Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Trends in Immunology

CellPress
REVIEWS

Special Issue: Molecular Mechanisms of Immunity

Review

AID in Antibody Diversification: There and Back Again

Yuqing Feng,[1,4] Noé Seija,[2,3,4] Javier M. Di Noia,[2,3,*] and Alberto Martin[1,*]

Activation-Induced cytidine Deaminase (AID) initiates affinity maturation and isotype switching by deaminating deoxycytidines within immunoglobulin genes, leading to somatic hypermutation (SHM) and class switch recombination (CSR). AID thus potentiates the humoral response to clear pathogens. Marking the 20th anniversary of the discovery of AID, we review the current understanding of AID function. We discuss AID biochemistry and how error-free forms of DNA repair are co-opted to prioritize mutagenesis over accuracy during antibody diversification. We discuss the regulation of DNA double-strand break (DSB) repair pathways during CSR. We describe genomic targeting of AID as a multilayered process involving chromatin architecture, *cis*- and *trans*-acting factors, and determining mutagenesis – distinct from AID occupancy at loci that are spared from mutation.

Highlights

Subverted base excision repair (BER) and mismatch repair (MMR) pathways act concertedly to generate antibody sequence diversity during SHM.

In CSR, DNA DSBs are repaired by the nonhomologous end-joining pathway involving the 53BP1–Rif1–Shieldin axis, and by an alternative end-joining pathway involving HMCES (5-Hydroxymethylcytosine binding, ES-cell-specific) that binds and protects resected DSB ends.

Genomic targeting of AID appears to be multilayered, with inbuilt redundancy, but robust enough to ensure that most of the genome is spared from AID activity.

Cis elements and genome topology act together with trans-acting factors involved in transcription and RNA processing to determine AID activity at specific Ig regions. Other loci sharing genomic and transcriptional features with the Ig are collaterally targeted during SHM and CSR.

AID as a Central Factor to Generating Antibody Repertoires

Antibodies are essential to clearing viral, bacterial, and parasitic infections, as well as neutralizing toxins. The ability of Activation-Induced cytidine Deaminase (AID) to enhance antibody affinity through **somatic hypermutation (SHM)** (see Glossary) and alter the effector function of antibodies through **class switch recombination (CSR)** is vital to these tasks (see later). In addition, the generation of a long-lasting antigen-specific antibody memory response is crucial to establishing immunity against pathogens, and AID plays a key role in this endeavor by engraving memories of past pathogen experience through genetic modifications of immunoglobulin (*Ig*) genes. Here, we review the most essential findings on AID since its discovery two decades ago by providing a timely update on the molecular mechanisms of AID in B cell antibody production. As this review was written during the period that the world was engulfed by the COVID-19 pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-COV-2), we hope it will provide background to inspire vaccine design and the generation of therapeutic antibodies against SARS-CoV-2, as well as inevitable future pathogens.

General Features of Secondary Antibody Diversification Mechanisms

Vertebrates have evolved elegant genetic mechanisms to produce a repertoire of antibodies capable of recognizing and neutralizing a potentially unlimited number of foreign pathogens. Prior to antigen exposure, site-specific recombination of germline *Ig* variable (V), diversity (D), and joining (J) segments during B cell development provides the primary antibody repertoire [1]. The processes of SHM, **gene conversion (GCV)**, and CSR further diversify this antibody repertoire upon B cell-cognate antigen interactions. In antigen-activated B cells, SHM introduces point mutations in the *Ig* V region of both Ig heavy (*Igh*) and light chains to enable the process of **affinity maturation** in the **germinal center** of secondary lymphoid organs. GCV is an alternative mechanism of altering antibody specificity and affinity in select vertebrate species, including rabbits and birds. GCV involves the production of single-stranded (ss) DNA breaks or DSBs within the rearranged V(D)J region, triggering a homologous recombination-like process that repairs the lesion using *Ig* V pseudogenes. This produces templated sequence alterations within the V region.

[1]Department of Immunology, University of Toronto, Toronto, ON, Canada
[2]Institute de Recherches Cliniques de Montréal, Montréal, QC, Canada
[3]Molecular Biology Programs, Department of Medicine, University of Montreal, Montréal, QC, Canada
[4]Joint first authors.

*Correspondence:
javier.di.noia@ircm.qc.ca (J.M. Di Noia) and alberto.martin@utoronto.ca (A. Martin).

