HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2014 February 01.

Published in final edited form as:

Nature. 2013 August 1; 500(7460): 89–92. doi:10.1038/nature12299.

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

Ritu Kumar[1], **Lauren DiMenna**[2], **Nadine Schrode**[3], **Ting-Chun Liu**[1], **Philipp Franck**[1], **Silvia Muñoz-Descalzo**[3], **Anna-Katerina Hadjantonakis**[3], **Ali A. Zarrin**[4], **Jayanta Chaudhuri**[2], **Olivier Elemento**[5], and **Todd Evans**[1,*]

[1]Department of Surgery, Weill Cornell Medical College, New York, NY 10065

[2]Immunology Program, Sloan-Kettering Institute, Memorial Sloan Kettering Cancer Institute, New York, NY 10065

[3]Developmental Biology Program, Sloan-Kettering Institute, Memorial Sloan Kettering Cancer Institute, New York, NY 10065

[4]Genentech, South San Francisco, CA 94080

[5]Institute for Computational Biomedicine, Weill Cornell Medical College, New York, NY 10065

Abstract

The activation-induced cytidine deaminase enzyme (AID) is required for somatic hyper-mutation and class switch recombination at the immunoglobulin locus[1]. In GC-B cells, AID is highly expressed, with inherent mutator activity that helps generate antibody diversity[2]. However, AID may also regulate gene expression epigenetically by directly deaminating 5-methylcytosine (5mC) in concert with base excision repair to exchange cytosine[3]. This pathway promotes gene demethylation, thereby removing epigenetic memory. For example, AID promotes active demethylation of the genome in primordial germ cells[4]. However, different studies have suggested either a requirement[5] or a lack of function[6] for AID promoting pluripotency in somatic nuclei following fusion with embryonic stem cells (ESCs). We tested directly whether AID regulates epigenetic memory, by comparing the relative ability of cells lacking AID to reprogram from a differentiated cell type to an induced pluripotent stem cell (iPSC). 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 the pluripotent state. The genome of *AID*-null cells remains hyper-methylated in reprogramming cells, and hyper-methylated genes associated with pluripotency fail to be stably up-regulated, including many MYC target genes. Recent studies identified a late step of reprogramming associated with methylation status[7], and implicated a

*Author for correspondence: Todd Evans, 1300 York Ave. LC704, tre2003@med.cornell.edu, phone: 212-746-9485, fax: 212-746-7378.

Author Contributions

RK conceived the study, carried out experiments, and wrote the manuscript. LD, NS, T-CL, PF, and SM-D carried out experiments. AAZ and A-KH provided essential reagents and expertise. JC conceived the study and wrote the manuscript. OE conceived the study, carried out computational and informatics analyses, and wrote the manuscript. TE conceived the study and wrote the manuscript.

The authors have no competing interests.

secondary set of pluripotency network components[8]. 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 modeled in transcription factor (TF)-induced reprogramming[9]. Reprogrammed iPSC clones initially retain methylation patterns that may reflect an epigenetic memory of the source lineage[10,11]. To test directly the function of AID in reprogramming, we prepared tail-tip fibroblasts (TTFs) from *AID* knockout or control sibling mice. The starting fibroblasts appeared morphologically identical (Fig.S1a), and were transduced equivalently with lentivirus encoding the four human "Yamanaka" reprogramming TFs[12]: OCT4, KLF4, SOX2 and cMYC (OKSM, Fig.S1b). Expression levels 72 hours post-infection of exogenous human OCT4 (prior to induction of the endogenous murine gene) were equivalent, showing that lack of AID does not impact expression vector function (Fig.S1c). Exogenous TF expression was subsequently equivalently repressed in wildtype and *AID*-null transduced cells (Fig.S1d). Using qRT-PCR assays, transcript levels for *AID* were not reliably detected above background in wildtype TTFs. However, after one week of reprogramming, transcripts were readily measured in wildtype cells, increasing as much as 10-fold during the reprogramming process (Fig.S2).

Initial reprogramming steps include induction of proliferation and change in the morphology of fibroblasts to smaller and rounder cells[13]. We discovered that cells lacking AID are initially hyper-responsive to reprogramming factors. Change in morphology was more rapid in $AID^{-/-}$ cells, beginning at 2 days post-transduction. After 4 days, $AID^{-/-}$ cells are rounded and smaller than $AID^{+/+}$ cells (Fig.1a). A higher fraction of $AID^{-/-}$ cells stain positive for SSEA1, an early marker for pluripotency[14] (Fig.1b). At day 7, more *AID*-null cells express NANOG compared to controls (Fig.1c), correlating with modestly higher transcript levels for several pluripotency genes prior to the first week of reprogramming (Fig.1d). Consistent with the molecular data, the *AID*-null cells consistently generated at 2 weeks more early colonies than wildtype cells (6-fold more colonies on average over six independent experiments). Thus, *AID*-null fibroblasts are hyper-responsive to reprogramming, even though growth curves for uninfected wildtype and *AID*-null cells were indistinguishable (Fig.S3a). 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 prior to transduction, this hyper-responsiveness was no longer seen (Fig.S3b), suggesting that passive removal of epigenetic marks through DNA replication can normalize the initial response to OKSM.

Although there are more of them, no obvious differences were observed at 2 weeks in morphologies of iPSC-like colonies derived from $AID^{-/-}$ cells compared with $AID^{+/+}$ cells, and both stain positive for the pluripotency marker NANOG (Fig.S4). However, by four weeks the colonies derived from $AID^{-/-}$ fibroblasts are 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 four

weeks most colonies differentiated with few NANOG+ cells (Fig.2c). This highly reproducible phenomenon was never observed for colonies derived from *AID*$^{+/+}$ cells, which retained iPSC morphology and NANOG expression through 4 weeks (Fig.2b–c). The colonies derived from *AID*$^{-/-}$ cells showed a progressive decline in the frequency of cells expressing SSEA1 (Fig.2d) and OCT4 (Fig.2e). The same results were observed using mouse embryonic fibroblasts (MEFs), as the AID-null cells failed to maintain a pluripotent iPSC phenotype (Fig.S5). The stabilization of pluripotency was equally effective in wildtype cells (and ineffective in *AID*-null cells) regardless of whether constitutively expressed or doxycycline-inducible OSKM cassettes were used (Fig.S6a).

