AID stabilizes stem-cell phenotype by removing epigenetic memory of pluripotency genes

Ritu Kumar1, Lauren DiMenna2, Nadine Schroe3, Ting-Chun Liu1, Philipp Franck1, Silvia Muñoz–Descalzo3, Anna–Katerina Hadjantonakis1, Ali A. Zarrin4, Jayanta Chaudhuri2, Olivier Elemento5 & Todd Evans1

The activation-induced cytidine deaminase (AID; also known as AICDA) enzyme is required for somatic hypermutation and class switch recombination at the immunoglobulin locus1. In germinal-centre B cells, AID is highly expressed, and has an inherent mutator activity that helps generate antibody diversity2. However, AID may also regulate gene expression epigenetically by directly deaminating 5-methylcytosine in concert with base-excision repair to exchange cytosine2. This pathway promotes gene demethylation, thereby removing epigenetic memory. For example, AID promotes active demethylation of the genome in primordial germ cells. However, different studies have suggested either a requirement3 or a lack of function4 for AID in promoting pluripotency in somatic nuclei after fusion with embryonic stem cells. Here we tested directly whether AID regulates epigenetic memory by comparing the relative ability of cells lacking AID to reprogram from a differentiated murine cell type to an induced pluripotent stem cell. We show that AID-null cells are transiently hyper-responsive to the reprogramming process. Although they initiate expression of pluripotency genes, they fail to stabilize in the pluripotent state. The genome of AID-null cells remains hypermethylated in reprogramming cells, and hypermethylated genes associated with pluripotency fail to be stably upregulated, including many MYC target genes. Recent studies identified a late step of reprogramming associated with methylation status5, and implicated a secondary set of pluripotency network components6. AID regulates this late step, removing epigenetic memory to stabilize the pluripotent state.

The path to pluripotency involves multiple steps that can be more or less efficient, and this can be modelled in transcription-factor-induced reprogramming7. Reprogrammed induced pluripotent stem cell (iPSC) clones initially retain methylation patterns that may reflect an epigenetic memory of the source lineage8,9. To test directly the function of AID in reprogramming, we prepared tail-tip fibroblasts (TTFs) from Aid-knockout or control sibling mice. Before reprogramming, the fibroblasts appeared morphologically identical (Supplementary Fig. 1a), and were transduced equivalently with lentivirus encoding the four human ‘Yamanaka’ reprogramming transcription factors12: OCT4, KLF4, SOX2 and cMYC (OKSM; Supplementary Fig. 1b). Expression levels of exogenous human OCT4 72 h after infection (before induction of the endogenous murine gene) were equivalent, showing that lack of AID does not affect expression vector function (Supplementary Fig. 1c). Exogenous transcription factor expression was subsequently equivalently repressed in wild-type and Aid-null transduced cells (Supplementary Fig. 1d). Using quantitative polymerase chain reaction with reverse transcription (RT–qPCR) assays, transcript levels for Aid were not reliably detected above background in wild-type TTFs. However, after 1 week of reprogramming, transcripts were readily measured in wild-type cells, increasing as much as tenfold during the reprogramming process (Supplementary Fig. 2).

Initial reprogramming steps include the induction of proliferation and a change in the morphology of fibroblasts to smaller and rounder cells13. We found that cells lacking AID are initially hyper-responsive to reprogramming factors. The change in morphology was more rapid in Aid−/− cells, beginning at 2 days after transduction. After 4 days, Aid−/− cells were rounded and smaller than Aid+/+ cells (Fig. 1a). A higher fraction of Aid−/− cells stained positive for SSEA1, an early marker for pluripotency44 (Fig. 1b). At day 7, more Aid-null cells expressed NANOG compared with controls (Fig. 1c), correlating with modestly higher transcript levels for several pluripotency genes before the first week of reprogramming (Fig. 1d). Consistent with the molecular data, at 2 weeks the Aid-null cells consistently generated more early colonies than wild-type cells (sixfold more colonies on average over six independent experiments). Thus, Aid-null fibroblasts are hyper-responsive to reprogramming, even though the growth curves for uninjured wild-type and Aid-null cells were indistinguishable (Supplementary Fig. 3a). Rather, the enhanced expression of pluripotency markers in Aid−/− cells suggests that AID normally helps to stabilize the differentiated state, creating a barrier to the initial process of reprogramming. When the Aid−/− fibroblasts were passaged before transduction, this hyperresponsiveness was no longer seen (Supplementary Fig. 3b), suggesting