CSR is a region-specific deletional recombination event whereby the exons encoding for the IgM constant region within the *Igh* are replaced with a downstream constant region without changing the antigen-binding specificity, thereby altering the effector functions of the antibody. While all three processes are distinct genetic transactions, they share some common features, such as transcription of their target sequence [2], and preferential mutation of RGYW/WRCY motifs (where W=A/T, R=A/G, Y=C/T), wherein the underlined G/C is mutated [3]. While not demonstrated until the discovery of AID, these and other specific similarities suggested a common inception to each of these processes.

Discovery of AID and Its Molecular Mechanism in Antibody Diversification

AID was identified by Tasuku Honjo's group in 1999 [4], and shown to be essential for SHM and CSR in both mice and humans in 2000 [5,6], and later for GCV in chicken **DT40 B lymphoma cell line** [7]. **Variable lymphocyte receptor** diversification in ancient jawless vertebrates also depends on an AID-like cytidine deaminases named CDA1 and CDA2 [8], indicating that diversification of antigen receptors by deamination is an ancestral mechanism in vertebrates. AID was originally proposed to be an RNA editing enzyme due to sequence homology to its paralog, the RNA-cytosine deaminase apolipoprotein B mRNA editing catalytic polypeptide 1 (**APOBEC 1**) [4]. However, the DNA deamination model which proposes that AID promotes antibody diversity by deaminating deoxycytidine (dC) to deoxyuridine (dU) within *Ig* genes is currently supported by significant experimental evidence [9].

Purified AID preferentially mutates dC within the WRCY motif [10,11]. As this motif is preferentially mutated during SHM *in vivo* [12], this result provides strong evidence supporting the DNA deamination model for AID function. Biochemical data suggest that AID is catalytically inefficient [13,14]. The low catalytic rate has been proposed to act as an evolutionary safeguard to limit the mutagenic activity of AID [11]. Moreover, the accessibility of AID's catalytic pocket appears to be a key determinant of AID's catalytic rate [15].

AID has been suggested to act on ssDNA by inducing multiple deaminations per binding event [10]. A single-molecule resolution Förster resonance energy transfer (FRET) study visualized the ssDNA scanning motion of AID, and suggested that the enzyme was able to move bidirectionally while remaining bound to the same ssDNA [16]. This behavior of AID on ssDNA may play a role in enhancing mutagenic repair by overwhelming faithful DNA repair (see later).

Modeling AID on pre-existing structures of APOBEC enzymes [15] and crystallizing truncated AID variants [17,18] have provided valuable structural details. X-ray structure analysis with human monomeric AID reveals that AID prefers to bind and deaminate structured substrates, such as G-quadruplex (G4) structures that can form at the switch (S) region, over linear ssDNA substrates [18–20] (Box 1). However, as such structures are rare, especially over diverse *Ig* V exons, it is likely that other structures are also acted upon by AID, such as short patches of melted DNA [21]. These X-ray structure studies suggested that AID contains a bifurcated nucleic acid binding surface [15,18]: one required for catalysis, and another positively charged surface in AID helix 6, thought to play a role in recognizing structured substrates. A similar bifurcated nucleic acid binding surface has been identified in APOBEC3H, mediating dimerization [22]. Besides facilitating substrate binding, specific arginine residues in AID helix 6 may underpin a licensing mechanism that links AID to transcription elongation and warrants productive targeting [23]. Nevertheless, biochemical, biophysical and structural approaches have provided valuable knowledge on AID targeting mechanisms during SHM and CSR (see later). A detailed summary of the structure–function description of AID can be found in [9].

