Immunoglobulin Class Switch Recombination Is Initiated by Rare Cytosine Deamination Events at Switch Regions

Ahrom Kim,^{a*} Li Han,^a Kefei Yu^a

^aDepartment of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan, USA

ABSTRACT Activation-induced cytidine deaminase (AID) initiates immunoglobulin (Ig) class switch recombination (CSR), somatic hypermutation (SHM), and gene conversion by converting DNA cytosines to uracils at specific genomic regions. In this study, we examined AID footprints across the entire length of an engineered switch region in cells ablated for uracil repair. We found that AID deamination occurs predominantly at WRC hot spots (where W is A or T and R is A or G) and that the deamination frequency remains constant across the entire switch region. Importantly, we analyzed monoallelic AID deamination footprints on both DNA strands occurring within a single cell cycle. We found that AID generates few and mostly isolated uracils in the switch region, although processive AID deaminations are evident in some molecules. The frequency of molecules containing deamination on both DNA strands at the acceptor switch region correlates with the class switch efficiency, raising the possibility that the minimal requirement for DNA double-strand break (DSB) formation is as low as even one AID deamination event on both DNA strands.

KEYWORDS DNA repair, R-loop, activation-induced cytidine deaminase, base excision, class switch recombination, mismatch repair, somatic hypermutation

At the mammalian immunoglobulin heavy chain (IgH) locus, the rearranged V_H region is followed by tandem constant (C) regions, of which only the V_H -proximal C region is expressed (1–4). Switching from the default expression of IgM (encoded by the μ constant region) to other isotypes (e.g., IgG, IgE, or IgA) occurs in antigen-stimulated B cells through an intrachromosomal deletion known as class switch recombination (CSR). CSR allows a B cell to express a different class of Ig (defined by the IgH C region) while maintaining the same antigen-binding specificity of the Ig. This is critical for the optimal elimination of infection in a humoral immune response because the C region mediates the effector function of Ig. Each C region is preceded by a large zone (2 to 10 kb) of repetitive DNA known as the switch (S) region, where DNA double-strand breaks (DSBs) are generated as requisite intermediates during CSR. A single CSR event involves the generation and joining of DSBs in the donor (e.g., μ) and acceptor (e.g., γ , ϵ , or α) S regions (1–4). Because DSBs are some of the most severe forms of DNA damage, dysregulated CSR can lead to oncogenic chromosomal translocations and cancer.

CSR is initiated by activation-induced cytidine deaminase (AID) binding to transcribed S regions and catalyzing DNA cytosine deamination that converts cytosines to uracils. Uracil in DNA can be either recognized as base damage and repaired by the base excision repair (BER) pathway or recognized as U-G mismatches and repaired by the mismatch repair (MMR) pathway (1–4). Apparently, both pathways are involved in the processing of AID-generated uracils, as the genetic ablation of components in either pathway (e.g., UNG2, APE1, MSH2, MSH6, MLH1, PMS2, and Exo1, etc.) inhibits CSR to various degrees (1–4). Intriguingly, BER and MMR in most cell types are error-free

Citation Kim A, Han L, Yu K. 2020.

Immunoglobulin class switch recombination is initiated by rare cytosine deamination events at switch regions. *Mol Cell Biol* 40:e00125-20. <https://doi.org/10.1128/MCB.00125-20>.

Copyright © 2020 American Society for Microbiology. All Rights Reserved.

Address correspondence to Kefei Yu, yuke@msu.edu.

* Present address: Ahrom Kim, Department of Biological Sciences, Auburn University, Auburn, Alabama, USA.

Received 26 March 2020

Returned for modification 14 May 2020

Accepted 1 June 2020

Accepted manuscript posted online 8 June 2020

Published 29 July 2020

processes, but the processing of AID-generated uracils in B cells is definitively error prone, resulting in DSBs in S regions and point mutations in V regions. How these two pathways are diverted in B cells to affect opposite outcomes (error-free repair versus mutations and DSBs) remains an incompletely answered question.

AID is a single-strand-specific deaminase (5–9). The absolute requirement for transcription for somatic hypermutation (SHM) at V regions and CSR at S regions is generally viewed as a necessity to separate the two DNA strands to make them accessible to AID (8, 10). However, transcription itself is not sufficient to explain why the Ig V and S regions are premier targets for AID. While the uniqueness of V region transcription (for targeting AID during SHM) remains unclear, a unique feature of S region transcription has been well documented. Mammalian S regions are rich in G-cluster motifs on the nontemplate strand, thereby making a G-rich RNA transcript, which is postulated to promote transcription-dependent R-loop formation (11). R-loops consist of a displaced single-stranded nontemplate strand and an RNA-DNA hybrid formed between the RNA transcript and the template DNA strand. AID may directly target the single-stranded DNA (ssDNA) within the R-loop or bind the S region transcript that acts as a guide RNA to bring AID to the S region by base pairing (12). Theoretically, SHM requires deamination on only one DNA strand (13), whereas CSR likely requires lesions on both DNA strands since CSR requires the generation of DSBs. Thus, it is critical to understand how uracils are distributed across the two DNA strands at S regions during typical CSR. Because of this gap in knowledge, our understanding of how switch region uracils lead to DSBs is limited and largely speculative.

