LETTERS

Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency

Christian Popp[1]*, Wendy Dean[1]*, Suhua Feng[2]*, Shawn J. Cokus[2], Simon Andrews[3], Matteo Pellegrini[2], Steven E. Jacobsen[2,4] & Wolf Reik[1]

Epigenetic reprogramming including demethylation of DNA occurs in mammalian primordial germ cells (PGCs) and in early embryos, and is important for the erasure of imprints and epimutations, and the return to pluripotency[1–9]. The extent of this reprogramming and its molecular mechanisms are poorly understood. We previously showed that the cytidine deaminases AID and APOBEC1 can deaminate 5-methylcytosine *in vitro* and in *Escherichia coli*, and in the mouse are expressed in tissues in which demethylation occurs[10]. Here we profiled DNA methylation throughout the genome by unbiased bisulphite next generation sequencing[11–13] in wild-type and AID-deficient mouse PGCs at embryonic day (E)13.5. Wild-type PGCs revealed marked genome-wide erasure of methylation to a level below that of methylation deficient (*Np95*$^{-/-}$, also called *Uhrf1*$^{-/-}$) embryonic stem cells, with female PGCs being less methylated than male ones. By contrast, AID-deficient PGCs were up to three times more methylated than wild-type ones; this substantial difference occurred throughout the genome, with introns, intergenic regions and transposons being relatively more methylated than exons. Relative hypermethylation in AID-deficient PGCs was confirmed by analysis of individual loci in the genome. Our results reveal that erasure of DNA methylation in the germ line is a global process, hence limiting the potential for transgenerational epigenetic inheritance. AID deficiency interferes with genome-wide erasure of DNA methylation patterns, indicating that AID has a critical function in epigenetic reprogramming and potentially in restricting the inheritance of epimutations in mammals.

In plants, active demethylation occurs widely (including in imprinted genes) and is carried out by 5-methylcytosine (5meC) glycosylases such as Demeter and Demeter-like proteins[14–16]. This class of enzymes does not seem to exist in mammalian genomes, but instead we found that the cytosine deaminases AID and APOBEC1 can deaminate 5meC both *in vitro* and in *E. coli*[10], indicating that deamination of 5meC followed by T:G base excision repair by glycosylases such as TDG or MBD4 might be an equivalent pathway for demethylation of DNA. Recently it has been shown that co-expression of AID and MBD4 in zebrafish embryos can lead to demethylation of DNA, lending support to the idea that demethylation in animals might proceed by deamination followed by base excision repair[17]. These results raise the question of whether naturally occurring demethylation in PGCs or early embryos involves deaminases or base excision repair.

We previously found that AID is expressed in PGCs and in early embryos at a time when demethylation occurs[10]. Methylation in some single-copy genes, differentially methylated regions in imprinted genes, and long interspersed nuclear element (LINE) 1 repeats has been shown to be substantially erased in PGCs between E11.5 and E13.5, whereas methylation of intracisternal A particles (IAPs) was significantly more resistant to erasure[6–9,18]. The extent of demethylation of the genome as a whole in PGCs is unknown. To understand the dynamics of erasure on a genome-wide scale we carried out unbiased sequencing of bisulphite-treated DNA by next generation sequencing (bisulphite next generation sequencing; BS-Seq), which accurately quantifies whole-genome methylation levels[11,13]. We were able to scale down by 20-fold the construction and analysis of Illumina Solexa libraries to as little as 150 ng of genomic DNA, enabling for the first time a genome-wide view of methylation in fully reprogrammed PGCs at E13.5, isolated by cell sorting of *Oct4-Gfp*-positive germ cells (Fig. 1). Both male and female PGCs were strikingly hypomethylated, with a median methylation level of 16.3% and 7.8%, respectively (for comparison, methylation-deficient *Np95*$^{-/-}$ embryonic stem (ES) cells had a methylation level of 22%). It is interesting to note that female PGCs had a lower methylation level than male PGCs; this is consistent with female ES cells having less methylation than male ones[19], indicating that the proposed effects of two active X chromosomes on decreasing genome methylation might operate *in vivo* as well as in pluripotent cell lines derived from inner cell mass (ICM) cells or PGCs. Methylation levels in fetus, ES cells and sperm were high (73.2–85%), whereas those in placenta were intermediate (42.3%), consistent with results of previous experiments in which individual gene sequences or transposons were analysed. Notably, genome-wide hypomethylation has recently been reported in *Arabidopsis* endosperm in comparison to the embryo (although the difference, 6%, is considerably smaller than the one that we found here in the mammalian system), a tissue that supports embryo development and nutrition in seed plants in analogy to the function of the placenta in mammals[15,16].