Glossary

Affinity maturation: phenomenon of increasing antibody affinity towards pathogens or antigens, days and weeks after the initial infection.
APOBEC 1: a member of a group of AID paralog enzymes that belong to the superfamily of cytidine deaminases and fulfill a diverse range of physiological functions, including lipid metabolism, immunity against foreign DNA and RNA viruses, and inhibition of retrotransposons.
CH12F3 B cells: clone derived from the CH12 murine B-cell lymphoma line that undergoes inducible AID expression and IgM to IgA CSR upon stimulation with CD40 ligand, interleukin (IL)-4, and transforming growth factor-β.
Class switch recombination (CSR): DNA recombination process initiated by AID that leads to isotype switching in B cells that typically progress from bearing IgM to IgG, IgA, or IgE.
Cohesin complex: multisubunit protein complex that outlines the chromatin architecture in different processes, such as DNA loop extrusion, recombination-based DNA repair, and chromatid cohesion during mitosis.
DIVAC elements: enhancers and enhancer-like DNA sequences originally found in *Ig* loci. DIVACs, contain a variety of transcription factor binding sites thought to cooperatively act to promote AID targeting by an unknown mechanism.
Divergent transcription: a process defined by RNAPII transcription in opposite directions, commonly observed as transcripts from DNA sequences upstream from promoters potentially regulating adjacent coding sequences, and which are typically degraded by the RNA exosome.
DT40 B lymphoma cell line: tumoral chicken B cell line, derived from an avian leukosis virus-induced bursal lymphoma that undergoes constitutive *Ig* V diversification by AID-mediated gene conversion and SHM.
Enhancer RNA (eRNA): type of ncRNA transcribed from enhancer elements, commonly degraded by the RNA exosome and the nonsense mediate decay pathway. eRNAs exhibit different poorly understood functions, and act mostly through the recruitment of regulatory proteins.
Gene conversion (GCV): homologous recombination-like process initiated by AID, responsible for short-track

Box 1. Role of Nucleic Acid Structures in Targeting AID

The transcriptional units encoding the different constant domains of *Igh* have revealed roles for RNA structures and RNA processing in CSR and AID targeting. These elements are transcribed from a cytokine-driven promoter, which produces a sterile transcript to allow AID access. The primary transcript contains a large intron encompassing the S region (1–12 kb), made up of repeats rich in WRCY motifs. The G-rich composition of the S regions enables the formation of secondary structures such as DNA–RNA hybrids producing R loops and G4 quadruplexes during transcription [20]. One model envisages that cotranscriptional formation of DNA G4 on the nontemplate strand ssDNA stabilizes R loops on the template strand, while processing of these R loops by the RNA exosome exposes ssDNA and permits deamination of both strands [20].

Others have proposed a role for R loops in CSR resolution, rather than on attracting AID activity [20]. AID shows high activity on G4 or branched DNA substrates in biochemical assays, allegedly due to a bifurcated substrate-binding surface that can recognize ssDNA overhangs on these structures [18]. G4 DNA additionally promotes AID oligomerization *in vitro*, which would facilitate DSB formation by increasing the deamination density at the S regions [18]; however, this remains to be tested *in vivo*.

Another model proposes that AID binds to G4-RNA formed in the debranched S-region intron, which would then guide AID back to the DNA [104], presumably by base complementarity. This model could account for the link of CSR to RNA splicing [110] and the interaction of AID with splicing factors such as heterogeneous nuclear ribonucleoproteins or polypyrimidine tract-binding proteins [9]. However, the mechanism by which G4-RNA guides and hands off AID to the DNA is not clear. The unwinding of G4-RNA in S-region transcripts by the DEAD-box RNA helicase 1 (DDX1) to allow the formation of DNA:RNA R-loop hybrids at the *Igh* by base complementarity was recently proposed as a mechanism [105]. However, ablation of DDX1 reduced CSR only by 25% when analyzed per cell division, indicating that this mechanism is not essential for CSR. Additional aspects of this model remain to be demonstrated, including evidence of AID transfer from RNA to DNA *in vivo*, and whether this transfer participates in targeting AID for SHM at the *Ig* V or at off-target loci.

Co-opted DNA Repair Pathways Involved in SHM

Uracils introduced in DNA are normally repaired faithfully, restoring the original DNA sequence (Figure 1). This is not the case during SHM, GCV, or CSR, whereby the initiating AID-induced lesion promotes further mutagenesis that benefits affinity maturation by providing a wide range of amino acid changes, or causes DNA breaks that trigger recombination events. Why faithful repair of DNA uracils is subverted during SHM has been the subject of intense investigation. The majority of the uracils in the genome are faithfully repaired by the base excision repair (BER) pathway: excision of uracils by uracil DNA glycosylase (UNG); incision at the resulting abasic sites by apurinic/apyrimidinic endonuclease 1; processing of the abasic sites by DNA polymerase β, followed by sealing of the final nick by DNA ligase 3 [24]. Uracils escaping detection by UNG or single-strand-selective monofunctional uracil-DNA glycosylase 1 (SMUG1) can be copied over by the replication machinery, producing transition mutations at G:C base pairs (Figure 1) [25,26]. However, during SHM, the later steps of BER are circumvented, for reasons that are not fully clear (see Outstanding Questions). Replication across the noninstructive abasic site by the translesional DNA polymerases REV1 and Pol η generates both transition and transversion mutations at G:C pairs [27,28] (Figure 1). Indeed, mice deficient in UNG (*Ung*$^{-/-}$) exhibit reduced transversion mutations at G:C pairs in *Ig* V regions relative to wild-type (WT) mice, with a minor impact on mutagenesis at A:T base pairs [29].

