa1     |
| NM_001053 | Gata3     | NM_006529 | Tgfb1    | NM_006544 | Hoxa1     |

Transcription regulator activity

| Ref Seq ID | Gene Name | Ref Seq ID | Gene Name | Ref Seq ID | Gene Name |
|-----------|-----------|-----------|-----------|-----------|-----------|
| NM_010254 | Dlx2      | NM_008849 | Hoxa1     | NM_001057 | Mx3      |
| NM_008849 | Hoxa1     | NM_001057 | Mx3      | NM_0159407 | Dlg6     |
| NM_178068 | Gata3     | NM_00159407 | Dlg6    | NM_178068 | Gata3     |
| NM_00159407 | Dlg6     | NM_008849 | Hoxa1     | NM_001057 | Mx3      |

Regulation of apoptosis

| Ref Seq ID | Gene Name | Ref Seq ID | Gene Name | Ref Seq ID | Gene Name |
|-----------|-----------|-----------|-----------|-----------|-----------|
| NM_007523 | Bak1      | NM_013598 | Kit       | NM_007523 | Bak1      |
| NM_007523 | Bak1      | NM_013598 | Kit       | NM_007523 | Bak1      |
| NM_007523 | Bak1      | NM_013598 | Kit       | NM_007523 | Bak1      |

Homebox, conserved site

| Ref Seq ID | Gene Name | Ref Seq ID | Gene Name | Ref Seq ID | Gene Name |
|-----------|-----------|-----------|-----------|-----------|-----------|
| NM_006544 | Hoxa1     | NM_001053 | Hox1      | NM_006544 | Hoxa1     |
| NM_0159407 | Dlg6     | NM_001053 | Hox1      | NM_006544 | Hoxa1     |
| NM_008849 | Hoxa1     | NM_001053 | Hox1      | NM_006544 | Hoxa1     |
Figure S4

Dnmt3L
+/+
Dnmt1
+/+
Dnmt3L
+/+
Ezh2

Dnmt3L
-/-
Dnmt3L
-/-
Dnmt3L
-/-
Ezh2

Input
IP: α-Dnmt3L

Dnmt3L

Ezh2
Figure S5

Panel A: Image showing testis tissue.

Panel B: Bar graph showing testes mass (mg) compared between Dnmt3LA/+ and Dnmt3LA/- genotypes. The bars are labeled with **** indicating statistical significance.

Panel C: Bar graph showing sperm number (x10^6) compared between Dnmt3LA/+ and Dnmt3LA/- genotypes. The bars are labeled with *** indicating statistical significance.
Figure S6

A

|          | Testes | $\text{Dnmt3L}^{+/+}$ | $\text{Dnmt3L}^{+/+}$ |
|----------|--------|------------------------|------------------------|
| Dnmt3L   | MEF    |                        |                        |
| Dnmt3L   |        |                        |                        |
| Dnmt3L   |        |                        |                        |
| Thy1-    |        |                        |                        |
| Thy1+    |        |                        |                        |
| Thy1-    |        |                        |                        |
| Thy1+    |        |                        |                        |

Dnmt3L

Tubulin

B

|          | 1 dpp Testes | Dnmt3L $^{+/+}$ PFC | Dnmt3L $^{+/+}$ OFB | Dnmt3L $^{+/+}$ PFC | Dnmt3L $^{+/+}$ OFB |
|----------|--------------|---------------------|--------------------|---------------------|--------------------|
|          | MEF          |                     |                    |                     |                    |
|          |              |                     |                    |                     |                    |