We next carried out a more detailed analysis of the patterns of methylation erasure in PGCs in various genomic elements (Figs 2 and 3). We assessed levels of methylation of promoters, exons, introns and intergenic regions (Fig. 2), as well as those of various transposon families (Fig. 3 and Supplementary Fig. 1; BS-Seq reads of transposons and repeats with identical sequence are eliminated bioinformatically, but BS-Seq reads of transposon family members that map uniquely, owing to unique sequence variation, can be assessed). In highly methylated tissues (ES, sperm, fetus), methylation was similarly high in exons (above 70%), and was considerably lower in promoters (35–40%), consistent with a recent whole-genome BS-Seq analysis in human ES cells[13]. Introns and retrotransposons were particularly highly methylated in sperm (approximately 10% higher than in the fetus or ES cells), supporting the notion that *de novo* methylation in male germ cells, before resumption of mitosis after birth, is important for suppressing transposon mobility in the germ line. In comparison

[1]Laboratory of Developmental Genetics and Imprinting, The Babraham Institute, Cambridge CB22 3AT, and Centre for Trophoblast Research, University of Cambridge, Cambridge CB2 3EG, UK. [2]Department of Molecular Cell and Developmental Biology, University of California at Los Angeles, Los Angeles, California 90095, USA. [3]Bioinformatics Group, The Babraham Institute, Cambridge CB22 3AT, UK. [4]Howard Hughes Medical Institute, University of California at Los Angeles, Los Angeles, California 90095, USA.
*These authors contributed equally to this work.

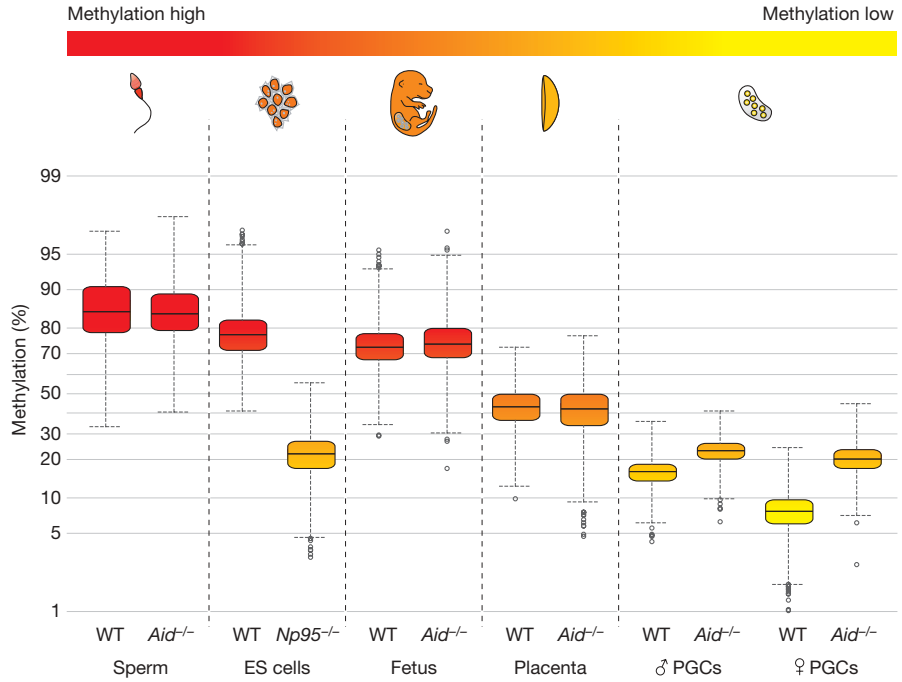

Figure 1 | Genome-wide BS-Seq reveals global hypomethylation in PGCs dependent on AID. Tissues and cells analysed by BS-Seq are shown in a gradient from red to yellow illustrating methylation levels from high to low, respectively. BS-Seq reads were analysed using windows of 250 kb across the whole genome. Boxes show the range of the 25–75th quartiles of the data; the line in the middle the median value. Whiskers show either highest and lowest values (if there are no outliers) or upper and lower confidence intervals. Outliers are shown as circles. Placenta, fetal carcass and PGCs were collected at E13.5. Note the global hypomethylation in PGCs, and the substantially higher levels of methylation in *Aid*$^{-/-}$ PGCs. WT, wild type.