Figure 1 | Cells lacking AID are initially hyper-responsive to transcription-factor-based reprogramming. a, Aid+/+ and Aid−/− fibroblasts after 4 days of OKSM transduction. Note the more rounded appearance of null cells. Original magnification, ×10. b, Cells positive for SSEA1 after 4 days of OKSM transduction, determined by flow cytometry. c, Immunostaining with anti-OCT4 and anti-NANOG antibodies after 1 week of OKSM transduction. Original magnification, ×40. d, Relative transcript levels of pluripotency genes determined by qPCR after 4 days of OKSM transduction. For each gene, transcript levels were normalized to Aid+/+ cells set to a value of 1. Data represent the mean ± standard error of the mean (s.e.m.) from three independent experiments (**P < 0.01, ***P < 0.001).
that passive removal of epigenetic marks through DNA replication can normalize the initial response to OKSM.

Although there were more cells, no obvious differences were observed at 2 weeks in the morphologies of iPSC-like colonies derived from Aid\(^{-/-}\) cells compared with Aid\(^{+/+}\) cells, and both stained positive for the pluripotency marker NANOG (Supplementary Fig. 4). However, by 4 weeks the colonies derived from Aid\(^{-/-}\) fibroblasts were flattened with less defined edges (Fig. 2a). We tracked individual iPSC-like colonies and observed many Aid\(^{-/-}\) colonies that appeared pluripotent at 3 weeks but showed a differentiated morphology at 4 weeks (Fig. 2b). At 3 weeks, the Aid\(^{-/-}\) colonies showed a ‘patchy’ NANOG pattern, and by 4 weeks most colonies were differentiated, with few NANOG\(^{+}\) cells (Fig. 2c). This highly reproducible phenomenon was never observed for colonies derived from Aid\(^{+/+}\) cells, which retained their iPSC morphology and NANOG expression throughout the 4 weeks (Fig. 2b, c). The colonies derived from Aid\(^{-/-}\) cells showed a progressive decline in the frequency of cells expressing SSEA1 and OCT4 (and ineffective in maintaining a pluripotent iPSC phenotype (Supplementary Fig. 5). The stabilization of pluripotency was equally effective in wild-type cells and in Aid\(^{-/-}\) clones that were stable beyond p10 retained pluripotent morphology (Supplementary Fig. 8).

Using either wild-type or Aid\(^{-/-}\) cells, the expression of OKSM using a doxycycline-inducible vector for only 6 days was insufficient to generate pluripotent colonies. OKSM expression for 9, 12, 21 or 28 days was sufficient to generate pluripotent colonies, showing a ‘dose response’ in the wild-type fibroblasts, but was equally ineffective for generating stable pluripotent cells from Aid\(^{-/-}\) cells (Fig. 3c) (Fig. 2d). In an independent set of experiments, wild-type and Aid\(^{-/-}\) clones were picked and either maintained in culture as colonies or passaged three times (p3) all the clones, regardless of genotype, retained OKSM expression and generated embryoid bodies (EBs) that differentiate into cells positive for smooth muscle actin (mesoderm), βIII tubulin (ectoderm) and α-fetoprotein (endoi), whereas the Aid\(^{-/-}\) cells differentiated. With passaging, 60% of the Aid\(^{-/-}\) clones were able to maintain a pluripotent morphology (Supplementary Fig. 7).