Short Exposure

Dnmt3a

Dnmt3a2

Long Exposure

Dnmt3a

Dnmt3a2

Short Exposure

Dnmt3L

Long Exposure

Dnmt3L

Tubulin
Supplemental Figure Legends

Figure S1 (related to Figure 1): Generation of $Dnmt3L^{A/A}$ animals and methylation analysis. A Targeting strategy used to introduce and detect the $D124A$ mutation at the endogenous $Dnmt3L$ locus. Identification of targeted clones involved digestion with BamHI (B) and SpeI (S) restriction enzymes. B Diagnostic Southern blot using the probe indicated in A showing the detection of the targeted allele. C Agarose gel electrophoresis of PCR products used to distinguish between wild type ($Dnmt3L^{+/+}$) and mutant ($Dnmt3L^{A/A}$) alleles. As well as introducing the A to C transversion resulting in the $D124A$ mutation, the targeting construct also introduced a silent A to C transversion resulting in the generation of an $Mscl$ restriction site. PCR across this site followed by restriction digestion of by $Mscl$ generates either an undigested fragment from the $Dnmt3L^{+}$ allele (337bp,) or two digested fragments from the $Dnmt3L^{A}$ allele (193bp and 144bp). D Sequencing traces of PCR amplicons across the D124A mutation site using genomic DNA extracted from tail biopsies. Note the presence of 2 peaks in the left panel indicative of heterozygosity for the $D124A$ mutation. The right panel indicates the animal is homozygous for the $D124A$ mutation. E Bar graphs indicating percentage methylation at CG and CH contexts in bisulphite-converted amplicons. Results indicate reduced CG and non-CG methylation at IAP and L1Tf retrotransposons. F Percentage CG and CH methylation in different compartments indicated. Data indicate that the $D124A$ mutation causes reduced methylation in all compartments analysed. G Methylation-sensitive Southern Blot data on DNA purified from 1dpp testes. As controls for un- and hypomethylated DNA, DNA extracted from $Dnmt1^{+/-}$ and $Dnmt1^{+/-}Dnmt3a^{+/-}Dnmt3b^{+/-}$ (TKO) ES cells was also
Dnmt1<sup>−/−</sup> ES cells are deficient for the maintenance methyltransferase Dnmt1; TKO cells are deficient for all known methyltransferases. Tail-biopsy DNA (somatic) from wild-type animals was used as a normally methylated control. Note both the presence and intensity of the lower molecular weight signal in the lanes containing DNA from 1dpp Dnmt3L<sup>+/+</sup> testes DNA. Bisulphite sequencing data from FACS-purified spermatogonia from 1dpp testes, different regions are as indicated. Note the presence of unmethylated, partially methylated and fully methylated sequences in DNA from Dnmt3L<sup>−/−</sup> animals. IAP methylation percentages are 81% (Dnmt3L<sup>+/+</sup>) and 53% (Dnmt3L<sup>−/−</sup>). LINE1-Tf methylation percentages are 88% (Dnmt3L<sup>+/+</sup>) and 66% (Dnmt3L<sup>−/−</sup>). Western blot data examining expression of Dnmt3a, Dnmt3b and Dnmt3L in RIPA extracts from 1dpp testes and other cells as indicated. e16.5 material indicates testes. Levels of tubulin and Cdc2 expression were used as loading controls. The data indicates that the D124A mutation does not affect expression of Dnmt3 family members.

**Figure S2 (related to Figure 3): Characterisation of Dnmt3L<sup>−/−</sup> tubules displaying normal and abnormal development.** A Example of a normal tubule with the various spermatogenic stages present. B Stage XI tubules, however the tubule on the right lacks diplotene spermatocytes (area marked by line). C Tubule with cells blocked at two stages; region I indicates germ cells that developed no further than the zygotene stage (stage IV). On the opposite side of the same tubule in region II, development is blocked at the round spermatid stage. D Partially normal tubule that has all developmental
stages (region I). However in region II, despite the presence of elongating spermatids, round spermatids are absent. **E** Tubule containing mainly young pachytene spermatocytes. Despite the presence of elongating spermatids, no round spermatids are present. **F** Stage IV tubule where majority of spermatocytes are undergoing apoptosis, although pachytene spermatocytes are also present. **G** Top panel indicates gates used to distinguish different spermatogenic stages by flow cytometry based on forward scatter height (FSC-H) and side scatter height (SSC-H) as described by Malkov et al. (Malkov et al., 1998). These were then used to position gates based on GFP expression. Briefly, testicular cells are subdivided into 7 populations; R1 and R2 represent cells at G2 of mitosis as well as primary spermatocytes undergoing prophase I. R3 and R4 represent diploid cells, which includes somatic (Sertoli, Leydig and peritubular cells) as well as germ cells (spermatogonia at mitotic G1, pre-leptotene spermatocytes prior to premeiotic S phase and secondary spermatocytes between meiosis I and II). R5 represents elongated and condensed spermatozoa, and R6 and R7 represent early round, elongating and condensing spermatids. **H** Representative FACS-density plots of testicular germ cells. Panel I shows a representative distribution of spermatogenic cells in wild-type animals. Panel II and III are both plots of cells from Dnmt3L^A/A^ animals demonstrating either a normal wild type like distribution (Panel II) or severe reductions (Panel III). **I** Low (5x) and high (40x) magnification micrographs of 1dpp testes from Oct4-GiP^+^ and Dnmt3L^A/A^ Oct4-GiP^+^ animals stained with anti-GFP antibody. Nuclei were counterstained with DAPI. Bar graphs represent average number of GFP^+^ prospermatogonia per seminiferous tubule (Dnmt3L^+/+^ = 5.09, Dnmt3L^A/A^ =
5.45). 100 tubules counted per genotype. No significant difference was observed (exact p value: 0.8240). The data indicates that the number of seeding prospermatogonia in seminiferous tubules is unaffected in $Dnmt3L^{A/A}$ animals.

