Acta Biochim Biophys Sin 2022, 54(6): 759–766
https://doi.org/10.3724/abbs.2022070
Advance Access Publication Date: 9 June 2022
Review

Review

AID function in somatic hypermutation and class switch recombination

Kefei Yu*

Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA
*Correspondence address. Tel: +1-517-884-5354; E-mail: yuke@msu.edu

Received 3 January 2022 Accepted 2 March 2022

Abstract

Activation-induced cytidine deaminase (AID) initiates somatic hypermutation of immunoglobulin (Ig) gene variable regions and class switch recombination (CSR) of Ig heavy chain constant regions. Two decades of intensive research has greatly expanded our knowledge of how AID functions in peripheral B cells to optimize antibody responses against infections, while maintaining tight regulation of AID to restrain its activity to protect B cell genomic integrity. The many exciting recent advances in the field include: 1) the first description of AID's molecular structure, 2) remarkable advances in high throughput approaches that precisely track AID targeting genome-wide, and 3) the discovery that the cohesion-mediate loop extrusion mechanism [initially discovered in V(D)J recombination studies] also governs AID-medicated CSR. These advances have significantly advanced our understanding of AID's biochemical properties *in vitro* and AID's function and regulation *in vivo*. This mini review will discuss these recent discoveries and outline the challenges and questions that remain to be addressed.

Key words AID, somatic hypermutation, class switch recombination, mismatch repair, base excision repair, error prone DNA repair

Introduction

Activation-induced cytidine deaminase (AID) initiates somatic hypermutation (SHM) of the immunoglobulin (Ig) gene variable (V) regions and class switch recombination (CSR) of the heavy (H) chain constant regions [1–5]. AID was discovered in a subtractive cDNA screening for *de novo* factors induced upon cytokine stimulation of a mouse B cell line (CH12F3) to undergo CSR [6]. Disruption of AID gene in mouse results in a complete loss of CSR and SHM [1] while transgenic expression of AID confers ectopic SHM and CSR activity [7,8]. AID mutations resulting in defective CSR cause type II hyper-IgM syndrome (HIGM2) in human patients [2].

AID belongs to a family of polynucleotide cytosine deaminases that has diverse functions in RNA editing (Apobec1), Ig gene diversification (AID) and host innate immunity against retroviruses (Apobec3 family members) [4,9]. Named after its limited homology to Apobec1, AID was initially thought to function in SHM and CSR by editing mRNA [10]. However, preponderant genetic and biochemical evidence soon made it clear that AID deaminates cytosines on DNA (converting cytosine to uracil) rather than on RNA [11–19]. Uracil in DNA triggers a cascade of DNA repair events that ultimately leads to mutations at Ig V regions during SHM and DNA double strand breaks (DSBs) at IgH switch (S) regions during CSR [3–5,20]. This short review focuses on some recent advances on AID structure, biochemical property, and AID-associated DNA repair events. Readers are encouraged to read excellent reviews that are more comprehensive on other aspects of AID function and regulations [3–5,21,22].

Error-prone Repair of AID-Generated Uracil

AID-generated uracil can be recognized as a base damage and processed by the uracil base excision repair (BER) pathway, or recognized as a U:G mismatch and processed by the mismatch repair (MMR) pathway. Neither pathway is B cell-specific and both pathways normally result in error-free repair. However, in antigen-stimulated B cells, repair of AID-generated uracil leads to error-prone outcomes at Ig V regions (mutations) and S regions (DSBs). MMR in germinal center B cells deviates from its canonical form by using error-prone DNA polymerases, primarily polymerase eta, during the gap filling step [23,24]. Ubiquitylation of PCNA plays an important role of recruiting polymerase eta during this process [25]. BER is also altered in activated B cells because of the induced expression of FAM72a, an uracil DNA glycosylase (UDG or UNG) antagonist. FAM72a interacts with nuclear uracil N-glycosylase (UNG2) and triggers its proteasome-mediated degradation. At-

tenuating UNG2-mediated BER thwarts high fidelity uracil repair and promotes error-prone outcomes [26,27].

SHM and CSR are quite distinct processes (point mutations at V regions versus DSBs at S regions), but the underlying mechanisms are fundamentally similar. While SHM at Ig V regions are mostly point mutations, a small percentage of alleles contains insertion or deletion that is characteristic of a DSB-mediated process [28,29]. Conversely, although DSBs are the prominent intermediates in CSR, point mutations are often found near switch junctions [30], and at the Sμ region in CSR-stimulated B cells [29,31–33]. On the other hand, SHM and CSR are independent of each other. IgM (unswitched) B cells can accumulate SHM at Ig V regions and isotype switched B cells can have unmutated V sequences. In addition, several AID mutations selectively affect one but not the other process [34], suggesting mechanistic differences between the two processes.

AID Structure and Substrate Preferences

Although AID is a small protein (24 kDa), working with recombinant AID has been notoriously difficult. Full length AID expresses poorly and tends to aggregate, severely hinders its structural or biochemical characterizations *in vitro*. Still, even with aggregated recombinant AID, several important properties of AID were discovered. AID deaminates cytosines on single stranded (ss) DNA [16–19], but not on double stranded (ds) DNA, RNA, or RNA:DNA hybrid [17]. The ssDNA specificity is consistent with the fact that AID's function is tightly associated with transcription [35,36]. Transcription separates the two DNA strands, at least transiently, to provide the ssDNA substrate for AID. However, transcription itself is insufficient to explain why AID targets predominantly V and S regions while sparing majority of the transcribed regions in the B cell genome. Another significant discovery is that AID preferably deaminates cytosines in the WRC (W = A or T; R = A or G) motif [19,37] which is complementary to the RGYW SHM hotspots previously identified by sequencing myriads of V regions from hybridomas [38].