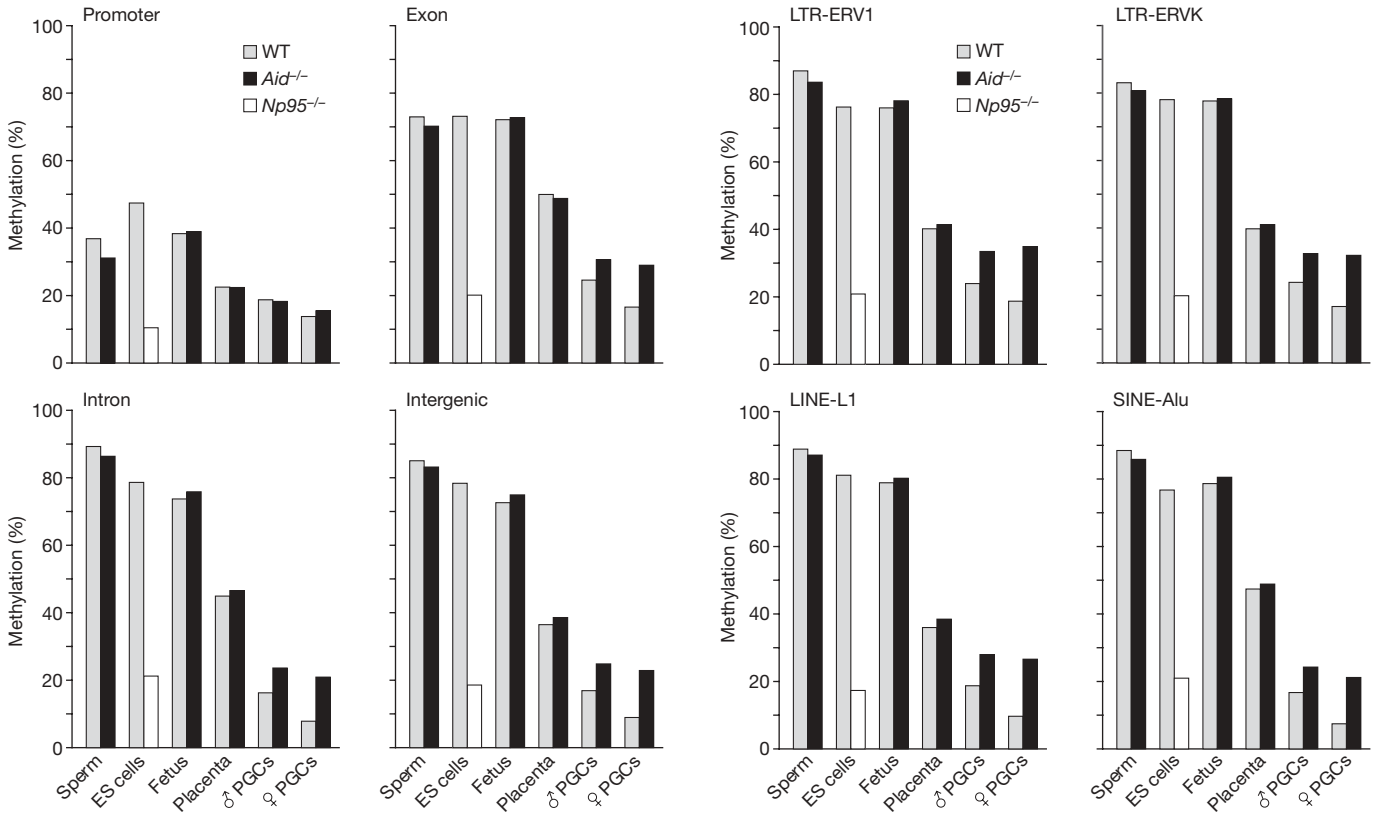

Figure 2 | Erasure of DNA methylation in different genomic elements in PGCs. Methylation levels in promoters, exons, introns and intergenic regions in ES cells, *Np95*$^{-/-}$ ES cells and various tissues of C57BL/6J and *Aid*$^{-/-}$ knockout mice are shown based on ratios of methylated to unmethylated BS-Seq reads. Placenta, fetal carcass and PGCs were all collected at E13.5. Note the particularly pronounced effect of AID deficiency on methylation of introns and intergenic regions.

