HHS Public Access

Author manuscript

Nat Immunol. Author manuscript; available in PMC 2014 January 01.

Published in final edited form as:

Nat Immunol. 2013 July ; 14(7): 749–755. doi:10.1038/ni.2616.

A comprehensive analysis of AID's effects on the transcriptome and methylome of activated B cells

Eric L. Fritz[1,2], **Brad R. Rosenberg**[1,3], **Kenneth Lay**[1,2], **Aleksandra Mihailović**[4], **Thomas Tuschl**[4], and **F. Nina Papavasiliou**[1]

[1]Laboratory of Lymphocyte Biology, The Rockefeller University, 1230 York Avenue, New York, NY, 10065

[2]The Rockefeller Graduate Program, The Rockefeller University, 1230 York Avenue, New York, NY, 10065

[3]John C. Whitehead Presidential Fellows Program, The Rockefeller University, 1230 York Avenue, New York, NY, 10065

[4]Laboratory of RNA Molecular Biology, Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, NY, 10065

Abstract

Beyond its well-characterized functions in antibody diversification, AID can catalyze off-target DNA damage and has been hypothesized to edit RNA and mediate DNA demethylation. To comprehensively examine AID's effects on the transcriptome and DNA methylome, we performed RNA-Seq and reduced-representation bisulfite sequencing (RRBS) on *Aicda*$^{-/-}$, wild-type and AID-overexpressing activated B cells. These analyses confirmed AID's known role in immunoglobulin isotype switching, while also demonstrating that it has little other effect on gene expression. Additionally, no evidence of AID-dependent mRNA or miRNA editing could be detected. Finally, RRBS data failed to support a role for AID in the regulation of DNA methylation. Thus, despite evidence of its additional activities in other systems, antibody diversification appears to be AID's sole physiological function in activated B cells.

Germinal center B cells undergo two processes of immunoglobulin diversification: somatic hypermutation (SHM), in which mutations are introduced into immunoglobulin (Ig) genes, and class-switch recombination (CSR), in which genomic constant regions are recombined to encode antibodies of different isotypes. Both of these processes require activation-induced cytidine deaminase (AID)[1,2], which generates C-to-U lesions at the Ig loci that are

Correspondence should be addressed to F.N.P. (papavasiliou@rockefeller.edu) or E.L.F.(efritz@rockefeller.edu)..

Author contributions E.L.F., B.R.R. and F.N.P designed experiments and analyses; E.L.F. and K.L. cultured cells; K.L. generated VSG spikes; A.M. prepared miRNA-Seq libraries under the supervision of T.T.; E.L.F. performed all other experiments and analyzed the data; E.L.F., B.R.R. and F.N.P. wrote the manuscript.

Competing financial interests The authors declare no competing financial interests.

Accession codes SRA: raw mRNA-Seq and trimmed miRNA-Seq and RRBS reads, SRP020240.

resolved to generate point mutations or double-stranded DNA breaks in the cases of SHM and CSR, respectively (reviewed in[3]).

Despite over a decade of intense study, a number of outstanding questions regarding AID remain. The first regards its range of substrates. Based on its homology to the well-characterized RNA editing enzyme APOBEC1, AID was initially thought to deaminate an RNA substrate[1]. Since then, roles for AID in CSR and SHM have been convincingly shown to proceed by a DNA-editing mechanism (reviewed in[4]), and no AID activity has been detected on RNA substrates *in vitro*[5–7]. However, AID has been shown to bind mRNA[7,8], to modulate translation[9], and to edit the genome of hepatitis B virus[10], leading to the hypothesis that it could edit RNA *in vivo* in conjunction with an unknown sequence motif or cofactor.

Additionally, AID has been shown to act on non-immunoglobulin loci in B cells[11,12]. These off-target events can result in the mutation of oncogenes and tumor suppressors genes[13,14], and have been implicated in oncogenic chromosomal translocations in B cell lymphomas[15–19]. AID expression has also been reported in a number of different non-B cell tumor types, where it has been hypothesized to contribute to malignant progression[20]. In support of such off-target activities, ChIP-Seq experiments indicated AID occupancy at a large number of promoters beyond the Ig loci[21]. However, the transcriptional consequences of these off-target activities for the typical B cell have not been determined.

Furthermore, numerous experiments in a variety of organisms have suggested that AID might also play a role in active DNA demethylation and corresponding gene regulation[22–28]. Because AID is active on 5-methylcytosine (5mC) in DNA[22], it has been hypothesized that, following AID-catalyzed 5mC:T deamination, active demethylation could occur via repair of the resulting T:G mismatch through error-free DNA repair[29], a model supported by reports linking AID to DNA demethylation in a variety of systems[23–26,30]. However, these findings are puzzling in light of the low abundance of AID expressed in many of these systems and biochemical preference of AID for C over 5mC as a substrate[31,32]. Additionally, it is clear that AID is not absolutely required for active demethylation at least during development as indicated by the viability and fertility of *Aicda*$^{-/-}$ mice[1].

To answer the outstanding questions regarding AID activity on RNA and its capacity to mediate DNA demethylation, we sought to systematically characterize the effects of AID by performing a combined transcriptome and methylome analysis of AID-deficient, wild-type and AID-overexpressing activated murine B cells. These analyses clearly demonstrated AID activity at the Ig loci, but did not reveal additional AID-dependent transcriptional differences. Furthermore, we were unable to detect AID-dependent RNA editing activity, strongly suggesting that, despite initial models, AID does not act as an RNA editor in B cells. Finally, our analysis of the activated B cell methylome revealed no DNA methylation differences attributable to AID, even when AID was overexpressed. Overall, our data suggest that, at least within B cells, AID acts near-exclusively in its well-described role in Ig diversification.