The first AID-related crystal structure was obtained from an AID variant (a hybrid of AID and Apobec3A) that maintains AID's preference for WRC and the ability to support SHM [39]. Later, through elegant protein engineering, the Wu lab produced a monomeric recombinant AID (AIDmono) that harbors two functional-compatible mutations at the protein surface. AIDmono is much more active and exhibits high affinity for G-quadruplex (G4) or branched DNA substrates by binding to the ssDNA tails in these structures [40]. Most importantly, a crystal structure was obtained after introduction of three additional mutations that further reduce aggregation tendency [40]. The crystal structure does not contain information about the DNA substrate, but the substrate binding grove is clearly observable [40]. A co-crystal structure with a dCMP at the catalytic center was obtained. That structure predicted a potential steric hindrance if the dCMP were replaced by CMP [40], providing a good explanation for why AID deaminates cytosine on DNA, but not on RNA. Remarkably, a separate positively charged surface, which the authors called "assistant patch", was identified in the structure adjacent to the substrate channel [40]. The organization of the substrate channel and assistant patch suggests a bifurcated substrate binding mode that nicely explains AID's preferences for the projected ssDNA tails in G4 and branched DNA (Figure 1). Interestingly, arginine residues (e.g., R171, R174, R178) in the assistant patch were found, in a separate study, to be critical for AID's translocation from gene promoter proximal region to the gene body, a process that was coined "licensing" [41]. These arginine residues that are important for "licensing" are essential for AID's function in CSR and SHM [41]. One of these mutations (R174S) has been previously found in HIGM2 human patients [42]. In the Wu study [40], G4 was found to be the most favorable substrate for AID. AID does not bind to the G4 core but to the adjacent ssDNA tails. Whether G4 structure exists *in vivo* (*e.g.*, at transcribed S region) remains a point of debate. The Y-shape branched substrates, however, are probably abundant *in vivo* at places like stalled transcription bubbles or sites that form DNA secondary structures. In essence, the mode of binding of AID to G4 and branched substrates is the same.

The AID crystal structure didn't explain why AID favors WRC motif, owing to the lack of information about the DNA substrate in the structure [39]. Molecular modeling using the bound dCMP as the anchor predicted a conformational rearrangement that is necessary to accommodate the substrate at the −1 and −2 positions but not at the +1 and +2 positions [40]. This strongly suggests that the −1 and −2 positions are more important than the +1 and +2 positions for the binding of DNA substrate. Among the three loops that form the substrate channel, the β4-α4 loop (a.a. 113–123) has been previously found to specify substrate sequence preferences [43,44]. However, substrate preferences likely involve additional parts of the molecule. For example, S43P, a mutation found in HIGM2 patients, changes AID site preference from WRC to GGC [45], suggesting that this residue is also involved in substrate specification. It is interesting that AID's substrate binding affinity (Kd) appears to be sequence-independent as measured by electrophoretic mobility shift assays (EMSA) [40,46], neither does it distinguish RNA versus DNA [40]. This is difficult to understand at the moment, but it is possible that the binding measured by EMSA only reflects ionic stacking between the highly positively charged AID (+11) and the negatively charge nucleic acid rather than an authentic enzyme-substrate interaction. Solving co-crystal structures of AID with its substrates will address this gap of knowledge.

Elucidating the 3D structure of AID is a major advance in the field, which greatly changes our views of AID's cognate substrates *in vivo*. The bifurcated binding mechanism implies that AID may bind to and deaminate many DNA secondary structures that expose multiple ssDNA tracts. This could have important implications in the understanding of recurrent translocation sites that are vulnerable to AID-inflicted damages [47]. The assistant patch of AID could also be occupied by RNA [48], which may play important context-dependent regulatory roles (Figure 1).

AID Deamination at Ig S Regions

CSR is mediated by the generation and joining of DSBs between two Ig S regions (donor and acceptor) [20,49,50]. Mammalian S regions are long, repetitive, and highly enriched for G-rich pentamers such as GGG(G/C)T and GAGCT on the non-template strand [20,49,50]. Transcription through S regions tends to form R-loop, a secondary structure that consists of an RNA:DNA hybrid between the RNA transcript and the DNA template strand, and a displaced non-template DNA strand [51]. The G/C skewness between the two DNA strands renders the RNA transcript highly G-rich. This feature is thought to promote R-loop formation due to the extraordinary thermodynamic stability of a duplex formed by a G-rich RNA and a

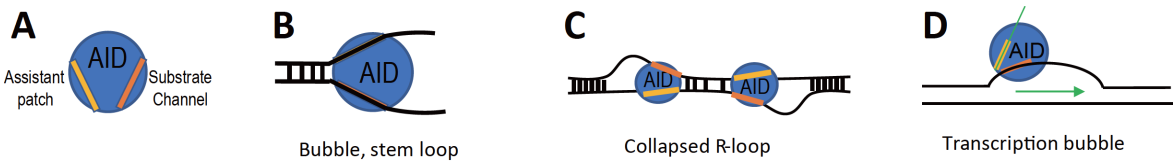

Figure 1. High affinity binding of AID to structured substrates (A) AID contains a substrate channel and an assistant patch that allow bifurcated substrate binding. (B) AID binds to branched DNA in a bubble, stem loop or other DNA secondary structures. (C) AID binds to a putative structure called collapsed R-loop that could be formed upon RNase H treatment of R-loop, causing switch repeats to misalign. (D) AID binds to a transcription bubble where the displaced RNA transcript occupies assistant patch.

C-rich DNA [52]. The GAGCT pentamer contains overlapping AID hotspots (AGCT) that are thought to be the primary targets for AID during CSR [53].

How AID deaminates S regions that results in DSBs has been a long-standing question. It is generally assumed that AID must deaminate both DNA strands whereas SHM may only require deamination on one strand [54]. Elucidating the abundance and positions of AID-generated uracils on both strands of S regions is crucial for understanding the mechanism of DSB formation. Because uracils are not repaired in cells deficient in both UNG2 and MSH2 [14], AID footprints can be captured from these cells by PCR and DNA sequencing to locate the positions of C to T transitions [15]. However, most endogenous S regions are practically impossible to PCR across the entire lengths [15]. Recently, AID footprint analysis was performed in the UNG2 and MSH2 double knockout CH12F3 B cell line that had been engineered to harbor a short core Sα region (1.1 kb). This core Sα allows robust PCR across its entire length while confers about 50% of WT CSR efficiency [33]. Footprint analysis from a bulk population of these cells revealed several interesting features. First, AID deamination occurs predominantly (80%–90%) at WRC motifs [33]. Second, the frequency of AID deamination does not drop across the core Sα [33], which contrasts to the exponential decrease of SHM frequency across the V regions. R-loops are known to initiate and terminate randomly in S regions [51]. This feature of R-loop fits the randomness of deamination across the entire S region. Third, AID targets both strands with similar frequency (slightly more deamination on the non-template strand), suggesting that deamination does not occur directly within the R-loop structure. Either R-loops must be processed by cellular RNase H [20] and/or helicases [55] to expose the template strand, or deamination mostly occurs at the edges of R-loop where branched DNA is available [20]. In the latter case, R-loop may serve primarily as a roadblock to stall transcription. Finally, AID deaminates S regions at very low frequency [33].

