NAR Molecular Medicine, 2025, **2**, ugaf003
https://doi.org/10.1093/narmme/ugaf003
Advance access publication date: 28 February 2025
Standard Article

Nucleotide level mapping of uracils in murine heavy chain switch regions shows correlation between uracilation and positions of switch junctions created during class-switch recombination

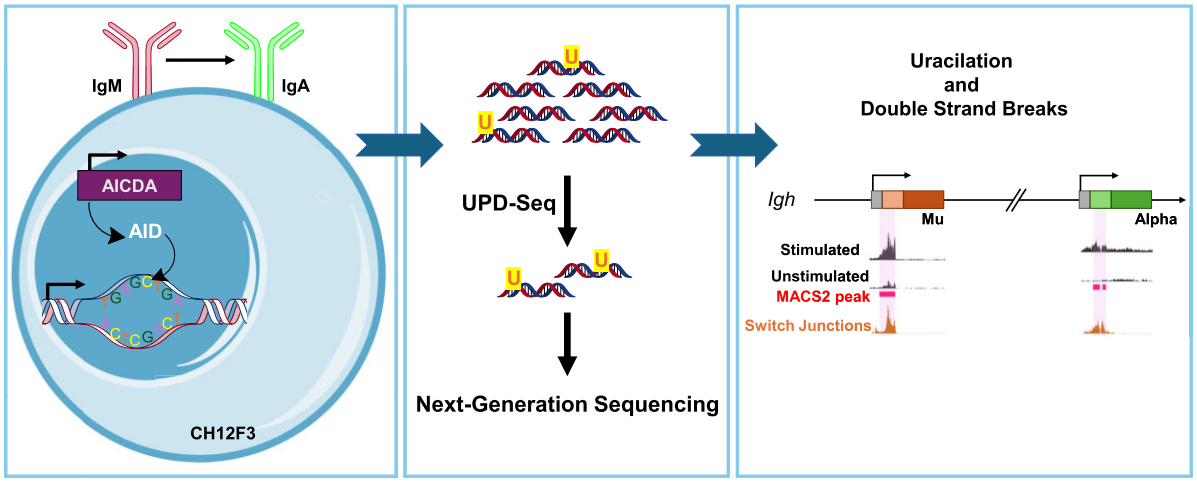
Rukshana Mohamad-Ramshan[1], Ramin Sakhtemani[2], Yasha Butt[1], Michael S. Lawrence[2,3], Ashok S. Bhagwat[1,4,*]

[1]Department of Chemistry, Wayne State University, Detroit, MI 48202, United States
[2]Massachusetts General Hospital Krantz Family Center for Cancer Research, Boston, MA 02129, United States
[3]Broad Institute of MIT and Harvard, Cambridge, MA 02142, United States
[4]Department of Biochemistry, Microbiology and Immunology, Wayne State University School of Medicine, Detroit, MI 48201, United States
[*]To whom correspondence should be addressed. Email: axb@chem.wayne.edu

Abstract

Introduction of uracils in specific regions within immunoglobulin genes by the activation-induced deaminase (AID) promotes mutations and double-strand breaks (DSBs). Although uracils are repaired through multiple DNA repair pathways, previous work has used mutations or DSBs as proxies for uracils and not mapped the uracils directly. We mapped uracils in the Ig heavy chain gene, *Igh*, in a murine cell line, CH12F3, undergoing class-switch recombination (CSR) using the uracil pull-down and sequencing technique. These cells undergo IgM-to-IgA switch upon expression of AID but do not undergo somatic hypermutation. We mapped uracils in cells defective in uracil repair and show that AID introduces high levels of uracils only in parts of switch-mu and switch-alpha regions and not in constant regions, the *Igh* variable region or the light chain genes. Furthermore, the peaks of uracilation match the previously determined distribution of switch junctions, which are representative of DSBs that cause isotype switching. This work confirms that AID creates uracils in both DNA strands and shows that there is a direct correlation between uracil creation and DSBs in the relevant switch regions. We evaluate proposed mechanisms of CSR in light of these findings and show that mapping uracils provides a fresh perspective on CSR.

Graphical abstract



Introduction

When mice or humans are exposed to infectious agents, their B lymphocytes utilize two genetic processes, somatic hypermutation (SHM) [1] and class-switch recombination (CSR) [2], to achieve antibody maturation. In SHM, the variable regions of the heavy chain (*Igh*) and light chains (*Igl* and *Igk*)

Received: February 20, 2025. **Editorial Decision:** February 24, 2025. **Accepted:** February 27, 2025
© The Author(s) 2025. Published by Oxford University Press on behalf of NAR Molecular Medicine.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

acquire all six possible base substitutions at high frequencies, while in CSR the μ constant segment (Cmu) of the heavy chain is replaced with one of the other constant segments through region-specific double-strand breaks (DSBs) followed by recombination. These processes create a wide repertoire of antibodies from which affinity maturation selects antibodies with high affinity toward infection agent-derived antigens with appropriate effector function determined by the new constant segment.

A DNA-cytosine deaminase, activation-induced deaminase (AID), is essential for both SHM and CSR [3]. It converts cytosines in single-stranded DNA to uracil [4–7] promoting mutations [3] and strand breaks [8]. Mammalian cells have multiple enzymes that can excise uracils from U•G pairs leading to the restoration of cytosines through base-excision repair (BER), the principal of these being uracil-DNA glycosylase, UNG [9]. Although replication of uracils created by AID should only cause C:G-to-T:A mutations, SHM generates all possible base substitution mutations [10]. Thus, excision of uracils by UNG followed by incomplete or error-prone BER is proposed to respectively cause strand breaks or base substitutions that fuel CSR and SHM [11]. Additionally, noncanonical mismatch repair of U•G pairs may also cause mutations and strand breaks that contribute to both SHM and CSR [12, 13].

Among the many unknowns regarding the molecular mechanisms of SHM and CSR is how the mutations and strand breaks caused by AID are localized in the variable regions of the Ig genes and the switch regions within *Igh*. Gene transcription is required for both SHM and CSR [14, 15] and AID associates with RNA polymerase II [16]. However, association of AID to a transcribed gene is not sufficient for its mutator activity [17, 18]. The activity of AID within a transcription bubble may be influenced by multiple factors including transcriptional pausing [19, 20], presence of "licensing factors" [21], RNA or DNA secondary structure [22–24], and RNA degradation [25], among others [26]. Additionally, the uracils created in some genomic regions may be subjected to error-free repair, while uracils in other regions may be subject to error-prone repair or no repair [27]. This complexity in uracil creation and repair makes it necessary that one must follow the fate of uracils created by AID to understand the mechanisms of SHM and CSR. Despite this need, nearly all previous studies have used mutations, strand breaks, or translocation end-points as proxies for uracils. These approaches only determine the consequences of uracils after various repair- and replication-related DNA transactions have taken place and cannot determine the history of uracil processing in Ig genes. As a result, doubts are still being expressed whether uracil creation in DNA by AID is the driving force behind antibody maturation [28].

The only previous study that directly addressed the role of uracil creation by AID in cells undergoing antibody maturation converted uracils to strand breaks using UNG and the abasic-site endonuclease, APE1, and detected the breaks using Southern blots [29]. Using this technique, they confirmed AID-dependent uracil accumulation in restriction fragments in genomes of chicken B cells undergoing SHM and gene conversion, and murine B cells undergoing SHM and CSR. They found that murine splenic B cells when stimulated *ex vivo* accumulated uracils in the switch μ (Smu) region of the *Igh* gene, that the uracils mostly originated from cytosines and were roughly equally distributed in the two strands of *Igh* [29]. However, these investigators did not map uracils across the Ig genes or correlate the uracils to mutational hotspots in SHM or DSBs involved in CSR.

We use below an unbiased approach to mapping uracils in a murine cell line undergoing CSR focusing mainly on the *Igh* gene. Using uracil pull-down and sequencing technique (UPD-seq) [30], we show that the uracils are highly localized within *Igh* and correlate well with the known locations of DSBs that cause CSR.

Materials and methods

Murine cell line growth

The mouse B-cell lymphoma cell line, WT CH12F3, and its $UNG^{-/-}$ and $AID^{-/-}$ knockout derivatives were kindly provided by Dr Kefei Yu (Michigan State University). These cells were cultured in RPMI 1640 media (Cytiva) supplemented with 8% fetal bovine serum (FBS, Hyclone), 2 mM L-glutamine, 100 units/ml penicillin–streptomycin, and 50 μM β-mercaptoethanol (Sigma–Aldrich). The cells were maintained at 37°C with 5% CO_2 in a humidified incubator in tissue culture-treated T-25 cm^2 or T-75 cm^2 flasks (US Scientific). Cell counts and viability were determined using trypan blue staining (Sigma) and a TC20 cell automated counter (Bio-Rad).

CIT stimulation and isotype switching assay

To stimulate the cells for isotype switching, they were cultured at a cell density of 50 000 cells/ml in media supplemented with 1 μg/ml of purified anti-mouse CD40 (eBiosciences), 10 ng/ml of recombinant mouse interleukin 4 (IL-4; R&D Systems), and 1 ng/ml of recombinant human transforming growth factor β (TGF-β; R&D Systems). The stimulated cells were then incubated at 37°C with 5% CO_2 in a humidified incubator for 72 h.

Following 72 h CIT stimulation, one million cells were transferred into 12 × 75 mm round-bottom polystyrene Falcon test tubes (Corning), and the cells were harvested by centrifugation at 500 *g* for 3 min at 4°C. They were resuspended in 1 ml of wash buffer (1× phosphate buffered saline (PBS), 0.5% fetal bovine serum (FBS), and 0.02% NaN_3) and harvested again using the same centrifugation conditions. Next, the cells were stained with 0.5 μg/ml fluorescein isothiocyanate (FITC)-conjugated anti-mouse IgA antibody (BD Biosciences) and incubated for ∼30 min at 4°C in the dark. Following incubation, the cells were washed and resuspended in 500 μl of wash buffer.

Fluorescence-activated cell sorting (FACS) analysis was performed using the LSR II Flow Cytometer (BD Biosciences) at the Microscopy, Imaging, and Cytometry Resources Core at Wayne State University. Isotype switching was quantified as the percentage of IgA-positive cells. Dead cells were excluded from the assay by labeling them with 7-AAD (BD Biosciences).

RNA extraction and RT-PCR

Total RNA was extracted from CH12F3 cells using TRIzol™ (Invitrogen) reagent and the manufacturer's protocol. Briefly, one million cells were harvested by centrifugation, and 150 μl of TRIzol™ reagent was added. After resuspension, the mixture was incubated for 5 min. Subsequently, 30 μl of chloroform was added per 150 μl of TRIzol™ reagent, and the mixture was incubated for an additional 2 min. The sample was

then centrifuged for 15 min at 12 000 *g* at 4°C. The aqueous phase containing the RNA was transferred to a new tube, and 75 µl of isopropanol was added per 150 µl of TRIzol™ reagent to precipitate the RNA. After centrifugation, the pellet was washed with 150 µl of 75% ethanol, air dried, and resuspended in RNase-free water.

The reverse transcription of RNA was carried out using oligo d(T)23 primers and the Protoscript® II First Strand cDNA synthesis kit (NEB). Expression levels of murine AID were determined by amplifying the copy DNA (cDNA) using the PowerUp SYBR Green master mix (Applied Biosystems). The gene- or region-specific primers used for reverse transcriptase polymerase chain reaction (RT-PCR) are listed in Supplementary Table S2. The reaction conditions were as follows: 95°C for 10 min, followed by 45 cycles of extension at 50°C for 15 s and 60°C for 1 min. Gene expression levels were determined relative to murine TATA-box binding protein (TBP), which served as the internal control. Amplification was performed using the 7500 Fast Real-Time PCR System (Applied Biosystems), and the data were analyzed using the ΔC_t method.