Results

AID has little effect on gene expression

To determine the effects of AID on the B cell transcriptome, we performed 100-nt, single-end RNA-Seq on polyadenylated RNA from naïve splenic B cells of *Aicda*$^{-/-}$, wild-type and AID–miR-155T transgenic mice[33], which contain an AID-GFP fusion gene that lacks the miR155 target sequence in its 3′UTR, resulting in AID overexpression. Isolated B cells were stimulated *ex vivo* with lipopolysaccharide (LPS), interleukin 4 (IL-4), and CD40-specific antibody for 72 h. These conditions induce strong expression of AID and appreciable frequencies of CSR to IgG1 (14% for WT, 31% for AID–miR-155T) (Supplementary Fig. 1). We prepared RNA-Seq libraries using a standard protocol with the addition of two sets of "spike in" control transcripts: the ERCC panel of RNAs[34] (to ascertain the lower limit of detection for gene expression), and 5 "pre-edited" RNAs derived from *Trypanosoma brucei* variant surface glycoprotein (VSG) genes with a single C-to-T change introduced at a frequency of 50% (to serve as a positive control for the detection of RNA editing).

To determine the lower bound of accurate transcript measurement, the abundance of ERCC spikes was calculated. Estimates of abundance were extremely well correlated for each pair of samples for transcripts with > 5 fragments per kilobase of transcript per million reads mapped (FPKM – a relative measure of transcript abundance) (Supplementary Fig. 2). This analysis demonstrates that even poorly expressed transcripts are well covered in these RNA-Seq datasets.

Overall, the gene expression profiles were extremely similar for the three samples, with a Pearson correlation coefficient > 0.995 for each pairwise comparison (Fig. 1a and Supplementary Fig. 3a,b). Within these expression profiles, AID itself was clearly expressed at the expected level for each dataset. But besides AID, the short list of genes that display > 2 fold-change differences between conditions (Supplementary Table 1) was composed of elements predominantly annotated as pseudogenes in other references, or elements with RNA-Seq coverage characteristic of mismapped reads derived from paralogous transcripts. In addition, few transcripts that displayed a > 2 fold difference did so for more than one of the binary comparisons, suggesting that these apparent differences were the result of noise rather than authentic AID-dependent effects. These findings are concordant with those from a lower-depth RNA-Seq comparison of *Aicda*$^{-/-}$ and wild-type under slightly different conditions (36 nt paired-end sequencing, and IL-4 plus anti-CD-40 stimulation) (Supplementary Fig. 3c).

We also used the RNA-Seq data to estimate the abundance of different transcript isoforms. Expression analysis at the isoform level shows a high degree of similarity between *Aicda*$^{-/-}$, wild-type and AID–miR-155T samples, with the exception of Ig transcripts (Fig. 1b and Supplementary Fig. 3d,e). Again, a high degree of similarity was also found in the lower-depth RNA-Seq runs (Supplementary Fig. 3f). While these results do not exclude the possibility of AID-dependent changes in gene expression in B cells, they demonstrate that if such changes exist outside the *Igh* locus, they are likely too small to be physiologically relevant.

The notable exception to this trend was transcripts derived from the *Igh* locus. By including a manually curated annotation for the *Igh* locus (Supplementary Table 2), we were able to estimate relative frequencies of CSR by comparing the abundance of J_H-C spliced transcripts between samples for each Ig isotype. As expected, nearly all J_H-C spliced transcripts for the *Aicda*$^{-/-}$ sample were IgM or IgD, while the wild-type and AID–miR-155T samples both displayed appreciable amounts of IgE, IgG1, IgG2a, and IgG3-derived transcripts (Fig.1c). Furthermore, apparent CSR frequencies were higher for AID–miR-155T than for wild-type for each secondary isotype, consistent with previous reports[33].

To determine how CSR frequencies quantified by RNA-Seq analysis compare to standard measurements, we compared the fraction of IgG1$^+$ cells as determined by flow cytometry and the abundance of J_H-$C_{\gamma 1}$ transcripts as a fraction of all J_H-C transcripts (Supplementary Fig. 3g). A clear linear relationship was observed, albeit with slope not equal to 1, which likely represents differing per-cell levels and/or sequencing efficiencies of different isotype transcripts. Determining these correction factors for each isotype should allow RNA-Seq to be used as quantitative tool for assaying absolute frequencies of CSR.

AID has no effect on V_H segment usage

AID deficiency has been associated with autoimmune disease and a skewed usage pattern of V_H segments in both humans[35] and mice[36]. We sought to determine whether the influence of AID on V_H gene usage is apparent in our RNA-Seq data sets. Because our data was derived from naïve B cells non-specifically stimulated to undergo CSR *ex vivo*, any differences in V_H gene usage between samples ought to mirror the *in vivo* repertoire prior to affinity maturation.

We used our RNA-Seq data to derive FPKM values for the entire set of V_H segments; these provided a measure of relative segment usage between samples. This transcript abundance analysis revealed only minor differences in V_H transcript abundance between *Aicda*$^{-/-}$, wild-type and AID–miR-155T samples. Importantly, none of these exhibited a clear relationship with AID expression (Supplementary Table 3), strongly suggesting that the pattern of V_H usage was unaffected by AID. This result is in contrast to the pattern of V_H usage in newly emigrant B cells in AID-deficient humans[21,35], suggesting that the influence of AID over V_H usage occurs after establishment of the primary repertoire and probably reflects the dynamics between AID-mediated affinity maturation and B cell survival, rather than a role in early B cell development.

AID-dependent RNA editing is not detectable

We have previously demonstrated the utility of paired wild-type and specific deaminase-deficient comparative RNA-Seq for identifying RNA editing events[37]. Because AID has long been hypothesized to act on RNA[1,8], we applied a similar analysis workflow to our B cell RNA-Seq data. Single nucleotide variants (SNVs) were considered candidate editing sites if they had greater than 30× read coverage with at least 20% apparent C-to-T editing, were a minimum distance from a non-C-to-T SNV (1 kb if using reference, 10 kb if not), were not significantly strand-biased, were not located in regions that were not isogenic between the mice used, and did not occur in the *Aicda*$^{-/-}$ sample. Parallel analyses were

performed for positions within reference exons (to achieve the lowest possible background) and for all positions (to include positions in transcripts not found in the reference annotation).

To estimate the fraction of the transcriptome covered by this technique, the amount of coverage by base was computed for ERCC spikes with various FPKM values (Fig. 2a). More than 90% of bases of the transcripts with FPKM ≥ 10 had at least 30× coverage, while those with FPKM = 8 had about 60% of bases with sufficient depth to be interrogated for editing by our strategy. For the B cell RNA-Seq datasets, the top ~6600 transcripts have FPKM ≥ 10 and are thus well covered by the RNA editing analysis. Additionally, the editing event in the exogenous "pre-edited" VSG RNA with FPKM = 23 was detected, but that with FPKM = 7 was not. Thus this analysis strategy is appropriate for identifying editing events in moderately to highly expressed transcripts.