Mismatch repair (MMR) also mediates repair of uracil lesions, detected as U:G mispairs. Replication-associated canonical MMR (cMMR) is traditionally a high-fidelity process that preserves genomic stability. cMMR begins with the recognition of a base mispair by the MutSα heterodimer complex (MSH2/MSH6), followed by the recruitment of MutLα complex (MLH1/PMS2) and the activation of the endonuclease activity of PMS2, and removal of a track of the DNA strand by exonuclease EXO1 [30,31]. The resynthesis of the ssDNA gap by high-fidelity replicative polymerases restores the original DNA sequences. Instead, during SHM, a subverted MMR pathway termed noncanonical MMR (ncMMR) using the translesion DNA polymerase η is largely responsible for mutations at A:T base pairs [31–37] (Figure 1). Mutagenic ncMMR is largely linked to AID

sequence changes within *Ig* V regions in specific organisms (e.g., birds); it results in altered antibody specificity or increased antibody affinity.
Germinal center: specific structure within secondary lymphoid organs that is largely responsible for selecting B cells carrying high affinity antibodies towards specific antigens.
Karyopherins: family of proteins, largely importins α and β; mediate nuclear import of other proteins.
Off-target AID activity: AID-dependent mutations or DNA breaks outside *Ig* loci; detectable in normal and cancerous B cells, can produce translocations fusing *Ig* genes and other genomic regions with potential oncogenic consequences.
Peyer's patch: lymphoid follicles present in the small intestine; specialized in immunological surveillance of the intestinal lumen by means of constant sampling of pathogens and subsequent elimination through a local immune response.
Ramos B cells: human B cell line, derived from Burkitt's lymphoma; constitutively expresses AID and undergoes low frequency *Ig* V SHM.
RNA exosome: multiprotein complex comprising a nine-subunit core and two additional subunits with 3'→ 5' RNAse activity. It degrades different types of ncRNAs as well as aberrant mRNAs , either in the nucleus or cytoplasm.
Somatic hypermutation (SHM): mutational process initiated by AID that occurs within *Ig* V regions; it alters antibody affinity towards antigens.
Superenhancer: cluster of physically interconnected enhancers, rich in transcription factors and specific epigenetic marks, driving high gene expression; usually regulate cell identity genes.
Transcriptional stalling: regulated process whereby the RNAPII complex slows down within the gene body. It is regulated by mechanisms that are partially similar to the transcriptional pausing that takes place ~50 base pairs downstream from the transcription start site in many genes in higher eukaryotes.
Variable lymphocyte receptor (VLR): antigen receptors enriched with leucine-rich repeat domains in jawless vertebrates such as lampreys. By contrast, the structural unit of antigen receptors in jawed vertebrates (ranging from cartilaginous fishes to mammals) is the Ig-fold domain.