Figure 3 | Erasure of DNA methylation in different classes of transposable elements in PGCs. Methylation levels of different classes of transposons in ES cells, *Np95*$^{-/-}$ ES cells, and various tissues of C57BL/6J and *Aid*$^{-/-}$ knockout mice are shown based on ratios of methylated to unmethylated BS-Seq reads. Placenta, fetal carcass and PGCs were all collected at E13.5. Note that LTR-ERV1 and LTR-ERVK elements retain more methylation in PGCs than any other repeat family.

to the highly methylated tissues, the greatest loss of methylation in PGCs (and in placenta) was observed within introns, intergenic regions and repeats, followed by exons, and then promoters. Hence, demethylation in PGCs is indeed global and encompasses genic, intergenic and transposon sequences. Sequences that retained the highest levels of methylation in the face of reprogramming were long terminal repeat (LTR)-ERV1 and LTR-ERVK elements (approximately 10% of the latter are IAPs). PGCs at the endpoint of reprogramming have therefore attained a unique epigenetic state, with genome-wide demethylation of DNA, and loss of the repressive histone marks H3K9me2 and H2A/H4R3me2 together with H2AZ, as well as loss of the active histone mark H3K9ac[9,20]. Thus, an epigenetic ground state that is depleted both of key activating and repressive epigenetic marks seems to be uniquely characteristic of reprogrammed PGCs. We carried out a biological replicate of the BS-Seq experiment on E13.5 PGCs isolated by a different method (see Supplementary Methods), which indeed replicated qualitatively all conclusions with the *Oct4-Gfp* fluorescence-activated cell sorted (FACS) PGCs, whereas the baseline was shifted to higher levels of methylation, reflecting either a greater contamination with somatic gonadal cells or epigenetic heterogeneity within germ cells that are not *Oct4-Gfp* positive (Supplementary Fig. 2).

We next examined the effects of AID deficiency on erasure of methylation in PGCs. We introgressed the *Oct4-Gfp* transgene into $Aid^{-/-}$ (also called $Aicda^{-/-}$) knockout mice[21] (both on a C57BL/6J inbred genetic background). PGCs from $Aid^{-/-}$ knockout mice showed substantially higher levels of methylation than wild type ones; this difference was more pronounced in female than in male mice (Figs 1–3), in which the mean methylation level was more than 2.5-fold higher in the $Aid^{-/-}$ knockout than the wild type. AID-deficient PGCs were particularly more methylated than wild-type PGCs in introns and transposons (more than a 3-fold difference in LINE-L2 and in short interspersed nuclear element (SINE)-B4 elements; Supplementary Fig. 1), followed by exons, whereas no differences were found in promoters. We were unable to analyse PGCs by BS-Seq before genome-wide erasure (E10.5, 11.5) because the numbers of cells that can be obtained are not currently sufficient for this technique. Previous analyses of individual gene loci and transposons found that the methylation profile of PGCs before erasure is similar to that of embryonic somatic cells[6–9,18], whereas another study suggested that the overall levels are slightly lower[20]. Hence, whereas $Aid^{-/-}$ PGCs are hypermethylated in comparison to wild type ones, a significant level of demethylation also occurs in $Aid^{-/-}$ PGCs, the extent of which depends on the precise genomic levels before erasure. Notably, an effect of AID deficiency on genome-wide methylation levels was only found in primordial germ cells, and not in the fetus, the placenta, or in sperm (Figs 1–3). These observations were again qualitatively replicated in an independent set of BS-Seq experiments using an alternative method of isolation of $Aid^{-/-}$ PGCs (Supplementary Fig. 2). Indeed, the ratios of the percentages of methylation between wild-type and $Aid^{-/-}$ PGCs were very similar when comparing the data sets obtained with the two different methods of isolation (Figs 2 and 3 and Supplementary Fig. 2).

The depth of sequencing in our experiments does not currently allow the comparison of methylation levels of individual gene loci between the various tissues analysed. However, given the global differences observed between wild-type and $Aid^{-/-}$ PGCs we compared methylation levels in randomly selected genomic loci (R1 and R2 are located in intergenic regions, R3 in the first intron of *Foxo1*, and R4 in the seventh exon of *Xirp2*), promoters or differentially methylated regions of genes including imprinted ones that are known to become demethylated during PGC development (*Dazl*, *H19*, *Lit1* (also called *Kcnq1ot1*)) and retrotransposons by Sequenom bisulphite assay (Fig. 4 and Supplementary Fig. 3). Where methylation differences were observed it was always the $Aid^{-/-}$ PGCs that were more highly methylated, confirming in an independent assay that $Aid^{-/-}$ PGCs remained relatively hypermethylated. This assay also confirmed that