Using either wildtype or *AID*-null cells, expression of OKSM using a doxycycline-inducible vector for only 6 days was insufficient to generate pluripotent colonies. Expression for 9, 12, 21, or 28 days was sufficient to generate pluripotent colonies, showing a “dose-response” with the wildtype fibroblasts, but was equally ineffective for generating stable pluripotent cells from *AID*-null cells at the 4 week time-point (Fig.S6b,c). When AID was expressed along with OKSM during initial reprogramming stages, this failed to rescue pluripotency (not shown). However, when OKSM-transduced *AID*-null cells were secondarily infected with an AID-expressing retrovirus after one week (Fig.S7a,b), there was a partial but significant rescue of pluripotency markers stabilized at the 4 week time-point. Since the rescue was not complete we cannot rule out additional molecular mechanisms working with AID for stabilization of pluripotency. Expression of AID in wildtype cells further enhanced the stable expression of pluripotency markers at 4 weeks (Fig.S7c). Since *Nanog* is syntenic with the *AID* locus, an independent set of experiments used fibroblasts derived from an *AID* knockout strain in which the *neoR* gene was deleted[15]. These “clean” knockout cells also failed to derive stable iPSC colonies (not shown).

It was previously reported that iPSCs derived from different cell types retain epigenetic memory of their somatic phenotype[10,11]. We tested if genome replication through passaging could stabilize a pluripotent phenotype even in the absence of AID. We generated iPSC-like colonies and isolated at three weeks 13 *AID*$^{+/+}$ and 12 *AID*$^{-/-}$ colonies (clones) that appeared pluripotent by morphology. After three passages (p3) all the clones, regardless of genotype, retained iPSC-like morphology and stained positive for OCT4 and NANOG (Fig. 3a). However, between p7 and p10, 5 out of 12 *AID*$^{-/-}$ clones failed to retain pluripotency and differentiated (Fig.3b). All 13 AID$^{+/+}$ colonies and the other 7 AID$^{-/-}$ clones that were stable beyond p10 retained pluripotency through p50. All iPSC clones that retained morphology beyond p10 formed embryoid bodies equivalently and could differentiate into cells expressing smooth muscle actin (mesoderm), βIII tubulin (ectoderm) and alpha-fetoprotein (endoderm), regardless of *AID* genotype (Fig.3c). In an independent set of experiments, wildtype and *AID*-null clones were picked and either maintained in culture as colonies or passaged. The non-passaged wildtype-derived clones maintained a pluripotent morphology, while the *AID*-null cells differentiated. With passaging, 60% of the *AID*-null clones were able to maintain a pluripotent morphology (Fig.S8).

Thus, although AID is not essential for reprogramming, an important transition occurs around 3 weeks that is assisted by AID to stabilize the pluripotent phenotype. We hypothesized that the previously described DNA demethylating role of AID might be at least

partially responsible for this phenotypic stabilization. We profiled the epigenome by carrying out reduced-representation bisulfite sequencing (RRBS) in *AID*-null and control cells after 3 weeks of reprogramming. We observed global DNA hyper-methylation in the *AID*-null cells (Fig.S9a), considering either total CpGs or differentially methylated regions (DMRs). DNA hyper-methylation occurs preferentially near RGYW motifs (Fig.S9b), the characteristic DNA-targeting sites for AID (45% of the hyper-methylated CpGs). Strikingly, as AID targets gene bodies, the hyper-methylated regions are enriched in gene bodies, even though RRBS is biased toward capturing CpG-rich promoter and enhancer regions (Fig.S9c).

Transcript profiles comparing starting fibroblasts from $AID^{+/+}$ and $AID^{-/-}$ embryos are closely matched (Fig.S10a). In contrast, many genes fail to be up-regulated in the mutant cells during reprogramming (Fig.4a,b;Fig.S10b). The genes that fail to be up-regulated during reprogramming are highly enriched ($p<1e-10$) in the gene-set displaying hyper-methylation in *AID*-null cells (Fig.4c). Of note, genes that are induced in wildtype and *AID*-null cells are also enriched in hyper-methylated genes ($p<1e-14$), suggesting that many genes overcome the loss of AID during reprogramming. Focused on the hyper-methylated genes, there is a failure to up-regulate in *AID*-null cells a set of secondary pluripotency genes, including *Rex1*, *Gdf3*, *Dnmt3L*, *Cbx7*, *Zfp296*, *Dnmt1*, *Apobec1*, and *Tet3* (Fig.4d). This expression data correlates well with the RBBS data, which was validated by MassArray bisulfite sequencing (Fig.S11), and is consistent with a failure of this downstream network to be demethylated at late stages of reprogramming. Comparing the gene-set of hyper-methylated under-expressed genes to public ESC ChIP-seq datasets revealed a dramatic enrichment for cMYC target genes in the reprogramming cells lacking AID (Fig.S12). According to qPCR data there was no difference in *cMyc* transcript levels in wildtype or *AID*-null cells at any point of the reprogramming process (not shown). Furthermore, enhanced levels of cMYC by retroviral transduction starting at one week of reprogramming failed to stabilize pluripotency in *AID*-null cells (Fig.S13), suggesting that cMYC access to binding key target pluripotency genes (due to hyper-methylation) 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. Hyper-methylation patterns in the *AID*-null clones were highly consistent with data obtained from bulk colony analysis. Based on genome-wide DNA methylation profiles (with the genome binned for 100 kb regions), *AID*-null or wildtype clones cluster according to genotypes (Fig.S14a). Regions that were hyper-methylated in the *AID*-null bulk analysis tend to be hyper-methylated in the *AID*-null clones (Fig.S14b). In fact, 66% of bulk hyper-methylated DMRs were also hyper-methylated in DNA derived from the isolated clones ($p<1e-151$). Methylation differences found previously for the secondary pluripotency genes were largely validated by MassArray bisulfite sequencing of DNA (Fig.S14c), and most of the hyper-methylated genes were expressed in the *AID*-null clones at significantly lower levels (Fig.S14d).