Uracil quantification assay

The uracil quantification assay was performed as described previously [31]. Briefly, genomic DNA (5 µg) was digested with HaeIII (NEB) and purified using phenol:chloroform extraction followed by ethanol precipitation. The digested DNA was then incubated with 10 mM *O*-allylhydroxylamine (Sigma–Aldrich) at 37°C for 1 h. Subsequently, the DNA was treated with *Escherichia coli* uracil DNA-glycosylase (UNG) at 37°C for 30 min to excise uracils creating abasic sites. The DNA was further incubated with 2 mM AA6 [32] for 1 h at 37°C to tag abasic sites generated from the excision of uracils. The AA6-tagged DNA was then labeled with 1.7 µM DBCO-Cy5 by shaking the reaction mixture at 37°C in the dark for 2 h. The fluorescently labeled DNA was purified using the DNA Clean and Concentrator Kit (Zymo Research) and transferred onto a positively charged Zeta Probe membrane (Bio-Rad) using a Bio-Dot apparatus (Bio-Rad). The membrane was scanned using a Typhoon 9210 phosphor imager (GE Healthcare). Images were analyzed using the ImageJ software, and fluorescence intensity was converted to uracil counts using the CJ236 standard curve, as previously described [31].

UPD-qPCR and UPD-seq

DNA sonication and pull-down were performed as previously described [30]. Briefly, genomic DNA was dissolved in 1× TE buffer (200 ng/µl) and sonicated to produce 500 bp fragments using a Covaris S2 sonicator at the Proteomics Core, Wayne State University. The fragmented DNA was incubated with 10 mM AA7 (Sigma–Aldrich) at 37°C for 1 h to block pre-existing abasic sites. The DNA was then treated with *E. coli* uracil DNA-glycosylase (UNG) at 37°C for 30 min to excise uracils and create new abasic sites. Subsequently, the DNA was incubated with 2 mM ssARP [30] for 1 h at 37°C to tag abasic sites generated from the excision of uracils. The ssARP has a biotin functionality, which was used for pull-down using Dynabeads MyOne Streptavidin C1 magnetic beads (Invitrogen). The beads were washed with 2× DNA binding and wash buffer (B&W buffer: 10 mM Tris–HCl, pH 7.5; 1 mM EDTA; 2 M NaCl; 0.05% Tween 20), and the DNA was bound to the beads by incubating for 20 min. DNA-bound beads were separated using a magnetic stand (DynaMag, Invitrogen), and the supernatant containing unbound DNA was removed. The beads were washed five times with 1× B&W buffer to remove impurities and resuspended in 100 mM dithiothreitol (DTT) for 10 min at 37°C. The beads were then placed on the magnetic stand, and the supernatant containing the eluted DNA was purified by ethanol precipitation. The DNA pellet was dissolved in 1× TE buffer, and the concentration was measured using a Qubit™ 4 Fluorometer (Invitrogen).

The switch region Sµ and constant region Cγ3 were amplified from CH12F3 WT DNA using Phusion U Hot Start DNA Polymerase (Thermo Scientific) for standard curve preparation, with the Sµ and Cγ3 nested primers listed in Supplementary Table S2. Polymerase chain reaction (PCR) products were confirmed on an agarose gel (Sµ: 295 bp; Cγ3: 297 bp) and purified using the GeneJet PCR purification kit (Thermo Fisher). Two standard curves were prepared using a series of 5-fold dilutions. The reaction mixture was prepared in a final volume of 25 µl, containing pull-down DNA from each sample from simulated and unstimulated UNG$^{-/-}$ CH212F3 cells, specific primer pairs, and PowerUp SYBR Green Master Mix (Applied Biosystems). Nuclease-free water was used for the reactions and for sample dilutions. The reaction conditions were 95°C for 10 min, followed by 45 cycles of extension at 50°C for 15 s, and finally 60°C for 1 min. The copy number of each DNA sample was calculated using the appropriate calibration curve and the respective C_t values. The qPCR assay was performed using the 7500 Fast Real-Time PCR System (Applied Biosystems).

The same pull-down DNA was also used for Illumina DNA library preparation. During the library preparation step, adaptors were ligated to the ends of DNA. DNA libraries were prepared using the TruSeq DNA Nano library preparation kit (Illumina) according to the manufacturer's protocol, targeting an average fragment size of 700 bp. Prior to sequencing, the libraries were amplified using adapter-specific primers and 8 cycles of PCR. The quality of the prepared libraries was assessed using the Agilent 4200 TapeStation or the Agilent Bioanalyzer at the Genomics Core Facility, Wayne State University. All libraries were pooled in 2 nM equimolar quantities for multiplexed sequencing. The prepared libraries were loaded onto one Illumina NovaSeq 6000 SP 300 flow cell and sequenced in a 2 × 150 bp paired-end format using the SP 300 cycle reagent cartridgeat the Genomics Core Facility, Wayne State University.

Alignment and quality filters

The raw sequencing data were quality-controlled using FastQC (v0.11.8; http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The sequencing reads were aligned to the MGSCv37 (mm9) genome using bwa (v0.7.15, [33]) with default parameters. Samtools (V.1.4.1, [34]) was used to convert the alignments to BAM format. SortSam and MarkDuplicates [Picard tools (v2.17.1), https://gatk.broadinstitute.org/hc/en-us] were used to sort reads and mark duplicate reads. When needed, Sambamba (V.0.4.7, [35]) was used to merge the sequences from multiple sequencing lanes. A custom script was used to filter out duplicate reads, as well as reads with mapping quality <10 [36], and reads with an alignment score <80% of read length. A bash script (available at github.com/rayanramin/AID.cry-UPDSeq) was used to identify reads that contain at least one mismatch at a C:G po-

sition. Integrated Genome Viewer (IGV V.2.10.3) [37] was used to manually inspect a sample of aligned reads for quality control.

Uracilation peak calling

The original high quality sequencing reads (QC score > 30) and mapping quality-filtered alignments were used to call enrichment peaks via MACS2 (v2.0.10, [38]). MACS2 was run in the paired mode—sequence alignments from the unstimulated samples were used as the control sample and the sequences from stimulated samples were taken as the treatment samples.

Analysis of base substitutions in UPD-seq

The bam-readcount utility [39] was used to extract the counts of aligned reads at each position in the IgH gene. The table of nucleotide-level coverage was imported into R and analyzed further to compare the frequencies of base changes in different regions.

R (v4.2.3; https://cran.r-project.org/) with the packages Rsamtools (v2.18.0, [40]), GenomicRanges (v1.54.1) and IRanges (v2.36.0) [41], and dplyr (v1.1.2; https://github.com/tidyverse/dplyr) were used to import sequencing reads aligning to various IgH region; these reads were analyzed to extract the sequence change information. These changes were filtered to mask known SNPs, as well as clonal events where the allele frequency of the "mutation" is >0.5. The C-to-T and G-to-A single substitutions in high mapping quality reads in MACS2 peaks found in Smu and Salpha were identified, and the numbers of such changes in NC, WRCY, and non-WRCY (nWRCY) sequence contexts were determined. Additionally, the numbers of adjacent C-to-T and G-to-A double sequence changes were determined and the distances between the adjacent changes were determined. These numbers were normalized to the number of available sequence motifs within the sequences (as relevant) and the number of mapped reads in the IgH region for each experiment. The frequencies are presented per 10 kbp. The code for these calculations will be available at Github.

Deep sequencing

Genomic DNA was extracted from WT, UNG$^{-/-}$, and AID$^{-/-}$ CH12F3 cells. The S_{μ} and $C_{\gamma 3}$ regions were amplified using a nested PCR approach. Outer PCR reactions were performed with Phusion U Hot Start DNA Polymerase (Thermo Scientific) using the primers listed in Supplementary Table S2. The 50 µl reactions were performed as follows: 1× incubation for 5 min at 98°C, followed by 10 cycles of 10 s at 98°C, 30 s at 64°C, 1 min at 72°C, and finally 5 min at 72°C. Nested PCR was then carried out using the PCR amplicons from the previous reaction as a template using the primers listed in Supplementary Table S2. The 50 µl of nested PCR reactions were as follows: 1× incubation for 5 min at 98°C, followed by 20 cycles of 10 s at 98°C, 30 s at 64°C, 15 s at 72°C, and finally 5 min at 72°C. The PCR products were purified, and DNA libraries were prepared using Illumina TruSeq DNA Nano following the manufacturer's protocol. The libraries were subsequently sequenced on the Illumina MiSeq platform at the Genomics Core Facility, Wayne State University.

Sequence reads were analyzed using a custom pipeline described previously [42]. Briefly, paired-end reads were joined using the fastq-join tool, considering reads with at least a 10 bp overlap and a mismatch rate ≤ 8%. The joined reads were aligned to the reference sequence (mm9) using bwa with default parameters. Position-wise statistics were generated and reads with a mapping quality of at least 60 and a minimum base quality of 30 were used for mutational analysis. For each sample, BAM files were converted to TSV format using the Java-based utility sam2tsv (http://lindenb.github.io/jvarkit/Sam2Tsv.html). A custom AWK script was then used to generate per-read summary files for point mutations. The source code is available at https://github.com/rayanramin/CH12-UPDSeq.

Sequence repeat analysis

The distribution of various sequence repeats within the µ (chr12:114656190–114666269) and α (chr12:114492990–114503549) segments of the mouse genome (mm9 reference) was determined using IGV and downloaded to an Excel spreadsheet. The sequences examined include WRC (where W is A or T), WRCY (where R is a purine and Y is a pyrimidine), AGCT, GGG, GGGG, and the putative G quadruplex-forming sequences GGGGTGAGCTGAGCTGAGCT (#1) and GAGCTGAGCTGGGGTGAGCT (#2). The Smu and Salpha sequences were divided in 60 bp windows ,and the number of occurrences of each repeat was counted within each window. These counts are presented in Supplementary Table S4.

Switch junction analysis

The visualization of genomic data within the *Igh* and *Igl* loci used the Integrative Genomics Viewer (IGV V.2.10.3, [37]) incorporating coverage files from UPD-seq as well as HTGTS and DRIP-seq datasets, represented in BAM, BIG WIG, BED, and BEDGRAPH formats. The coordinates of all promoters, switch, constant, and regulatory regions within the mouse IgH locus, as well as the coordinates for IgL regions, have been previously described [43].

The high-throughput genome-wide translocation and sequencing (HTGTS) data deposited in the NCBI Gene Expression Omnibus (GEO) under accession number GSE1069228 [44] were used to identify DSBs within the Smu and Salpha in CH12F3 cells. The coordinates of switch junctions within the mouse IgH locus were extracted, and the frequency of these switch junctions was calculated at each nucleotide within the locus, then binned in 60 bp sliding windows. BEDGRAPH files were subsequently generated and visualized using IGV. Additionally, HTGTS data deposited at the UCSC Genome Browser development site (http://52.20.87.23/public/Dis3_project/, [45]) were used to identify DSBs within the Smu in mouse B cells. Previous studies that reported switch junctions determined by targeted PCR sequencing in CH12F3 cells or C57BL/6 murine splenocytes [46–52] are listed in Supplementary Table S3. DSB frequency was determined at the nucleotide level within the S_{μ} and S_{α} loci, and the junctions were binned in 60 bp windows and converted to BEDGRAPH files for visualization using IGV.

R-loop analysis

R-loops within the Smu region of the murine B cells and the Salpha region of the CH12F3 cells have been mapped using a bisulfite-based assay [53, 54]. The sequencing files from the study by Huang *et al.* [53] were kindly provided by Dr Kefei Yu (Michigan State University), and we selected

14 of these files for analysis. The long stretches of C-to-T changes in each of these sequences were taken to represent the single-stranded regions defining an R-loop. These sequences and the single-stranded regions therein are in Supplementary Sequence File S1. In Salpha, the extent of the R-loops was determined based on the PCR primers used in this study for DNA amplification [54]. One end of each R-loop was near one of the PCR primers and the other end was estimated based on the relevant illustration in the paper (Supplementary Sequence File S1, [54]). These R-loop sequences were then mapped onto the murine genomic sequence (mm9). The DRIP-seq data regarding R-loops in Smu in mouse B cells were obtained from the UCSC Genome Browser (http://52.20.87.23/public/Dis3_project/, [45]). The Bigwig files of these data from Dis3 ± mice were downloaded and visualized in IGV.