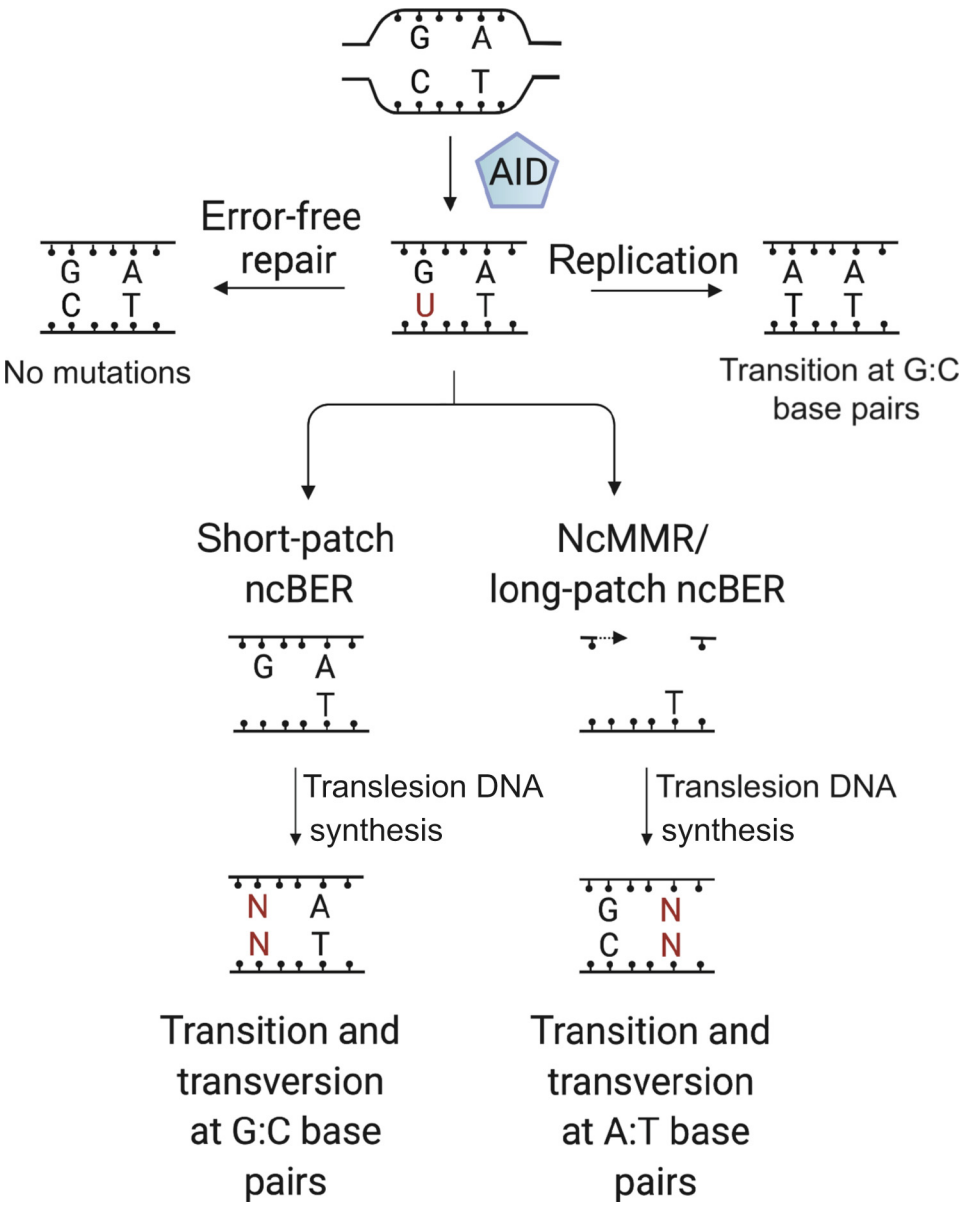

Figure 1. Multifaceted Uracil Lesion Processing in Somatic Hypermutation (SHM). AID-Induced dU:dG mismatches can be resolved by various pathways, leading to a diverse mutation spectrum at G:C and A:T base pairs in immunoglobulin variable (*Ig* V) regions. AID-induced uracil lesions can be repaired faithfully by canonical BER and MMR, leading to no SHM. Replicating over uracils by replicative polymerases generates transition mutations at G:C base pairs. UNG, and to a lesser extent, SMUG1, from the BER pathway, detect and excise uracils from DNA, creating abasic sites. Replicating over abasic sites by translesion synthesis polymerase from short-patch ncBER pathway generates transition and transversion mutations at G:C base pairs. Noncanonical MMR pathway or long-patch BER can excise single-stranded DNA surrounding either the dU:dG mismatch, or the abasic site, respectively, to initiate patch excision that leads to the engagement of translesion synthesis polymerases, generating both transition and transversion mutations at A:T base pairs. Abbreviations: AID, Activation-induced Cytidine Deaminase; ncBER, noncanonical base excision repair; ncMMR, noncanonical mismatch repair; SMUG1, single-strand selective monofunctional uracil DNA glycosylase; UNG, uracil DNA glycosylase. This figure was created using BioRender (https://biorender.com/).

lesions, as non-AID targets are faithfully repaired by MMR [38]. Recent work by a number of groups has shed light on ncMMR, whereby the UNG/SMUG1 and MMR pathways converge and work collaboratively to facilitate A:T and G:C mutagenesis [39–43]. This view is supported by findings that both UNG and ncMMR are able to function in the G1 phase of the cell cycle [41,42], where AID activity peaks [44]. Hence, the introduction of an abasic site within the EXO1 excision track induces translesional repair and mutations at A:T base pairs. This can arise in a number of ways: EXO1 excises the wrong strand of an AID-induced G:U mismatch, thereby exposing dU on the ssDNA track to uracil excision by UNG or SMUG1. Alternatively, AID disrupts cMMR by deaminating a dC within the EXO1 excision track, triggering UNG/SMUG1 to create an abasic site that leads to mutagenic translesional repair.