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, since 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[16]. A recent report suggested that loss of AID impacts TF-induced reprogramming only by an early acute shRNA-dependent depletion[17], but did not report the "iPSC" phenotype beyond three weeks, when we show *AID*-null cells eventually fail to stabilize pluripotency. 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 (not shown), and qPCR analysis validates significant up-regulation of fibroblast-associated genes in the *AID*-null samples (Fig.S15), 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, since *AID* knockout strains have small litters. We found that ESC lines could be established from *AID*-null blastocysts, and these appeared by morphology and staining patterns to be normal (Fig.S16). However, the efficiency of ESC line derivation was markedly reduced (17/28 embryos, 60.7%) compared with congenic wildtype embryos (21/24, 87.5%,), $p<0.03$ by chi-squared analysis.

While there is much promise for the use of iPSCs for disease modeling and cellular therapies[18], there remains concern whether iPSC genomes are damaged through the process of reprogramming[19–21]. AID with 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 toward parental lineage fate. Furthermore, if AID-independent reprogramming lessens the mutation load, it could provide a safer strategy to generatie iPSCs for cellular therapies.

METHODS

Summary

TTFs were prepared using 0.2% collagenase; MEFs were prepared from E13.5 embryos. Intracellular immuno-staining was performed after fixing the cells with 4% paraformaldehyde. Surface staining was done without fixing the cells. cDNA was prepared using Superscript III and qPCR performed (primers listed in Table ST1) using SYBR-green and a Roche LightCycler 480 II. All data were from at least three independent experiments. Enhanced RRBS was performed as described[22] and bisulfite reads aligned to the bisulfite-converted mm9 using Bismark[23]. We achieved very high coverage by sequencing one full Illumina lane (~200M 51bp reads) per sample. For bulk colonies, the wildtype and AID-null samples (at 3 weeks) were analyzed with RRBS that covered 1.96M and 1.82M CpGs, respectively, with at least 10X coverage (with >1.62M CpGs in common between the two runs). 1.52M and 1.59M CpGs had ≥20X coverage in wildtype and $AID^{-/-}$, respectively. 1.17M and 1.32M CpGs had ≥50X coverage, respectively. Differentially methylated CpGs were identified using the Fisher exact test with correction for Benjamini-Hochberg multiple testing. DMRs were defined as containing at least 5 differentially methylated CpGs and total

methylation difference more than 10%. Paired-end RNA-seq libraries were constructed as described[24] 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. 1.5-fold changes with FPKM>5 in at least one condition were used to derive differentially expressed genes. MassArray EpiTyper analysis was performed on selected regions identified as hyper-methylated by RRBS. PCR primers (Table ST2) were designed to probe amplicons using EpiDesigner (http://www.epidesigner.com/). Bisulfite conversion and MassArray analysis were performed as previously described[22]. Metadata for all RRBS are given in Table ST3. Methylation differences were calculated at each CpG and boxplot analysis performed at each interrogated region. All statistical analyses were performed using the R software package. All RNA-seq and RRBS data is available at GEO, accession GSE46700.

Mice

AID-deficient ($AID^{-/-}$) mice were a gift from T. Honjo. Some of the reprogramming experiments were performed on the $AID^{-/-}$ and $AID^{+/+}$ littermates 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 7 $AID^{-/-}$ and 7 $AID^{+/+}$ age-matched mice (4 littermates and 3 non-littermates, each). We observed similar results when comparing littermates 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 washing with PBS. For preparation of mouse embryonic fibroblasts (MEFs) $AID^{+/-}$ males and $AID^{+/-}$ females were time mated and E13.5 embryos were collected. Embryos were washed several times and extra care was taken to prevent maternal cross contamination. After removing the head and all internal organs embryos were minced using a razor blade/scalpel. Minced tails or embryos were washed twice with DMEM (Cellgro) containing 10% FBS (Gemini Bio Products), 1 mM L-glutamine (Cellgro), 100 U/ml penicillin (Cellgro) and 10 μg/ml streptomycin (Cellgro) and dissociated using 0.2% collagenase from Clostridium histolyticum (Sigma) at 37°C for 4 hours (tails) or 2 hours (embryos) with continuous shaking. Dissociated cells were filtered through a 70 μm cell strainer (BD falcon) and washed two times with DMEM containing 10% FBS and plated on 0.1% gelatin (Cellgro) coated tissue culture plates. Fibroblasts were cultured in DMEM containing 10% FBS, 1 mM L-Glutamine, 100 U/ml penicillin and 100 μg/ml streptomycin and media was replaced daily. After passaging two times, fibroblasts were frozen in media containing 10% DMSO, 40% FBS and 50% DMEM. Generally, passage two (P2) fibroblasts were used for the reprogramming experiments, unless indicated otherwise. Genotype was confirmed by PCR. Primers for wildtype, F: 5'-GGTCCCAGTCTGAGATGTA, R: 5'-CAACGTGGCGTCCAAACAGGC; for knockout, F: 5'-CTGCCAAACCTGATGTCTTGA, R: 5'-AACCAAGCCTATGCCTACAGC.