The most interesting observations came after capturing footprints of individual S regions where deaminations on both DNA strands are illustrated. AID deamination appears to be rare and separated far from each other [33]. Concurrent deamination at AGCT site was indeed observed but in only one of the 16 molecules [33]. Several molecules contain deaminations of two or more consecutive GAGCT pentamers [33], reflecting a degree of processivity. Interestingly, the processive deamination events are restricted to only one strand [33]. Finally, the percentage of molecules that harbor deamination on both strands correlates with the CSR efficiency when these cells are complemented with UNG2 and MSH2 cDNAs, and stimulated under the same condition used for footprint analysis [33], suggesting that DSB formation indeed requires both DNA strands to be deaminated.

The prevalent hypothesis for S region DSB formation is that AID deaminates AGCT sites concurrently on both strands [53], such that a DSB is formed by two nicks across each other that are generated by (UNG2) and apurinic/apyrimidinic endonuclease (APE1; Figure 2). Mutagenesis study of switch region sequence motifs confirmed the importance of overlapping AID hotspots for CSR [53]. However, AID footprint analysis surprisingly showed that concurrent deamination at AGCT site is an infrequent event, suggesting that this mechanism may not be the dominant one for S region DSBs as previously thought.

The rare and distally located deamination event by AID in S regions does provide a good explanation for the importance of MMR factors in CSR [14,56–62]. MMR is a gap repair that involves Exo1-mediated strand excision. Distal nicks on different strands can initiate convergent strand excision that results in DSB (Figure 2). This hypothesis was initially raised to explain the absolute requirement of MSH2 for CSR in mouse B cells that are deleted of the majority of the Sμ repeats [63,64]. It was thought that the loss of Sμ repeats results in a loss of concurrent deamination mechanism, and MMR is needed because deamination of the remaining sequences at the Sμ region only results in distal nicks [65]. The latest AID footprint analysis seems to suggest that even in the presence of abundant GAGCT pentamers, MMR is likely still the major mechanism for DSB formation. In this regard, nicks generated by UNG2 and APE1 mainly serve as the entry points for Exo1, not directly forming DSB. Because Exo1 deficiency results in a more severe CSR defect than that of MSH2-deficiency, it is likely that Exo1 can be loaded in MSH2-dependent as well as MSH2-independent manners. In an artificial system where distal nicks (up to 200 bp apart) are generated in Sμ and Sα by a Cas9 nickase in AID-KO CH12F3 cells, cells can switch from IgM to IgA, albeit at very low efficiencies [66], suggesting that distal nicks can indeed be processed into DSBs. It will be interesting to test whether DSB formation under this condition requires MMR factors, and whether the absence of U:G mismatch (due to AID-KO) is a culprit for the low efficiency of switching.

The scarcity of overlapping AID footprints at AGCT sites came as a surprise, given the functional importance of AGCT sites in CSR [53]. However, it does make some sense upon further consideration. If concurrent deamination at AGCT sites were the dominant mechanism for DSB formation, it would be difficult to explain several known features of CSR. First, switch junctions are not focused at AGCT sites [30]. If DSBs always initiate at AGCT sites, extensive end processing must be proposed to explain the random distribution of switch junctions within the switch repeats [30]. However, extensive end processing is probably detrimental to CSR as evidenced by the severe CSR defect in 53bp1-KO B cells [67,68] that display excessive end resection [69–71]. Second, it does not explain why MMR factors are required for efficient CSR, even when switch repeats are present. Third, it seems inconsistent with the

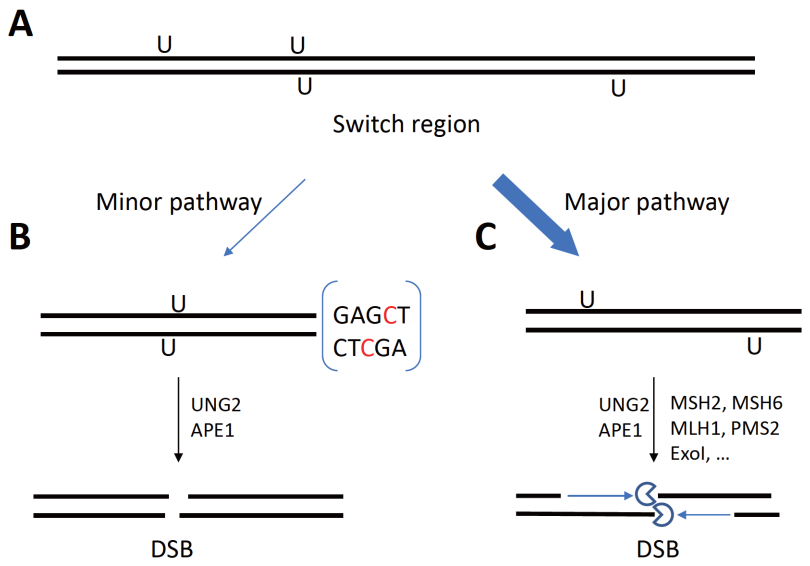

Figure 2. Model of DSB formation at switch regions (A) AID generates rare and distally spaced uracils at switch regions. (B) A minor pathway of DSB formation is through concurrent deamination at AGCT sites that contain overlapping AID hotspots. Close nicks produced by UNG2 and APE1 result in DSB. (C) The major pathway of DSB formation requires both BER and MMR factors. BER produces nicks that serve as loading sites for Exo1. The MSH2/6 heteroduplex binds to a U:G mismatch and facilitates Exo1 loading at nicks to initiate 5′–3′ strand excision. If a nick is located at the 3′ of the U:G mismatch, endonuclease activity of MLH1/PMS2 is required to generate a nick at the 5′ of the mismatch. Convergent Exo1-mediated strand excision results in DSB formation.