To facilitate AID footprint delineation on a single-molecule basis, a number of genetic modifications are required. First, the experiment must be performed in *UNG2*^{−/−} *MSH2*^{−/−} cells such that uracil is not repaired and can be detected as a C-to-T mutation following PCR amplification (14). Second, the S region must be engineered to be short enough to allow PCR across its entirety while still supporting relatively efficient CSR. Third, the cell must be monoallelic for the S region that is being examined. This is because each AID-targeted allele can potentially yield two different PCR products. CSR occurs biallelically in stimulated B cells, with each cell containing two IgH alleles that can potentially yield four different PCR products (top and bottom strands of each allele). When performing PCR from a pool of BER/MMR-ablated cells, although one can differentiate top- and bottom-strand-derived PCR products (based on C>T versus G>A substitutions), it would not be feasible to determine which PCR products originate from the same allele. Introducing all of these modifications in a mouse model would be difficult (15, 16) but is relatively straightforward in a B cell line called CH12F3, which is capable of robust cytokine-dependent CSR *in vitro* (17). A subclone of CH12F3 (CH12F3.2A [here called CH12F3]) has a productive (successful VDJ rearrangement) IgH allele in the germ line configuration and a nonproductive allele (DJ join) that already underwent $S\mu$ - $S\alpha$ recombination (18). Therefore, this cell line is monoallelic for germ line $S\mu$ or $S\alpha$ and suitable for the delineation of AID footprints on a single-molecule basis.

In this study, AID footprints were analyzed across the entire length of an engineered $S\alpha$ region from either a pool of cells or individual cell clones. We found that AID deaminations at $S\alpha$ during CSR are rare and mostly distally spaced. The frequency of molecules harboring AID deamination on both DNA strands at $S\alpha$ correlates with the CSR frequency, suggesting that DSB formation requires deamination on both DNA strands, and the minimal requirement might be as low as one deamination event on each strand. The implications of these findings are discussed.

RESULTS

Genetic manipulations required for AID footprint analyses in CH12F3 cells. To delineate AID deamination (termed AID footprints), several requisite genetic modifications were performed in the CH12F3 cell line. First, the endogenous $S\alpha$ region was shortened to allow PCR amplification while still supporting relatively efficient CSR. Previously, we replaced endogenous $S\alpha$ in CH12F3 cells with a 1.1-kb “core” using

FIG 1 CRISPR-mediated disruption of MSH2 and UNG2 in CH12F3 cells. (A) CRISPR target site in the MSH2 gene. (B) Loss of MSH2 protein in MSH2^{-/-} cells shown by Western blotting. (C) Reduced CSR in MSH2^{-/-} cells. Cells were stimulated with CIT (anti-CD40 plus IL-4 and TGF-β1) for 72 h. (D) CRISPR target site in the UNG gene. (E) Loss of uracil repair activity in the UNG2^{-/-} cell extract. (F) Loss of CSR in UNG2^{-/-} cells. Cells were stimulated with CIT for 72 h. FSC, forward scatter.

recombinase-mediated cassette exchange (19). This 1.1-kb core *S_α* represents the most uniformly repeated region of *S_α* and supports CSR to ~30 to 40% of the wild-type (WT) level (19). In this study, the delineation of AID footprints was performed mostly on this core *S_α*. Second, the *UNG2* and *MSH2* genes were disrupted individually or in combination in WT as well as *S_α*^{core} cells by CRISPR/Cas9-mediated genome editing (Fig. 1). MSH2 disruption (Fig. 1A) was screened by Western blotting (Fig. 1B) and confirmed by sequencing of both alleles for CRISPR-mediated deletions. We found that MSH2 deficiency alone results in an ~3-fold reduction in CSR efficiency (Fig. 1C), consistent with previous studies showing that MSH2 is required for maximal CSR efficiency (20, 21). *UNG2* disruption (Fig. 1D) was screened by an activity assay (Fig. 1E) followed by Sanger sequencing confirmation of the frameshift deletions on both alleles. Although UNG2 deficiency alone in CH12F3 cells virtually abolishes CSR (Fig. 1F), AID footprint analyses were performed in cells deficient for both UNG2 and MSH2 (to ensure that AID-generated uracil is not repaired) and with the endogenous *S_α* region replaced by the 1.1-kb core. The resulting cell line is named *S_α*^{core}-dKO here.