**Figure S3 (related to Figure 4):** List of genes deregulated in different GO classes in $Dnmt3L^{A/A}$ spermatogonia. **A** List of genes upregulated. **B** List of genes downregulated.

**Figure S4 (related to Discussion):** Absence of detectable Ezh2 in complex with Dnmt3L in ES cells. Western blot of whole cell lysates before and after immunoprecipitation with anti-Dnmt3L antibody, probed with either anti-Dnmt3L or Ezh2 antibodies. 7aabb indicates $Dnmt3A^{-/-}Dnmt3B^{-/-}$ ES cells.

**Figure S5 (related to Discussion):** Complete penetrance of sterility phenotype caused by hemizygosity of $D124A$ mutation. **A** 10x micrograph image of H&E stained section of paraformaldehyde-fixed and OCT-embedded testes from a $Dnmt3L^{A/-}$ male animal (Age, 11 weeks). Image shown is representative of eight testes sections analyses from eight mutant animals. Note that all seminiferous tubules are defective, analogous to $Dnmt3L$-null testes. **B** Bar graph showing testes mass and sperm counts from $Dnmt3L^{A/+}$ and $Dnmt3L^{A/-}$ (n = 8, Age: 8-12 weeks) animals. Graphs show mean +/- S.D. Exact p values: p<0.0001 (testes mass); p = 0.001 (sperm count). Note that all $Dnmt3L^{A/-}$ animals were azoospermic.
Figure S6 (related to Discussion): Lack of detectable Dnmt3L protein expression in SSC and juvenile prefrontal cortex. A Western blot data showing absence of detectable Dnmt3L protein expression in RIPA extracts from FACS-purified Thy1+ and Thy1− SSCs purified from 9dpp Dnmt3L<sup>A/A</sup> animals. Note abundant expression of Dnmt3L protein in 1dpp whole testes. Expression of tubulin was used as a loading control. B Western blot data showing absence of detectable Dnmt3L protein expression in RIPA extracts from prefrontal cortex (PFC) and olfactory bulb (OB) extracted from wild type (46dpp) and Dnmt3L<sup>A/A</sup> (33dpp) animals. Note abundant expression of Dnmt3L protein in 1dpp whole testes and ES cells. Expression of tubulin was used as a loading control.
Supplemental Experimental Procedures

Methylation-sensitive Southern blot

Kit purified genomic DNA was digested using HpaII restriction endonuclease for six hours before separation by gel electrophoresis. Following resolution, DNA was depurinated, denatured and neutralized before overnight transfer onto Nytran membrane. The next day, DNA was cross-linked by UV light before incubating in Prehybridisation Solution for a minimum of two hours at 65°C. Radio labeled probes were added to membranes and incubated over night at 65°C. The next day, membranes were washed to remove unbound probe before detection of radioactive signal using a PhosphorImager screen.

Bisulphite sequencing

Kit-purified DNA from FACS-purified prospermatogonia extracted from 1dpp testes was bisulphite converted using the EZ DNA Methylation Lightning Kit (Zymo Research) according to the manufacturers protocol. Post conversion, DNA was used for PCR. Amplicons were resolved by gel electrophoresis and DNA was purified using the GeneJET Gel Extraction Kit (Thermofisher) prior to cloning using the CloneJET PCR Cloning Kit (Thermofisher) according to the manufacturers protocol before transformation of cloned products into heat-shock competent bacteria. The next day, colony PCRs were performed (25µl volume) and clones containing amplicons of the expected sizes were identified by gel electrophoresis. Exo-SAP was added to the remaining
volume of positive clones before subjecting to Sanger sequencing. Analysis of bisulphite data was performed using QUMA software (http://quma.cdb.riken.jp).

Antibodies

For western blot, the following antibodies were used: anti-Dnmt3a rabbit polyclonal (SC-20703, Santa Cruz Biotechnologies), anti-Dnmt3b mouse monoclonal (SC-52922, Santa Cruz Biotechnologies), anti-Cdc2 (PSTAIRE) rabbit polyclonal (SC053, Santa Cruz Biotechnologies), anti-Dnmt3L rabbit polyclonal (12309, Cell Signaling Technology), anti-Tubulin mouse monoclonal (T6199, Sigma-Aldrich), anti-Ezh2 mouse monoclonal (3147, Cell Signaling Technology). For purification of Thy1⁺ SSCs, APC-conjugated anti-CD90.2 (Thy1.2) (53-2.1, Biolegend) was used.