requirement for FAM72a to antagonize UNG2 to achieve fully efficient CSR [26,27]. Finally, it will be difficult to explain why a catalytically crippled mutant UNG can rescue CSR in UNG-KO cells despite providing only very low UNG activity in these cells [72–75]. Taking all these into consideration, it appears more likely that cleavage at AGCT sites only accounts for a small percentage of S region DSBs. The majority of DSBs are likely formed by the MMR mechanism (Figure 2). UNG is essential, but few are needed to generate nicks to load the Exo1. So, why are AGCT sites important for CSR? One explanation could be that having lots of AGCT sites at S regions increases the chances of cytosine deamination on both DNA strands because both strands have abundant AID hotspots.

HIGM2 is caused by AID mutations that are defective for CSR. These mutations are found throughout the AID gene body [76], affecting AID's enzyme activity, protein folding and/or stability. Several AID mutants maintain a significant level of deaminase activity [34,45]. One such mutation, the C-terminal truncation of AID (hereafter called AIDΔC) [42], has been most extensively studied. The C-terminus (~10 a.a.) of AID comprises a nuclear exporting signal that directs AID to the cytoplasm where majority of the cellular AID accumulates [77,78]. AIDΔC is an active deaminase with a specific activity even higher than the WT [79]. Although AIDΔC fails to mediate CSR in AID-KO B cells, it can mutate the 5′-Sμ region in these cells [31], or induce gene conversion in DT40 cells [80], suggesting that the mutation only affects CSR, but not SHM. In human patients, this mutation has a dominant negative (DN) phenotype on CSR [42]. The DN effect has been recapitulated *in vitro* in cultured mouse B cells [80–82], but not yet in animal models. AIDΔC expresses poorly in mammalian cells likely due to its primary residence in the nucleus where it is vulnerable to proteosome-mediated degradation [77]. The molecular basis of the DN effect is still elusive. If AID must function in CSR as a multimer, then one bad subunit ruins the entire complex. However, there has been insufficient evidence to support this hypothesis. Alternatively, AIDΔC might produce aberrant DNA damages that are difficult to repair. One study reported interactions of the AID C-terminus with UNG2, MSH2 and MSH6, and argued that such interactions may influence the subsequent DNA repair [81]. Another study demonstrated excessive end processing and reduced level of end-joining factors at S regions in AIDΔC-expressing cells [80]. Both studies revealed that deaminase activity is required for the DN effect of AIDΔC [80,81], which is consistent with the aberrant DNA damage hypothesis.

AID Deamination at Ig V Regions

Unlike S regions, V region sequences do not have features that promote R-loop formation. Although protein-dependent short stretches of ssDNA has been reported at Ig V regions [83], how much they contribute to SHM is unclear. Moreover, SHM of V regions likely only requires AID deamination on one of the two DNA strands [54]. This may be an important factor why certain AID mutations differentially affect SHM and CSR [34]. V regions are privileged sites for SHM regardless of the local sequences [29], but the reason for this privilege is unclear. Searching for *cis*-acting elements for AID recruitment has been an ever-lasting effort [84–88]. Ig enhancers, not Ig promoters, contain elements that specify SHM activity [89]. However, Ig enhancers contain a myriad of transcription factor sites, and it has been difficult to pinpoint which ones are more important than others and why.

A lingering question in the field is why *ex vivo* stimulated mouse spleen B cells accumulate few V region mutations despite a high level of AID expression and robust CSR [90,91]. It is known that the 5′-Sμ region accumulates mutations in these *ex vivo* stimulated cells [90], suggesting that the SHM machinery is intact. One study found that SHM activity coincides with the presence of suppressor of Ty 5 homolog (Spt5) [92], a transcription elongation factor that is associated with paused RNA polymerase II. In *ex vivo* stimulated B cells, Spt5 is found at 5′-Sμ but not V regions [92], whereas in germinal

center B cells, Spt5 is found to associate with the V regions [92]. Therefore, Spt5 appears to be a good marker for SHM activity at different regions [92]. Interestingly, deep sequencing analysis revealed a very low level of V region mutations in *ex vivo* stimulated B cells, but the mutation profile nevertheless matches that of the germinal center B cells [29]. These data imply a quantitative rather than a qualitative difference of AID targeting to the V regions. Similarly, cytokine-stimulated CH12F3 cells undergo robust CSR to IgA, but do not mutate their V_H regions [32]. AID in CH12F3 cells is primarily located to the Sμ region and not to the V_H region [32]. Even when AID is retargeted to the V_H region by a trick of knocking down a splice factor (PTBP2) that is required for S region association of AID, the V_H region is still not mutated [32]. Therefore, AID occupancy does not directly correlate with SHM activity. This additional layer of regulation of SHM is still a mystery.

Historically, both SHM and CSR were thought to occur in centroblasts in germinal centers. However, a recent study provided strong evidence that CSR occurs predominantly in B cells when they make the initial contact with T cells prior to the germinal center formation [93]. In contrast, SHM occurs inside the germinal center. This makes the differential AID targeting during CSR and SHM more palpable, as the two processes occur at different stages of B cell development. Bcl6 is highly expressed in germinal center B cells and functions as a master regulator of germinal center reactions. Bcl6 gene disruption in DT40 cells results in the loss of AID expression and SHM [94]. However, forced expression of AID does not rescue SHM in Bcl6 deficient cells [94], nor does complementing the cells with a Bcl6 transgene that partially restores Bcl6 expression. A change of gene expression profile was noted, suggesting a change of cell identity that could not be easily reversed [94]. It is currently unknown which Bcl6 target genes are needed to restore SHM in those cells. CH12F3 cells do not express Bcl6 (KY, unpublished), likely a reason why CH12F3 cells do not mutate V regions.