To verify that this strategy can effectively identify RNA editing events, the pipeline was applied to sets of RNA-Seq data from macrophages lacking the *bona fide* RNA editor *Apobec1* and a paired wild-type control (Hamilton *et al*., manuscript in preparation). Many unique C-to-T events were found in the wild-type sample that were absent in the *Apobec1*$^{-/-}$ (Fig. 2b). When the reciprocal comparison was made, very few appropriate SNVs occurred in the *Apobec1*$^{-/-}$ sample that were absent in the wild-type. The small number of such events, which are likely due to mismapping or genomic differences between the mice from which the cells were derived, demonstrates the low background of the technique. The ratio of unique editing events in the deaminase-negative sample to those in the deaminase-positive sample yields an implied false positive rate (IFPR) for identification of the candidate editing events.

Application of the editing site identification pipeline to the B cell RNA-Seq data revealed very few candidate editing sites in the AID–miR-155T and wild-type samples that were absent in the *Aicda*$^{-/-}$ sample (Fig. 2c). Approximately equal numbers of candidate sites were found in the reciprocal comparison, resulting in IFPRs of > 75% for each condition. In contrast, the IFPR for Apobec1-dependent editing in macrophages was 7–8%. RNA-Seq mapping data for the few candidate AID-dependent editing events were visually inspected, and in each case these events were adjudged to be false positives because of one of the following criteria: near-threshold distance from non-C-to-T SNVs, C-to-T mismatches also occurring in the *Aicda*$^{-/-}$ sample, or complete absence of apparent editing in the AID–miR-155T sample. Thus, while we can clearly detect Apobec1-catalyzed mRNA editing using this method, we failed to detect AID-dependent editing of moderately or highly expressed polyadenylated RNAs.

AID has also been implicated in the mutation of miRNA genes[17] as well as the editing of miRNAs themselves[9,38]. We therefore sought to characterize the effects of AID deficiency on the miRNAome. Small RNA-Seq was performed on samples from *Aicda*$^{-/-}$ naïve B cells stimulated in culture and retrovirally transduced to overexpress AID and GFP or GFP alone for 24 or 48 h. For both time points, miRNA abundance was well correlated between the samples (Pearson correlation coefficient > 0.95) (Fig. 3a). Thus AID appears to have no effect on mature miRNA abundance.

To determine whether AID acts as an editor of miRNAs, we employed an RNA editing detection strategy similar to that used for mRNAs. Briefly, positions with at least 10% mismatches and at least 10× depth were identified and classified by type of base change. C-to-T changes were among the least common base alterations observed (Fig. 3b). For each base change type, including C-to-T, the frequency observed was approximately the same for each sample. These likely represent sequencing error, mismapping, or other types of RNA editing that do not depend on AID. For instance, A-to-G alterations constitute the most common change observed for all samples, and are possibly the results of ADAR-catalyzed editing[39–42].

To look more specifically at potential AID-dependent editing, unique mismatched sites were analyzed. These were defined as events that did not occur in the corresponding sample at the same time point while using a lower depth threshold. Again, C-to-T sites were among the least frequent (Fig. 3c). Only one C-to-T site for each of the AID-overexpressing samples fulfilled these criteria. These sites do not appear to be authentic RNA editing as they have insufficient depth in the GFP-overexpressing sample to compare, and they occur at nearly 100%, suggesting a genomic SNV or mismapping. Thus, if AID-catalyzed editing of miRNAs occurs, it does so at a rate that is well below the background of the sequencing protocol, and is therefore unlikely to be physiologically significant.

AID-dependent DNA methylation changes are undetectable by RRBS

Given the numerous results suggesting a role for AID in DNA demethylation, we sought to look for differences in the methylomes of *Aicda*$^{-/-}$, wild-type and AID–miR-155T B cells. To obtain genome-scale methylation data while minimizing the level of uninformative sequencing, we used reduced-representation bisulfite sequencing (RRBS)[43]. By performing an MspI digest and size selection before library preparation and bisulfite conversion, RRBS allows sequencing coverage to be focused on the CpG-rich, genic portion of the genome. Our coverage data reflects this enrichment (64% of promoters and 85% of CpG islands with ≥ 100 individual CpG measurements), despite low overall genomic coverage (2% of 1 kb windows with ≥ 100 individual CpG measurements) (Supplementary Fig. 4). Overall, nearly 950,000 CpGs had at least 10× coverage in all 3 samples and were used for further analysis.

To determine whether AID has a gross effect on the B cell DNA methylome, the distribution of methylation frequency for various genomic features was compared for the three AID genotypes. For each genomic feature type analyzed (1 kb windows, individual CpGs, CpG islands and promoters), there was no apparent difference in DNA methylation distributions associated with AID expression (Fig. 4). The mean methylation frequency for each set of features was highly similar for each genotype. For each feature type, the expected bimodal distribution of methylation was observed, with proportionally more 1 kb windows near-fully methylated than the other feature types.

To assess more subtle differences in DNA methylation by AID expression, methylation frequencies of individual features for each pair of samples were compared. For each feature set analyzed, methylation frequencies between samples were very strongly correlated (Fig. 5a,b and Supplementary Fig. 5a–d). For 1 kb windows, a Pearson r = 0.997–0.998 was

observed for each pair; in comparison, the maximum reported *r* for methylation values of 1 kb windows for biological replicates of cells in the hematopoietic lineage is 0.997 (ref.[44]).

The high degree of correlation of methylation values did not exclude the possibility that sampling noise could mask small numbers of true AID-dependent changes in methylation. To determine if this was the case, we sought to independently verify the largest apparent AID-dependent decreases in DNA methylation identified by RRBS. To this end, we used the Sequenom Epityper system to assay regions with > 20% methylation in the *Aicda*$^{-/-}$ sample than in the wild-type as measured by RRBS, along with randomly selected regions with < 10% difference in methylation between samples. High quality data for all 3 samples was obtained for 8 apparently AID-dependent hypermethylated CpGs and 18 similarly methylated CpGs that were also well covered by RRBS.