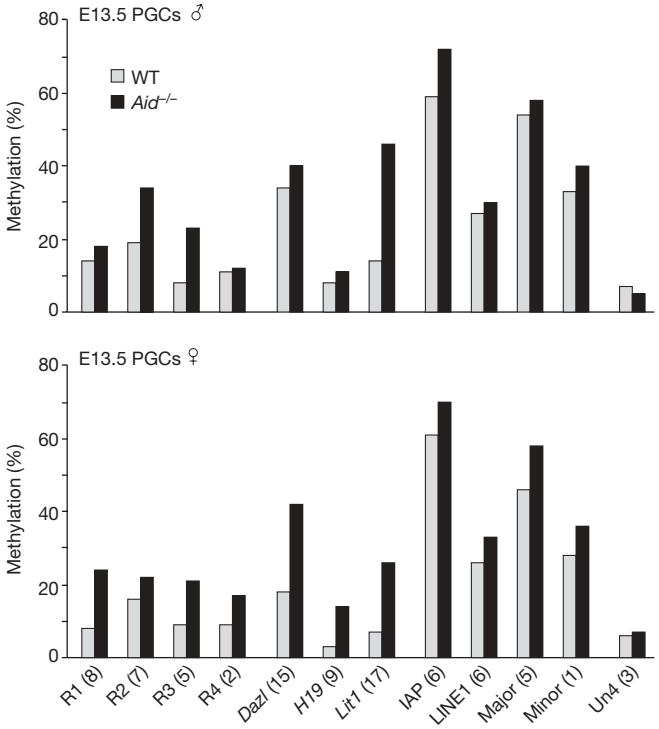

Figure 4 | Analysis of DNA methylation of individual genomic loci in E13.5 PGCs by Sequenom MassArray. Methylation levels of individual genomic loci (R1–4, randomly selected sequences, located in intergenic regions (R1, R2), the first intron of *Foxo1* (R3) and the seventh exon of *Xirp2* (R4); the *Dazl* amplicon is located in the promoter region and the *H19* and *Lit1* amplicons are located in the differentially methylated regions) and of transposons and major and minor satellite repeats in wild-type and $Aid^{-/-}$ E13.5 PGCs are shown based on Sequenom MassArray analysis. The number of CpGs analysed for each region is stated in brackets. Un4, unmethylated control located in the *Hoxc* cluster. Note substantially increased levels of methylation in many genomic loci in $Aid^{-/-}$ PGCs.

IAPs retained relatively high levels of methylation in PGCs (note that IAPs are a subfamily of LTR-ERVK elements and that the absolute values of methylation cannot therefore be compared between the Sequenom and BS-Seq assays). We also analysed methylation by Sequenom bisulphite assay in E12.5 PGCs, and again found relative hypermethylation of several loci in $Aid^{-/-}$ in comparison to wild-type PGCs (data not shown).

More detailed analyses at single CpG resolution of regions that were found to be hypermethylated in $Aid^{-/-}$ PGCs (Fig. 4) revealed that predominantly there was a homogeneous increase in methylation across the whole of the regions analysed (Supplementary Fig. 3).

Our results show that the vast majority of genomic DNA methylation is erased during normal development of PGCs. In mammals, this limits the scope for inheritance of epigenetic marks (based on methylation) across generations, which if it occurred would be of potential evolutionary significance as well as affecting disease risk in humans[22]. Indeed, in the mouse the only two well-documented examples of transgenerational inheritance involve alleles of genes with an insertion of an IAP element whose DNA methylation alters the expression of the neighbouring gene[22], consistent with our observation that LTR retrotransposons, including IAPs, are the genomic elements that are most resistant to erasure. This is in marked contrast to seed plants in which reprogramming of DNA methylation does not seem to occur in germ cells themselves[23], and in which stable inheritance of epialleles across generations is more common[24]. Our observation that erasure of DNA methylation is less efficient in $Aid^{-/-}$ PGCs (especially in the female germ line) indicates that the extent to which epigenetic information is heritable in mammals might be under regulation by genetic factors. In crosses between AID-deficient and wild-type parents, we observed

significant effects on growth of offspring at birth as well as on litter size. Notably, AID-deficient mothers did not regulate the size of their offspring in an inverse relationship to litter size, as wild-type mothers do (Supplementary Fig. 4). Second, litter sizes from both AID-deficient mothers and fathers in crosses with wild-type animals were significantly larger than in wild-type or homozygous $Aid^{-/-}$ crosses (Supplementary Fig. 4). Detailed epigenetic analyses need to be carried out on mature AID-deficient germ cells (particularly oocytes) and offspring to examine if heritable epimutations are indeed the basis for these significant reproductive phenotypes in $Aid^{-/-}$ knockout mice. It cannot be excluded that kinetics of methylation reprogramming are shifted in AID-deficient animals or that further reprogramming steps occur after E13.5 in $Aid^{-/-}$ knockout PGCs, which could further modify the differences observed here. This requires a substantially increased depth of sequencing on much reduced numbers of cells, which awaits further technical improvements and cost savings in BS-Seq technology.