ChIP-Seq Metagene Analysis

ChIP-Seq for Suz12 and H3K27me3 was performed in E14 ES cells with the antibodies ab12073 and ab6002 (Abcam), respectively, using methods reported previously (Kanhere et al., 2012). ChIP-Seq data for Ezh2 was downloaded from GEO (GSM327668, (Ku et al., 2008). Sequence reads were aligned using Bowtie (default settings). Wig files were generated by calculating tag density in 10bp windows and normalized to reads per million total reads using in-house R scripts. The data for ChIP sequencing runs were then background corrected (input for Suz12 ChIP, H3 for H3K27me3 ChIP)
and average-binding profiles across TSSs (for all genes, up-regulated genes and down-regulated genes in \textit{Dnmt3L}^{A/A} SSCs) plotted using a custom R script. PRC2 target genes were selected as those bound by both Suz12 and Ezh2 in mESC according to the data of Ku et al., (Ku et al., 2008). Those associated with H3K27me3 in mESC as defined by both Mikkelsen et al. (Mikkelsen et al., 2008) and Lienert et al. (Lienert et al., 2011).

Analysis of repeat proximity to differentially expression genes

The TSS locations of all genes and repeats were extracted and downloaded from the mm9 build of the mouse reference genome from the UCSC genome browser (http://genome.ucsc.edu). Repeats were filtered into 3 groups (LINE, LTR and SINE), using the repeat classifications provided by RepeatMasker (http://www.repeatmasker.org). For each TSS, the distance to the closest repeat was calculated using bedtools (closestBed option, bedtools v2.18.1) (http://bedtools.readthedocs.org/en/latest/). This algorithm calculates the distance from TSS to the ends of the repeats and reports on the closest distance. Using R, box plots were plotted of distances to closest repeats for genes that were either up- or down-regulated in \textit{Dnmt3L}^{A/A} SSCs. Closest distances for all genes were also plotted as a control. Significance was calculated using the Wilcoxon Rank Sum test, comparing average repeat distances for up or down-regulated genes to average repeat distance for all genes.
MethylC-seq library preparation and sequencing

200 ng of genomic DNA (gDNA) along with 5 ng of spiked in lambda DNA was sonicated in 130 µl nuclease free water to 200 bp fragments using a Covaris S-2 sonicator (Covaris Inc.). Large fragment removal was accomplished by mixing magnetic purification beads (MPB - see below) with sonicated gDNA at a volume ratio of 0.6X and incubation for 10 minutes at room temperature. After 10 minutes samples were moved to a magnetic plate to separate beads from solution. Supernatant was removed from beads and added to a volume of MPB at a 0.8 ratio compared to original gDNA solution volume to create a total MPB to gDNA volume ratio of 1.4; tubes containing beads and large fragment DNA were discarded. Samples were then cleaned per described purification methods (See below).

Fragmented gDNA from sonication was repaired using End-It DNA End-Repair Kit (Epicentre) according to manufacturer's instructions. A-tails were added to blunt ended fragments during a 30 minute 37°C incubation using Klenow 3’-5’ exonuclease and dA-Tailing Buffer (New England Biolabs). Methylated NEXTflex DNA adapters (Bioo Scientific) were ligated via 16 hour, 16°C incubation with T4 DNA ligase and ligation buffer (New England Biolabs). Post ligation, samples underwent two clean up procedures and then were subject to bisulfite conversion using a MethylCode kit, as per manufacturer's instructions. Libraries were amplified with eight cycles of PCR using Kapa HiFi Uracil+ Hotstart DNA Polymerase (Kapa Biosystems).

MPB were mixed with samples at a volume ratio of 1.4:1 prior to adapter
ligation and 1.0:1.0 post adapter ligation. To wash, MPB solution was mixed with samples by pipetting and left at room temperature for 10 minutes. After, samples were moved to magnets to isolate magnetic beads from solution. Supernatant was removed and beads were washed twice with 80% ethanol. Samples were removed from magnets and DNA was eluted of beads into 10mM Tris-HCl pH 8.0. and left at room temperature for 10 minutes. Sequencing was performed on an Illumina NextSeq 500.