For the candidate differentially methylated CpGs, the methylation frequencies for the *Aicda*$^{-/-}$ sample as determined by RRBS were generally much higher than as determined by Epityper (Fig. 5c). In contrast, these methods gave similar values for the wild-type and AID–miR-155T samples. The fact that the apparent AID-dependent hypermethylated CpGs fail to validate and yield uniformly lower methylation values when assayed by an independent technique strongly suggests that the RRBS-derived values for these CpGs represent overestimates of the true population mean methylation frequency, due to noise or an artifact of the method. Additionally, the set of CpGs with similar methylation values between samples by RRBS displayed excellent agreement in methylation frequencies as determined by the two methods (Fig. 5d), demonstrating that RRBS as performed here yields accurate methylation values for well-covered CpGs. Taken together, these results suggest that the most extremely hypermethylated CpGs in the *Aicda*$^{-/-}$ sample were not a result of an authentic AID-dependent process.

Finally, in an attempt to locate any subtle but biologically meaningful AID-dependent changes in DNA methylation, the fold-changes in gene expression by RNASeq were compared to the changes in promoter methylation for each pair of samples (Fig. 5e and Supplementary Fig. 5e,f). Each pairwise comparison between genotypes showed that these variables were uncorrelated ($|r| < 0.03$ in all cases), suggesting that the observed modest differences in DNA methylation were not associated with changes in gene expression and therefore were unlikely to be physiologically relevant. As a whole, these results demonstrate that no candidate for consistent AID-dependent loss of methylation in B cells can be identified by RRBS.

Discussion

Numerous studies have proposed roles for AID in RNA editing and DNA demethylation. Here we sought to examine these hypotheses in activated B cells by using genome-wide methods to comprehensively characterize the effects of AID on the transcriptome and methylome. Toward this goal, we have performed mRNA-Seq, miRNA-Seq and RRBS on material from *Aicda*$^{-/-}$, wild-type and AID-overexpressing cells. The resulting data allowed for RNA expression analyses at the gene-, isoform-, and miRNA-levels, detection of RNA

editing via a validated pipeline, and determination of quantitative methylation values for nearly 1 million CpGs.

The capacity of AID to edit RNA *in vivo* has remained an open question for over a decade[1]. The advent of high-throughput RNA sequencing has made possible the comprehensive characterization of enzyme-specific RNA editing[37], finally allowing this issue to be resolved. We employed a bioinformatic pipeline that can successfully identify APOBEC1-dependent RNA editing to query the transcriptome of B cells for AID-dependent RNA editing. Within moderately to well-expressed transcripts, and with a lower limit of detection of 20% editing frequency, we were not able to detect AID-dependent editing. Assuming that low efficiency editing of poorly expressed transcripts is unlikely to be a crucial determinant of a mechanism as efficient as CSR or SHM, our data imply that AID is unlikely to exert its functions in antibody diversity through RNA editing.

Our comparative RNA-Seq data also allowed for analysis of the population-level transcriptional consequences of AID activity. One notable aspect of our findings was the lack of AID-dependent changes in gene expression, despite the known ability of AID to catalyze hypermutation[12,21] and translocations[17,18] at loci throughout the genome in B cells, especially when over-expressed. While certain varieties of AID-dependent DNA damage clearly have severe effects on B cell function[16], the lack of AID-dependent differences in gene expression or DNA methylation that we have observed suggests that such biologically meaningful off-target AID activity at a given locus is rare in absolute terms. Therefore, while AID can act at many places in the genome, at the population level its only observable activities at the transcriptional level are at the Ig loci. We note that none of the analyses presented here are direct measurements of the frequency of off-target AID activity or AID occupancy in the genome *per se*; while measurements of SHM and translocation frequency are theoretically possible from mRNA-Seq data, they are below the limit of detection for the depth of sequencing presented here.

Finally, AID has been proposed to act in DNA demethylation in a variety of systems[23–28]. Despite previous data showing AID occupancy throughout the genome[21] and DNA demethylation concurrent with B cell activation[45], we observed no AID-dependent demethylation events in activated B cells. There are a number of possible explanations for this discrepancy. It is possible that AID does not catalyze DNA demethylation at all in B cells because cofactors required for AID-dependent DNA demethylation are not present in B cells, or that B cells possess some other factor that prevents deleterious demethylation in the presence of AID. Alternatively, AID-dependent demethylation in B cells could be obscured by rapid re-methylation, for example by DNMT1. It is also possible that the lack of a dramatic DNA methylation phenotype in AID-deficient mammals as compared to *D. rerio*[24] could be related to differing deamination efficiencies on 5mC among these species[46]. Regardless, if AID can act as a DNA demethylase, it does so below the limits of detection of our analyses, in an extremely targeted manner at loci not covered by RRBS, or in a cell type-specific manner (and not in *ex vivo* stimulated murine B cells). Overall, the lack of AID-dependent differences in all of our analyses, excepting the case of Ig isotype usage, strongly suggests that the sole function of AID in B cells is its well-characterized role in the initiation of CSR and SHM.

In addition to confirming a narrow role for AID action in B cells, the analyses presented here also demonstrate the broad capabilities of high throughput RNA sequencing for the study of B cell biology. In addition to its standard usage as a gene expression assay, its utility in detecting RNA editing, CSR frequency and V_H segment usage demonstrate its flexibility as a tool for characterizing populations of B cells.