Thus, although AID is not essential for reprogramming, an important transition occurs at around 3 weeks that is assisted by AID to stabilize the pluripotent phenotype. We considered whether the previously described DNA-demethylating role of AID might be at least partially responsible for this phenotypic stabilization. We profiled the genome by carrying out reduced-representation bisulphite sequencing (RRBS) in Aid\(^{-/-}\) and control cells after 3 weeks of

**Figure 2** | AID stabilizes pluripotency. a, Aid\(^{-/-}\) cells lose ES-cell-like morphological characteristics 4 weeks after OKSM transduction. b, Aid\(^{-/-}\) iPSC-like colonies progressively lose this phenotype. wk, weeks. c, Mutant cells lose NANOG expression. a-c, Original magnification, \(\times 10\). d, e, Cells that stained positive for SSEA1 (d) or OCT4 (e) were measured by flow cytometry after 1, 2, 3 or 4 weeks of OKSM transduction; n = 3 independent experiments, error bars denote standard deviation. NS, not significant.
reprogramming. We observed global DNA hypermethylation in the Aid-null cells (Supplementary Fig. 9a), when we considered either total CpGs or differentially methylated regions (DMRs). DNA hypermethylation occurs preferentially near RGW motifs (Supplementary Fig. 9b), the characteristic DNA-targeting sites for AID (45% of the hypermethylated CpGs). Notably, as Aid targets gene bodies, the hypermethylated regions are enriched in gene bodies, even though RRBS is biased towards capturing CpG-rich promoter and enhancer regions (Supplementary Fig. 9c).

Transcript profiles comparing the original, pre-transduced fibroblasts from Aid+/+ and Aid−/− embryos closely match (Supplementary Fig. 10a). By contrast, many genes fail to be upregulated in the mutant cells during reprogramming (Fig. 4a, b and Supplementary Fig. 10b). The genes that fail to be upregulated during reprogramming are highly enriched (P < 1 × 10−16) in the gene set showing hypermethylation in Aid-null cells (Fig. 4c). Of note, genes that are upregulated during reprogramming in both wild-type and Aid-null cells are also enriched in hypermethylated genes (P < 1 × 10−14), suggesting that many genes overcome the loss of AID during reprogramming. Among the set of hypermethylated genes, in Aid-null cells there is a failure to upregulate a set of secondary pluripotency genes, including Rex1 (also known as Zfp42), Gdf3, Dnmt3L, Cbx7, Zfp296, Dnmt1, Apobec1 and Tet3 (Fig. 4d). This expression data correlates well with the RRBS data, which was validated by MassArray bisulphite sequencing (Supplementary Fig. 11), and is consistent with a failure of this downstream network to be demethylated at late stages of reprogramming. Comparing the gene set of hypermethylated underexpressed genes to public embryonic stem (ES)-cell chromatin immunoprecipitation and massively parallel DNA sequencing (ChIP-seq) data sets revealed a marked enrichment of cMYC target genes in reprogramming cells lacking AID (Supplementary Fig. 12). According to qPCR data there was no difference in cMYC transcript levels in wild-type or Aid-null cells at any point in the reprogramming process (data not shown). Furthermore, enhanced levels of cMYC, achieved by retroviral transduction starting at 1 week of reprogramming, failed to stabilize pluripotency in Aid-null cells (Supplementary Fig. 13), suggesting that cMYC access to key target pluripotency genes (due to hypermethylation) rather than cMYC levels per se may be a limiting step for stabilization of the network.

Individual iPSC-like clones were isolated at 2 weeks and expanded to evaluate global methylation patterns for each individual clone. Hypermethylation patterns in the Aid-null clones were highly consistent with data obtained from bulk colony analysis. On the basis of genome-wide DNA methylation profiles (with the genome binned for 100 kb regions), Aid-null or wild-type clones cluster according to genotype (Supplementary Fig. 14a). Regions that were hypermethylated in the Aid-null bulk analysis tended to be hypermethylated in the Aid-null clones (Supplementary Fig. 14b). In fact, 66% of bulk hypermethylated DMRs were also hypermethylated in DNA derived from the isolated clones (P < 1 × 10−151). Methylation differences found previously for the secondary pluripotency genes were largely validated by MassArray bisulphite sequencing of DNA (Supplementary Fig. 14c), and most of the hypermethylated genes were expressed in the Aid-null clones at significantly lower levels compared with levels in wild-type clones (Supplementary Fig. 14d).