FIG 2 Frequencies and distributions of AID footprints at pre-S μ or S α regions. Pie charts show the frequency of AID deamination. The number in the pie center indicates the number of molecules sequenced. Each slice indicates the proportion of sequences containing various numbers of mutations. The graphs on the right side of the pie charts show the distribution of AID footprints across the switch regions. Each vertical line indicates the percentage of C-to-T changes at that position. Lines above or below the horizontal line indicate deamination on the nontemplate or template strand, respectively. Red indicates an AID hot spot, and blue indicates a non-hot spot. (A) S α region from cells stimulated for 1 week. (B) S α region from cells stimulated for 4 weeks. (C) Pre-S μ region from cells stimulated for 1 week. (D) Pre-S μ region from cells stimulated for 4 weeks.

Frequency and distribution of AID deamination at S regions. The engineered core S α region made it possible to delineate AID footprints across an entire switch region. To accumulate enough events for an overview of the frequency and distributions of AID footprints across the S α region, S α^{core} -dKO cells were stimulated with CIT (anti-CD40 plus interleukin-4 [IL-4] and transforming growth factor β 1 [TGF- β 1]) for either 1 or 4 weeks (with the necessary splitting of confluent cultures), followed by PCR, cloning, and sequencing of core S α from cell pools.

As can be seen, after 1 week of stimulation, every sequenced molecule contains at least one C-to-T conversion (AID footprint) (Fig. 2). Some molecules accumulate more than 10 (Fig. 2). The mutation frequency in the core S α region after 1 week of stimulation was $6.8 \times 10^{-3}/\text{bp}$ and increased to $1.3 \times 10^{-2}/\text{bp}$ after 4 weeks (Fig. 2). Deamination occurred predominantly at WRC sites (where W is A or T and R is A or G) (89% for 1 week and 80% for 4 weeks) (Fig. 2), which are known AID hot spots. Deamination at non-WRC sites increased slightly over time, possibly due to a gradual depletion of WRC sites. We estimated that CH12F3 cells double every 12 to 16 h when cultured in CIT medium (17, 22). Therefore, for a week, there are 10 to 14 generations. Based on this estimation, we calculate that less than one (~ 0.7 to 0.9) AID deamination event occurs per cell per generation within this core S α target sequence. This number is likely an overestimation since early conversions are propagated into daughter cells and accumulate over generations. Although perhaps not every cell in the population responds to stimuli, the above-described estimation nevertheless indicates a very low frequency of AID deamination at the acceptor S region during CSR.

Deamination on the template strand is roughly equal to that on the nontemplate strand (48% versus 51% at 1 week and 44% versus 56% at 4 weeks, respectively). Also, AID footprints appear evenly distributed across the entire length of the S region (Fig. 2),

FIG 3 Strategy for AID footprint analysis of individual S regions. (A) AID-mediated cytosine deamination generates uracil in DNA and converts a normal C-G base pair to a U-G mispair. Cell division separates the mispair, in the absence of uracil repair, into daughter cells, resulting in two different “species” (one derived from the top strand and the other derived from the bottom strand). Sanger sequencing of the PCR product from each cell clone results in a double peak on the chromatogram at this position. (B) Representative chromatograms showing double peaks as an indication of PCR product heterogeneity from individual cell clones. (Left) Double peak of C and T reflecting AID deamination on the nontemplated DNA strand. (Right) Double peak of G and A reflecting AID deamination on the template DNA strand. Sequence heterogeneity was confirmed by cloning of the PCR product followed by Sanger sequencing.

in contrast to SHM, where the mutation frequency rises sharply ~ 150 bp downstream of the transcription initiation site and then decreases exponentially over a 1- to 2-kb region (23, 24).

The donor $S\mu$ region was not engineered to facilitate complete PCR amplification. To assess deamination at $S\mu$, a small region at the beginning of $S\mu$ (called pre- $S\mu$) was examined as a surrogate for $S\mu$ (25–27). The mutation frequencies at pre- $S\mu$ were $5.6 \times 10^{-3}/\text{bp}$ for week 1 and $1.3 \times 10^{-2}/\text{bp}$ for week 4 (Fig. 2), similar to those of core $S\alpha$. Also similar to core $S\alpha$, AID deamination at the pre- $S\mu$ region was roughly equal between the two DNA strands (Fig. 2). There was a slightly higher number of deamination events at non-WRC sites (25% after 1 week and 27% after 4 weeks of stimulation) at pre- $S\mu$. These results indicate that AID targets both DNA strands and deaminates cytosines predominantly at WRC motifs *in vivo*.