A recent study reported the instigation of V region SHM by retroviral transduction of AID in CH12F3 or *ex vivo* stimulated B cells, but only if the endogenous AID gene were disrupted [95]. The authors proposed an unknown SHM-specific factor that is sequestered by the basal level of AID protein in unstimulated cells. This explanation is not very satisfactory, as there is very little AID protein expressed without stimulation. Therefore, whatever unknown factor that is sequestered by AID must exist at an even lower level. It cannot be ruled out, however, that the basal level AID may somehow affect the expressions of other genes or cell identity that renders the cell non-permissive for V region SHM. This surprising discovery may have created a valuable experimental platform for further investigations of transacting factors that are essential for SHM.

AID Off-targets

AID-mediated off-target mutations are threats to genome integrity, which may explain why more than 90% of human lymphomas are of B cell origin [96]. It has been well known that AID targets are not restricted to Ig genes, as SHM-like mutations accumulate at proto-oncogenes such as *Bcl6* and others in germinal center and neoplastic B cells [89,97,98]. A large-scale DNA sequencing study revealed that as many as 25% of germinal center expressed gene are subject to AID damages [91]. Of these, some are deaminated by AID but manage to avoid mutation thanks to high fidelity DNA repair. This study revealed two levels of protection (targeting and repair) against AID-induced damages. Genome-wide chromatin immunoprecipitation using an anti-AID antibody identified 5910 genes that are occupied by AID in *ex vivo* stimulated B cells [99]. These studies all pointed to a promiscuous nature of AID targeting. As expected, AID off-target activities are linked to open chromatins and active transcription. Genome-wide studies also identified super enhancers (SEs) as prominent AID off-targets [100,101]. SEs are clusters of enhancers that are hyper-acetylated, highly active, engaging long-range interactions, and controlling cell type-specific gene expressions [102]. It should be noted that only a small percentage of SEs are AID targets. Therefore, neither high level transcription nor SEs are sufficient to specify where SHM accumulates. Divergent and convergent transcriptions have also been described as features that favor AID-targeting [101,103]. Contradictory as it may seem, divergent and convergent transcriptions probably represent the same topological mechanism of promoting ss DNA, depending on the timing of transcription [104].

Despite great interests and intensive research efforts, the targeting mechanism for AID remains obscure. Not a single or simple combination of motifs or features can be used for reliable prediction of AID off-targets. No single characteristic was found to be sufficient or essential for recruiting AID. It seems that a very large number of factors collectively control the where-about of AID and SHM in the B cell genome. Recent CSR studies proposed a concept called "recombination center" as a key element of the chromatin loop extrusion mechanism that governs CSR [105–107]. It is tempting to speculate a similar mechanism for SHM such that AID may be enriched (physically and/or functionally) at discrete sub-nuclear locations and that access to these locations confers SHM activity.

Disclosure statement

The author apologizes for any relevant work that is not cited.

Funding

This work was supported by the grants from National Institutes of Health (Nos. R01AI139039 and R21AI126359 to K.Y.).

Conflict of Interest

The author declares that he has no conflict of interest.

References

1. Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 2000, 102: 553–563
2. Revy P, Muto T, Levy Y, Geissmann F, Plebani A, Sanal O, Catalan N, *et al.* Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the hyper-IgM syndrome (HIGM2). *Cell* 2000, 102: 565–575
3. Hwang JK, Alt FW, Yeap LS. Related mechanisms of antibody somatic hypermutation and class switch recombination. *Microbiol Spectr* 2015, 3: MDNA3-0037-2014
4. Siriwardena SU, Chen K, Bhagwat AS. Functions and malfunctions of mammalian DNA-cytosine deaminases. *Chem Rev* 2016, 116: 12688–12710
5. Methot SP, Di Noia JM. Molecular mechanisms of somatic hypermutation and class switch recombination. *Adv Immunol* 2017, 133: 37–87
6. Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO, Honjo T. Specific expression of activation-induced cytidine

deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J Biol Chem* 1999, 274: 18470–18476
7. Okazaki I, Kinoshita K, Muramatsu M, Yoshikawa K, Honjo T. The AID enzyme induces class switch recombination in fibroblasts. *Nature* 2002, 416: 340–345
8. Martin A, Bardwell PD, Woo CJ, Fan M, Shulman MJ, Scharff MD. Activation-induced cytidine deaminase turns on somatic hypermutation in hybridomas. *Nature* 2002, 415: 802–806
9. Wedekind JE, Dance GSC, Sowden MP, Smith HC. Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. *Trends Genet* 2003, 19: 207–216
10. Honjo T, Kinoshita K, Muramatsu M. Molecular mechanism of class switch recombination: linkage with somatic hypermutation. *Annu Rev Immunol* 2002, 20: 165–196
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
12. Di Noia J, Neuberger MS. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* 2002, 419: 43–48
13. Rada C, Williams GT, Nilsen H, Barnes DE, Lindahl T, Neuberger MS. Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr Biol* 2002, 12: 1748–1755
14. Rada C, Di Noia JM, Neuberger MS. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol Cell* 2004, 16: 163–171
15. 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: 2085–2094
16. Dickerson SK, Market E, Besmer E, Papavasiliou FN. AID mediates hypermutation by deaminating single stranded DNA. *J Exp Med* 2003, 197: 1291–1296
17. Bransteitter R, Pham P, Scharff MD, Goodman MF. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc Natl Acad Sci USA* 2003, 100: 4102–4107
18. Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E, Alt FW. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* 2003, 422: 726–730
19. Yu K, Huang FT, Lieber MR. DNA substrate length and surrounding sequence affect the activation-induced deaminase activity at cytidine. *J Biol Chem* 2004, 279: 6496–6500
20. Yu K, Lieber MR. Current insights into the mechanism of mammalian immunoglobulin class switch recombination. *Crit Rev Biochem Mol Biol* 2019, 54: 333–351
21. Chandra V, Bortnick A, Murre C. AID targeting: old mysteries and new challenges. *Trends Immunol* 2015, 36: 527–535
22. Ghorbani A, Quinlan EM, Larijani M. Evolutionary comparative analyses of DNA-editing enzymes of the immune system: from 5-dimensional description of protein structures to immunological insights and applications to protein engineering. *Front Immunol* 2021, 12: 642343
23. Zeng X, Winter DB, Kasmer C, Kraemer KH, Lehmann AR, Gearhart PJ. DNA polymerase η is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. *Nat Immunol* 2001, 2: 537–541
24. Rogozin IB, Pavlov YI, Bebenek K, Matsuda T, Kunkel TA. Somatic mutation hotspots correlate with DNA polymerase η error spectrum. *Nat Immunol* 2001, 2: 530–536
25. Roa S, Avdievich E, Peled JU, MacCarthy T, Werling U, Li Kuang F, Kan R, *et al.* Ubiquitylated PCNA plays a role in somatic hypermutation and class-switch recombination and is required for meiotic progression. *Proc Natl Acad Sci USA* 2008, 105: 16248–16253
26. Feng Y, Li C, Stewart JA, Barbulescu P, Seija Desivo N, Álvarez-Quilón A, Pezo RC, *et al.* FAM72A antagonizes UNG2 to promote mutagenic repair during antibody maturation. *Nature* 2021, 600: 324–328
27. Rogier M, Moritz J, Robert I, Lescale C, Heyer V, Abello A, Martin O, *et al.* Fam72a enforces error-prone DNA repair during antibody diversification. *Nature* 2021, 600: 329–333
28. Goossens T, Klein U, Küppers R. Frequent occurrence of deletions and duplications during somatic hypermutation: implications for oncogene translocations and heavy chain disease. *Proc Natl Acad Sci USA* 1998, 95: 2463–2468
29. Yeap LS, Hwang JK, Du Z, Meyers RM, Meng FL, Jakubauskaitė A, Liu M, *et al.* Sequence-intrinsic mechanisms that target AID mutational outcomes on antibody genes. *Cell* 2015, 163: 1124–1137
30. Dunnick W, Hertz GZ, Scappino L, Gritzmacher C. DNA sequences at immunoglobulin switch region recombination sites. *Nucleic Acids Res* 1993, 21: 365–372
31. Barreto V, Reina-San-Martin B, Ramiro AR, McBride KM, Nussenzweig MC. C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. *Mol Cell* 2003, 12: 501–508
32. Matthews AJ, Husain S, Chaudhuri J. Binding of AID to DNA does not correlate with mutator activity. *J Immunol* 2014, 193: 252–257
33. 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:
34. Shinkura R, Ito S, Begum NA, Nagaoka H, Muramatsu M, Kinoshita K, Sakakibara Y, *et al.* Separate domains of AID are required for somatic hypermutation and class-switch recombination. *Nat Immunol* 2004, 5: 707–712
35. Ramiro AR, Stavropoulos P, Jankovic M, Nussenzweig MC. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nat Immunol* 2003, 4: 452–456
36. Pham P, Malik S, Mak C, Calabrese PC, Roeder RG, Goodman MF. AID–RNA polymerase II transcription-dependent deamination of IgV DNA. *Nucleic Acids Res* 2019, 47: 10815–10829
37. Pham P, Bransteitter R, Petruska J, Goodman MF. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature* 2003, 424: 103–107
38. Rogozin IB, Kolchanov NA. Somatic hypermutagenesis in immunoglobulin genes. *Biochim Biophys Acta (BBA) - Gene Structure Expression* 1992, 1171: 11–18
39. Pham P, Afif SA, Shimoda M, Maeda K, Sakaguchi N, Pedersen LC, Goodman MF. Structural analysis of the activation-induced deoxycytidine deaminase required in immunoglobulin diversification. *DNA Repair* 2016, 43: 48–56
40. Qiao Q, Wang L, Meng FL, Hwang JK, Alt FW, Wu H. AID recognizes structured DNA for class switch recombination. *Mol Cell* 2017, 67: 361–373.e4
41. Methot SP, Litzler LC, Subramani PG, Eranki AK, Fifield H, Patenaude AM, Gilmore JC, *et al.* A licensing step links AID to transcription elongation for mutagenesis in B cells. *Nat Commun* 2018, 9: 1248
42. Imai K, Zhu Y, Revy P, Morio T, Mizutani S, Fischer A, Nonoyama S, *et al.* Analysis of class switch recombination and somatic hypermutation in patients affected with autosomal dominant hyper-IgM syndrome type 2. *Clin Immunol* 2005, 115: 277–285
43. Wang M, Rada C, Neuberger MS. Altering the spectrum of immunoglobulin V gene somatic hypermutation by modifying the active