Altogether, our results show that active demethylation through AID-dependent cytosine deamination is important for stabilizing a genetic network controlling stem-cell phenotype. AID is not essential for reprogramming, because cells lacking AID can form stable pluripotent iPSCs, either through a passive demethylation process facilitated by DNA replication, or perhaps through compensation from related members of the APOBEC family. A recent study suggested that loss of AID affects transcription-factor-induced reprogramming only by an early acute short hairpin RNA (shRNA)-dependent depletion, but did not report the ‘iPSC’ phenotype beyond 3 weeks, when we show that Aid-null cells eventually fail to stabilize as pluripotent. Unless passaged, Aid-null cells always fail to reprogram. RNA-seq profiles of Aid-null cells after 4 weeks of reprogramming cluster with fibroblast samples (data not shown), and qPCR analysis validates significant upregulation of fibroblast-associated genes in the Aid-null samples (Supplementary Fig. 15), consistent with an epigenetic memory for fibroblast fate that fails to be fully removed during reprogramming in
the absence of AID. The function of AID for transition to a stable pluripotent phenotype may be relevant during embryogenesis, as AID-knockout strains have small litters. We found that ES-cell lines could be established from Aid-null blastocysts, and these appeared by morphology and staining patterns to be normal (Supplementary Fig. 16). However, the efficiency of ES-cell-line derivation was markedly reduced (17/28 embryos, 60.7%; P < 0.03 by chi-squared analysis).

Although there is much promise for the use of iPSCs for disease modelling and cellular therapies18, there remains concern about whether iPSC genomes are damaged through the process of reprogramming19–21. AID, which has natural mutator activity, is activated during reprogramming, and mediates demethylation in gene bodies including secondary pluripotency genes that encode proto-oncogenes. Although inefficient, we showed that iPSCs could be generated in the absence of AID, removing epigenetic memory marks through an AID-independent mechanism. Retaining epigenetic memory, in the absence of AID, might be useful for promoting efficiency of differentiation towards parental lineage fate. Furthermore, if AID-independent reprogramming lessens the mutation load, it could provide a safer strategy for the generation of iPSCs for cellular therapies.

METHODS SUMMARY

TTFs were prepared using 0.2% collagenase; MEFs were prepared from embryonic day (E)13.5 embryos. Intracellular immunostaining was performed after fixing the cells with 4% paraformaldehyde. Surface staining was done without fixing the cells. Complementary DNA was prepared using Superscript III and qPCR was performed (primers listed in Supplementary Table 1) using SYBR green and a Roche LightCycler 480 II. All data were from at least three independent experiments. Enhanced RRBS was performed as described22 and bisulphite reads were aligned to the bisulphite-converted mm9 genome using Bismark23. We achieved very high coverage by sequencing one full Illumina lane (1.62 M CpGs in common between the two runs).

For bulk colonies, the wild-type and Aid-null samples (at 3 weeks) were analysed with RRBS that covered 1.96 M and 1.82 M CpGs, respectively, with at least 10× coverage (with >1.62 M CpGs in common between the two runs). 1.52 M and 1.59 M CpGs had ≥20× coverage in wild-type and Aid-/− cells, respectively. 1.17 M and 1.32 M CpGs had ≥50× coverage, respectively. Differentially methylated CpGs were identified using the Fisher exact test with Benjamin–Hochberg correction for multiple testing. DMRs were defined as containing at least five differentially methylated CpGs and a total methylation difference of more than 10%. Paired-end DNA-seq libraries were constructed as described44 and sequenced using an Illumina HiSeq2000. Reads were aligned to mm9 using TopHat and gene expression levels were quantified using Cufflinks, using upper-quartile and GC-content normalizations. Fold changes of 1.5 with fragments per kilobase of transcript per million fragments mapped (FPKM) >5 in at least one condition were used to derive differentially expressed genes. MassArray EpiTYPER analysis was performed on selected regions identified as hypermethylated by RRBS. PCR primers (Supplementary Table 2) were designed to probe amplicons using EpiDesigner (http://www.epidesigner.com/). Bisulphite conversion and MassArray analysis were performed as previously described22. Metadata for all RRBS experiments are given in Supplementary Table 3. Methylation differences were calculated at each CpG and boxplot analysis was performed at each interrogated region. All statistical analyses were performed using the R software package.