Production of lentiviruses and retroviruses

Reprogramming virus was produced by transfection of a single lentiviral stem cell cassette (STEMCCA) in HEK293T cells using polyethylenimine (PEI) (Polysciences). STEMCCA (constitutive) expresses all four human factors (OCT4, KLF4, SOX2, cMYC), using a single Elongation Factor 1α (EF1α) promoter within a lentiviral vector using a combination of 2A peptide and internal ribosome entry site (IRES) strategies. Both constitutive and doxycycline-inducible STEMCCA vectors were used for the reprogramming studies. For the doxycycline-inducible system, fibroblasts were treated with doxycycline (1 μg/ml) starting one day after transduction for 2 weeks. Lentiviruses were produced using a five-plasmid transfection system with STEMCCA transfected together with four expression vectors encoding the packaging proteins Gag-Pol, Rev, Tat, and the G protein of the vesicular stomatitis virus (VSV). Five vectors were incubated in DMEM with 0.06% PEI for 20 minutes and then transferred to the plates containing 80% confluent HEK293T cells. Transfection was performed in FBS (10%) containing media. After 5 hours media was replaced with fresh media. Viral supernatant was collected after 48 hours and spun at 3000 rpm to remove dead cells and filtered through a sterile syringe filter with 0.45 μm polyethersulfone membrane (VWR international). Ectopic expression of wildtype AID or catalytically mutant AID was achieved using retroviral vectors as described previously[25]. Retroviruses were produced by co-transfecting the vector with the packaging plasmid pCL-ECO 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 six hours, 1 ml of viral supernatant containing 8 μg/ml polybrene (Millipore) was added to the fibroblast containing wells. For every experiment, $AID^{-/-}$ and $AID^{+/+}$ fibroblasts were infected at the same time with the identical titer of virus, prepared for each experiment from the same batch. Viral supernatant was removed after 12 hours of infection and fresh media containing DMEM, 10% FBS, 1 mM L-glutamine, 100 U/ml penicillin and 10 μg/ml streptomycin was added. After one 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^{-4} M monothioglycerol (MTG; Sigma), 1 mM L-Glutamine, 100 U/ml penicillin and 10 μg/ml streptomycin. Cells were maintained in the same media. In some of the experiments cells were transferred to mitotically inactivated mouse embryonic fibroblast feeders after two days of transduction. Approximately 30,000 to 50,000 cells were transferred to one well of a feeder-containing 6 well plate. Isolated induced pluripotent stem cell (iPSC) colonies were always cultured on feeders. All the cells were kept at 37°C in a humidified environment at 5% CO_2.

Immuno-staining

Immuno-staining was performed on fixed cells (4% PFA in BBS 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 4C at the

following dilutions: anti-Nanog (eBiosciences 14-5761, 1:100, 5μg/ml) and Anti-Oct4 (Santa Cruz 5279, 1:100, 2 μg/ml). Cells were washed and blocked in BBT-BSA and then incubated with Alexa-conjugated secondary antibodies (1:500, from Molecular Probes). Vectashield-DAPI was used as 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 hour and then stained in suspension. SSEA1 (Santa Cruz, 1:100) staining was performed on unfixed cells. Cells were analyzed on a BD-Accuri C6 flow cytometer (BD Biosciences) using CFlow Plus software.

RT-qPCR

Cells were trypsinized and collected in Trizol reagent (Life technologies). Total RNA extraction was done using the RNeasy Mini Kit (Qiagen). The cDNA was synthesized using the SuperScript III First-Strand Synthesis System (Invitrogen). Quantitative RT-PCR (qPCR) was performed on three experimental replicates using SYBR green Master I (Roche). Data was generated on a LightCycler 480 II (Roche) and analyzed 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 ST1).

In vitro differentiation

For embryoid body (EB) formation iPSCs were passaged two times on feeder free gelatin coated culture dishes. 5×10^5 iPSCs were plated on a 10 cm low attachment dish. EBs were cultured for 6 days in MES media without LIF, and media was changed every other day. For mesodermal differentiation EBs were replated on gelatin-coated dishes and cultured for another 7 days in the same media. For neural differentiation, EBs were cultured in media containing DMEM F-12 (Cellgro), 0.5% N2 (Gibco) and 0.5% B27 (Gibco) supplements, 1 mM L-glutamine, 1% non essential amino acids (Gibco) and 1.5×10^{-4} M monothioglycerol for another 3 days, after which EBs were seeded onto gelatin coated dishes with the same media, plus 10 μM retinoic acid (Sigma) for another 4 days. For endoderm differentiation, EBs were replated onto gelatin-coated dishes in MES media lacking LIF but containing 0.5% FBS and 50 ng/ml Activin (R&D) for seven days. For the all differentiation cultures, media was changed daily.

ESC derivation from blastcyst stage mouse embryos

AID$^{-/-}$ males and *AID*$^{-/-}$ females (or congenic wildtype pairs as controls) were time mated and E3.5 blastocyst staged embryos were collected in M2 media according to standard protocols[26]. Embryos were placed on mitotically inactivated mouse embryo fibroblasts (MEFs[27]) in knockout DMEM (Gibco) containing 15% KSR (knockout serum replacement, Gibco), 2 mM L-Glutamine, 0.1 mM β-mercaptoethanol, 0.1 mM non-essential amino acids, 1 mM sodium pyruvate, 100 U/ml penicillin and 100 μg/ml streptomycin, 1000 U/ml 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 penicillin and 100 μg/ml streptomycin, 1000 U/ml 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 ESC-like colonies were observed and passaged at 70% confluency. Stable ESC lines were validated by passaging a minimum of 10 times without loss of pluripotent morphology or marker expression.

Genomic analysis

RRBS was performed as described[22] and bisulfite reads were aligned to the bisulfite-converted mm9 using Bismark[28]. Differentially methylated CpGs were identified using the Fisher exact test with correction for Benjamini-Hochberg multiple testing. We defined differentially methylated regions (DMRs) as regions containing at least 5 differentially methylated CpGs, where contiguous differentially methylated CpGs are separated by 250 bp or less, and for which the total methylation change between wildtype 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 = 4081 bp. For the lower number of hypomethylated DMRs, we observed the following statistics: average = 229 bp, median = 188 bp, minimum = 19 bp, maximum = 1250 bp. Paired-end RNA-seq libraries were constructed as previously described[29] 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. 2-fold 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 RBBS data is provided in Supplemental Table ST3.