Methods

Mice and cell culture

Mice used for 100-nt mRNA-Seq, RRBS, and small RNA-Seq were 9-week-old males, and those used for 36-nt mRNA-Seq were 6-week-old females. *Aicda*$^{-/-1}$ and AID–miR-155T[33] mice have been previously described. Naïve splenic B cells were purified by negative selection with anti-CD43 magnetic beads (Miltenyi Biotec) and plated at a concentration of 10^6 cells/ml in RPMI with glutamine (Gibco) supplemented with 10% FBS, 1× pen/strep, 2 mM L-glutamine, 50 μM β-mercaptoethanol, 5 ng/ml IL-4 (Sigma),1 μg/ml anti-CD40 (eBiosciences, clone HM40-3), with or without 25 μg/ml LPS (Sigma). For small RNA-Seq, *Aicda*$^{-/-}$ B cells were activated for 1 d and then infected with pMX-AID-IRES-GFP or pMX-GFP + pCL-Eco derived particles. GFP$^+$ cells were sorted 1 d and 2 d after infection. All mice were bred and maintained under specific pathogen-free conditions at the Rockefeller University Animal Care Facility and all procedures involving mice were approved by The Rockefeller University Institutional Animal Care and Use Committee.

mRNA-Seq

RNA was extracted from B cells with Trizol (Invitrogen). Libraries for Illumina sequencing were prepared as previously described[37], with the following modification: poly-A$^+$ RNA was purified using Sera Mag oligo-dT magnetic beads (Thermo). Prior to fragmentation, ERCC spikes (Ambion, 500 nl of 1:100 dilution per sample), and VSG spikes (30–500 pg/sample) were added. Following ligation of Illumina TruSeq adapters, cDNA was separated by agarose gel electrophoresis and fragments with a size of 300–350 bp were excised for analysis.

Single-end 100-nt sequencing was performed on the Illumina HiSeq2000, yielding 74–82 million reads per sample. Reads were mapped to the Ensembl reference transcriptome (release 63) supplemented with *Igh* transcripts with $J_{1–4}$-C and I-C junctions explicitly added for each isotype and to the NCBI37 reference genome supplemented with ERCC and VSG sequences using Tophat v2.0.3 (ref. [47]) with the parameters “--b2-sensitive --no-novel-juncs”. Values for gene expression were calculated with Cuffdiff v.2.0.2 (ref.[48]), using the Ensembl gene set supplemented with *Igh* transcripts and a masking file with annotated mitochondrial genes, rRNA and tRNA genes and pseudogenes. V segment usage was calculated with Cuffdiff using an annotation derived by mapping all mouse V_H segments IMGT[49] to chromosome 12 from NCBI37 with Bowtie v0.12.7[50] and manually curating the resulting hits.

Lower-coverage 36-nt paired-end RNA-Seq was performed by a similar protocol, with the following modifications: no exogenous spikes were used, Illumina PE adapters were used,

sequencing was performed on an Illumina Genome Analyzer IIx, yielding 19–28 million paired-end reads per sample.

Small RNA-Seq

Libraries were prepared as previously described[51]. Two bases were trimmed from the 3′ end, and trimmed reads were aligned with Bowtie v0.12.7 against NCBI37 using the parameters "-l 15 -v 2 --best --strata -m 1". Reads overlapping with sequences annotated in miRbase 18 were quantified with SeqMonk.

RNA editing analysis

Filters described in the text were implemented with custom Python scripts, available upon request. Candidate editing sites were inspected with IGV[52].

Reduced-representation bisulfite sequencing

Genomic DNA was prepared from B cells using the Qiagen DNEasy kit. For each sample, 500 ng DNA was digested with 2 μl MspI for 18 h and purified by phenol/chloroform extraction. Following end repair, adenylation, and ligation of Illumina methylated adapters, products with size 200–350 bp were purified on agarose gels. Bisulfite conversion was performed twice with the Qiagen Epitect kit, and the library was amplified using Pfu Cx hotstart polymerase (Agilent).

Sequencing on the Illumina HiSeq2000 yielded 47–54 million 50-nt reads per sample. Reads were trimmed using TrimGalore v0.2.2 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) 3 times sequentially with the parameters "-a AGATCGGAAGAGC", then "-s 8 -a CGGTTCAG", then "-s 8 AGCAGGAA". Trimmed reads were aligned to the mouse genome (NCBI37) using Bismark[53] v0.7.4 with Bowtie[50] v0.12.7 with parameters "-l 20". Analysis was performed using Seqmonk (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/). Each position was used for analysis if it was covered at least 10x for all of the samples, and features were used if they contained at least 3 such positions. Features were defined using the Ensembl reference. Promoters were defined as −5 kb to +1 kb from the TSS.

Epityper assay

Primers were designed using the Epidesigner tool (http://www.epidesigner.com/). Epityper assays were performed by the Weill Cornell Medical College Epigenomics Core.

Statistical analysis

All statistical analysis was performed using R software[54]. Graphics were generated using the ggplot2 package[55].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank all members of the Papavasiliou lab for helpful discussion, C. Hamilton and D. Schulz for comments on the manuscript, C. Mason (Weill Cornell Medical College) for providing ERCC spikes, G. Hovel-Miner for providing VSG constructs, K. Velinzon for cell sorting, S. Dewell for sequencing advice, and Y. Li (Weill Cornell Medical College) for assistance with Epityper assays. This work was supported by grants from the NCI (#CA098495) and the Starr Cancer Consortium (#I4-A447). T.T. is a Howard Hughes Medical Institute investigator, and work in his laboratory was supported by a grant from the Starr Cancer Foundation and NIH.