Full Methods and any associated references are available in the online version of the paper.

Received 14 August 2012; accepted 13 May 2013.

Published online 26 June 2013.

1. Muramatsu, M. et al. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell 102, 553–563 (2000).

2. Chaudhuri, J. et al. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. Nature 422, 726–730 (2003).

3. Bhutani, N., Burns, D. M. & Blau, H. M. DNA demethylation dynamics. Cell 146, 866–872 (2011).

4. Popp, C. et al. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature 463, 1101–1105 (2010).

5. Bhutani, N. et al. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature 463, 1042–1047 (2010).

6. Foshey, K. M. et al. Embryonic stem cells induce pluripotency in somatic cell fusion through biphasic reprogramming. Mol. Cell 46, 159–170 (2012).

7. Polo, J. M. et al. A molecular roadmap of reprogramming somatic cells into iPSCs. Cell 151, 1617–1632 (2012).

8. Buganim, Y. et al. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell 150, 1209–1222 (2012).

9. Meissner, A. Epigenetic modifications in pluripotent and differentiated cells. Nature Biotechnol. 28, 1079–1088 (2010).

10. Kim, K. et al. Epigenetic memory in induced pluripotent stem cells. Nature 467, 285–290 (2010).

11. Polo, J. M. et al. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nature Biotechnol. 28, 848–855 (2010).

12. Sommer, C. A. et al. Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. Stem Cells 27, 543–549 (2009).

13. Smith, Z. D., Nachman, L., Regen, A. & Meissner, A. Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. Nature Biotechnol. 28, 521–526 (2010).

14. Brambrink, T. et al. Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. Cell Stem Cell 2, 151–159 (2008).

15. Sun, Y. et al. Critical role of activation induced cytidine deaminase in experimental autoimmune encephalomyelitis. Autoimmunity 46, 157–167 (2013).

16. Corticello, S. G. The AID/APOBEC family of nucleic acid mutators. Genome Biol. 9, 229 (2008).

17. Bhutani, N. et al. A critical role for AID in the initiation of reprogramming to induced pluripotent stem cells. FASEB J. 27, 1107–1113 (2012).

18. Robinson, D. A. & Daley, G. Q. The promise of induced pluripotent stem cells in research and therapy. Nature 481, 295–305 (2012).

19. Gore, A. et al. Somatic coding mutations in human induced pluripotent stem cells. Nature 471, 63–67 (2011).

20. Hussein, S. M. et al. Copy number variation and selection during reprogramming to pluripotency. Nature 471, 58–62 (2011).

21. Mayshar, Y. et al. Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. Cell Stem Cell 7, 521–531 (2010).

22. Akalin, A. et al. Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. PLoS Genet. 8, e1002781 (2012).

23. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulite-Seq applications. Bioinformatics 27, 1571–1572 (2011).

24. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods 5, 621–628 (2008).

Supplementary information is available in the online version of the paper.

Acknowledgements We thank G. Mostoslavsky for the gift of reprogramming vectors. A. Melnick provided advice and consultation. We are grateful to T. Honjo for AID-knockout mice, and N. Pavasantivong for assistance in designing genotyping PCR primers. The Epigenomics Core Facility of Weill Cornell Medical College carried out the bisulphite sequencing and provided consultation. This study was supported by National Institutes of Health grant HL056182 (T.E.), AI072194 (J.C.) and National Science Foundation CAREER grant 0949564 (O.E.).

Author Contributions P.K. conceived the study, carried out experiments, and wrote the manuscript. J.D., N.S., T.-C.L., P.F. and S.M.-D. carried out experiments. A.A.Z. and A.K.H. provided essential reagents and expertise. J.C.O. conceived the study and wrote the manuscript. O.E. conceived the study, carried out computational and informatics analyses, and wrote the manuscript. T.E. conceived the study and wrote the manuscript.