Alignment and analysis of MethylC-seq data

Raw FASTQ files were trimmed for adapter a sequence, preprocessed to remove low quality reads and aligned to the mm10 reference genome as previously described in (Lister et al., 2011). Inefficiencies in the sodium bisulfite conversion reaction are calculated by measuring the fraction of methylated base calls detected in the spiked-in Lambda control DNA (which is unmethylated). This non-conversion rate is used as the null hypothesis using a binomial test to determine if a cytosine is methylated and the resulting p-values are corrected for multiple testing using Benjamini-Hochberg with an FDR cut off of 5%. For Dnmt3L^{+/+} and Dnmt3L^{A/A}, a total of 408,932,744 (~15x genome coverage) and 299,546,426 (~11x genome coverage) reads were uniquely and non-clonally aligned to the reference genome, respectively. The sodium bisulphite conversion rates for the Dnmt3L^{+/+} and Dnmt3L^{A/A} MethylC-seq libraries were 99.56 % and 99.52%, respectively. Data for somatic mouse tissues were obtained from the following study (Hon et al., 2013).
Calculation of methylation levels

All methylation levels that were calculated from WGBS data were determined using the weighted methylation levels described in (Schultz et al., 2012)

Differential gene expression analysis in prospermatogonia

RNA-seq data from e16.5 prospermatogonia was previously described by Seisenberger et al. (Seisenberger et al., 2012) and obtained from the European Nucleotide Archive (ERP001953). To identify PSG specific transcripts, RNA-seq data from mouse lung fibroblasts (GEO Accession number GSM521651) was used for comparison. A list of highly and lowly expressed genes was obtained using the Genestack Platform (http://platform.genestack.org). A binomial test was used to determine if genes were differentially expressed between e16.5 prospermatogonia and mouse embryonic lung fibroblast and the resulting p-values were corrected for multiple testing using the Benjamini and Hochberg correction with an FDR cut off of 5%. All raw data, processed files and final outputs are available on Genestack Platform and can be shared upon request for viewing or reproducing the analyses reported.

Flow cytometric analysis of GFP+ germ cells

Single cell suspensions of mouse testicular cells enriched for seminiferous epithelium cells (principally germ and Sertoli cells) were isolated from age-
matched wild-type and Dnmt3L\textsuperscript{A/A} animals carrying the Oct4-GiP reporter using the method described by Garcia and Hofmann (Garcia and Hofmann, 2012). Live cells were identified using the Annexin V: PE Apoptosis detection kit (BD Biosciences), according to the manufacturers protocol. Cells were analysed on a BD Fortessa Flow Cytometer (BD Biosciences) and analysis performed using FlowJo Cytometric Analytical software (FlowJo).

GFP staining of 1dpp testes sections

1dpp/newborn animals were sacrificed and testes were harvested from male animals and fixed and frozen in OCT as previously described. Sections mounted on slides were fixed for 10 minutes at room temperature with 4% PFA, followed by two washes in PBS + 0.1% Triton X-100 (PBT). Slides were then blocked in Blocking Solution (PBT + 1% BSA + 0.15% Glycine) for one hour at room temperature before addition of anti-GFP antibody in Blocking Solution and incubating overnight at 4\(^{\circ}\)C. The next day, slides were washed with PBT several times before incubation with secondary antibody (details) and incubated in the dark at room temperature for one hour. Slides were then washed several times with PBT before mounting. Analysis and image capture was performed on a Zeiss Axio Observer Z1 fluorescent microscope.
Supplemental References

Hon, G.C., Rajagopal, N., Shen, Y., McCleary, D.F., Yue, F., Dang, M.D., and Ren, B. (2013). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. Nature genetics 45, 1198-1206.

Kanhere, A., Hertweck, A., Bhatia, U., Gokmen, M.R., Perucha, E., Jackson, I., Lord, G.M., and Jenner, R.G. (2012). T-bet and GATA3 orchestrate Th1 and Th2 differentiation through lineage-specific targeting of distal regulatory elements. Nature communications 3, 1268.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS genetics 4, e1000242.

Lienert, F., Mohn, F., Tiwari, V.K., Baubec, T., Roloff, T.C., Gaidatzis, D., Stadler, M.B., and Schubeler, D. (2011). Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells. PLoS genetics 7, e1002090.

Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S., et al. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature 471, 68-73.

Malkov, M., Fisher, Y., and Don, J. (1998). Developmental schedule of the postnatal rat testis determined by flow cytometry. Biology of reproduction 59, 84-92.

Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature 454, 49-55.

Schultz, M.D., Schmitz, R.J., and Ecker, J.R. (2012). 'Leveling' the playing field for analyses of single-base resolution DNA methylomes. Trends in genetics : TIG 28, 583-585.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Molecular cell 48, 849-862.