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, San Diego, CA) as previously described[30]. Briefly 1 μg of DNA was treated with sodium bisulfite using the EZ methylation kit (Zymo-Research, Irvine, CA). 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 analyzer. Amplicons probed are given in Supplemental Table ST2.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Gustavo Mostoslavsky for the gift of reprogramming vectors. Ari Melnick provided advice and consultation. We are grateful to Tasuku Honjo for knockout AID mice, and Nina Papavasiliou for assistance in designing genotyping PCR primers. The Epigenomics Core Facility of Weill Cornell Medical College carried out the bisulfite sequencing and provided consultation. This study was supported by NIH grant HL056182 (TE), and NSF CAREER grant 1054964 (OE).

References

1. Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell. 2000; 102:553–563. [PubMed: 11007474]
2. Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E, Alt FW. Transcriptiontargeted DNA deamination by the AID antibody diversification enzyme. Nature. 2003; 422:726–730. [PubMed: 12692563]
3. Bhutani N, Burns DM, Blau HM. DNA demethylation dynamics. Cell. 2011; 146:866–872. [PubMed: 21925312]
4. Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, et al. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature. 2010; 463:1101–1105. [PubMed: 20098412]
5. Bhutani N, Brady JJ, Damian M, Sacco A, Corbel SY, Blau HM. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature. 2010; 463:1042–1047. [PubMed: 20027182]
6. Foshay KM, Looney TJ, Chari S, Mao FF, Lee JH, Zhang L, et al. Embryonic stem cells induce pluripotency in somatic cell fusion through biphasic reprogramming. Mol Cell. 2012; 46:159–170. [PubMed: 22445485]
7. Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, et al. A molecular roadmap of reprogramming somatic cells into iPS cells. Cell. 2012; 151:1617–1632. [PubMed: 23260147]
8. Buganim Y, Faddah DA, Cheng AW, Itskovich E, Markoulaki S, Ganz K, et al. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell. 2012; 150:1209–1222. [PubMed: 22980981]
9. Meissner A. Epigenetic modifications in pluripotent and differentiated cells. Nat Biotechnol. 2010; 28:1079–1088. [PubMed: 20944600]
10. Kim K, Doi A, Wen B, Ng K, Zhao R, Cahan P, et al. Epigenetic memory in induced pluripotent stem cells. Nature. 2010; 467:285–290. [PubMed: 20644535]
11. Polo JM, Liu S, Figueroa ME, Kulalert W, Eminli S, Tan KY, et al. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nat Biotechnol. 2010; 28:848–855. [PubMed: 20644536]
12. Sommer CA, Stadtfeld M, Murphy GJ, Hochedlinger K, Kotton DN, Mostoslavsky G. Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. Stem Cells. 2009; 27:543–549. [PubMed: 19096035]
13. Smith ZD, Nachman I, Regev A, Meissner A. Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. Nat Biotechnol. 2010; 28:521–526. [PubMed: 20436460]
14. Brambrink T, Foreman R, Welstead GG, Lengner CJ, Wernig M, Suh H, et al. Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. Cell Stem Cell. 2008; 2:151–159. [PubMed: 18371436]
15. Sun Y, Peng I, Senger K, Hamidzadeh K, Reichelt M, Baca M, et al. Critical role of activation induced cytidine deaminase in experimental autoimmune encephalomyelitis. Autoimmunity. 2013; 46:157–167. [PubMed: 23167594]
16. Conticello SG. The AID/APOBEC family of nucleic acid mutators. Genome Biol. 2008; 9:229. [PubMed: 18598372]

17. Bhutani N, Decker MN, Brady JJ, Bussat RT, Burns DM, Corbel SY, et al. A critical role for AID in the initiation of reprogramming to induced pluripotent stem cells. Faseb J. 2012
18. Robinton DA, Daley GQ. The promise of induced pluripotent stem cells in research and therapy. Nature. 2012; 481:295–305. [PubMed: 22258608]
19. Gore A, Li Z, Fung HL, Young JE, Agarwal S, Antosiewicz-Bourget J, et al. Somatic coding mutations in human induced pluripotent stem cells. Nature. 2011; 471:63–67. [PubMed: 21368825]
20. Hussein SM, Batada NN, Vuoristo S, Ching RW, Autio R, Narva E, et al. Copy number variation and selection during reprogramming to pluripotency. Nature. 2011; 471:58–62. [PubMed: 21368824]
21. Mayshar Y, Ben-David U, Lavon N, Biancotti JC, Yakir B, Clark AT, et al. Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. Cell Stem Cell. 2010; 7:521–531. [PubMed: 20887957]
22. Akalin A, Garrett-Bakelman FE, Kormaksson M, Busuttil J, Zhang L, Khrebtukova I, et al. Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. PLoS Genet. 2012; 8:e1002781. [PubMed: 22737091]
23. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011; 27:1571–1572. [PubMed: 21493656]
24. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008; 5:621–628. [PubMed: 18516045]

Methods References

25. Vuong BQ, Lee M, Kabir S, Irimia C, Macchiarulo S, McKnight GS, et al. Specific recruitment of protein kinase A to the immunoglobulin locus regulates class-switch recombination. Nat Immunol. 2009; 10:420–426. [PubMed: 19234474]
26. Nagy, A.; Gertsenstein, M.; Vintersten, K.; Behringer, R. Manipulating the mouse embryo: a laboratory manual. 3rd edn. Cold Spring Harbor Press; 2003.
27. Joyner, AL. Gene targeting: a practical approach. 2nd edn. Oxford University Press; 2000.
28. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011; 27:1571–1572. [PubMed: 21493656]
29. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008; 5:621–628. [PubMed: 18516045]
30. Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G, et al. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci U S A. 2005; 102:15785–15790. [PubMed: 16243968]