Results

AID expression causes greater accumulation of uracils and mutations in S_μ but not $C_{\gamma 3}$

We chose murine B-cell tumor-derived cell line, CH12F3, for our studies because it undergoes an IgM-to-IgA isotype switch at high frequencies following stimulation of cells by a cocktail of cytokines referred to as CIT [55]. Following stimulation of CH12F3 cells with CIT, AID expression increased 12- and 16-fold in wild-type (WT) and $UNG^{-/-}$ cells, respectively (Fig. 1A) and corresponding genomic uracil levels increased 31% and 49% (Fig. 1B). As reported previously [56], the uracilation levels were higher in $UNG^{-/-}$ cells compared to WT cells (Fig. 1B) and the stimulation caused a large increase in IgM-to-IgA switching in WT cells, but not in $UNG^{-/-}$ cells (Fig. 1C and Supplementary Fig. S1A; [57]). Isotype switching was undetectable in $AID^{-/-}$ cells (Supplementary Fig. S1B). The modest increase in genomic uracils even in $UNG^{-/-}$ cells suggests that cytosine deamination due to AID is restricted to only a limited number of genomic regions. This experiment was repeated, and the results showed similar trends (Supplementary Fig. S2A–C).

As described previously [30], we replaced uracils in genomic DNA from the two cell lines with a tag containing biotin and pulled down the DNA fragments using streptavidin beads. To confirm that the switch region DNA had been damaged by AID, we performed qPCR on the pull-down DNA using primers specific for Smu, with $C_{\gamma 3}$ (Cgamma3) region serving as the negative control (Supplementary Fig. S3A). This UPD-qPCR showed that the Smu region was enriched in the pull-down even without cell stimulation, and the amount of Smu-specific pull-down increased substantially following stimulation (Fig. 1D and Supplementary Fig. S2D). Interestingly, the amount of pull-down DNA from Cgamma3 decreased following stimulation and the decrease was statistically significant in both independent experiments (Fig. 1D and Supplementary Fig. S2D). The reason for this decrease is not unclear. Thus, it appears that AID-mediated cytosine deaminations increase in Smu following cellular stimulation but decrease in Cgamma3.

To confirm that AID was creating uracils in Smu [58], we amplified segments of Smu from genomic DNA (without pull-down) from $AID^{-/-}$, WT, and $UNG^{-/-}$ stimulated cells using PCR and subjected it to deep sequencing. The frequency of all mutations and those at C:G pairs in Smu was higher in WT cells compared to $AID^{-/-}$ background and were even higher in $AID^{+/+}$ $UNG^{-/-}$ cells (Fig. 1E). In contrast, the mutation frequency at Cgamma3 was low and remained the same in different genetic backgrounds suggesting that it represented the background mutation frequency for the experiment (Fig. 1E). Together these results show that stimulation of WT and $UNG^{-/-}$ CH12F3 cells results in increased AID expression, increased total genomic uracil levels, and increased uracils and mutations in Smu, but not Cgamma3.

AID creates uracilation peaks in S_μ and S_α, but not in the rest of *Igh* or V_k

To determine the distribution of uracils in the *Igh* gene during CSR, we performed UPD-seq [30] on genomes of $UNG^{-/-}$ CH12F3 cells following CIT stimulation and compared the uracil distribution across the murine *Igh* gene to the unstimulated control. The pulled-down DNA fragments, which should contain uracils, were mapped to the murine genome and their frequency of occurrence in the *Igh* gene is shown in Fig. 2A. The increase in uracilation following CIT stimulation was found concentrated only in Smu and S_α (Salpha) (Fig. 2A). It was highest within Smu increasing from a maximum of ~70 sequencing reads in unstimulated cells to greater than five times that value in stimulated cells. The increase in uracilation following stimulation begins around the mu promoter and drops sharply before the end of Smu (Fig. 2B). There was also uracil accumulation in Salpha, but the magnitude of increase in uracilation in this region following cell stimulation was smaller (3.7-fold, Fig. 2A and B).

The MACS2 algorithm [38] was used to find significant increases in read coverage as a result of stimulation and these peaks of uracilation are shown in Fig. 2B. This analysis showed a broad uracilation peak in Smu and a narrower peak in Salpha (Fig. 2B). Conversely, there was only scattered uracilation in other regions within *Igh*, and none showed much increase following CIT stimulation (Fig. 2A). MACS2 identified no gain peaks in any regions within *Igh* other than Smu and Salpha (Fig. 2A and B; and Supplementary Fig. S4A and B). Notably, there was little accumulation of uracils in the five other switch regions and MACS2 did not identify peaks in any of these switch regions (Fig. 2A). Thus, both the switch regions involved in isotype switching acquire uracils, while the switch regions within *Igh* where mutations or strand breaks do not occur did not show an increase in uracilation. This same general pattern of uracilation increase in Smu and Salpha following stimulation without much increase in other *Igh* regions was also seen in two biological replicates (Supplementary Fig. S4A and B). Following CIT stimulation there were overlapping MACS2 uracilation peaks in Smu in all three biological replicates, while overlapping peaks in Salpha were found in two of the replicates (Supplementary Fig. S4B). In the third replicate, there was an increase in Salpha following CIT stimulation, but no MACS2 peak (Supplementary Fig. S4B). This establishes a direct correlation between AID expression, uracil accumulation in only the relevant switch regions and IgM-to-IgA switching.

AID expressed in CH12F3 cells does not localize to the *Igh* variable region [17] and does not cause SHM [59]. However, it is not known whether there is any accumulation of uracils in VDJ or VJ regions of the Ig genes of this cell line following stimulation. To answer this question, we mapped UPD-seq reads to these regions and subjected them to MACS2 analysis. As seen in Fig. 2B and D, and Supplementary Fig. S4B and

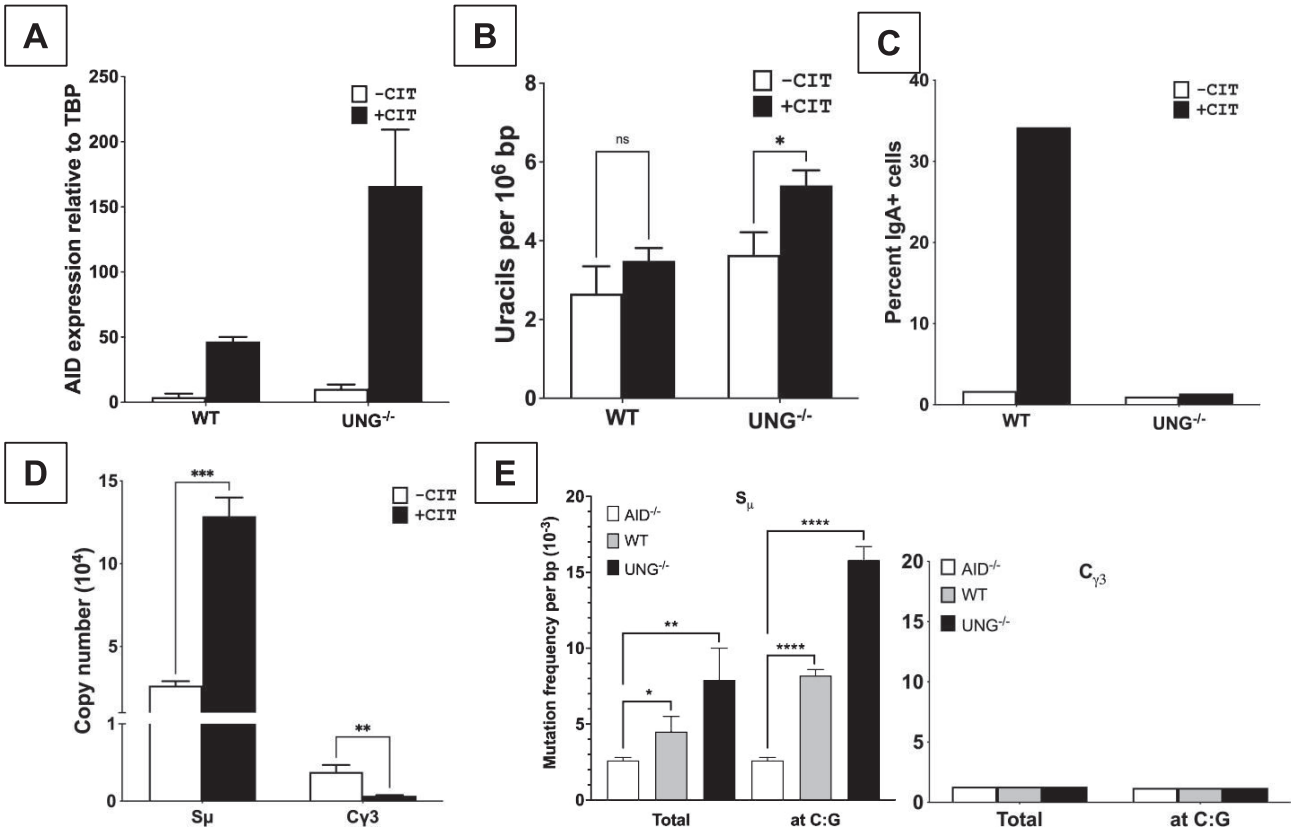


Figure 1. Characterization of CH12F3 cells stimulated to undergo CSR. "-CIT" are cells prior to stimulation and "+CIT" are cells after a 3-day CIT treatment. In panels (A), (B), and (D), the data are from three technical replicates and are presented as mean ± standard deviation. Statistical significance was determined using a two-tailed Student's *t*-test. ns, not significant, $P > 0.05$; "*"- $P < 0.05$; "**"- $P < 0.001$; "***"- $P < 0.0001$. (**A**) Quantification of AID expression using RT-qPCR of RNA isolated from WT and UNG$^{-/-}$ cells. (**B**) Quantification of genomic uracils from DNA isolated from WT and UNG$^{-/-}$ cells. (**C**) Quantification of IgM-to-IgA switching frequency in one biological replicate (Rep 0) using the flow cytometry. Flow cytometry data from three biological replicates are shown in Supplementary Fig. S1. (**D**) The number of copies of Sμ and Cγ3 regions in pull-down DNA from UNG$^{-/-}$ CH12F3 determined using qPCR. (**E**) Quantification of mutation frequencies in Sμ (left) and Cγ3 (right) regions in WT, UNG$^{-/-}$, and AID$^{-/-}$ DNA at day 3 after CIT treatment, for all mutations and mutations at C:G base pairs. Mean and standard deviations from three biological replicates are shown. *P*-values were determined using one-tailed *t*-test and are indicated using asterisks as described above.

C, the uracilation levels remained low even after CIT stimulation in the variable region in *Igh* and the rearranged variable segment in *Igk* (note that the scale for the variable regions is 0 to 200, while that for Smu is 0 to 400). Examination of the whole *Igl* gene also did not reveal much uracil accumulation anywhere in the gene (not shown), but we could not map the uracilation in the transcriptionally active variable region because the VJ rearrangement in this gene has not been characterized in CH12F3. Regardless, MACS2 did not find uracilation gain peaks in either *Igk* or *Igl* genes (Fig. 2D and Supplementary Fig. S4C, and not shown). Thus, the regions in Ig genes that are known not to acquire hypermutations also do not accumulate uracils suggesting that a lack of uracilation—and not error-free repair of uracils—is responsible for the lack of SHM in CH12F3 cells.

It has been proposed that isotype switching may occur either through a sequential mechanism involving an intermediate switching step or a direct mechanism that lacks an intermediate step. IgM-to-IgE switching has been proposed to follow the former mechanism with an intermediate switching step of IgM to IgG1 [60–63]. The lack of detectable uracilation in switch regions between Smu and Salpha (Fig. 2A) suggests that isotype switching in CH12F3 cells does not occur by a sequential mechanism.