Together with work just published in which it was shown that AID is required in ES cells to demethylate pluripotency genes during reprogramming of a somatic genome by cell fusion[25], our study shows that AID has a role in epigenetic reprogramming in mammals. Off-target effects of AID have recently been described in the immune system[26]; in this respect it is perhaps not surprising that AID has also evolved roles outside of the immune system. AID might exert its substantial effect on genome-wide demethylation of DNA in PGCs through its established function as a cytosine[27] or 5meC (refs 10, 28) deaminase. This can be tested genetically by examining the base excision repair pathways that are expected to be downstream of AID, including TDG and MBD4. Of particular interest is a recent report in zebrafish indicating that AID and MBD4 are involved co-operatively in demethylation of DNA[17]; this link might also involve GADD45, which has been implicated previously in demethylation of DNA. Notably, the effect of AID deficiency on methylation in PGCs was considerably more pronounced than that of deficiency of the 5meC glycosylase Demeter in the *Arabidopsis* endosperm[15,16]. Alternatively, AID might be required more indirectly as an essential component in the pathway that regulates erasure of methylation in PGCs. Other deaminases such as APOBEC1–3 might also have roles in demethylation of DNA, especially as our results show that AID deficiency does interfere with, but does not abrogate, erasure of methylation in PGCs. It is interesting to note that in plants Demeter is specifically required for erasure of methylation in imprinted genes, whereas Demeter-like genes are responsible for more general reprogramming of methylation patterns[14]. Other pathways to deamination of 5meC, such as one involving the *de novo* methyltransferases DNMT3A and DNMT3B, are less likely to operate in PGCs, as DNMT3A is not expressed and DNMT3B is excluded from the nucleus at the time of erasure of methylation[6]. Finally, our results do not exclude the existence of other pathways to demethylation of DNA in mammals, including oxidation of 5meC by the recently described TET family of 5meC hydroxylases[29,30], or removal of 5meC by glycosylases.

METHODS SUMMARY

Mouse tissues, including male and female PGCs at E13.5, were isolated from C57BL/6J mice, C57BL/6J mice transgenic for *Oct4-Gfp*, or $Aid^{-/-}$ knockout mice bred into a C57BL/6J background for seven generations during this study. The *Oct4-Gfp* transgene was subsequently bred into $Aid^{-/-}$ knockout mice. PGCs were isolated on a FACS-Aria cell sorter by sorting for green fluorescence; the isolated cell populations were >98% pure. DNA was isolated, bisulphite converted, and prepared for Illumina Solexa libraries. Each Illumina Solexa library was sequenced in a single end read, except for *Oct4-Gfp* isolated PGC libraries which were sequenced in two single end reads each; subsequently, a published highly customized software package was used to carry out Gaussian base-calling and sequence alignment for bisulphite-converted reads against the mouse genome[11]. On average, around 1.5 million aligned 27-bp reads (5.4 million 50-bp reads for the PGC libraries) were obtained for each library. For methylation analysis, bases 6 to 22 in the 27-bp reads (bases 15 to 41 in the 50-bp reads) were used, and CpGs were base called as methylated or unmethylated, respectively. Genome-wide averages of DNA methylation of individual samples, or averages of methylation in promoters, exons, introns, intergenic regions, and different classes of transposons, were bioinformatically determined. For Sequenom MassArray, bisulphite-converted DNA was amplified and subjected to quantitative analysis of methylation by mass spectrometry.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 November 2009; accepted 15 January 2010.
Published online 22 January 2010.