AID-Initiated DSB Formation and Repair during CSR

DSB formation in the Ig S region is a prerequisite for CSR. Ig S regions are highly enriched with AGCT motifs, a palindromic variant of the WRCY motif (Figure 2A) [45]. Similar to SHM, the generation of DSBs during CSR also co-opts the activities of UNG and MMR, as *Ung*$^{-/-}$ *Msh2*$^{-/-}$ mice show abolished CSR relative to WT mice [46]. MMR- and UNG-induced nicks and gaps on opposite DNA strands are converted into DSBs in donor and acceptor S regions, which are synapsed and ligated by DSB repair mechanisms, completing the CSR process. DSB repair during CSR primarily relies on canonical nonhomologous end joining (NHEJ), which promotes DSB end joining with no or minimal sequence microhomology. Alternative end joining (Alt-EJ) plays a minor role in CSR and utilizes short regions of microhomology between both DNA ends to join DSBs [47] (Figure 2B). DSB repair by NHEJ involves limited DNA end processing, whereas increased resection and microhomology usage are two features of repair by Alt-EJ. As such, mechanisms that regulate DNA end resection promote usage of one DSB repair pathway over the other (see later). The structure of the AID-induced DSB has also been suggested to provide a decision of which DSB repair pathway to use [48,49]. The stochastic activity of AID on both strands of DNA leads to a mixed spectrum of DSB structures in the S region. Accordingly, staggered DSBs are the preferred substrates for Alt-EJ, whereas blunt DSBs are preferentially repaired by the NHEJ pathway [49]. Indeed, CSR serves as an attractive physiological model for studying various DNA repair pathways at the same time, as all of BER, MMR, NHEJ, and Alt-EJ play key roles in DSB formation and repair.

NHEJ Cascade

The Ku70/80 complex, XRCC4, and Ligase 4 (Lig4) represent the core NHEJ factors based on their exclusive role in end joining. Upon AID-induced S-region DSB formation, the Mre11–Rad50–Nbs1 (MRN) sensor–-kinase complex binds DSBs and recruits the ataxia telangiectasia-mutated (ATM) kinase to the break site. ATM phosphorylates numerous substrates, leading to cell cycle arrest, as well as phosphorylating histone H2AX, which stimulates a cascade of events leading to the recruitment of 53BP1 to chromatin flanking the DSB [50,51]. 53BP1 promotes NHEJ and CSR by preventing DSB end resection [52] through the recruitment of Rif1 [53–56] and Shieldin, a protein complex composed of Shld1, Shld2, Shld3, and Rev7 [57–62] (Figure 2B). How the Shieldin complex prevents DNA end resection is still not clear, but it requires the binding of Shld2 to ssDNA exposed at DSBs [63] (see Outstanding Questions). After end processing, Lig4 and XRCC4 ligate the two DNA ends [64].

Alt-EJ

DSBs generated during CSR are also repaired by the Alt-EJ pathway [47]. Factors that are implicated in Alt-EJ include CtIP [65,66], which seems to have a role in end processing, XRCC1, ligase 1, and ligase 3, which are involved in end joining [67,68]. HMCES (5-Hydroxymethylcytosine