When UPD-seq was performed on two independent batches of stimulated AID$^{-/-}$ CH12F3 cells, the levels of uracilation in Smu and Salpha were quite low (Fig. 2C) as would be expected. We also determined uracilation in the *Igh* gene in WT cells and found low levels of uracils regardless of whether the cells were stimulated or not (Supplementary Fig. S4D). This shows that UNG-mediated excision of uracils removes most of the uracils created by AID.

Switch junctions mapped using HTGTS match uracilation peaks

If uracils are necessary for the creation of DSBs that lead to switching [11], how well does the distribution of uracils match the switch junctions? To answer this question, we obtained locations of switch junctions determined in previous studies and compared them with the distribution of uracils determined by UPD-seq. The switch junction studies were divided into two groups: those using high-throughput, genome-wide translocation sequencing (HTGTS) [64], and those using targeted PCR amplification followed by Sanger sequencing (PCR sequencing). The PCR sequencing studies used specific primer pairs, one for Smu and the other for Salpha, to amplify the junctions, and each report contained sequences of only a few

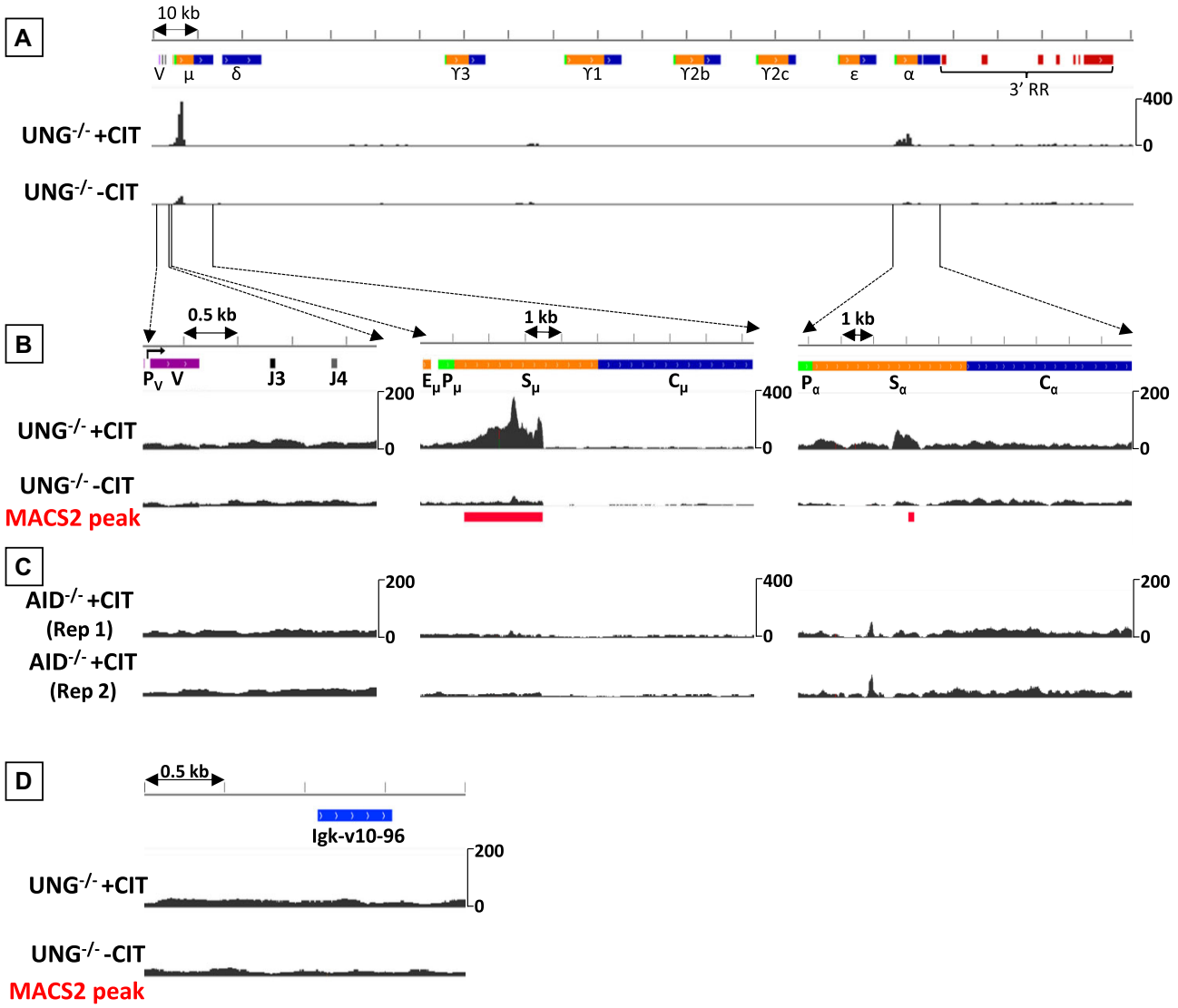


Figure 2. Uracilation in *Ig* genes. (**A**) Distribution of sequencing reads from one biological replicate ("Rep 0") across the rearranged *Igh* gene. UPD-seq coverage of sequencing reads from the unstimulated (-CIT) and stimulated (+CIT) cells are shown. The number of sequencing reads covering each position in *Igh* is shown. The scale is 0 to 400. Top line- The magenta bar highlights the rearranged VDJ variable gene (V), followed by grey bars highlighting unrearranged J regions. Different switch regions (μ, δ, γ3, γ1, γ2b, γ2c, ε, and α) are shown in orange, the corresponding constant segments are shown in blue and the 3′ *Igh* regulatory region (3′ RR) is marked in red. The promoters are indicated in green, and the variable region is shown in purple. This covers chromosome 12 coordinates: (right to left) 114450000–114670000 in BL/6 mouse genome (mm9). Variable region (VDJ): 114666523–114668769, μ segment: 114656710–11466 832, and α segment: 114492979–114503315. (**B**) Expanded views of UPD-seq coverage shown in part A for the variable region, the μ segment and the α segment. The uracilation gain peaks in the stimulated cells relative to unstimulated control identified by MACS2 are shown as red bars below the maps. The rest of the color code is same as in panel (A). P_V, P_μ, and P_α are promoters; E_μ is a collection of enhancers; J3 and J4 are the J segments within the intron between V and the constant segment C_μ; and C_α is the constant segment for the IgA isotype. The scale is 0 to 200 for the V and α regions and 0 to 400 for the μ region. (**C**) UPD-Seq tracks of CH12F3 AID$^{-/-}$ at day 3 following CIT treatment from two independent experiments (Rep1 and Rep2) within the *Igh*, VDJ, μ, and α regions. UPD-Seq tracks in are coverage profiles of high-quality sequencing reads (quality score > 30). (**D**) UPD-seq coverage of the *Igk* rearranged variable segment *Igk*-v10-96. This covers chromosome 6 coordinates: 68581500–68583500 in the BL/6 mouse genome (mm9). The scale is 0 to 200. No MACS2 peaks were found.

switch junctions. To get a more representative picture of where the junctions are located, we combined the results from several PCR sequencing studies [46–52]. In contrast, the HTGTS studies used Roche-454 [64] or Illumina [45] sequencing and mapped the sequencing reads to the mouse genome, reporting hundreds of switch junctions. One of the HTGTS study used IgM-to-IgA switching in CH12F3 cells [44], while the other studied *ex vivo* stimulated murine splenocytes that switched from IgM to principally to IgG1 [45]. In the latter case, we used data from only the Smu side of the junctions for *Dis3*$^{+/-}$ mice [45]. To smooth out stochastic variation in the positions of switch junctions, we added together the junctions in 60-bp windows across the two switch regions.

When the locations of HTGTS junction sequences were compared to the accumulation of uracilated DNA fragments, there were remarkable similarities in the two distributions. In Smu, both the switch junctions and uracilation increase from the start of the switch region peaking around its middle and then decrease rapidly to background levels before the 3′ edge of the Smu core (Fig. 3A). Additionally, the switch junctions

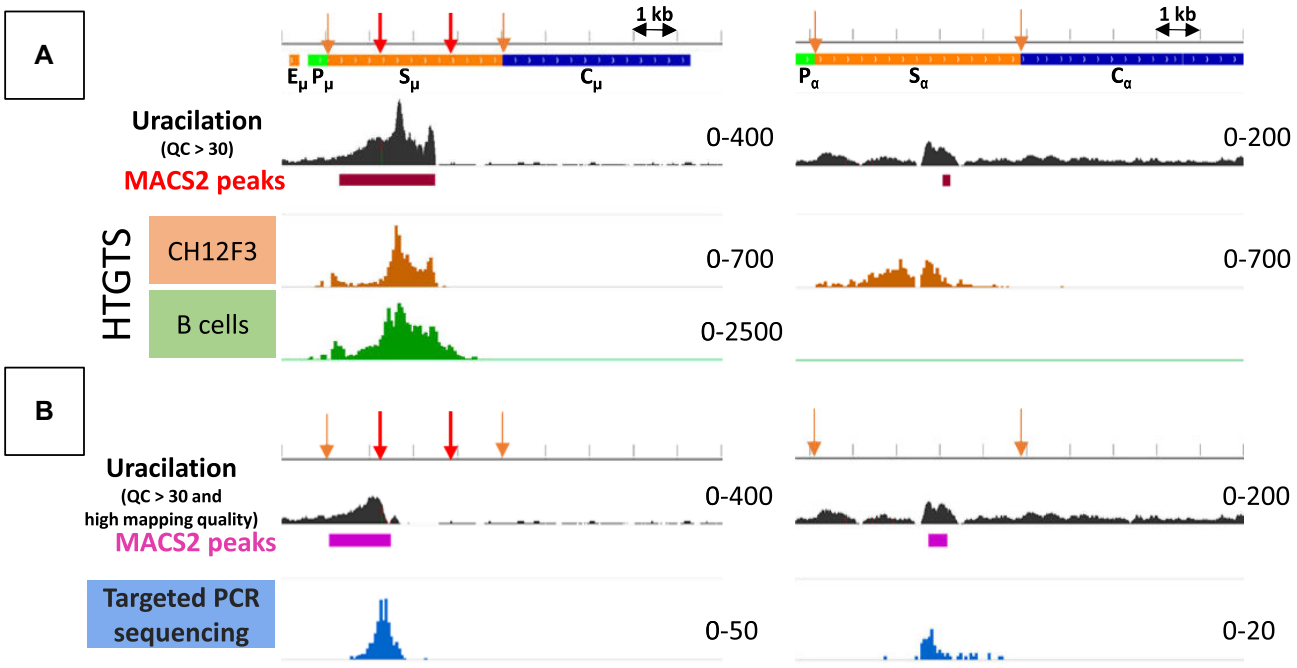


Figure 3. Comparison of distributions of uracilation and switch junctions The top line in each part shows the Mu and Alpha regions. Smu and Salpha region boundaries are shown using orange downward arrows, and the Smu and Salpha core region boundaries are shown using red downward arrows. (**A**) Comparison of distribution of high-quality sequencing reads from UPD-seq with switch junction positions in Smu and Salpha regions determined using HTGTS. The range for each region is shown on the right. The MACS2 peaks are shown as dark red boxes below uracilation distribution. (**B**) Comparison of distribution of high-quality well-mapped sequencing reads from UPD-seq with switch junction positions in Smu and Salpha regions determined using Sanger sequencing (targeted PCR sequencing). The MACS2 peaks are shown as purple boxes below the UPD-seq data. The switch junctions were binned using 60 bp windows and the range in the number of sequencing reads or junctions is shown next to each plot.

in CH12F3 cells showed the double-peak shape seen in all three Smu uracilation profiles (Fig. 3A and Supplementary Fig. S4B). There were multiple uracilation humps in Salpha, two of which roughly matched with the two peaks in the junction sequence profile, and MACS2 identified one of them as a peak (Fig. 3A). In the biological replicate labeled Rep 1, MACS2 found two peaks (Supplementary Fig. S4B) very similar to those in the switch junctions (Fig. 3A).