1. Reik, W., Dean, W. & Walter, J. Epigenetic reprogramming in mammalian development. *Science* **293,** 1089–1093 (2001).
2. Sasaki, H. & Matsui, Y. Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nature Rev. Genet.* **9,** 129–140 (2008).
3. Oswald, J. *et al.* Active demethylation of the paternal genome in the mouse zygote. *Curr. Biol.* **10,** 475–478 (2000).
4. Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Demethylation of the zygotic paternal genome. *Nature* **403,** 501–502 (2000).
5. Dean, W. *et al.* Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. *Proc. Natl Acad. Sci. USA* **98,** 13734–13738 (2001).
6. Hajkova, P. *et al.* Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* **117,** 15–23 (2002).
7. Lee, J. *et al.* Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. *Development* **129,** 1807–1817 (2002).
8. Yamazaki, Y. *et al.* Reprogramming of primordial germ cells begins before migration into the genital ridge, making these cells inadequate donors for reproductive cloning. *Proc. Natl Acad. Sci. USA* **100,** 12207–12212 (2003).
9. Hajkova, P. *et al.* Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452,** 877–881 (2008).
10. Morgan, H. D., Dean, W., Coker, H. A., Reik, W. & Petersen-Mahrt, S. K. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J. Biol. Chem.* **279,** 52353–52360 (2004).
11. Cokus, S. J. *et al.* Shotgun bisulfite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452,** 215–219 (2008).
12. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454,** 766–770 (2008).
13. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462,** 315–322 (2009).
14. Gehring, M., Reik, W. & Henikoff, S. DNA demethylation by DNA repair. *Trends Genet.* **25,** 82–90 (2009).
15. Gehring, M., Bubb, K. L. & Henikoff, S. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science* **324,** 1447–1451 (2009).
16. Hsieh, T. F. *et al.* Genome-wide demethylation of *Arabidopsis* endosperm. *Science* **324,** 1451–1454 (2009).
17. Rai, K. *et al.* DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase and gadd45. *Cell* **135,** 1201–1212 (2008).
18. Lane, N. *et al.* Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* **35,** 88–93 (2003).
19. Zvetkova, I. *et al.* Global hypomethylation of the genome in XX embryonic stem cells. *Nature Genet.* **37,** 1274–1279 (2005).
20. Seki, T. *et al.* Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development* **134,** 2627–2638 (2007).
21. Muramatsu, M. *et al.* Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potent RNA editing enzyme. *Cell* **102,** 553–563 (2000).
22. Whitelaw, N. C. & Whitelaw, E. Transgenerational epigenetic inheritance in health and disease. *Curr. Opin. Genet. Dev.* **18,** 273–279 (2008).
23. Slotkin, R. K. *et al.* Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136,** 461–472 (2009).
24. Teixeira, F. K. *et al.* A role for RNAi in the selective correction of DNA methylation defects. *Science* **323,** 1600–1604 (2009).
25. Bhutani, N. *et al.* Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* doi:10.1038/nature08752 (21 December 2009).
26. Robbiani, D. F. *et al.* Aid produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. *Mol. Cell* **36,** 631–641 (2009).
27. Neuberger, M. S., Harris, R. S., Di Noia, J. & Petersen-Mahrt, S. K. Immunity through DNA deamination. *Trends Biochem. Sci.* **28,** 305–312 (2003).
28. Larijani, M. *et al.* Methylation protects cytidines from Aid-mediated deamination. *Mol. Immunol.* **42,** 599–604 (2005).
29. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324,** 929–930 (2009).

30. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324,** 930–935 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Morgan for his contributions to some of the early analysis of *Aid*$^{-/-}$ mice, A. Segonds-Pichon for help with statistical evaluation, and J. Hetzel for assisting in preparing the Illumina Solexa libraries and their sequencing. We also thank S. Petersen-Mahrt, C. Rada and F. Santos for advice and discussions. C.P. was a Boehringer-Ingelheim predoctoral Fellow. S.F. is a Howard Hughes Medical Institute Fellow of the Life Science Research Foundation. S.E.J. is an investigator of the Howard Hughes Medical Institute. This work was supported by BBSRC, MRC, EU NoE The Epigenome, and CellCentric (to W.R.), and by HHMI, NSF Plant Genome Research Programme, and NIH (to S.E.J.).

Author Contributions C.P. and W.D. isolated tissue samples and PGCs, assessed the purity of the samples and prepared DNA. C.P. undertook genetic crosses, determined weights of mouse pups and carried out Sequenom EpiTYPER analysis. S.F. constructed bisulphite libraries and did Illumina Solexa sequencing. S.J.C., S.A. and M.P. carried out mapping, base-calling and computational analyses. C.P., W.D., S.F., S.J.C., S.A., M.P., S.E.J. and W.R. analysed data. C.P., W.D., S.F., S.E.J. and W.R. designed experiments; S.E.J. and W.R. designed and directed the study. C.P. and W.R. wrote the manuscript.

Author Information All sequencing files have been deposited in GEO under accession code GSE19960. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.E.J. (jacobsen@ucla.edu) or W.R. (wolf.reik@bbsrc.ac.uk).

METHODS

Mice and isolation of tissue and DNA samples. Mice deficient for AID have been described previously[21] and were provided by T. Honjo. These were backcrossed for seven generations into the C57BL/6J strain during the course of this study. C57BL/6J mice or C57BL/6J mice carrying an *Oct4-Gfp* transgene were used as controls throughout. The *Oct4-Gfp* transgene was bred into the $Aid^{-/-}$ knockout following backcrossing into the C57BL/6J background. Epididymal sperm was collected from fertile adult males. PGCs were isolated on a FACS-Aria cell sorter by sorting for green fluorescence; the isolated cell populations were >98% pure. Placentas and carcasses were taken from fetuses used for PGC collection. Genomic DNA from wild-type (E14) and $Np95^{-/-}$ mouse ES cells was provided by A. Clarke.