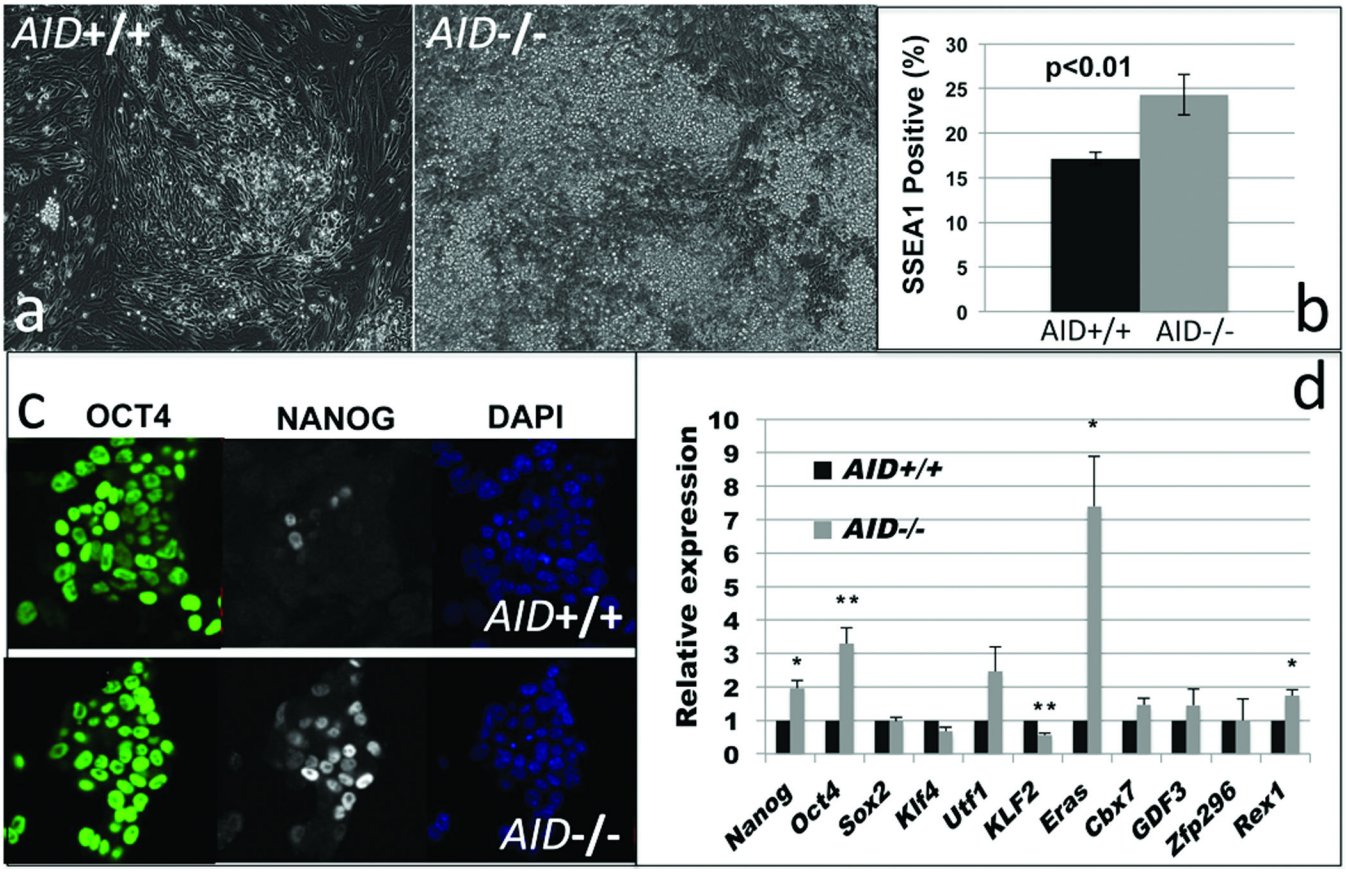

Figure 1. Cells lacking AID are initially hyper-responsive to TF-based reprogramming

a, $AID^{+/+}$ and $AID^{-/-}$ fibroblasts after 4 days of OKSM transduction. Note the rounded appearance of more null cells. **b,** Cells positive for SSEA1 after 4 days of OKSM transduction, by flow cytometry. **c,** Immuno-staining with anti-OCT4 and anti-NANOG antibodies after 1 week of OKSM transduction. **d,** Relative transcript levels of pluripotency genes 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 from three independent experiments (*p<0.05, **p<0.01).