Author Information All RNA-seq and RRBS data have been deposited in the Gene Expression Omnibus under accession GSE46700. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.E. (tre2003@med.cornell.edu).
METHODS

Summary. Intracellular immunostaining was performed after fixing the cells with 4% paraformaldehyde. Surface staining was done without fixing the cells. All data were from at least three independent experiments. Enhanced RRBS was performed as described23 and bisulphite reads were aligned to the bisulphite-converted mm9 genome using Bismark24. We achieved very high coverage by sequencing one full Illumina lane (~200 M, 51 bp reads) per sample. For bulk colonies, the wild-type and Aid- null samples (at 3 weeks) were analysed with RRBS that covered 1.96 M and 1.82 M CpGs, respectively, with at least 10X coverage (with >1.62 M CpGs in common between the two runs). 1.52 M and 1.59 M CpGs had ≥20X coverage in wild-type and Aid-/- cells, respectively. 1.17 M and 1.32 M CpGs had ≥50X coverage, respectively. Differentially methylated CpGs were identified using the Fisher exact test with Benjamin–Hochberg correction for multiple testing. DMRs were defined as containing at least five differentially methylated CpGs and a total methylation difference of more than 10%. Paired-end RNA-seq libraries were constructed as described24 and sequenced using an Illumina HiSeq2000. Mass-Array Epityper analysis was performed on selected regions identified as hyper-methylated by RRBS. PCR primers (Supplementary Table 2) were designed to probe amplicons using EpiDesigner (http://www.epidesigner.com/). Bisulphite conversion and MassArray analysis were performed as previously described25. Metadata for all RRBS experiments are given in Supplementary Table 3. Methylation percentages were calculated at each CpG and boxplot analysis was performed at each interrogated region. All statistical analyses were performed using the R software package.

Mice. Aid-deficient (Aid-/-) mice were a gift from T. Honjo. Some of the reprogramming experiments were performed on the Aid-/- and Aid-/-/ mice derived by crossing Aid-/- mice, and for some of the experiments wild-type BALB/c mice were purchased from the Jackson Laboratory. Reprogramming experiments were performed on tail fibroblasts obtained from seven Aid-/- and seven Aid-/-/ age-matched mice (four littersmates and three non-littermates, each). We observed similar results when comparing littersmates or the purchased mice. Mice or embryos used for MEFs were genotyped by PCR. All animals were maintained according to the guidelines for animal welfare of the Memorial Sloan-Kettering Research Animal Resource Center.

Preparation of fibroblasts. For preparation of tail fibroblasts either whole tails or tail tips were collected from age-matched adult Aid-/- and Aid-/-/ mice (age range, 3–8 weeks). Tails or tail tips were minced using a razor blade/scalpel after BALB/c mice were purchased from the Jackson Laboratory. Reprogramming seven Aid-/- Aid-/-/ mice was replaced with fresh media. Viral supernatant was collected after 48 h and spun at 3,000 r.p.m. to remove dead cells and filtered through a sterile syringe filter with 0.45 μm polyethersulfone membrane (VWR international). Ectopic expression of wild-type AID or catalytically mutant AID was achieved using retroviral vectors as described previously26. Retroviruses were produced by co-transfecting the vector with the packaging plasmid pCL-EOC in HEK293T cells. The cMYC overexpression experiments were achieved using retroviral vector pMXs-c-Myc (Addgene, 13375). pMXs-c-Myc was co-transfected with vector encoding VSVG in GP2-293 cells to collect functional virus.