site of AID. *J Exp Med* 2010, 207: 141–153
44. Kohli RM, Abrams SR, Gajula KS, Maul RW, Gearhart PJ, Stivers JT. A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase. *J Biol Chem* 2009, 284: 22898–22904
45. Pham P, Smolka MB, Calabrese P, Landolph A, Zhang K, Zhou H, Goodman MF. Impact of phosphorylation and phosphorylation-null mutants on the activity and deamination specificity of activation-induced cytidine deaminase. *J Biol Chem* 2008, 283: 17428–17439
46. Larijani M, Petrov AP, Kolenchenko O, Berru M, Krylov SN, Martin A. AID associates with single-stranded DNA with high affinity and a long complex half-life in a sequence-independent manner. *Mol Cell Biol* 2007, 27: 20–30
47. Liu D, Lieber MR. The mechanisms of human lymphoid chromosomal translocations and their medical relevance. *Crit Rev Biochem Mol Biol* 2022, 57: 227–243
48. Liu D, Goodman MF, Pham P, Yu K, Hsieh CL, Lieber MR. The mRNA tether model for activation-induced deaminase and its relevance for ig somatic hypermutation and class switch recombination. *DNA Repair* 2022, 110: 103271
49. Chaudhuri J, Alt FW. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat Rev Immunol* 2004, 4: 541–552
50. Yu K, Lieber MR. Nucleic acid structures and enzymes in the immunoglobulin class switch recombination mechanism. *DNA Repair* 2003, 2: 1163–1174
51. Yu K, Chedin F, Hsieh CL, Wilson TE, Lieber MR. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat Immunol* 2003, 4: 442–451
52. Ratmeyer L, Vinayak R, Zhong YY, Zon G, Wilson WD. Sequence specific thermodynamic and structural properties for DNA.cntdot.RNA duplexes. *Biochemistry* 1994, 33: 5298–5304
53. Han L, Masani S, Yu K. Overlapping activation-induced cytidine deaminase hotspot motifs in Ig class-switch recombination. *Proc Natl Acad Sci USA* 2011, 108: 11584–11589
54. Faili A, Aoufouchi S, Guéranger Q, Zober C, Léon A, Bertocci B, Weill JC, *et al.* AID-dependent somatic hypermutation occurs as a DNA single-strand event in the BL2 cell line. *Nat Immunol* 2002, 3: 815–821
55. Lim J, Giri PK, Kazadi D, Laffleur B, Zhang W, Grinstein V, Pefanis E, *et al.* Nuclear proximity of mtr4 to RNA exosome restricts DNA mutational asymmetry. *Cell* 2017, 169: 523–537.e15
56. Stavnezer J, Guikema JEJ, Schrader CE. Mechanism and regulation of class switch recombination. *Annu Rev Immunol* 2008, 26: 261–292
57. Ehrenstein MR, Neuberger MS. Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. *EMBO J* 1999, 18: 3484–3490
58. Bardwell PD, Woo CJ, Wei K, Li Z, Martin A, Sack SZ, Parris T, *et al.* Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1–mutant mice. *Nat Immunol* 2004, 5: 224–229
59. Schrader CE, Edelmann W, Kucherlapati R, Stavnezer J. Reduced isotype switching in splenic B cells from mice deficient in mismatch repair enzymes. *J Exp Med* 1999, 190: 323–330
60. Li Z, Scherer SJ, Ronai D, Iglesias-Ussel MD, Peled JU, Bardwell PD, Zhuang M, *et al.* Examination of Msh6- and Msh3-deficient mice in class switching reveals overlapping and distinct roles of MutS homologues in antibody diversification. *J Exp Med* 2004, 200: 47–59
61. Martin A, Li Z, Lin DP, Bardwell PD, Iglesias-Ussel MD, Edelmann W, Scharff MD. Msh2 ATPase activity is essential for somatic hypermutation at a-T basepairs and for efficient class switch recombination. *J Exp Med* 2003, 198: 1171–1178
62. Peron S, Metin A, Gardes P, Alyanakian MA, Sheridan E, Kratz CP, Fischer A, *et al.* Human PMS2 deficiency is associated with impaired immunoglobulin class switch recombination. *J Exp Med* 2008, 205: 2465–2472
63. Min IM, Schrader CE, Vardo J, Luby TM, D'Avirro N, Stavnezer J, Selsing E. The smu tandem repeat region is critical for ig isotype switching in the absence of Msh2. *Immunity* 2003, 19: 515–524
64. Min IM, Rothlein LR, Schrader CE, Stavnezer J, Selsing E. Shifts in targeting of class switch recombination sites in mice that lack μ switch region tandem repeats or Msh2. *J Exp Med* 2005, 201: 1885–1890
65. Eccleston J, Schrader CE, Yuan K, Stavnezer J, Selsing E. Class switch recombination efficiency and junction microhomology patterns in Msh2-, Mlh1-, and Exo1-deficient mice depend on the presence of mu switch region tandem repeats. *J Immunol* 2009, 183: 1222–1228
66. Ling AK, So CC, Le MX, Chen AY, Hung L, Martin A. Double-stranded DNA break polarity skews repair pathway choice during intrachromosomal and interchromosomal recombination. *Proc Natl Acad Sci USA* 2018, 115: 2800–2805
67. Manis JP, Morales JC, Xia Z, Kutok JL, Alt FW, Carpenter PB. 53BP1 links DNA damage-response pathways to immunoglobulin heavy chain class-switch recombination. *Nat Immunol* 2004, 5: 481–487
68. Ward IM, Reina-San-Martin B, Olaru A, Minn K, Tamada K, Lau JS, Cascalho M, *et al.* 53BP1 is required for class switch recombination. *J Cell Biol* 2004, 165: 459–464
69. Hakim O, Resch W, Yamane A, Klein I, Kieffer-Kwon KR, Jankovic M, Oliveira T, *et al.* DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. *Nature* 2012, 484: 69–74
70. Yamane A, Robbiani DF, Resch W, Bothmer A, Nakahashi H, Oliveira T, Rommel PC, *et al.* RPA accumulation during class switch recombination represents 5′–3′ DNA-end resection during the S–G2/M phase of the cell cycle. *Cell Rep* 2013, 3: 138–147
71. Bothmer A, Robbiani DF, Di Virgilio M, Bunting SF, Klein IA, Feldhahn N, Barlow J, *et al.* Regulation of DNA end joining, resection, and immunoglobulin class switch recombination by 53BP1. *Mol Cell* 2011, 42: 319–329
72. Begum NA, Kinoshita K, Kakazu N, Muramatsu M, Nagaoka H, Shinkura R, Biniszkiewicz D, *et al.* Uracil DNA glycosylase activity is dispensable for immunoglobulin class switch. *Science* 2004, 305: 1160–1163
73. Di Noia JM, Williams GT, Chan DTY, Buerstedde JM, Baldwin GS, Neuberger MS. Dependence of antibody gene diversification on uracil excision. *J Exp Med* 2007, 204: 3209–3219
74. (a)Stivers JT. Comment on "Uracil DNA glycosylase activity is dispensable for immunoglobulin class switch". *Science* 2004, 306: 2042; (b) author reply 2042
75. Unniraman S, Fugmann SD, Schatz DG. UNGstoppable switching. *Science* 2004, 305: 1113–1114
76. Durandy A, Revy P, Imai K, Fischer A. Hyper-immunoglobulin M syndromes caused by intrinsic B-lymphocyte defects. *Immunol Rev* 2005, 203: 67–79
77. Geisberger R, Rada C, Neuberger MS. The stability of AID and its function in class-switching are critically sensitive to the identity of its nuclear-export sequence. *Proc Natl Acad Sci USA* 2009, 106: 6736–6741
78. Ito S, Nagaoka H, Shinkura R, Begum N, Muramatsu M, Nakata M, Honjo T. Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc Natl Acad Sci USA* 2004, 101: 1975–1980
79. Mu Y, Prochnow C, Pham P, Chen XS, Goodman MF. A structural basis

for the biochemical behavior of activation-induced deoxycytidine deaminase class-switch recombination-defective hyper-IgM-2 mutants. *J Biol Chem* 2012, 287: 28007–28016