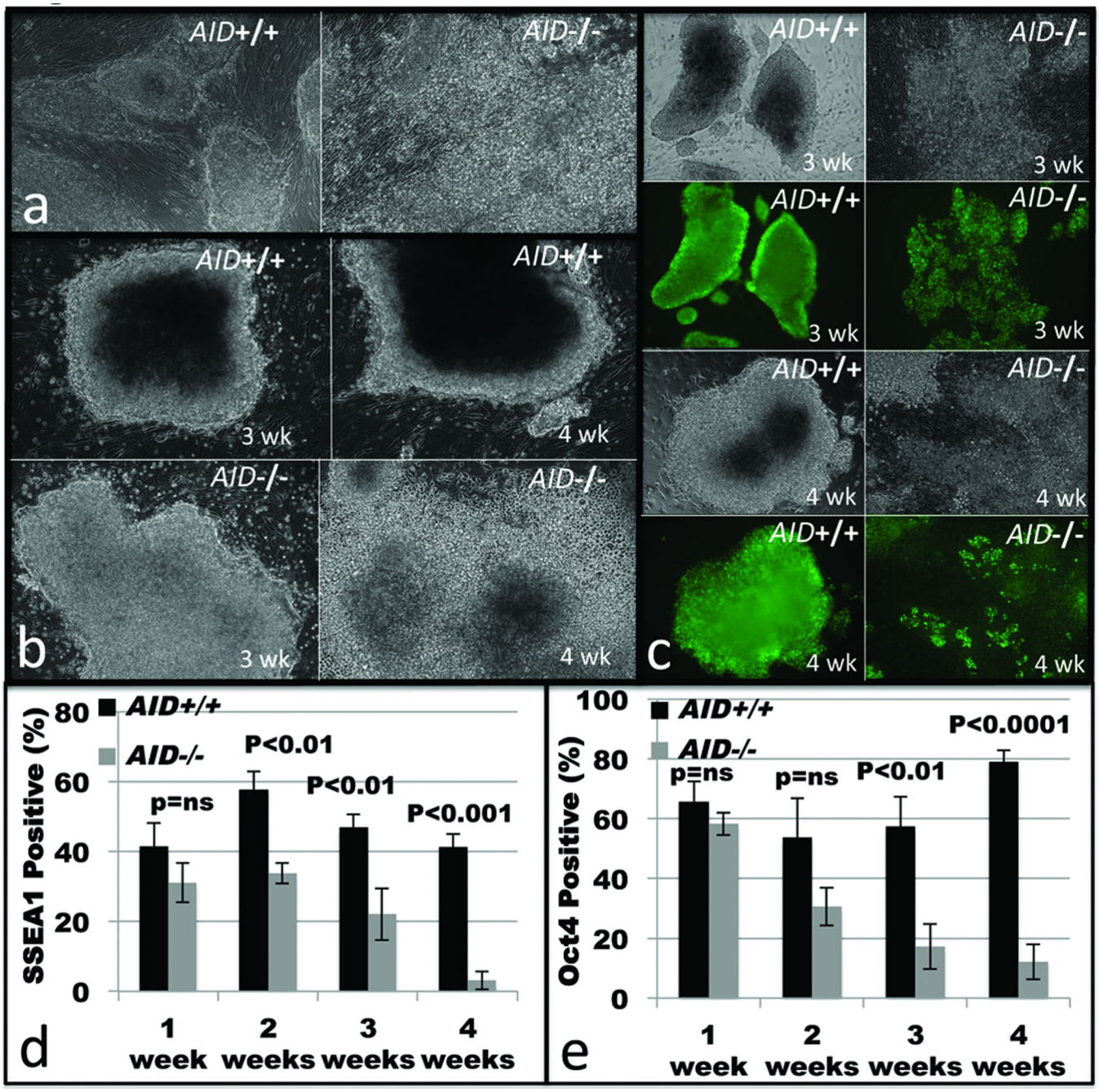

Figure 2. AID stabilizes pluripotency

a, *AID*$^{-/-}$ cells lose ESC-like morphological characteristics four weeks after OKSM transduction. **b**, *AID*$^{-/-}$ iPSC-like colonies progressively lose this phenotype. **c**, The mutant cells lose NANOG expression. **d,** Flow cytometry measured cells that stain positive for SSEA1 or **e,** OCT4 after 1, 2, 3 or 4 weeks of OKSM transduction; n=3 independent experiments, error bars denote standard deviation.