References

1. Muramatsu M, et al. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell. 2000; 102:553–563. [PubMed: 11007474]
2. Revy P, et al. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). Cell. 2000; 102:565–575. [PubMed: 11007475]
3. Gazumyan A, Bothmer A, Klein IA, Nussenzweig MC, McBride KM. Activation-induced cytidine deaminase in antibody diversification and chromosome translocation. Adv. Cancer Res. 2012; 113:167–190. [PubMed: 22429855]
4. Pavri R, Nussenzweig MC. AID targeting in antibody diversity. Adv. Immunol. 2011; 110:1–26. [PubMed: 21762814]
5. Bransteitter R, Pham P, Scharff MD, Goodman MF. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proc. Natl. Acad. Sci. USA. 2003; 100:4102–4107. [PubMed: 12651944]
6. Chaudhuri J, et al. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. Nature. 2003; 422:726–730. [PubMed: 12692563]
7. Dickerson SK, Market E, Besmer E, Papavasiliou FN. AID mediates hypermutation by deaminating single stranded DNA. J. Exp. Med. 2003; 197:1291–1296. [PubMed: 12756266]
8. Nonaka T, et al. Carboxy-terminal domain of AID required for its mRNA complex formation in vivo. Proc. Natl. Acad. Sci. USA. 2009; 106:2747–2751. [PubMed: 19196959]
9. Kobayashi M, et al. AID-induced decrease in topoisomerase 1 induces DNA structural alteration and DNA cleavage for class switch recombination. Proc. Natl. Acad. Sci. USA. 2009; 106:22375–22380. [PubMed: 20018730]
10. Liang G, et al. RNA editing of hepatitis B virus transcripts by activation-induced cytidine deaminase. Proc. Natl. Acad. Sci. USA. 2013 doi:10.1073/pnas.1221921110.
11. Wang CL, Harper RA, Wabl M. Genome-wide somatic hypermutation. Proc. Natl. Acad. Sci. USA. 2004; 101:7352–7356. [PubMed: 15123833]
12. Liu M, et al. Two levels of protection for the B cell genome during somatic hypermutation. Nature. 2008; 451:841–845. [PubMed: 18273020]
13. Shen HM, Peters A, Baron B, Zhu X, Storb U. Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. Science. 1998; 280:1750–1752. [PubMed: 9624052]
14. Pasqualucci L, et al. BCL-6 mutations in normal germinal center B cells: evidence of somatic hypermutation acting outside Ig loci. Proc. Natl. Acad. Sci. USA. 1998; 95:11816–11821. [PubMed: 9751748]
15. Ramiro AR, et al. AID is required for c-myc/IgH chromosome translocations in vivo. Cell. 2004; 118:431–438. [PubMed: 15315756]
16. Robbiani DF, et al. AID Is Required for the Chromosomal Breaks in c-myc that Lead to c-myc/IgH Translocations. Cell. 2008; 135:1028–1038. [PubMed: 19070574]
17. Robbiani DF, et al. AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. Molecular Cell. 2009; 36:631–641. [PubMed: 19941823]

18. Klein IA, et al. Translocation-Capture Sequencing Reveals the Extent and Nature of Chromosomal Rearrangements in B Lymphocytes. Cell. 2011; 147:95–106. [PubMed: 21962510]
19. Chiarle R, et al. Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. Cell. 2011; 147:107–119. [PubMed: 21962511]
20. Pasqualucci L, et al. AID is required for germinal center-derived lymphomagenesis. Nat Genet. 2007; 40:108–112. [PubMed: 18066064]
21. Yamane A, et al. Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. Nat. Immunol. 2010; 12:62–69. [PubMed: 21113164]
22. Morgan HD, Dean W, Coker HA, Reik W, Petersen-Mahrt SK. Activation-induced Cytidine Deaminase Deaminates 5-Methylcytosine in DNA and Is Expressed in Pluripotent Tissues: IMPLICATIONS FOR EPIGENETIC REPROGRAMMING. J. Biol. Chem. 2004; 279:52353–52360. [PubMed: 15448152]
23. Bhutani N, et al. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature. 2010; 463:1042–1047. [PubMed: 20027182]
24. Rai K, et al. DNA Demethylation in Zebrafish Involves the Coupling of a Deaminase, a Glycosylase, and Gadd45. Cell. 2008; 135:1201–1212. [PubMed: 19109892]
25. Popp C, 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]
26. Bhutani N, et al. A critical role for AID in the initiation of reprogramming to induced pluripotent stem cells. FASEB J. 2013; 27:1107–1113. [PubMed: 23212122]
27. Rai K, et al. DNA Demethylase Activity Maintains Intestinal Cells in an Undifferentiated State Following Loss of APC. Cell. 2010; 142:930–942. [PubMed: 20850014]
28. Cortellino S, et al. Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair. Cell. 2011; 146:67–79. [PubMed: 21722948]
29. Fritz EL, Papavasiliou FN. Cytidine deaminases: AIDing DNA demethylation? Genes Dev. 2010; 24:2107–2114. [PubMed: 20889711]
30. Guo JU, Su Y, Zhong C, Ming G-L, Song H. Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain. Cell. 2011:1–22. doi:10.1016/j.cell.2011.03.022.
31. Nabel CS, et al. AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. Nat. Chem. Biol. 2012; 8:751–758. [PubMed: 22772155]
32. Wijesinghe P, Bhagwat AS. Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G. Nucl. Acids Res. 2012; 40:9206–9217. [PubMed: 22798497]
33. Teng G, et al. MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. Immunity. 2008; 28:621–629. [PubMed: 18450484]
34. Jiang L, et al. Synthetic spike-in standards for RNA-seq experiments. Genome Res. 2011; 21:1543–1551. [PubMed: 21816910]
35. Meyers G, et al. Activation-induced cytidine deaminase (AID) is required for B-cell tolerance in humans. Proc. Natl. Acad. Sci. USA. 2011; 108:11554–11559. [PubMed: 21700883]
36. Kuraoka M, et al. Activation-induced cytidine deaminase mediates central tolerance in B cells. Proc. Natl. Acad. Sci. USA. 2011; 108:11560–11565. [PubMed: 21700885]
37. Rosenberg BR, Hamilton CE, Mwangi MM, Dewell S, Papavasiliou FN. Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3' UTRs. Nat. Struct. Mol. Biol. 2011; 18:230–236. [PubMed: 21258325]
38. Kobayashi M, et al. Decrease in topoisomerase I is responsible for activation-induced cytidine deaminase (AID)-dependent somatic hypermutation. Proc. Natl. Acad. Sci. USA. 2011; 108:19305–19310. [PubMed: 22080610]
39. Yang W, et al. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. Nat. Struct. Mol. Biol. 2005; 13:13–21. [PubMed: 16369484]
40. Kawahara Y, et al. Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science. 2007; 315:1137–1140. [PubMed: 17322061]