Few mutations were found in nonstimulated cells (1/14,971 for core $S\alpha$ (c $S\alpha$) and 3/5,850 for pre- $S\mu$). Because they are exclusively C-to-T mutations, they are likely deamination products by a basal level of AID expression in unstimulated cells rather than PCR errors.

Single-molecule analyses of AID footprints. DSBs at donor and acceptor S regions are essential intermediates of CSR. The generation of DSBs requires AID-initiated lesions, likely on both DNA strands. Thus, it is critical to delineate how AID deaminations are distributed across both DNA strands of individual molecules. After AID deaminates an S region, the two DNA strands are no longer completely complementary; each can give rise to a different PCR product. In experiments where genomic DNA from a pool of cells is used as the PCR template, it is impossible to identify the original pair for each molecule. However, initial attempts to amplify S regions from single cells yielded very few PCR products. Therefore, we used an alternative approach (Fig. 3A): cells were stimulated for a brief period of time (24 h) before cytokine withdrawal and then individually seeded into 96-well plates to allow clones to develop in the absence of stimuli. The choice of 24 h is based on observations and previous results showing peak levels of AID expression at this time point (28, 29). Prolonged stimulation was

FIG 4 AID footprints of individual S_{α} regions. Each line represents a single molecule of the S_{α} region. Dots above or below the line indicate deamination on the nontemplate or template strand, respectively. Red indicates AID hot spots, and blue indicates non-hot spots. Numbers at the bottom are coordinates of nucleotide positions in the S_{α} region.

avoided to limit multiple rounds of AID-induced mutations. For each clone, the PCR product consists of two “species,” one derived from the template strand and the other derived from the nontemplate strand. In an individual cell, if no AID deamination has occurred or if a cell harboring deamination has replicated prior to single-cell seeding, the two species within a PCR are identical, as is the case for the majority of the clones analyzed (289/307). However, if even a single AID deamination event occurred, the two species will be different, resulting in a double peak(s) on a sequencing chromatogram (Fig. 3B).

Using this method, a total of 307 PCRs (from individual cell clones) were sequenced, 18 of which showed one or more double peaks (C/T or G/A) in the chromatogram, indicating the existence of two species in these PCRs. These 18 PCR products were cloned into a plasmid and then transformed into *Escherichia coli* to separate the two species (among different bacterial colonies). After Sanger sequencing of at least 12 bacterial colonies from each cloning, AID deamination profiles on both DNA strands from these 18 molecules were reconstituted (Fig. 4). In 16 of the 18 cases, both species were successfully recovered. In the other two cases, only WT sequences were recovered, suggesting that the initially observed double peaks were sequencing artifacts. No PCR product yielded more than two species after cloning and sequencing, supporting the fact that these were authentic clones. No mutations that were pure C-to-T or G-to-A conversions were found in any of the PCR products, suggesting that no AID deamination had occurred before single-cell seeding. All but one AID deamination occurred at a WRC hot spot motif (Fig. 4).

Out of the 16 molecules that harbor distinct AID deamination patterns between the two DNA strands, 8 molecules have deamination on only one of the strands (Fig. 4, molecules 2 through 5, 8, and 11 through 13), and the other 8 molecules contain

FIG 5 CSR efficiency of $S\alpha^{\text{core-dKO}}$ cells complemented with UNG2 and MSH2. (A) Representative histograms from CSR assays after 24 h of CIT stimulation. (B) Average CSR efficiency of $S\alpha^{\text{core-dKO}}$ cells complemented with UNG2 and MSH2 after 24 h of CIT stimulation. Error bars indicate standard deviations of the means from three independent experiments.

deamination on both strands (Fig. 4, molecules 1, 6, 7, 9, 10, and 14 through 16). It has been suggested that CSR requires proximate single-strand breaks on both DNA strands, at overlapping AID hot spots (i.e., WGCW); however, only one molecule shows evidence of such staggered deamination at an overlapping AID hot spot motif (Fig. 4, molecule 14). Most deamination events observed in this study are isolated and quite distal to each other, suggesting that the majority of the DSBs at acceptor S regions may result from distally located nicks. Clustered deamination events are observed in a number of molecules (Fig. 4, molecules 10, 11, and 14 through 16), evident for a limited degree of AID processivity (30, 31) *in vivo*. Of note, this clustering occurs mostly on one DNA strand.

There has been evidence that CSR initiates at $S\mu$ and secondarily involves downstream S regions (32), suggesting that cells harboring DSBs in acceptor switch regions may already have DSBs in the donor $S\mu$ region. In this regard, the frequency of DSBs at the acceptor switch regions may dictate CSR frequency. Interestingly, the frequency (2.6%) of molecules harboring mutations on both DNA strands at $S\alpha$ correlates well with the CSR efficiency ($2.7\% \pm 0.4\%$) when the same cells are complemented with UNG2 and MSH2 cDNAs and stimulated for 24 h (Fig. 5). These data suggest that the minimal requirement for DSB formation at the switch region may be as low as one uracil on each DNA strand.