Switch junctions determined using PCR sequencing match uracilation with high mapping quality

Both the UPD-seq and HTGTS techniques sequence DNA fragments with unknown ends and then map the obtained sequences to the genome. Even though we selected only high-quality sequencing reads (Q-score > 30), mapping the reads to the "core" of Smu is challenging because of the highly repetitive nature of this region (see below). To reduce this problem, we filtered out UPD-seq sequencing reads that had low mapping quality and confidence (hereafter called mapping quality, Supplementary Fig. S5). This dramatically changed the observed uracilation distribution.

Filtration for mapping quality eliminated most of the sequencing reads from the core Smu and moved the uracilation humps and associated MACS2 peak toward the 5′ side of Smu (Fig. 3B and Supplementary Fig. S4B). The post-filtration MACS2 peak begins at the start of Smu and ends near the 5′ edge of the core. In Salpha the effect of the mapping filter was subtler, only somewhat reducing the heights of uracilation peaks. Interestingly, MACS2 assigned a broader peak Salpha after the filtration (Fig. 3B), presumably because of the removal of low mapping quality reads prior to differential peak calling.

In both Smu and Salpha, the switch junctions determined by PCR sequencing matched better with the post-filtration MACS2 peaks than the pre-filtration peaks (Fig. 3A and B). Although the number of switch junctions determined using PCR sequencing is much smaller than the HTGTS, the distribution of the former set of switch junctions matched well with the post-filtration MACS2 peaks and the density of the junctions was similar to the density of uracilation within these peaks. Thus, mapping quality filtration creates uracilation data with greater mapping confidence that matches well with the intrinsically high-quality mapping of PCR-based sequencing of switch junctions.

Repetitive sequences and R-loops do not show a strong correlation with uracilation

The switch regions contain many repetitive sequences [65, 66], and we mapped the frequency of occurrence of six different repetitive sequences across Smu and Salpha using 60-bp windows (Fig. 4A). These included preferred targets of AID (WRC, WRCY, and AGCT), repeats that may promote R-loop formation (GGG and GGGG) and two G-quadruplex promoting sequences [22, 67]. While the Smu core had the highest concentration of all seven repeats, other parts of Smu also contained multiple repetitive sequences. There was good correlation between the frequency of repeats and uracilation in the Smu core, but the correlation was poor outside the core. WRC and WRCY sequences are found at high frequencies on the 3′ side of core Smu, but few UPD-seq reads map in this region. The correlation between repetitive sequences and uracilation is worse in Salpha. The WRC, WRCY and GGG repeats are scattered throughout Salpha and GGGG repeats are present mostly at the 3′ half of Salpha (Fig. 4A). However,

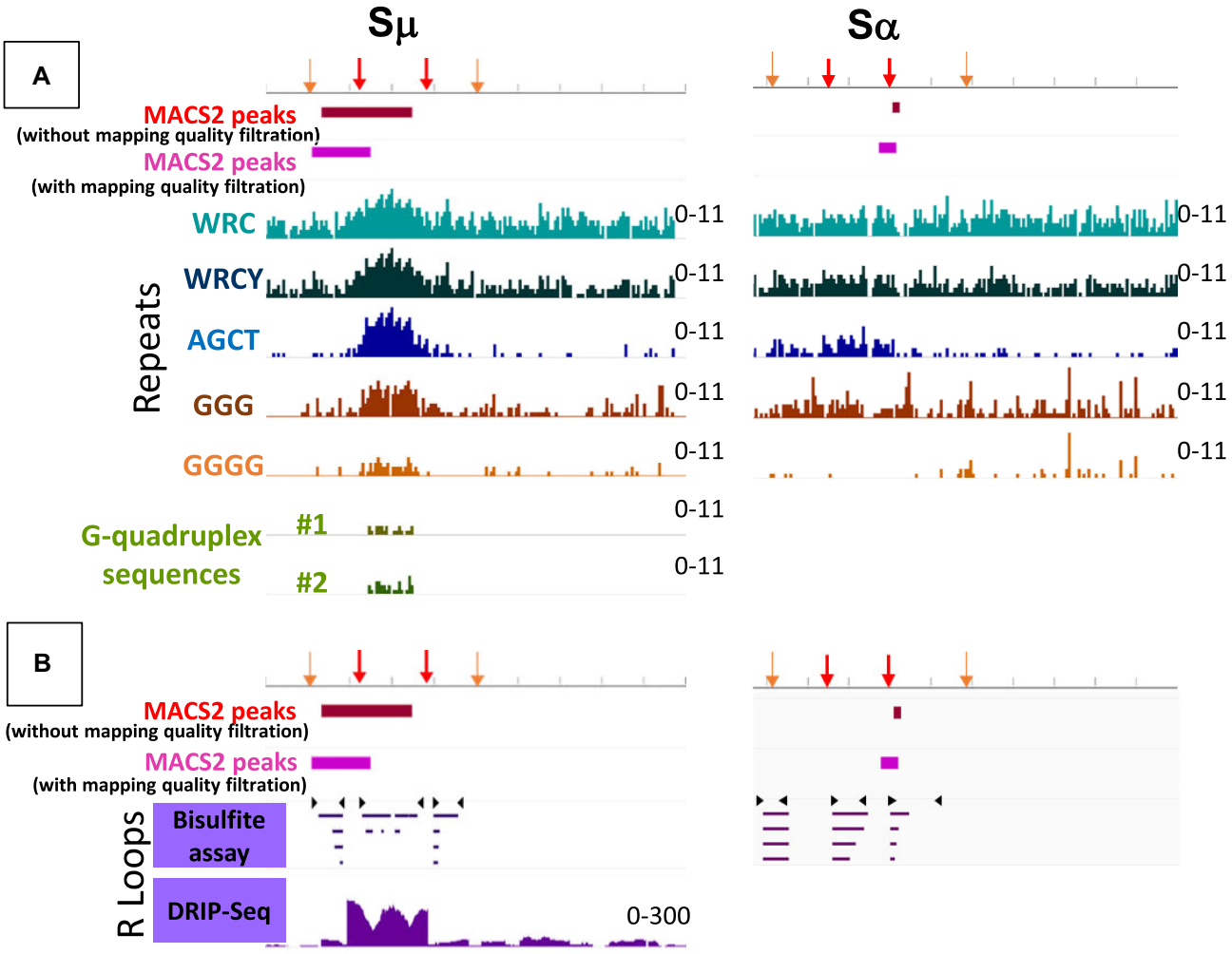


Figure 4. Relationship between uracilation, repetitive sequences and R-loops. The top line in each part shows the Mu and Alpha regions. Smu and Salpha region boundaries are shown using orange downward arrows, and the Smu and Salpha core region boundaries are shown using red downward arrows. The MACS2 peaks for first data set (Rep 0) are shown below. (**A**) Comparison of MACS2 peaks determined using UPD-seq reads, with and without prior filtration of low mapping-quality reads with repetitive sequences. The distribution of repeats of sequences WRC (W is A or T), WRCY (Y is a pyrimidine), AGCT, GGG, GGGG, and of the G quadruplex forming sequences GGGGTGAGCTGAGCTGAGCT [22] and GAGCTGAGCTGGGGTGAGCT (respectively #1 and #2 in the figure, [22, 67]) is shown in 60-bp windows across the Smu and Salpha regions. The range of the number of sequencing reads or repeats is shown next to each plot. (**B**) Comparison of coverage of MACS2 peaks determined using UPD-seq reads with and without prior filtration of low mapping quality reads with the distribution of R-loop forming regions analyzed in activated B cells (Smu) or CH12F3 cells (Salpha) using either bisulfite treatment followed by sequencing (top) or DRIP sequencing (bottom). For the bisulfite-based mapping, only representative loops are displayed to show the extent of coverage of the regions by R-loops. The primers used in the PCR amplification of DNA are indicated in black triangles.

MACS2 identifies uracilation peaks in only one or two narrow regions near the center of Salpha (Figs. 2B and 3B; and Supplementary Fig. S4B). The putative quadruplex forming sequences are found only within the Smu core, but uracilation extends to the 5′ side of the core (Fig. 4A). Furthermore, the quadruplex forming sequences are completely absent from Salpha (Fig. 4A). Thus, the occurrence of various repetitive sequences in Smu and Salpha does not fully explain uracilation of these regions.

The Smu and Salpha regions have been shown to form R-loops [53, 54], and an increase in the frequency of G-repeats promotes R-loop formation and increases isotype switching [68, 69]. R-loops contain single-stranded DNA that is thought to be a good substrate for cytosine deamination by AID [23]. To gauge the role of R-loop formation in uracilation by AID, we compared our uracilation maps with previously mapped R-loops in the Smu region of splenic B cells and Salpha in CH12F3 (Fig. 4B). The R-loops determined using bisulfite sensitivity [53] start ~200 bp downstream of the intronic promoter, cover the entire Smu core, and extend ~700 bp to the 3′ side of the core (Fig. 4B). Uracilation of Smu, as indicated by the MACS2 peak, has roughly the same 5′ edge as the R-loops, but ends several hundred base pairs before 3′ edge of the mapped R-loops. The R-loops mapped in Salpha region of CH12F3 cells [54] also do not correlate well with MACS2 uracilation peaks. While we find that enhanced uracilation due to AID expression occurs in the middle of the Salpha core, R-loops have been found on either side of the core sequence [54].

Leffleur *et al.* stimulated murine B cells *ex vivo* using IL-4 causing a switch from IgM to IgG1 and mapped R-loops in Smu using DRIP-seq [45]. The R-loops determined by DRIP-seq in *Dis3*$^{+/-}$ mice [45] also extended beyond the 3′ edge of the uracilation peaks, but did not extend as far as the peaks on

the 5′ side (Fig. 4B). The frequency of R-loops determined by DRIP-seq had a "W" shape which is not reflective of uracilation in Smu. Thus, while the mapped R-loops in Smu do cover most of the uracilated regions, the apparent presence of R-loops on the 3′ side of the Smu core and beyond, and the presence of R-loops on either side of Salpha core are not consistent with the UPD-seq results.

It is possible that during UPD-seq DNA in R-loops is not efficiently pulled down causing under-representation of this DNA in subsequent sequencing steps. We consider this to be unlikely because R-loops contain a single-stranded DNA (ssDNA) and an RNA:DNA hybrid and uracils in either structure should be highly susceptible to excision by Ung to create abasic sites that would react with ssARP. Uracils in ssDNA are better substrates for Ung than those in double-stranded DNA [70] and human Ung is known to excise uracils from the DNA:RNA hybrid intermediate formed during HIV replication resulting in degradation of the viral genome [71].

S_μ, but not S_α, contains many closely spaced uracils on opposite strands

When the conversion of cytosine to uracil in $UNG^{-/-}$ cells is followed by replication, the original C:G is changed to a U:A or a T:A pair. Additionally, during UPD-seq the pulled-down DNA fragments are amplified for NextGen sequencing and the Taq DNA polymerase used for the amplification inserts an adenine across from the abasic site creating a C:G-to-T:A change (Supplementary Fig. S3B and [30]). Thus a C:G-to-T:A change in the sequencing reads is evidence of a C to U conversion by AID in that base pair.

We cataloged all C:G-to-T:A changes in high mapping quality reads within MACS2 peaks and determined the strand in which the putative cytosine deamination occurred in each sequencing read. The results are shown in Fig. 5 and Table 1. The number of C:G-to-T:A changes in both Smu and Salpha peaks was higher following stimulation (Fig. 5A), and this is consistent with the increased pull-down in these regions (Fig. 1D). In both Smu and Salpha, the C:G-to-T:A changes were highest in the WRCY sequence context (Fig. 5A), which is consistent with the known sequence preference of AID [72]. In Smu, the percentage of sequence changes that were C:G-to-T:A increased in the Smu peak for all three replicates following stimulation and the ratio of C to T changes in the two DNA strands was close to 1 suggesting a lack of strong strand bias in the deaminations (Table 1 and Fig. 5B). The number of C:G-to-T:A changes in Salpha peaks was much smaller and hence those numbers are less reliable for determining percentage increases following stimulation or quantifying the strand bias in deamination. When a *t*-test was applied to the number of base changes in the two strands, the differences were not statistically significant (*P*-value > 0.5). The number of C:G-to-T:A changes in Cmu and Calpha was much smaller than in the corresponding switch regions and in most cases did not increase upon cell stimulation (Table 1). This is despite the fact that constant segments are several times larger than the MACS2 peaks in the switch regions. This shows that AID does not target the constant regions within *Igh*.