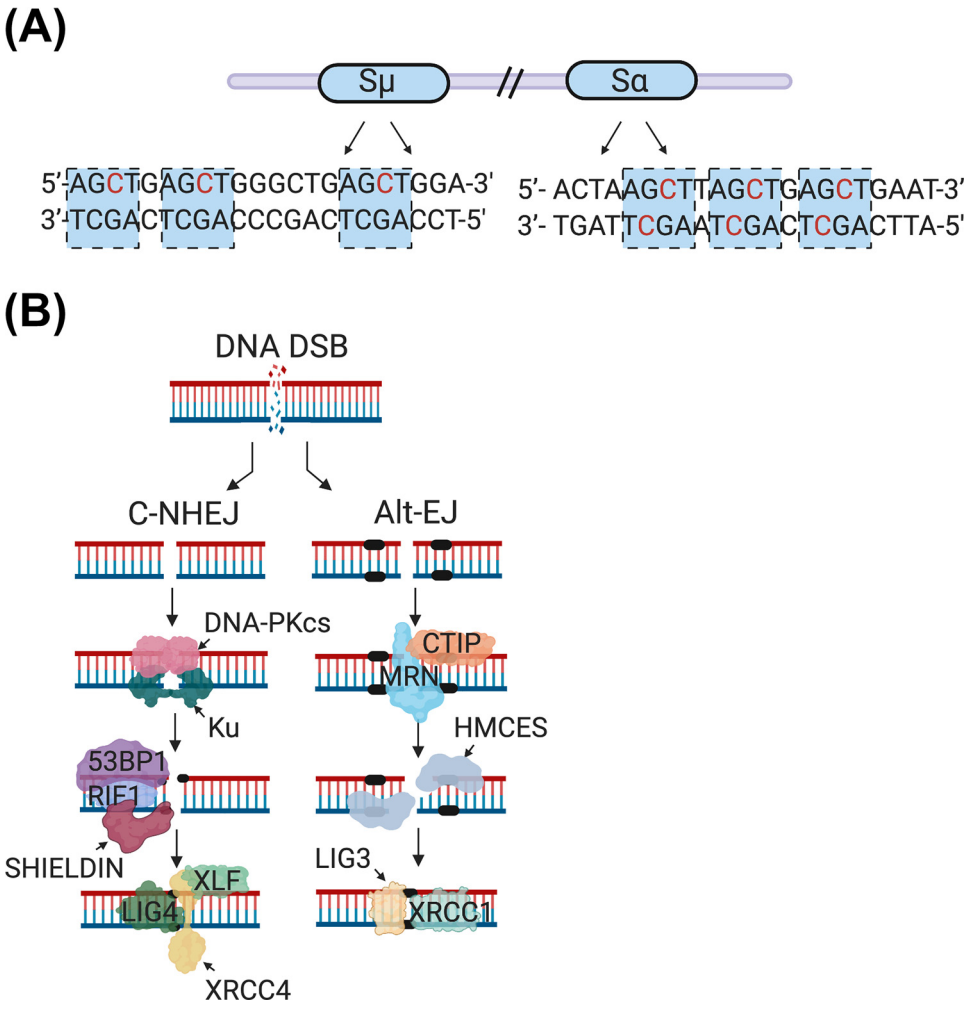

Figure 2. AID-Mediated DSB Formation within S Regions Drives CSR. (A) Sequence feature of S regions in mouse Ig heavy chain locus: S regions preceding different constant region exons are highly enriched with AID deamination hotspot 5′-AGCT-3′, a palindromic variant representing the canonical RGYW/WRCY sequences. Only Sμ and Sα are shown here for simplicity. (B) DSB ends are joined through NHEJ and Alt-EJ to complete CSR. NHEJ mediates DSB EJ with no or minimal microhomology, whereas Alt-EJ is facilitated by complementary base-paring interactions between microhomologous sequences present between the two synapsed S regions DSBs. Black bars represent homologous sequences near DSB ends. Abbreviations: 53BP1, p53-binding protein 1; AID, Activation-induced Cytidine Deaminase; Alt-EJ, alternative end joining; CSR, class switch recombination; CTIP, CtBP-interacting protein; DNA-PKcs, DNA-dependent protein kinase, catalytic subunit; DSB, double-strand break; ES cell specific; HMCES, 5-Hydroxymethylcytosine binding, ES-cell-specific; KU, Ku 70/80 complex; LIG3, ligase 3; LIG4, ligase 4; MRN, MRE11–RAD50–NBN complex; NHEJ, nonhomologous end joining; RIF1, replication timing regulatory factor 1; S, switch; SHIELDIN, protein complex composed of SHLD1, SHLD2, SHLD3, and REV 7; XLF, XRCC4-like factor; XRCC 1/4, X-ray repair cross-complementing protein 1/4. This figure was created using BioRender (https://biorender.com/).

binding, ES-cell-specific) acts as an abasic site sensor in ssDNA [69], and also functions in the Alt-EJ pathway during CSR [70]. HMCES-defective mouse primary B cells and **CH12F3 B cells** have shown reduced microhomology usage in switch junctions relative to WT cells [70]. This finding has been postulated to be mechanistically due to HMCES's ability to bind ssDNA overhangs, likely caused by strand resection of DSB at the S regions, through its SOS-response-associated-peptidase domain [71].