DISCUSSION

This study provides new insight into how AID facilitates deamination in S regions during CSR. First, AID targets template and nontemplate strands at similar frequencies. Because AID is an ssDNA deaminase, it was thought that AID would preferentially target the nontemplate strand (10), especially in the context of an R-loop structure (4). This might suggest that deamination occurs predominately after R-loops are processed by cellular enzymes that expose ssDNA on both DNA strands. One candidate for such an activity is the RNA exosome, which had already been implicated in CSR (33). The RNA exosome degrades myriad RNA species and may have direct and indirect roles in CSR (34, 35). Other candidates for such an activity would be RNase H enzymes. Mammals have two distinct RNase H enzymes, RNase H1 (single subunit) and RNase H2 (three subunits) (36). Although neither has been implicated in CSR, *in vitro*, *E. coli* RNase H1 can convert R-loops into structures termed “collapsed R-loops” that contain single-stranded regions on both DNA strands (as a result of a misalignment of switch repeats) (4, 11). A recent structural study demonstrated that AID has two distinct DNA-binding grooves, nicely explaining why AID has higher binding affinity and deaminase activity for branched or G4 DNA structures than linear DNAs (37). Collapsed R-loops contain numerous branched DNA structures; this could explain how AID deaminates both DNA strands (3). It also provides an explanation for the observation that although the primary sequences of switch regions are not highly evolutionarily conserved, the

repetitive nature of switch regions is uniformly observed in all species. Second, AID footprints distribute evenly across the entire length of an S region, coincident with the random distribution of recombination breakpoints across S regions. This feature is also consistent with the R-loop/collapsed R-loop model, since R-loops initiate and terminate randomly across the S region (11). In contrast, SHM is not evenly distributed over a distance (23, 24), suggesting that AID deaminates S and V regions via distinct mechanisms. However, S region deamination via an SHM-like mechanism (e.g., stalled transcription bubble) cannot be formally ruled out.

In this study, we examined AID deamination on individual molecules within a single cell cycle. Interestingly, most AID deamination events are distally located. This appears at odds with the widely conjectured mechanism by which DSBs form predominantly by deamination at overlapping AID hot spots (e.g., 5'-AGCT-3'), resulting in nicks just across from one another. Surprisingly, only 1 of 8 molecules harboring AID deamination on both strands displayed this pattern. One caveat is that distally located AID footprints on primary sequences may actually be spatially close in a collapsed R-loop. Additional secondary structures such as G4 quadruplexes might also contribute to bringing spatially distant sites into proximity, although there has been no clear evidence of G4 quadruplex formation in S regions in living cells. Previous studies proposed a model involving MMR-mediated strand excision as a means to create DSBs from distally located nicks, potentially explaining why MMR components are required for fully efficient CSR (2, 38, 39). Interestingly, it was shown recently that distal nicks generated by Cas9 nickase (Cas9n) can mediate "CSR-like" intrachromosomal deletions (40), suggesting that distal nicks are indeed capable of inducing DSB formation. However, there are intrinsic differences between Cas9n-generated nicks and uracil-induced nicks. DNA uracil has only one chance to induce strand breakage within a cell cycle because it will be either repaired via BER or replicated over by a DNA polymerase. In contrast, Cas9n will keep nicking the DNA even after the nicks are repaired (e.g., joined by DNA ligase I or III) as long as the guide/protospacer-adjacent motif (PAM) sequences remain intact. In addition, Cas9n unwinds the local duplex upon binding to the guide sequences. Thus, nicking by Cas9n does not directly mirror the nicking mechanism via DNA uracil repair.

MATERIALS AND METHODS

Reagents, cell culture, and class switch recombination assay. CH12F3 cells were cultured in RPMI 1640 medium supplemented with 10% (vol/vol) fetal bovine serum and 50 μ M β -mercaptoethanol. To induce CSR, CH12F3 cells were seeded at 5×10^4 cells/ml in culture medium containing 0.5 μ g/ml anti-CD40 antibody (Ab) (clone HM40-3; eBioscience), 5 ng/ml IL-4 (R&D Systems), and 0.5 ng/ml TGF- β 1 (R&D Systems) and grown for 72 h unless otherwise specified. Cells were stained with a fluorescein isothiocyanate (FITC) (or CF633, equivalent to allophycocyanin [APC])-conjugated anti-mouse IgA rat monoclonal antibody (clone C10-3; BD Biosciences) and analyzed on an LSR II flow cytometer (BD Biosciences). CSR efficiency was determined as the percentage of IgA-positive cells. Rat anti-AID (catalog number 14-5959-82; eBioscience) and mouse anti-MSH2 (catalog number sc-376501; Santa Cruz Biotechnology) monoclonal antibodies were used for Western blotting. Oligonucleotides were ordered from Millipore-Sigma and Integrated DNA Technologies.