41. Alon S, et al. Systematic identification of edited microRNAs in the human brain. Genome Res. 2012; 22:1533–1540. [PubMed: 22499667]
42. Vesely C, Tauber S, Sedlazeck FJ, von Haeseler A, Jantsch MF. Adenosine deaminases that act on RNA induce reproducible changes in abundance and sequence of embryonic miRNAs. Genome Res. 2012; 22:1468–1476. [PubMed: 22310477]
43. Meissner A, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. 2008 doi:10.1038/nature07107.
44. Bock C, et al. DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. Molecular Cell. 2012; 47:633–647. [PubMed: 22841485]
45. Shaknovich R, et al. DNA methyltransferase 1 and DNA methylation patterning contribute to germinal center B-cell differentiation. Blood. 2011; 118:3559–3569. [PubMed: 21828137]
46. Abdouni H, et al. Zebrafish AID is capable of deaminating methylated deoxycytidines. Nucl. Acids Res. 2013 doi:10.1093/nar/gkt212.
47. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009; 25:1105–1111. [PubMed: 19289445]
48. Trapnell C, et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature Biotechnol. 2013; 31:46–53. [PubMed: 23222703]
49. Lefranc M-P, et al. IMGT, the international ImMunoGeneTics information system. Nucl. Acids Res. 2009; 37:D1006–12. [PubMed: 18978023]
50. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25. [PubMed: 19261174]
51. Hafner M, et al. Barcoded cDNA library preparation for small RNA profiling by next-generation sequencing. Methods. 2012; 58:164–170. [PubMed: 22885844]
52. Robinson JT, et al. Integrative genomics viewer. Nature Biotechnol. 2011; 29:24–26. [PubMed: 21221095]
53. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011; 27:1571–1572. [PubMed: 21493656]
54. Team, RDC. R: A language and environment for statistical computing. (R Foundation for Statistical Computing, 2012). at <http://www.R-project.org/>
55. Wickham, H. ggplot2: elegant graphics for data analysis. Springer; 2009. at <http://had.co.nz/ggplot2/book>

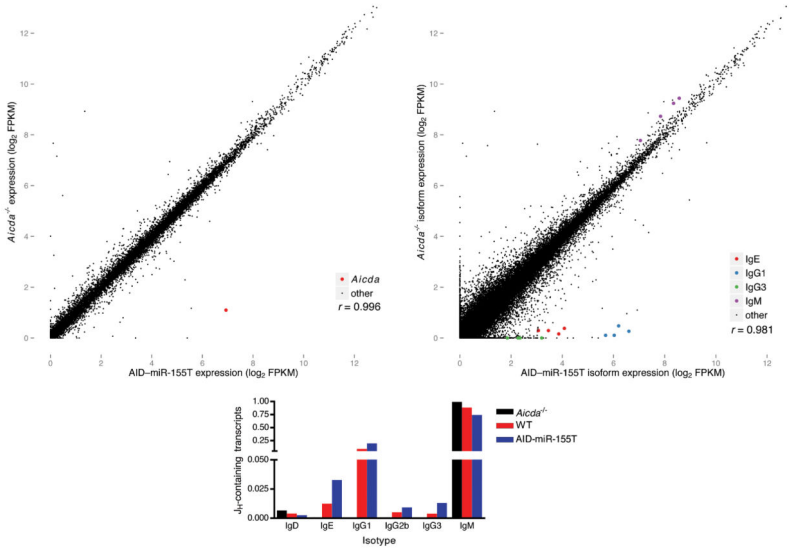

Figure 1. AID-dependent effects on the B cell transcriptome are limited to the *Igh* locus

(**a,b**) Pairwise comparisons of (**a**) gene-level and (**b**) transcript-level expression values from poly-A$^+$ RNA-Seq for *Aicda*$^{-/-}$ and AID-overexpressing activated B cells. *Aicda* is highlighted in (**a**) and *Igh*-derived transcripts are highlighted in (**b**). (**c**) Immunoglobulin isotype frequency estimated by abundance of J_H-C spliced transcripts. (n = 1 mouse per genotype ; *r* = Pearson's correlation coefficient)

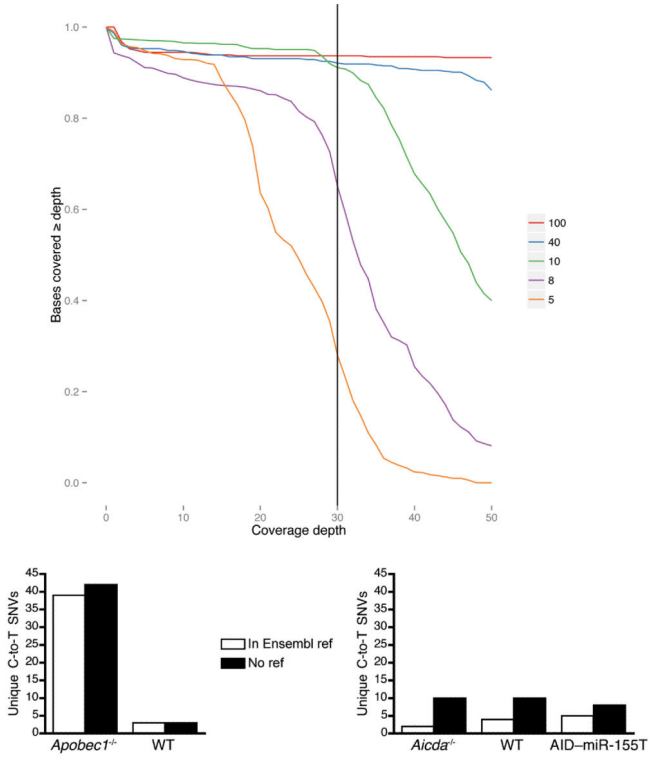

Figure 2. AID-dependent RNA editing events are not detectable by comparative RNA-Seq

(**a**) Cumulative per-base coverage depth for ERCC spike transcripts with various expression values (FPKM), illustrating the portion of the transcriptome accessible to the RNA editing analysis workflow for the AID-overexpressing 100bp mRNA-Seq. (**b,c**) Count of candidate C-to-T editing events for (**b**) wild-type and *Apobec1*$^{-/-}$ macrophages as a positive control for the analyses, and **(c)** *Aicda*$^{-/-}$, wild-type (WT), and AID-overexpressing B cells, by analysis pipeline. (n = 1 mouse per genotype)