Transduction with lentiviruses. Five-hundred-thousand fibroblasts were plated in one well of a gelatin-coated 6-well tissue culture plate. After 6 h, 1 μl of viral supernatant containing 8 μg ml-1 polybrene (Millipore) was added to the fibroblast-containing wells. For every experiment, Aid-/- and Aid-/-/ fibroblasts were infected at the same time with an identical titre of virus, prepared for each experiment from the same batch. Viral supernatant was removed after 12 h of infection and fresh media containing DMEM, 10% FBS, 1 mM l-glutamine, 100 U ml-1 penicillin and 10 μg ml-1 streptomycin was added. After 1 day media was replaced by mouse embryonic stem-cell (MES) media containing DMEM, 20% ES-cell compatible FBS (Gemini Bio Products), LIF (2% conditioned medium) and 1.5 × 10-5 M monoiodoacetate (MTG; Sigma), 1 mM l-glutamine, 100 U ml-1 penicillin and 10 μg ml-1 streptomycin. Cells were maintained in the same media. In some of the experiments cells were transferred to mitotically inactivated MEF feeders after 2 days of transduction. Approximately 30,000-50,000 cells were transferred to one well of a feeder-containing 6-well plate. Isolated iPSC colonies were always cultured on feeders. All the cells were kept at 37 °C in a humidified environment at 5% CO2.

Immunostaining. Immunostaining was performed on fixed cells (4% PFA in PBS with 1 mM CaCl2, 15 min) washed and blocked for 30 min in BBT-BSA buffer (BBS with 0.5% BSA, 0.1% Triton and 1 mM CaCl2). Cells with primary antibodies were incubated overnight at 4 °C in the following dilutions: anti-NANOG (1:1000, Santa Cruz 2561), 1:1000 (eBiosciences 14-5761, 1:1000, 5 μg ml-1) and anti-Oct4 (Santa Cruz 5279, 1:200, 2 μg ml-1). Cells were washed and blocked in BBT-BSA and then incubated with Alexa-conjugated secondary antibodies (1:500, from Molecular Probes). VECTASTAIN-DAPI was used as a mounting medium. Images were acquired using a Zeiss LSM 510-Meta confocal microscope or a Zeiss epifluorescence microscope with AxioVision software. For flow cytometric analysis cells were trypsinized, fixed with 4% PFA for 20 min, blocked for 1 h and then stained in suspension. SSEA1 (Santa Cruz, 1:100) staining was performed on unfixed cells. Cells were analysed on a BD-Accuri C6 flow cytometer (BD Biosciences) using CFlowPlus software.

RT-qPCR. Cells were trypsinized and collected in Trizol reagent (Life technologies). Total RNA extraction was done using the RNAeasy Mini Kit (Qiagen). The cDNA was synthesized using the SuperScript III First-Strand Synthesis System (Invitrogen). RT-qPCR was performed on three experimental replicates using SYBR green Master 1 (Roche). Data were generated on a LightCycler 480 II (Roche) and analysed using LightCycler 480 software. qPCR data was calculated based on the mean of three experimental replicates. All quantifications were normalized to an endogenous Gapdh control. The relative quantification value for each target gene was compared to the calibrator for that target gene. All the primers used for qPCR spanned introns (Supplementary Table 1).

In vitro differentiation. For embryoid body formation, iPSCs were passaged two times on feeder-free gelatin-coated culture dishes. 5 × 103 iPSCs were plated on a 10-cm low attachment dish. Embryo body cultures were kept in 6 days in MES media without LIF, and media was changed every other day. For mesodermal differentiation, embryo bodies were replated on gelatin-coated dishes and cultured for another 7 days in the same media. For neural differentiation, embryo bodies were cultured in media containing DMEM F-12 (Cambrex), 0.5% N2 (Gibco) and 0.5% B27 (Gibco) supplements, 1 mM l-glutamine, 1% non-essential amino acids (Gibco) and 1.5 × 10-5 M monoiodoacetate for another 3 days, after which embryo bodies were seeded onto gelatin-coated dishes with the same media, plus 2 μg ml-1 heparin (Sigma) for another 4 days. For endoderm differentiation, embryo bodies were replated onto gelatin-coated dishes in MES media lacking LIF but containing 0.5% FBS and 50 ng ml-1 Activin (R&D) for 7 days. For all differentiation cultures, media was changed daily.