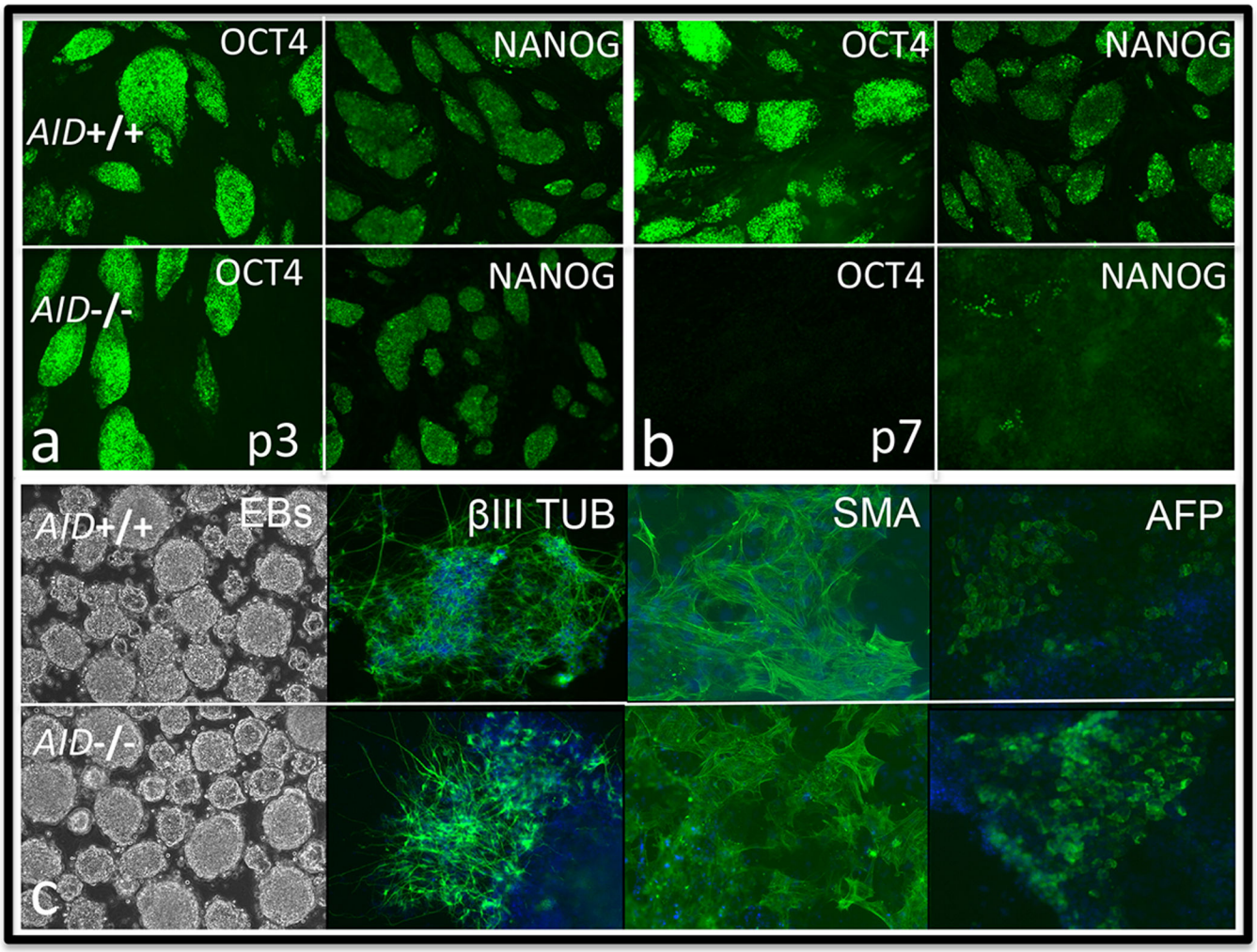

Figure 3. Cells lacking AID reprogram inefficiently, but those that do are pluripotent

a, Immuno-staining of iPSC colonies isolated 3 weeks after OKSM transduction and passaged 3 times or **b,** 7 times, before staining. In this example, the null cells failed to from stable iPSCs. **c,** Stable iPSC clones were passaged over 10 times. In this case (unlike in b) the passaged *AID*-null cells maintained NANOG expression (not shown) and generate EBs that differentiate into cells positive for smooth muscle actin (SMA), beta-III tubulin (βIII TUB) and alpha-fetoprotein (AFP).

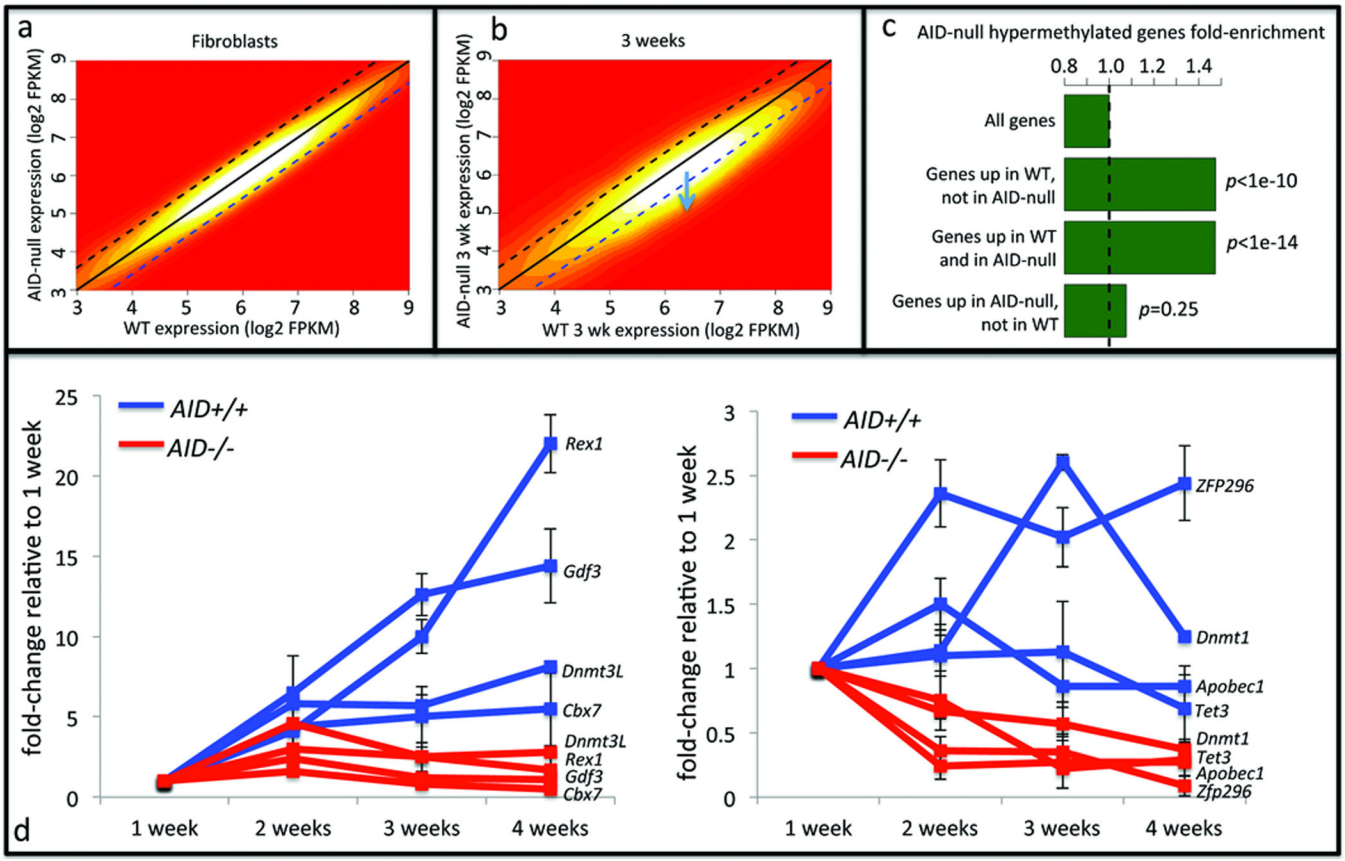

Figure 4. Cells lacking AID fail to activate expression of hyper-methylated pluripotency genes

a, In starting fibroblasts few differentially-expressed genes are off the diagonal. **b,** Many genes are under-expressed in *AID*-null cells 3 weeks following transduction. **c,** Fold-enrichment for hyper-methylated genes in *AID*-null cells. All p-values were calculated using the hypergeometric distribution. **d,** Transcript levels for hyper-methylated pluripotency genes during reprogramming comparing wildtype cells (blue lines) and *AID*-null cells (red lines). Data represent the mean +/− standard error from three independent experiments.