If a C-to-T change occurs close to a G-to-A change in the same sequencing read, it is indicative of two closely spaced uracils on opposite strands. These deoxyuridines could be converted to a DSB if UNG and an AP endonuclease were to act on them in a short span of time. As an illustration, if strand breaks occur on opposite strands separated by the sequence GGGGAGCT, a sequence frequently found within the switch regions, the melting temperature (T_m) of the nicked DNA would be ∼33°C and hence the two strands should separate under physiological conditions creating a DSB. In contrast, if the strand breaks are separated by GGGGTGAGCTAGCT, another common sequence in the switch regions, the T_m would be 58°C and the DNA should remain double-stranded. To locate potential DSBs due to paired cytosine deaminations on opposite strands, we identified C-to-T and G-to-A changes within in high-mapping-confidence UPD-seq reads and binned them based on the distance between the base changes. The results are shown in Supplementary Table S1 and Fig. 5C.

There were many C-to-T/G-to-A double changes a short distance apart in the Smu peak, especially following stimulation (Fig. 5C). There were nearly 600 such pairs within all the sequence reads that mapped in the Smu MACS2 peaks in the three replicates. In over 50 such pairs, the two sequence changes were <9 bp apart (Supplementary Table S1 and Fig. 5C). Thus, it is easy to visualize double deaminations coupled with the action of UNG and APE1 resulting in DSBs in Smu.

Such DSBs would be much less feasible in Salpha. There were no paired C-to-T/G-to-A double changes within MACS2 peaks in Salpha in unstimulated cells, and upon stimulation the number of such pairs rose to only 10% those seen in the Smu MACS2 peak (Supplementary Table S1 and Fig. 5C). Only four such pairs in MACS2 peaks in Salpha were <9 bp apart. Thus, it is likely that DSBs due to sequential action of UNG and AP endonuclease occur much less frequently in Salpha than Smu. Successful isotype switching from IgM to IgA requires that DSBs in Smu and Salpha occur nearly simultaneously and hence the paucity of potential DSBs in Salpha may limit CSR frequency unless other mechanisms increase the frequency of occurrence of DSBs in Salpha.

Discussion

We have shown here that uracils created by AID in a murine cell line undergoing CSR are largely concentrated in the two switch regions that are known to suffer DSBs during isotype switching. However, there are two reasons to think that the uracilation maps created in these studies are likely to underestimate the amount of uracils created by AID. First, Maul *et al.* [29] showed that the amount of UNG-sensitive DNA within Smu was highest one day post-stimulation and decreased thereafter. We used cells three days post-stimulation for UPD-seq because isotype switching culminates around this time allowing us to make a more direct comparison between uracil accumulation and IgM-to-IgA switching. It is possible that some of the uracils created by AID are lost through excision by DNA glycosylases that act as back-ups to UNG [9] or by replication-independent mismatch repair (MMR; [73, 74]). Thus it is likely that we will find even higher levels of uracilation in the Smu and Salpha regions in the first 24 h post-stimulation. Second, during UPD-seq the library DNA is amplified using the Taq polymerase for 8 cycles, during which the small chemical scar left behind during the enrichment step must be bypassed by the polymerase (Supplementary Fig. S3B). Although Taq polymerase does bypass this scar and insert adenine across from it, some stalling of the polymerase at the scar does occur [30]. Consequently, any DNA lacking uracils that may be contaminating the pulled-down DNA may

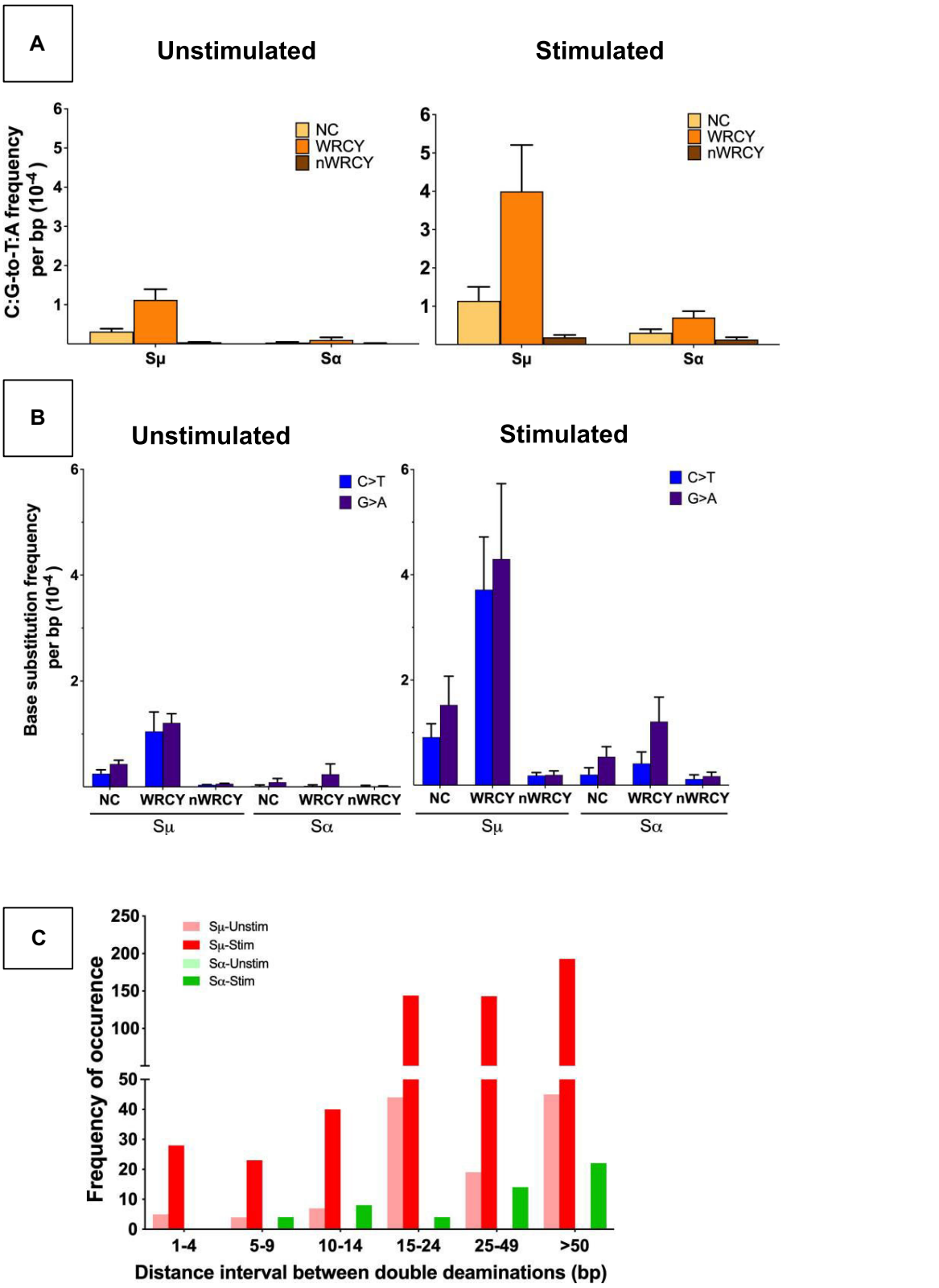


Figure 5. Frequency of single and double deamination events. The normalized frequencies of C-to-T and G-to-A changes and adjacent C-to-T/G-to-A double changes within MACS2 peaks from high mapping quality reads are presented. (**A**) C:G to T:A changes in different sequence contexts. Mean and standard deviation from three biological replicates are shown. N is any base, W is A or T, R is a purine, and Y is a pyrimidine. nWRCY counts cytosines in sequences other than WRCY. (**B**) The normalized frequencies of C-to-T and G-to-A changes in different sequence contexts. (**C**) The frequency of C-to-T/G-to-A double changes in MACS2 peaks from high mapping quality reads is plotted against distance between the deamination events. The double deamination events were binned as indicated on the *X*-axis. A G-to-A change immediately following a C-to-T change is counted as 1 bp apart; a G-to-A change and a C-to-T change separated by an unchanged base pair is counted as 2 bp apart, and so on.

Table 1. C:G to T:A changes within MACS2 peaks

Region	Length (bp)	Condition	Rep 0			Rep 1			Rep 2		
			C:G to T:A Count	C:G to T:A percentage[a]	NTS/TS*	C:G to T:A Count	C:G to T:A percentage[a]	NTS/TS[b]	C:G to T:A Count	C:G to T:A percentage[a]	NTS/TS[b]
Smu peak	1411	Unstimulated	165	65.2	0.99	133	61.6	0.82	151	52.2	1.1
		Stimulated	892	76.2	0.87	774	69.4	0.88	400	63.7	0.8
Salpha peak1	426	Unstimulated	15	68.2	14.00	15	83.3	1.50			
		Stimulated	138	80.7	1.88	87	56.1	2.11			
Salpha peak2	599	Unstimulated				11	47.8	0.22			
		Stimulated				208	53.9	0.54			
Cmu	4220	Unstimulated	7	9.9	2.50	7	5.83	6.00	7	4.0	1.33
		Stimulated	6	7.1	1.00	5	5.26	0.67	5	2.6	1.50
Calpha	5455	Unstimulated	17	9.4	0.55	10	6.29	0.43	18	5.0	1.57
		Stimulated	25	13.0	0.92	27	9.78	1.08	18	4.9	1.57

[a]C:G to T:A changes as percentage of all the base changes
[b]NTS is nontranscribed strand; TS is transcribed strand

amplify better than the DNA containing a scar and amplification of uracil-containing fragments may worsen as the number of uracils in DNA fragment increases. Thus, the switch region uracilation seen in our experiments is likely to be lower than the actual total amount of uracilation.

On the other hand, it could be argued that our decision to examine uracilation at 3 days post-stimulation may have allowed uracils to accumulate in the switch regions. This may create an overestimation of closely spaced C-to-T/G-to-A double changes suggestive of double deamination events that could cause DSBs. In WT cells if the uracils are rapidly repaired through BER they would not accumulate and the likelihood of creating closely spaced single-strand breaks on opposite strands may be small. This points to the importance of FAM72A in reducing uracil excision through degradation of UNG2 [56, 75] and allowing accumulation of uracils in the switch regions. Harvesting $UNG^{-/-}$ CH12F3 cells a few hours after AID expression should provide a truer picture of the frequency of C-to-T/G-to-A changes that could cause DSBs.

Highly repetitive sequences present two challenges for Illumina Next-Gen sequencing [76]. First, it is difficult to accurately determine the number of repeats within such sequences and second, the repetitive nature of the regions may prevent accurate mapping of the sequences to unique genomic locations [77]. We recognized that this problem also exists in comparing distribution of uracils determined by UPD-seq to previously determined positions of switch junctions in the switch regions. To circumvent this, we filtered out sequencing reads with low mapping confidence obtained during UPD-seq and compared the distribution of high mapping quality reads with switch junctions determined using PCR-based method. The PCR-based method amplifies pre-determined regions of the genome we show that their distribution correlates with peaks of high mapping quality reads from UPD-seq (Fig. 3B).

The UPD-seq results presented here have provided several insights into the molecular mechanism of CSR. First, the highly focused uracilation of the Smu and Salpha regions in $UNG^{-/-}$ cells eliminates the possibility that AID creates uracils in large swaths of the switch region, but U•G pairs are repaired by BER in all regions except Smu and Salpha. In other words AID targeting, and not the differential removal of uracils from different switch regions, confers isotype specificity in CSR. Differential repair of genomic regions explains why some non-Ig genes suffer SHM, while others escape it [27], but it does not explain the switching specificity of CSR.