Shotgun bisulphite sequencing and computational analysis of data. Genomic DNA extracted from various mouse tissues with the Qiagen blood and tissue kit were treated with sodium bisulphite and then used to generate Illumina/Solexa sequencing libraries as described previously[11], except that fewer cycles of PCR amplification were used (15 cycles instead of 18 cycles) to optimize the base composition of the libraries. For PGC samples, owing to the limited sources of tissue, the input DNA amount for libraries had to be reduced to as low as 150 ng. Therefore, the enzymatic reaction steps used for library construction (including reagents and adapters for PCR) were scaled down to accommodate the reduced input amount. On the other hand, more DNA template (in volume) was used in each PCR reaction and more duplicate PCR reactions were performed in parallel to obtain equivalent amounts of product as for the other libraries. The libraries were sequenced on an Ultra-high-throughput Illumina/Solexa 1G Genome Analyser following manufacturer's instructions. Initial sequencing data analysis was performed using version 0.3 of the Illumina/Solexa Analysis Pipeline; subsequently, a previously published highly customized software package was used to carry out Gaussian base-calling and sequence alignment for bisulphite-converted reads against the mouse genome[11]. Around 5.4 million aligned 50-bp reads were obtained for PGC libraries isolated by *Oct4-Gfp* FACS sorting, whereas sequencing of all other libraries yielded on average 1.5 million aligned 27-bp reads. For methylation status analysis, bases 15 to 41 in the 50-bp reads and bases 6 to 22 in the 27-bp reads were used, equalling a coverage of around 5.8% and 1%, respectively. Methylated cytosines were identified as cytosines (or guanines as appropriate) in sequencing reads aligned to genomic cytosines, whereas unmethylated cytosines were identified as thymines (or adenines as appropriate) in sequencing reads aligned to genomic cytosines. Bisulphite conversion efficiency was always above 95% as judged by conversion of cytosines in CHG and CHH contexts (data not shown). The mapped bisulphite sequences were split into three groups. Sequences not spanning a CpG were discarded, and separate lists were made for sequences showing complete methylation or complete demethylation. In the very small number of cases where the same sequence showed both methylation and demethylation it was added to both lists. Where there were multiple data sets for the same sample the methylated and unmethylated lists were merged. Analysis of the data was performed using SeqMonk (http://www.bioinformatics.bbsrc.ac.uk/projects/seqmonk). The methylated and unmethylated lists were merged together with the methylation status being encoded in the strand of the read (methylated = forward; unmethylated = reverse). A tile of 250-kb regions was overlaid on the genome and the methylation status of each tile was calculated. Tiles containing fewer than ten reads were discarded, as were tiles where there were five or more reads with exactly the same mapped position. The methylation status was calculated as the $\log_2$ ratio of the methylated:unmethylated counts. The distribution of values showed a normal distribution and a comparison between tissues was made using a boxwhisker plot which showed the median, upper and lower quartiles and extremities (median $\pm 2 \times$ interquartile range). Any values outside this range were plotted individually as outliers. To calculate the methylation levels in specific genomic regions (promoters, genes, introns, exons, transposon families) SeqMonk was used to generate probe regions using the Ensembl features from the annotated NCBIM37 genome as a template. Total counts of overlapping reads in all of these regions across the genome were made and a single methylated:unmethylated ratio was produced. The positions of all repeats in the NCBIM37 mouse genome were extracted from Ensembl and classified into families based on their annotation. A count was made of reads which overlapped with all of these repeat regions and these counts were combined across all members of each family. A single measure per family was then made of the $\log_2$ ratio of methylated:unmethylated reads. All repeat families shown are represented by more than 1,000 CpG containing reads in each data set.

Methylation analysis by Sequenom MassArray. DNA from FACS-sorted PGCs was extracted using the AllPrep DNA/RNA Micro Kit (Qiagen). The DNA was then treated with bisulphite reagent using the two-step modification procedure outlined in the Imprint DNA modification kit (Sigma). Primer pairs were designed using the MethPrimer program (http://www.urogene.org/methprimer/index1.html). A complete list of primers used for analysis is available on request (primers for IAPs were based on the consensus sequence of IAPLTR1a repeats which represent approximately 1.5% of the ERVK family). Amplification of the bisulphite-converted DNA and preparation of PCR products for quantitative analysis of methylation as detected by the MassArray system was according to the protocol provided by the manufacturer.