80. Zahn A, Eranki AK, Patenaude AM, Methot SP, Fifield H, Cortizas EM, Foster P, *et al.* Activation induced deaminase C-terminal domain links DNA breaks to end protection and repair during class switch recombination. *Proc Natl Acad Sci USA* 2014, 111: E988–97
81. Ucher AJ, Ranjit S, Kadungure T, Linehan EK, Khair L, Xie E, Limauro J, *et al.* Mismatch repair proteins and AID activity are required for the dominant negative function of C-terminally deleted AID in class switching. *J Immunol* 2014, 193: 1440–1450
82. Al Ismail A, Husain A, Kobayashi M, Honjo T, Begum NA. Depletion of recombination-specific cofactors by the C-terminal mutant of the activation-induced cytidine deaminase causes the dominant negative effect on class switch recombination. *Int Immunol* 2017, 29: 525–537
83. Ronai D, Iglesias-Ussel MD, Fan M, Li Z, Martin A, Scharff MD. Detection of chromatin-associated single-stranded DNA in regions targeted for somatic hypermutation. *J Exp Med* 2007, 204: 181–190
84. Kohler KM, McDonald JJ, Duke JL, Arakawa H, Tan S, Kleinstein SH, Buerstedde JM, *et al.* Identification of core DNA elements that target somatic hypermutation. *J Immunol* 2012, 189: 5314–5326
85. Buerstedde JM, Alinikula J, Arakawa H, McDonald JJ, Schatz DG. Targeting of somatic hypermutation by immunoglobulin enhancer and enhancer-like sequences. *PLoS Biol* 2014, 12: e1001831
86. Blagodatski A, Batrak V, Schmidl S, Schoetz U, Caldwell RB, Arakawa H, Buerstedde JM. A cis-acting diversification activator both necessary and sufficient for AID-mediated hypermutation. *PLoS Genet* 2009, 5: e1000332
87. Michael N, Martin TE, Nicolae D, Kim N, Padjen K, Zhan P, Nguyen H, *et al.* Effects of sequence and structure on the hypermutability of immunoglobulin genes. *Immunity* 2002, 16: 123–134
88. Michael N, Shen HM, Longerich S, Kim N, Longacre A, Storb U. The E box motif CAGGTG enhances somatic hypermutation without enhancing transcription. *Immunity* 2003, 19: 235–242
89. Storb U, Shen HM, Michael N, Kim N. Somatic hypermutation of immunoglobulin and non–immunoglobulin genes. *Phil Trans R Soc Lond B* 2001, 356: 13–19
90. Nagaoka H, Muramatsu M, Yamamura N, Kinoshita K, Honjo T. Activation-induced Deaminase (AID)-directed Hypermutation in the Immunoglobulin Sμ Region. *J Exp Med* 2002, 195: 529–534
91. Liu M, Duke JL, Richter DJ, Vinuesa CG, Goodnow CC, Kleinstein SH, Schatz DG. Two levels of protection for the B cell genome during somatic hypermutation. *Nature* 2008, 451: 841–845
92. Maul RW, Cao Z, Venkataraman L, Giorgetti CA, Press JL, Denizot Y, Du H, *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: 2297–2306
93. Roco JA, Mesin L, Binder SC, Nefzger C, Gonzalez-Figueroa P, Canete PF, Ellyard J, *et al.* Class-switch recombination occurs infrequently in germinal centers. *Immunity* 2019, 51: 337–350.e7
94. Williams AM, Maman Y, Alinikula J, Schatz DG. Bcl6 Is required for somatic hypermutation and gene conversion in chicken DT40 cells. *PLoS ONE* 2016, 11: e0149146
95. Liu J, Xiong E, Zhu H, Mori H, Yasuda S, Kinoshita K, Tsubata T, *et al.* Efficient induction of ig gene hypermutation in ex vivo–activated primary B cells. *J Immunol* 2017, 199: 3023–3030
96. Küppers R. Mechanisms of B-cell lymphoma pathogenesis. *Nat Rev Cancer* 2005, 5: 251–262
97. Shen HM, Peters A, Baron B, Zhu X, Storb U. Mutation of *BCL-6* Gene in Normal B Cells by the Process of Somatic Hypermutation of Ig Genes. *Science* 1998, 280: 1750–1752
98. Pasqualucci L, Migliazza A, Fracchiolla N, William C, Neri A, Baldini L, Chaganti RSK, *et al.* BCL-6 mutations in normal germinal center B cells: evidence of somatic hypermutation acting outside Ig loci. *Proc Natl Acad Sci USA* 1998, 95: 11816–11821
99. Yamane A, Resch W, Kuo N, Kuchen S, Li Z, Sun H, Robbiani DF, *et al.* Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat Immunol* 2011, 12: 62–69
100. Qian J, Wang Q, Dose M, Pruett N, Kieffer-Kwon KR, Resch W, Liang G, *et al.* B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell* 2014, 159: 1524–1537
101. Meng FL, Du Z, Federation A, Hu J, Wang Q, Kieffer-Kwon KR, Meyers RM, *et al.* Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* 2014, 159: 1538–1548
102. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 2013, 153: 307–319
103. Pefanis E, Wang J, Rothschild G, Lim J, Chao J, Rabadan R, Economides AN, *et al.* Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature* 2014, 514: 389–393
104. Pannunzio NR, Lieber MR. RNA polymerase collision versus DNA structural distortion: twists and turns can cause break failure. *Mol Cell* 2016, 62: 327–334
105. Zhang X, Zhang Y, Ba Z, Kyritsis N, Casellas R, Alt FW. Fundamental roles of chromatin loop extrusion in antibody class switching. *Nature* 2019, 575: 385–389
106. Zhang X, Yoon HS, Chapdelaine-Williams AM, Kyritsis N, Alt FW. Physiological role of the 3′IgH CBEs super-anchor in antibody class switching. *Proc Natl Acad Sci USA* 2021, 118: e2024392118
107. Yu K. An insulator that regulates chromatin extrusion and class switch recombination. *Proc Natl Acad Sci USA* 2021, 118: e2026399118