Two mechanisms have been proposed for the creation of strand breaks through deamination of cytosines in DNA (Supplementary Fig. S6) [11, 78–80]. One involves processing of uracils by UNG and AP endonuclease (UNG-only mechanism; Supplementary Fig. S6A). Alternately, MMR may recognize the U•G mispairs, create strand breaks in their vicinity and initiate long-patch repair (MMR-dependent mechanism, Supplementary Fig. S6B) [81]. When breaks in opposite strands occur close together or are brought into close proximity through the action of Exonuclease 1 during long-patch repair [82], this creates double-strand breaks (Supplementary Fig. S6B). The greater abundance of closely spaced C-to-T/G-to-A double sequence changes (Fig. 5C) in Smu suggests that the UNG-only mechanism may be sufficient to create frequent DSBs in Smu and trigger CSR. In contrast, the relative paucity of closely spaced C-to-T/G-to-A double sequence changes in Salpha (Fig. 5C and Table 1) suggests that creation of DSBs within this region is likely to rely much more on the MMR-dependent mechanism. Unlike the UNG-only mechanism (Supplementary Fig. S6A), the two cytosine deaminations that cause DSBs through the MMR-dependent mechanism can be several hundred base pairs apart and can be on the same DNA strand (Supplementary Fig. S6B) [73, 74].

A few studies have determined the footprint of AID by expressing it in uracil repair-deficient B cells, amplifying the relevant regions, cloning the product in *E. coli* and mapping C:G to T:A mutations using Sanger sequencing. Two such studies mapped mutations found in the "pre-Smu" region (DNA between Imu and the Smu core) and found that the mutations were more frequent near the Smu core than in Imu [83, 84], which is consistent with MACS2 analysis of high mapping quality reads (Fig. 3B and Supplementary Fig. S4B). One of these studies also mapped mutations in a shortened Salpha region within a CH12F3 derivative [83] and found that the mutations were spread evenly throughout the shortened Salpha [83]. The lack of a mutational peak in Salpha in this study may be due to the small sample size (~30 clones for each condition) and the long time period over which the cells were grown (1–4 weeks). This study also mapped mutations that accumulate in the shortened Salpha during the first 24 h after cell stimulation, and found that only one out of the 16 clones contained a pair of closely spaced C-to-T and G-to-A mutations [83, 84]. This is consistent with our findings regarding the paucity of double mutations in the Salpha region (Fig. 5C and Supplementary Table S1).

A variety of sequence repeats and non-B structures in the switch regions have been invoked to explain how AID may specifically target these regions and how the uracils are converted to DSBs [2, 23, 80, 85–88]. While our data does not eliminate the possibility that one or more of these features plays a role in the action of AID, none explains the selective enhancement of uracilation in only parts of the switch regions. In particular, these sequences and proposed structures do not explain why some regions within Smu and Salpha display poor uracilation. Either a creative synthesis of these hypotheses or a new hypothesis is needed to explain the uracilation data. Mapping genomic uracilation using UPD-seq should help in testing any future hypotheses regarding the molecular mechanisms of CSR and SHM.

Acknowledgements

A.S.B. would like to thank Dr. Patricia J. Gearhart (National Institute on Aging) for discussions about the project.

Author contributions: A.S.B. conceived and oversaw this research project, helped plan experimental work and data analysis, and wrote parts of the manuscript. R.M.-R. did most of the experimental work and helped with data analysis. She also wrote parts of the manuscript. R.S. did most of the bioinformatic data analysis and wrote parts of the manuscript. Y.B. did some of the experimental work. M.S.L. oversaw bioinformatic data analysis.

Supplementary data

Supplementary data is available at NAR Molecular Medicine online.

Conflict of interest

None declared.

Funding

The research in the Bhagwat lab was supported by NIH grants (R21AI144708, R21CA252858, and R21AI180183) and Wayne State University. Research in the Lawrence lab was supported by Rullo Family Innovation Award and NIH grant R01CA262874.

Data availability

The source code is released with a permanent DOI- On GitHub: https://github.com/rayanramin/CH12-UPDSeq and on Zenodo: https://doi.org/10.5281/zenodo.14853312. The DNA sequences are submitted to SRA with the BioProject : PRJNA1139184. These sequences are scheduled to be released on 31 August 2025 or upon publication (whichever comes first). There are 10 samples in this BioProject.

References

1. Ridani J, Barbulescu P, Martin A *et al.* In: Honjo T, Reth M, Radbruch A, Alt FW, Martin A (eds.), *Molecular Biology of B Cells*. San Diego,CA: Academic Press, 2024, 235–56.
2. Meng F-L, Nair L, Fernandez KC *et al.* In: Honjo T, Reth M, Radbruch A, Alt FW, Martin A (eds.), *Molecular Biology of B Cells*. San Diego, CA: Academic Press, 2024, 213–34.
3. Muramatsu M, Kinoshita K, Fagarasan S *et al.* Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 2000;**102**:553–63. https://doi.org/10.1016/S0092-8674(00)00078-7
4. Bransteitter R, Pham P, Scharff MD *et al.* Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc Natl Acad Sci USA* 2003;**100**:4102–7. https://doi.org/10.1073/pnas.0730835100
5. Chaudhuri J, Tian M, Khuong C *et al.* Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* 2003;**422**:726–30. https://doi.org/10.1038/nature01574
6. Dickerson SK, Market E, Besmer E *et al.* AID mediates hypermutation by deaminating single stranded DNA. *J Exp Med* 2003;**197**:1291–6. https://doi.org/10.1084/jem.20030481
7. Sohail A, Klapacz J, Samaranayake M *et al.* Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. *Nucleic Acids Res* 2003;**31**:2990–4. https://doi.org/10.1093/nar/gkg464
8. Rush JS, Fugmann SD, Schatz DG. Staggered AID-dependent DNA double strand breaks are the predominant DNA lesions targeted to S mu in Ig class switch recombination. *Int Immunol* 2004;**16**:549–57. https://doi.org/10.1093/intimm/dxh057
9. Visnes T, Doseth B, Pettersen HS *et al.* Uracil in DNA and its processing by different DNA glycosylases. *Phil Trans R Soc B* 2009;**364**:563–8. https://doi.org/10.1098/rstb.2008.0186
10. Allen D, Cumano A, Dildrop R *et al.* Timing, genetic requirements and functional consequences of somatic hypermutation during B-cell development. *Immunol Rev* 1987;**96**:5–22. https://doi.org/10.1111/j.1600-065X.1987.tb00506.x
11. Petersen-Mahrt SK, Harris RS, Neuberger MS. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature* 2002;**418**:99–104. https://doi.org/10.1038/nature00862
12. Phung QH, Winter DB, Cranston A *et al.* Increased hypermutation at G and C nucleotides in immunoglobulin variable genes from mice deficient in the MSH2 mismatch repair protein. *J Exp Med* 1998;**187**:1745–51. https://doi.org/10.1084/jem.187.11.1745
13. Rada C, Ehrenstein MR, Neuberger MS *et al.* Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity* 1998;**9**:135–41.https://doi.org/10.1016/S1074-7613(00)80595-6
14. Peters A, Storb U. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity* 1996;**4**:57–65. https://doi.org/10.1016/S1074-7613(00)80298-8
15. Stavnezer-Nordgren J, Sirlin S. Specificity of immunoglobulin heavy chain switch correlates with activity of germline heavy chain genes prior to switching. *EMBO J* 1986;**5**:95–102. https://doi.org/10.1002/j.1460-2075.1986.tb04182.x
16. Nambu Y, Sugai M, Gonda H *et al.* Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science* 2003;**302**:2137–40. https://doi.org/10.1126/science.1092481
17. Matthews AJ, Husain S, Chaudhuri J. Binding of AID to DNA does not correlate with mutator activity. *J Immunol* 2014;**193**:252–7. https://doi.org/10.4049/jimmunol.1400433
18. Maul RW, Cao Z, Venkataraman L *et al.* Spt5 accumulation at variable genes distinguishes somatic hypermutation in germinal center B cells from *ex vivo*-activated cells. *J Exp Med* 2014;**211**:2297–306. https://doi.org/10.1084/jem.20131512
19. Canugovi C, Samaranayake M, Bhagwat AS. Transcriptional pausing and stalling causes multiple clustered mutations by human activation-induced deaminase. *FASEB J* 2009;**23**:34–44. https://doi.org/10.1096/fj.08-115352
20. Pavri R, Gazumyan A, Jankovic M *et al.* Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* 2010;**143**:122–33. https://doi.org/10.1016/j.cell.2010.09.017
21. Methot SP, Litzler LC, Subramani PG *et al.* A licensing step links AID to transcription elongation for mutagenesis in B cells. *Nat*