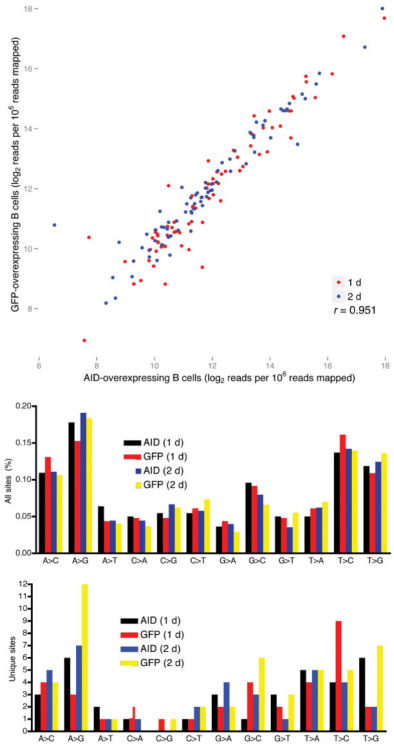

Figure 3. AID overexpression affects neither the abundance nor the editing of small RNAs

(**a**) Abundance of small RNA-Seq reads overlapping annotated miRNAs that make up at least 0.1% of any sample for B cells overexpressing AID or GFP for 1 d or 2 d, normalized to million reads mapped. Points are colored according to time. (**b,c**) Number of (**b**) total and (**c**) unique observed mismatches for each base change type in aligned small RNA-seq reads. (n = 1 mouse per timepoint; *r* = Pearson's correlation coefficient)

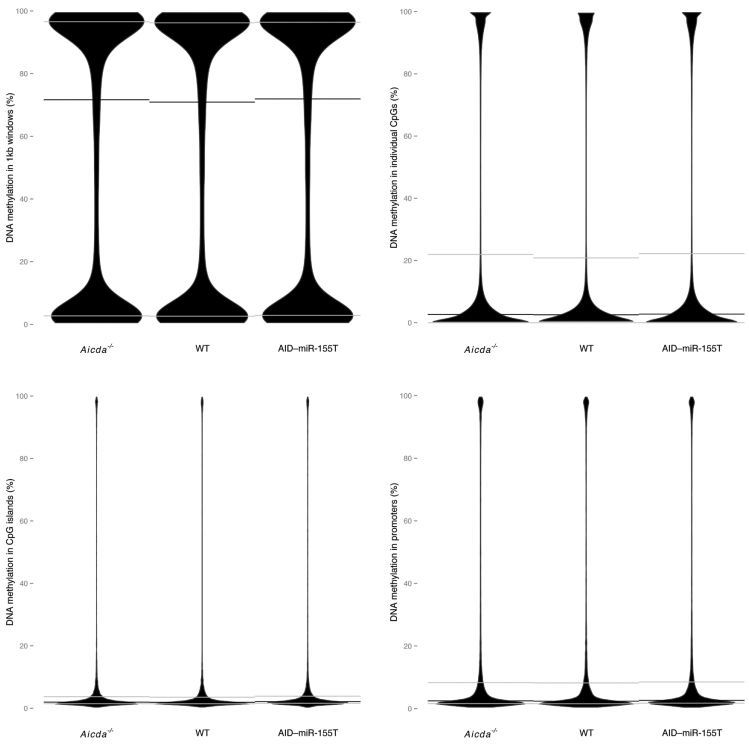

Figure 4. AID overexpression does not affect global DNA methylation in switching B cells

(**a–d**) Distribution of DNA methylation frequencies in activated B cells by AID expression as determined by RRBS for (**a**) 1 kb windows, (**b**) individual CpGs, (**c**) CpG islands, and (**d**) gene promoters (n = 1 mouse per genotype). Width along *x*-axis denotes relative frequency of features with given level of methylation. Black horizontal line is sample median; gray horizontal lines are first and third quartiles.

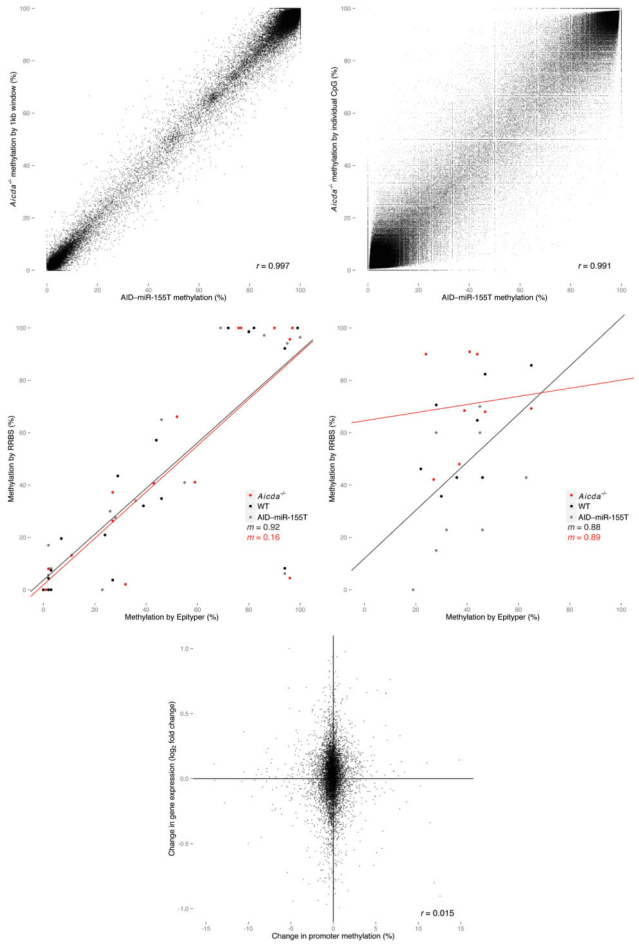

Figure 5. DNA methylation changes in activated B cells are not AID-dependent

(**a,b**) Pairwise comparisons of methylation frequency in AID-overexpressing and *Aicda*$^{-/-}$ B cells for (**a**) 1 kb windows and (**b**) all CpGs, as determined by RRBS. (**c,d**) Comparison of DNA methylation frequencies as determined by RRBS and Epityper for a random subset of CpGs with (**c**) > 20% greater methylation in *Aicda*$^{-/-}$ than in WT and (**b**) CpGs with < 10% difference between *Aicda*$^{-/-}$ and WT. Lines are linear fit for *Aicda*$^{-/-}$ or pooled WT and AID–miR-155T data. (**e**) Comparison of differences in gene expression and methylation in the associated promoters for AID–miR-155T and *Aicda*$^{-/-}$ B cells. (n = 1 mouse per genotype; *r* = Pearson's correlation coefficient)