ES-cell derivation from blastocyst-stage mouse embryos. Aid-/-/ males and Aid-/-/ females (or congenic wild-type pairs as controls) were timed mated and E3.5 blastocyst-stage embryos were collected in M2 media according to standard protocols27. Embryos were placed on mitotically inactivated MEFs27 in knockout DMEM (Gibco) containing 15% knockout serum replacement (KSR; Gibco), 2 mM l-glutamine, 0.1 mM β-mercaptoethanol, 0.1 mM non-essential amino acids, 1 mM sodium pyruvate, 1.5 × 10-5 M CHIR99021 (inhibitor of GSK3B), 1.000 Ui ml-1 LIF, 1 μM PD0325901 (ERK inhibitor) and 3 μM CHIR99021 (inhibitor of GSK3B). After embryos attached, media was replaced every other day. After 10–11 days an outgrowth could be observed and was dissociated in 0.25% trypsin/EDTA using a mouth-controlled drawn glass pipette. Trypsin was inactivated by adding knockout...
DMEM (Gibco) containing 15% ES-cell-compatible FBS, 2 mM l-glutamine, 0.1 mM β-mercaptoethanol, 0.1 mM non-essential amino acids, 1 mM sodium pyruvate, 100 U ml⁻¹ penicillin and 100 μg ml⁻¹ streptomycin, 1,000 U ml⁻¹ LIF, 1 μM PD0325901 and 3 μM CHIR99021. From this point onwards the culture was maintained in the same media on mitotically inactivated MEFs. Media was replaced every other day. After 3–5 days ES-cell-like colonies were observed and passaged at 70% confluency. Stable ES-cell lines were validated by passaging a minimum of ten times without loss of pluripotent morphology or marker expression.

Genomic analysis. RRBS was performed as described²² and bisulphite reads were aligned to the bisulphite-converted mm9 using Bismark²³. Differentially methylated CpGs were identified using the Fisher exact test with Benjamini–Hochberg correction for multiple testing. We defined DMRs as regions containing at least five differentially methylated CpGs, where contiguous differentially methylated CpGs are separated by 250 bp or less, and for which the total methylation change between wild-type and Aid-null cells is 10% or more (calculated using all CpGs within the considered region including those that were not called as differentially methylated).

The distribution of DMR lengths was as follows. For hypermethylated DMRs, average, 222 bp; median, 190 bp; minimum, 10 bp; maximum, 4,081 bp. For the lower number of hypomethylated DMRs, we observed the following statistics: average, 229 bp; median, 188 bp; minimum, 19 bp; maximum, 1,250 bp. Paired-end RNA-seq libraries were constructed as previously described²⁴ and sequenced using an Illumina HiSeq2000. Reads were aligned to mm9 using TopHat and gene expression levels were quantified using Cufflinks, using upper-quartile and GC-content normalizations. Twofold changes with FPKM > 5 in at least one condition were used to derive differentially expressed genes. All statistical analyses were performed using the R software package. The meta-data for all RRBS data are provided in Supplementary Table 3.

Quantitative DNA methylation analysis by mass spectrometry. The level of DNA methylation for specific genes was measured using a MALDI-TOF mass spectrometry based method (EpiTYPER; Sequenom) as previously described²⁸. Briefly, 1 μg of DNA was treated with sodium bisulphite using the EZ methylation kit (Zymo-Research). The treatment converts non-methylated cytosines into uracil, leaving methylated cytosines unchanged. PCR amplification, addition of SAP solution and Transcription/RNase A cocktails were performed according to the protocol provided by Sequenom and the mass spectra were quantified by the EpiTYPER analyser. Amplicons probed are given in Supplementary Table 2.

25. Vuong, B. Q. et al. Specific recruitment of protein kinase A to the immunoglobulin locus regulates class-switch recombination. Nature Immunol. 10, 420–426 (2009).
26. Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R. Manipulating the Mouse Embryo: A Laboratory Manual 3rd edn (Cold Spring Harbor, 2003).
27. Joyner, A.L. Gene Targeting: A Practical Approach 2nd edn (Oxford Univ. Press, 2000).
28. Ehrich, M. et al. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc. Natl Acad. Sci. USA 102, 15785–15790 (2005).