Commun 2018;**9**:1248. https://doi.org/10.1038/s41467-018-03387-6
22. Qiao Q, Wang L, Meng FL *et al.* AID recognizes structured DNA for class switch recombination. *Mol Cell* 2017;**67**:361–73. https://doi.org/10.1016/j.molcel.2017.06.034
23. Yu K, Lieber MR. Current insights into the mechanism of mammalian immunoglobulin class switch recombination. *Crit Rev Biochem Mol Biol* 2019;**54**:333–51. https://doi.org/10.1080/10409238.2019.1659227
24. Zheng S, Vuong BQ, Vaidyanathan B *et al.* Non-coding RNA generated following lariat debranching mediates targeting of AID to DNA. *Cell* 2015;**161**:762–73. https://doi.org/10.1016/j.cell.2015.03.020
25. Basu U, Meng FL, Keim C *et al.* The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell* 2011;**144**:353–63. https://doi.org/10.1016/j.cell.2011.01.001
26. Samaranayake M, Bujnicki JM, Carpenter M *et al.* Evaluation of molecular models for the affinity maturation of antibodies: roles of cytosine deamination by AID and DNA repair. *Chem Rev* 2006;**106**:700–19. https://doi.org/10.1021/cr040496t
27. Liu M, Duke JL, Richter DJ *et al.* Two levels of protection for the B cell genome during somatic hypermutation. *Nature* 2008;**451**:841–5. https://doi.org/10.1038/nature06547
28. Begum NA, Kobayashi M, Nagaoka H *et al.* In Honjo T, Reth M, Radbruch A, Alt FW, Martin A (eds.), *Molecular Biology of B Cells*. San Diego, CA: Academic Press, 2024, 257–308.
29. Maul RW, Saribasak H, Martomo SA *et al.* Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions. *Nat Immunol* 2011;**12**:70–6. https://doi.org/10.1038/ni.1970
30. Sakhtemani R, Senevirathne V, Stewart J *et al.* Genome-wide mapping of regions preferentially targeted by the human DNA-cytosine deaminase APOBEC3A using uracil-DNA pulldown and sequencing. *J Biol Chem* 2019;**294**:15037–51. https://doi.org/10.1074/jbc.RA119.008053
31. Siriwardena SU, Perera MLW, Senevirathne V *et al.* A tumor-promoting phorbol ester causes a large increase in APOBEC3A expression and a moderate increase in APOBEC3B expression in a normal human keratinocyte cell line without increasing genomic uracils. *Mol Cell Biol* 2019;**39**:e00238-18. https://doi.org/10.1128/MCB.00238-18
32. Wei S, Perera MLW, Sakhtemani R *et al.* A novel class of chemicals that react with abasic sites in DNA and specifically kill B cell cancers. *PLoS One* 2017;**12**:e0185010. https://doi.org/10.1371/journal.pone.0185010
33. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;**25**:1754–60. https://doi.org/10.1093/bioinformatics/btp324
34. Li H, Handsaker B, Wysoker A *et al.* The sequence alignment/Map format and SAMtools. *Bioinformatics* 2009;**25**:2078–9. https://doi.org/10.1093/bioinformatics/btp352
35. Tarasov A, Vilella AJ, Cuppen E *et al.* Sambamba: fast processing of NGS alignment formats. *Bioinformatics* 2015;**31**:2032–4. https://doi.org/10.1093/bioinformatics/btv098
36. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 2008;**18**:1851–8. https://doi.org/10.1101/gr.078212.108
37. Robinson JT, Thorvaldsdóttir H, Winckler W *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011;**29**:24–6. https://doi.org/10.1038/nbt.1754
38. Zhang Y, Liu T, Meyer CA *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;**9**:R137. https://doi.org/10.1186/gb-2008-9-9-r137
39. Khanna A, Larson DE, Srivatsan SN *et al.* Bam-readcount - rapid generation of basepair-resolution sequence metrics. arXiv, https://arxiv.org/abs/2107.12817, 27 July 2021, preprint: not peer reviewed.
40. Morgan M, Pagès H, Obenchain V *et al.* Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. R package version 2.22.0. 2024. https://bioconductor.org/packages/Rsamtools
41. Lawrence M, Huber W, Pagès H *et al.* Software for computing and annotating genomic ranges. *PLoS Comput Biol* 2013;**9**:1003118. https://doi.org/10.1371/journal.pcbi.1003118
42. Wu L, Shukla V, Yadavalli AD *et al.* HMCES protects immunoglobulin genes specifically from deletions during somatic hypermutation. *Genes Dev* 2022;**36**:433–50. https://doi.org/10.1101/gad.349438.122
43. Dong J, Panchakshari RA, Zhang T *et al.* Orientation-specific joining of AID-initiated DNA breaks promotes antibody class switching. *Nature* 2015;**525**:134–9. https://doi.org/10.1038/nature14970
44. Panchakshari RA, Zhang X, Kumar V *et al.* DNA double-strand break response factors influence end-joining features of IgH class switch and general translocation junctions. *Proc Natl Acad Sci USA* 2018;**115**:762–7. https://doi.org/10.1073/pnas.1719988115
45. Laffleur B, Lim J, Zhang W *et al.* Noncoding RNA processing by DIS3 regulates chromosomal architecture and somatic hypermutation in B cells. *Nat Genet* 2021;**53**:230–42. https://doi.org/10.1038/s41588-020-00772-0
46. Bardwell PD, Woo CJ, Wei K *et al.* Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1–mutant mice. *Nat Immunol* 2004;**5**:224–9. https://doi.org/10.1038/ni1031
47. Cook AJ, Oganesian L, Harumal P *et al.* Reduced switching in SCID B cells is associated with altered somatic mutation of recombined S regions. *J. Immunol* 2003;**171**:6556–64. https://doi.org/10.4049/jimmunol.171.12.6556
48. Han L, Yu K. Altered kinetics of nonhomologous end joining and class switch recombination in ligase IV-deficient B cells. *J Exp Med* 2008;**205**:2745–53. https://doi.org/10.1084/jem.20081623
49. Reina-San-Martin B, Difilippantonio S, Hanitsch L *et al.* H2AX is required for recombination between immunoglobulin switch regions but not for intra-switch region recombination or somatic hypermutation. *J Exp Med* 2003;**197**:1767–78. https://doi.org/10.1084/jem.20030569
50. Rivera-Munoz P, Soulas-Sprauel P, Le Guyader G *et al.* Reduced immunoglobulin class switch recombination in the absence of Artemis. *Blood* 2009;**114**:3601–9. https://doi.org/10.1182/blood-2008-11-188383
51. Sabouri Z, Okazaki I-m, Shinkura R *et al.* Apex2 is required for efficient somatic hypermutation but not for class switch recombination of immunoglobulin genes. *Int Immunol* 2009;**21**:947–55. https://doi.org/10.1093/intimm/dxp061
52. Schrader CE, Vardo J, Linehan E *et al.* Deletion of the nucleotide excision repair gene Ercc1 reduces immunoglobulin class switching and alters mutations near switch recombination junctions. *J Exp Med* 2004;**200**:321–30. https://doi.org/10.1084/jem.20040052
53. Huang FT, Yu KF, Balter BB *et al.* Sequence dependence of chromosomal R-loops at the immunoglobulin heavy-chain Sμ class switch region. *Mol Cell Biol* 2007;**27**:5921–32. https://doi.org/10.1128/MCB.00702-07
54. Kao YP, Hsieh WC, Hung ST *et al.* Detection and characterization of R-loops at the murine immunoglobulin Sα region. *Mol Immunol* 2013;**54**:208–16. https://doi.org/10.1016/j.molimm.2012.11.009
55. Nakamura M, Kondo S, Sugai M *et al.* High frequency class switching of an IgM+ B lymphoma clone CH12F3 to IgA+ cells. *Int Immunol* 1996;**8**:193–201. https://doi.org/10.1093/intimm/8.2.193
56. Feng Y, Li C, Stewart JA *et al.* FAM72A antagonizes UNG2 to promote mutagenic repair during antibody maturation. *Nature* 2021;**600**:324–8. https://doi.org/10.1038/s41586-021-04144-4
57. Rada C, Williams GT, Nilsen H *et al.* Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr Biol* 2002;**12**:1748–55. https://doi.org/10.1016/S0960-9822(02)01215-0

58. Nagaoka H, Muramatsu M, Yamamura N *et al.* 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* 2002;**195**:529–34. https://doi.org/10.1084/jem.20012144
59. Eto T, Kinoshita K, Yoshikawa K *et al.* RNA-editing cytidine deaminase Apobec-1 is unable to induce somatic hypermutation in mammalian cells. *Proc Natl Acad Sci USA* 2003;**100**:12895–8. https://doi.org/10.1073/pnas.2135587100
60. Yoshida K, Matsuoka M, Usuda S *et al.* Immunoglobulin switch circular DNA in the mouse infected with *Nippostrongylus brasiliensis*: evidence for successive class switching from mu to epsilon via gamma 1. *Proc Natl Acad Sci USA* 1990;**87**:7829–33. https://doi.org/10.1073/pnas.87.20.7829
61. Siebenkotten G, Esser C, Wabl M *et al.* The murine IgG1/IgE class switch program. *Eur J Immunol* 1992;**22**:1827–34. https://doi.org/10.1002/eji.1830220723
62. Mandler R, Finkelman FD, Levine AD *et al.* IL-4 induction of IgE class switching by lipopolysaccharide-activated murine B cells occurs predominantly through sequential switching. *J Immunol* 1993;**150**:407–18. https://doi.org/10.4049/jimmunol.150.2.407
63. Shen HM, Wuerffel R, Cantillo JF *et al.* Loop extrusion promotes an alternate pathway for isotype switching. *Cell Rep* 2021;**37**:110059. https://doi.org/10.1016/j.celrep.2021.110059
64. Chiarle R, Zhang Y, Frock RL *et al.* Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* 2011;**147**:107–19. https://doi.org/10.1016/j.cell.2011.07.049
65. Dunnick W, Hertz GZ, Scappino L *et al.* DNA sequences at immunoglobulin switch region recombination sites. *Nucleic Acids Res* 1993;**21**:365–72. https://doi.org/10.1093/nar/21.3.365
66. Kataoka T, Miyata T, Honjo T. Repetitive sequences in class-switch recombination regions of immunoglobulin heavy chain genes. *Cell* 1981;**23**:357–68. https://doi.org/10.1016/0092-8674(81)90131-8
67. Sen D, Gilbert W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* 1988;**334**:364–6. https://doi.org/10.1038/334364a0
68. Zhang ZZ, Pannunzio NR, Han L *et al.* The strength of an Ig switch region is determined by its ability to drive R loop formation and its number of WGCW sites. *Cell Rep* 2014;**8**:557–69. https://doi.org/10.1016/j.celrep.2014.06.021
69. Zhang ZZ, Pannunzio NR, Hsieh C-L *et al.* The role of G-density in switch region repeats for immunoglobulin class switch recombination. *Nucleic Acids Res* 2014;**42**:13186–93. https://doi.org/10.1093/nar/gku1100
70. Kumar NV, Varshney U. Inefficient excision of uracil from loop regions of DNA oligomers by *E.coli* uracil DNA glycosylase. *Nucl Acids Res* 1994;**22**:3737–41. https://doi.org/10.1093/nar/22.18.3737
71. Hansen EC, Ransom M, Hesselberth JR *et al.* Diverse fates of uracilated HIV-1 DNA during infection of myeloid lineage cells. *eLife* 2016;**5**:e18447. https://doi.org/10.7554/eLife.18447
72. Pham P, Bransteitter R, Petruska J *et al.* Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature* 2003;**424**:103–7. https://doi.org/10.1038/nature01760
73. Bregenhorn S, Kallenberger L, Artola-Boran M *et al.* Non-canonical uracil processing in DNA gives rise to double-strand breaks and deletions: relevance to class switch recombination. *Nucleic Acids Res* 2016;**44**:2691–705. https://doi.org/10.1093/nar/gkv1535
74. Peña-Diaz J, Bregenhorn S, Ghodgaonkar M *et al.* Noncanonical mismatch repair as a source of genomic instability in human cells. *Mol Cell* 2012;**47**:669–80. https://doi.org/10.1016/j.molcel.2012.07.006
75. Rogier M, Moritz J, Robert I *et al.* Fam72a enforces error-prone DNA repair during antibody diversification. *Nature* 2021;**600**:329–33. https://doi.org/10.1038/s41586-021-04093-y
76. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* 2012;**13**:36–46. https://doi.org/10.1038/nrg3117
77. Tørresen OK, Star B, Mier P *et al.* Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucleic Acids Res.* 2019;**47**:10994–1006. https://doi.org/10.1093/nar/gkz841
78. Chaudhuri J, Alt FW. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat Rev Immunol* 2004;**4**:541–52. https://doi.org/10.1038/nri1395
79. Kenter AL. Class-switch recombination: after the dawn of AID. *Curr Opin Immunol* 2003;**15**:190–8. https://doi.org/10.1016/S0952-7915(03)00018-9
80. Min IM, Selsing E. Antibody class switch recombination: roles for switch sequences and mismatch repair proteins. *Adv Immunol* 2005;**87**:297–328. https://doi.org/10.1016/S0065-2776(05)87008-7
81. Kadyrov FA, Dzantiev L, Constantin N *et al.* Endonucleolytic function of MutLalpha in human mismatch repair. *Cell* 2006;**126**:297–308. https://doi.org/10.1016/j.cell.2006.05.039
82. Genschel J, Bazemore LR, Modrich P. Human exonuclease I is required for 5' and 3' mismatch repair. *J Biol Chem* 2002;**277**:13302–11. https://doi.org/10.1074/jbc.M111854200
83. Kim A, Han L, Yu K. Immunoglobulin class switch recombination is initiated by rare cytosine deamination events at switch regions. *Mol Cell Biol* 2020;**40**:e00125-20. https://doi.org/10.1128/MCB.00125-20
84. Xue K, Rada C, Neuberger MS. The *in vivo* pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in msh2-/- ung-/- mice. *J Exp Med* 2006;**203**:2085–94. https://doi.org/10.1084/jem.20061067
85. Bello A, Hirth G, Voigt S *et al.* Mechanism and regulation of secondary immunoglobulin diversification. *Cell Cycle* 2023;**22**:2070–87. https://doi.org/10.1080/15384101.2023.2275397
86. Dézé O, Laffleur B, Cogné M. Roles of G4-DNA and G4-RNA in class switch recombination and additional regulations in B-lymphocytes. *Molecules* 2023;**28**:1159. https://doi.org/10.3390/molecules28031159
87. Methot SP, Di Noia JM. In: Alt FW (ed.), *Advances in Immunology*. Vol. **133**, San Diego, CA: Academic Press, 2017, 37–87.
88. Stavnezer J, Schrader CE. IgH chain class switch recombination: mechanism and regulation. *J Immunol* 2014;**193**:5370–8. https://doi.org/10.4049/jimmunol.1401849

Received: February 20, 2025. **Editorial Decision:** February 24, 2025. **Accepted:** February 27, 2025
© The Author(s) 2025. Published by Oxford University Press on behalf of NAR Molecular Medicine.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.