Knockout and genetic complementation of the *UNG2* and *MSH2* genes. Annealed oligonucleotides were cloned into pSpCas9(BB)-2A-Puro (Addgene plasmid 62988) to make CRISPR vectors as described previously (41). The targeting sequence for the MSH2 gene is 5'-GTCAAAGACATGTTGGAGTT-3', and the one for the UNG2 gene is 5'-GCTGGAAGCAGCAGCTGTGC-3'. CRISPR vector-transfected cells were allowed to recover for 48 h before being cloned in 96-well plates by limiting dilution. Cell clones were screened by Western blotting (MSH2) or an activity assay (UNG). Candidate clones were further analyzed by PCR cloning and Sanger sequencing to confirm the frameshift deletions on each allele.

A bicistronic expression vector (pEF-UNG2-2A-MSH2-ires-neo) where the UNG2 and MSH2 cDNAs were separated by a self-cleaving 2A peptide was constructed to complement the UNG2 and MSH2 deficiencies simultaneously in *UNG2*^{-/-} *MSH2*^{-/-} cells. G418-resistant stable transfectants were screened for UNG expression by a UNG activity assay and for MSH2 expression by Western blotting.

Uracil-DNA glycosylase assay. One million cells were lysed in 50 μ l of a solution containing 20 mM Tris-Cl (pH 7.5), 100 mM NaCl, 10 mM EDTA, and 0.5% Igepal CA-630 by freezing and thawing twice. After centrifugation for 10 min at $21,000 \times g$, the supernatant (cell extract) was collected. A uracil-DNA glycosylase assay was carried out in a 10- μ l reaction mixture containing 10 mM Tris-HCl (pH 7.4), 50 mM NaCl, 5 pmol of the fluorescein-labeled oligonucleotide substrate, and 2 μ l of the cell extract. The reaction mixture was incubated for 10 min at 37°C before sodium hydroxide was added to a final concentration of 0.1 M and incubated for 5 min at 95°C. The reaction mixture was mixed with an equal

volume of formamide and heated for 5 min at 95°C before being resolved on a 10% denaturing polyacrylamide gel. Gel images were acquired on a Typhoon 9500 phosphorimager (GE Life Technologies).

AID footprint analyses. For AID footprint analyses on cell pools, cells were seeded at 5×10^4 cells/ml in medium containing CIT and grown for either 1 or 4 weeks with subculturing (1:5) when the cell density reached 1×10^6 cells/ml. Switch regions were PCR amplified with the Phusion high-fidelity DNA polymerase (New England Biolabs) from genomic DNA extracted from cell pools, cloned, and sequenced.

For AID footprint analyses on individual molecules, cells were seeded at 5×10^4 cells/ml in medium containing CIT and grown for 24 h before being collected and washed twice with medium without cytokines (CIT withdrawal). Cells were seeded into the wells of 96-well plates at 50 cells per plate and examined visually under a microscope. After 10 days, clones were picked for PCR (by Phusion) and Sanger sequencing. PCR products showing double peaks on the chromatogram were cloned into a plasmid and transformed into *E. coli*. At least 12 colonies from each cloning were sequenced to separate the template- and nontemplate-strand-derived molecules coexisting in the PCR product (which is responsible for the appearance of double peaks on the chromatogram).

ACKNOWLEDGMENTS

This work is supported by NIH grants R21AI126359 and R01AI139039 to K.Y.

REFERENCES

- Chaudhuri J, Alt FW. 2004. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat Rev Immunol* 4:541–552. <https://doi.org/10.1038/nri1395>.
- Stavnezer J, Guikema JE, Schrader CE. 2008. Mechanism and regulation of class switch recombination. *Annu Rev Immunol* 26:261–292. <https://doi.org/10.1146/annurev.immunol.26.021607.090248>.
- Yu K, Lieber MR. 2019. Current insights into the mechanism of mammalian immunoglobulin class switch recombination. *Crit Rev Biochem Mol Biol* 54:333–351. <https://doi.org/10.1080/10409238.2019.1659227>.
- Yu K, Lieber MR. 2003. Nucleic acid structures and enzymes in the immunoglobulin class switch recombination mechanism. *DNA Repair (Amst)* 2:1163–1174. <https://doi.org/10.1016/j.dnarep.2003.08.010>.
- Branstetter R, Pham P, Scharff MD, Goodman MF. 2003. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc Natl Acad Sci U S A* 100:4102–4107. <https://doi.org/10.1073/pnas.0730835100>.
- Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E, Alt FW. 2003. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* 422:726–730. <https://doi.org/10.1038/nature01574>.
- Dickerson SK, Market E, Besmer E, Papavasiliou FN. 2003. AID mediates hypermutation by deaminating single stranded DNA. *J Exp Med* 197:1291–1296. <https://doi.org/10.1084/jem.20030481>.
- Sohail A, Klapacz J, Samaranyake M, Ullah A, Bhagwat AS. 2003. Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. *Nucleic Acids Res* 31:2990–2994. <https://doi.org/10.1093/nar/gkg464>.
- Yu K, Huang FT, Lieber MR. 2004. DNA substrate length and surrounding sequence affect the activation-induced deaminase activity at cytidine. *J Biol Chem* 279:6496–6500. <https://doi.org/10.1074/jbc.M311616200>.
- Ramiro AR, Stavropoulos P, Jankovic M, Nussenzweig MC. 2003. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nat Immunol* 4:452–456. <https://doi.org/10.1038/ni920>.
- Yu K, Chedin F, Hsieh CL, Wilson TE, Lieber MR. 2003. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat Immunol* 4:442–451. <https://doi.org/10.1038/ni919>.
- Zheng S, Vuong BQ, Vaidyanathan B, Lin JY, Huang FT, Chaudhuri J. 2015. Non-coding RNA generated following lariat debranching mediates targeting of AID to DNA. *Cell* 161:762–773. <https://doi.org/10.1016/j.cell.2015.03.020>.
- Faill A, Aoufouchi S, Gueranger Q, Zober C, Leon A, Bertocci B, Weill JC, Reynaud CA. 2002. AID-dependent somatic hypermutation occurs as a DNA single-strand event in the BL2 cell line. *Nat Immunol* 3:815–821. <https://doi.org/10.1038/ni826>.
- Rada C, Di Noia JM, Neuberger MS. 2004. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol Cell* 16:163–171. <https://doi.org/10.1016/j.molcel.2004.10.011>.
- Chen Z, Viboolsittiseri SS, O'Connor BP, Wang JH. 2012. Target DNA sequence directly regulates the frequency of activation-induced deaminase-dependent mutations. *J Immunol* 189:3970–3982. <https://doi.org/10.4049/jimmunol.1200416>.
- Shinkura R, Tian M, Smith M, Chua K, Fujiwara Y, Alt FW. 2003. The influence of transcriptional orientation on endogenous switch region function. *Nat Immunol* 4:435–441. <https://doi.org/10.1038/ni918>.
- Nakamura M, Kondo S, Sugai M, Nazarea M, Imamura S, Honjo T. 1996. High frequency class switching of an IgM+ B lymphoma clone CH12F3 to IgA+ cells. *Int Immunol* 8:193–201. <https://doi.org/10.1093/intimm/8.2.193>.
- Kim A, Han L, Santiago GE, Verdun RE, Yu K. 2016. Class-switch recombination in the absence of the IgH 3' regulatory region. *J Immunol* 197:2930–2935. <https://doi.org/10.4049/jimmunol.1600530>.
- Han L, Masani S, Yu K. 2011. Overlapping activation-induced cytidine deaminase hotspot motifs in Ig class-switch recombination. *Proc Natl Acad Sci U S A* 108:11584–11589. <https://doi.org/10.1073/pnas.1018726108>.
- Ehrenstein MR, Neuberger MS. 1999. Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. *EMBO J* 18:3484–3490. <https://doi.org/10.1093/emboj/18.12.3484>.
- Schrader CE, Edelmann W, Kucherlapati R, Stavnezer J. 1999. Reduced isotype switching in splenic B cells from mice deficient in mismatch repair enzymes. *J Exp Med* 190:323–330. <https://doi.org/10.1084/jem.190.3.323>.
- Han L, Yu K. 2008. Altered kinetics of NHEJ and CSR in DNA ligase IV-deficient B cells. *J Exp Med* 205:2745–2753. <https://doi.org/10.1084/jem.20081623>.
- Lebecque SG, Gearhart PJ. 1990. Boundaries of somatic mutation in rearranged immunoglobulin genes: 5' boundary is near the promoter, and 3' boundary is approximately 1 kb from V(D)J gene. *J Exp Med* 172:1717–1727. <https://doi.org/10.1084/jem.172.6.1717>.
- Rada C, Milstein C. 2001. The intrinsic hypermutability of antibody heavy and light chain genes decays exponentially. *EMBO J* 20:4570–4576. <https://doi.org/10.1093/emboj/20.16.4570>.
- Barreto V, Reina-San-Martin B, Ramiro AR, McBride KM, Nussenzweig MC. 2003. C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. *Mol Cell* 12:501–508. [https://doi.org/10.1016/S1097-2765\(03\)00309-5](https://doi.org/10.1016/S1097-2765(03)00309-5).
- Matthews AJ, Husain S, Chaudhuri J. 2014. Binding of AID to DNA does not correlate with mutator activity. *J Immunol* 193:252–257. <https://doi.org/10.4049/jimmunol.1400433>.
- Nagaoka H, Muramatsu M, Yamamura N, Kinoshita K, Honjo T. 2002. Activation-induced deaminase (AID)-directed hypermutation in the immunoglobulin Smu region: implication of AID involvement in a common step of class switch recombination and somatic hypermutation. *J Exp Med* 195:529–534. <https://doi.org/10.1084/jem.20012144>.
- Cortizas EM, Zahn A, Hajjar ME, Patenaude AM, Di Noia JM, Verdun RE.

2013. Alternative end-joining and classical nonhomologous end-joining pathways repair different types of double-strand breaks during class-switch recombination. *J Immunol* 191:5751–5763. <https://doi.org/10.4049/jimmunol.1301300>.
29. Kinoshita K, Harigai M, Fagarasan S, Muramatsu M, Honjo T. 2001. A hallmark of active class switch recombination: transcripts directed by I promoters on looped-out circular DNAs. *Proc Natl Acad Sci U S A* 98:12620–12623. <https://doi.org/10.1073/pnas.221454398>.
 30. Pham P, Malik S, Mak C, Calabrese PC, Roeder RG, Goodman MF. 2019. AID-RNA polymerase II transcription-dependent deamination of IgV DNA. *Nucleic Acids Res* 47:10815–10829. <https://doi.org/10.1093/nar/gkz821>.
 31. Pham P, Bransteitter R, Petruska J, Goodman MF. 2003. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature* 424:103–107. <https://doi.org/10.1038/nature01760>.
 32. Schrader CE, Bradley SP, Vardo J, Mochegova SN, Flanagan E, Stavnezer J. 2003. Mutations occur in the Ig Smu region but rarely in Sgamma regions prior to class switch recombination. *EMBO J* 22:5893–5903. <https://doi.org/10.1093/emboj/cdg550>.
 33. Basu U, Meng FL, Keim C, Grinstein V, Pefanis E, Eccleston J, Zhang T, Myers D, Wasserman CR, Wesemann DR, Januszyk K, Gregory RI, Deng H, Lima CD, Alt FW. 2011. The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell* 144:353–363. <https://doi.org/10.1016/j.cell.2011.01.001>.
 34. Pefanis E, Wang J, Rothschild G, Lim J, Chao J, Rabadan R, Economides AN, Basu U. 2014. Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature* 514:389–393. <https://doi.org/10.1038/nature13580>.
 35. Pefanis E, Wang J, Rothschild G, Lim J, Kazadi D, Sun J, Federation A, Chao J, Elliott O, Liu ZP, Economides AN, Bradner JE, Rabadan R, Basu U. 2015. RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell* 161:774–789. <https://doi.org/10.1016/j.cell.2015.04.034>.
 36. Cerritelli SM, Crouch RJ. 2009. Ribonuclease H: the enzymes in eukaryotes. *FEBS J* 276:1494–1505. <https://doi.org/10.1111/j.1742-4658.2009.06908.x>.
 37. Qiao Q, Wang L, Meng FL, Hwang JK, Alt FW, Wu H. 2017. AID recognizes structured DNA for class switch recombination. *Mol Cell* 67:361–373.e4. <https://doi.org/10.1016/j.molcel.2017.06.034>.
 38. Min IM, Schrader CE, Vardo J, Luby TM, D'Avirro N, Stavnezer J, Selsing E. 2003. The Smu tandem repeat region is critical for Ig isotype switching in the absence of Msh2. *Immunity* 19:515–524. [https://doi.org/10.1016/S1074-7613\(03\)00262-0](https://doi.org/10.1016/S1074-7613(03)00262-0).
 39. Min IM, Rothlein LR, Schrader CE, Stavnezer J, Selsing E. 2005. Shifts in targeting of class switch recombination sites in mice that lack mu switch region tandem repeats or Msh2. *J Exp Med* 201:1885–1890. <https://doi.org/10.1084/jem.20042491>.
 40. Ling AK, So CC, Le MX, Chen AY, Hung L, Martin A. 2018. Double-stranded DNA break polarity skews repair pathway choice during intra-chromosomal and interchromosomal recombination. *Proc Natl Acad Sci U S A* 115:2800–2805. <https://doi.org/10.1073/pnas.1720962115>.
 41. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8:2281–2308. <https://doi.org/10.1038/nprot.2013.143>.