AID Targeting

The mechanism by which AID preferentially mutates *Ig* genes is still incompletely understood (see Outstanding questions). This issue has implications for immunity as well as for the oncogenic mutations introduced by AID outside the *Ig* loci [72]. AID targeting entails successive steps of increasing specificity: entering the nucleus, associating quite extensively to chromatin, and mutating *Ig* loci at a high frequency, in addition to a small fraction of other genes at low (but significant) frequency.

Entering the Nucleus

Several mechanisms regulate the amount of AID protein per B cell present at steady state, as previously reviewed [72,73]. AID is a nuclear–cytoplasmic shuttling protein largely excluded from the nucleus at steady state. Restraining nuclear residence limits AID activity, which may balance *Ig* diversification and **off-target AID activity**. The small proportion of AID present in the nucleus during the cell cycle interphase reflects a dynamic equilibrium between opposing mechanisms.

AID is retained in the cytoplasm; a heat shock protein 90–DnaJ homolog subfamily A member 1 (HSP90–DNAJA1) complex first stabilizes AID [9]. This molecular chaperone stabilizes proteins until they acquire functional proficiency. For AID, this step is defined by its transition to another cytoplasmic complex that contains elongation factor 1α (eEF1A) and holds the bulk of functional AID, preventing its diffusion into the nucleus [74]. The maturation step required for this transition, and whether this process is regulated, remain unknown [74]. In addition, nuclear AID is either continuously exported by chromosomal maintenance 1 (CRM1) or targeted for proteasomal degradation [9].

Three mechanisms can allow AID into the nucleus. First, AID continuously enters the nucleus through a nuclear localization signal recognized by **karyopherins** [9]. Second, AID is trapped in the nucleus when the nuclear membrane reassembles at the end of mitosis [44]. This transient accumulation correlates with the time when more AID-induced mutations are observed, and is consistent with SHM and CSR being initiated in the G1 phase [72]. Third, AID can transiently and stochastically accumulate in the nucleus in short pulses [75]. This provocative observation was made in non-B cells (i.e., human fibrosarcoma line HT1080 and SV40-transformed fibroblast line GM639) overexpressing AID–GFP, so the pulses could not be correlated with SHM or CSR. Unlike AID pulses, AID nuclear import and entrapment during mitosis have both been confirmed in B cells by several groups [23,44,76–78]. Yet, it is unclear why a nuclear import mechanism for AID would evolve if passive trapping was sufficient to access the nucleus for SHM and CSR. While different AID fractions could conceivably access the nucleus by different pathways, this issue remains unknown, but certainly warrants further research.

Associating with Chromatin and Transcription

AID most preferentially mutates *Ig* genes but it also deaminates other genes (termed off-target genes) in normal B cells. In addition, the available data are consistent with AID occupying up to thousands of loci, of which just ~450 genes have been found mutated in normal B cells [23,79–82] (Boxes 2 and 3). These findings suggest the existence of AID susceptible and nonsusceptible loci [72]. If the occupancy of a locus by AID does not necessarily result in mutagenesis, what determines AID activity after it reaches chromatin?

Biochemical fractionation indicates that ~50% of nuclear AID is associated with chromatin in mouse primary B cells and the CH12F3 B cell line [23,83]. The phosphorylated form of AID at Ser38 (pS38), a protein kinase A substrate, is enriched in the chromatin-associated AID fraction [83], which allows its interaction with the ssDNA-binding replication protein A (RPA), boosting

Box 2. Chromatin Occupancy of AID

It is important to identify the loci occupied by AID to understand its targeting mechanisms. The only available analysis of endogenous AID chromatin occupancy was performed by ChIP-seq in primary activated mouse B cells, which identified >5000 loci occupied by AID [23,82]. Others reanalyzed the data and reported no AID-specific peaks [111], which was rebutted, indicating that AID ChIP-seq requires specific analyses [112]. ChiP-seq peaks are defined by signal enrichment over background. The signal of AID ChIP is typically low, as only ~5% of the B cell AID is chromatin-associated according to biochemical fractionation [23,83]. The background is primarily determined by the specificity of the antibody. The AID ChIP-seq study reported two experiments using different antibodies with similar results. They also used primary B cells from AID knockout mice as negative controls [23,82]. Chromatin fragmentation by mechanical shearing can be a source of nonspecific signals. Highly transcribed DNA regions are hypersensitive to sonication, resulting in an over-representation of these regions in the background. AID ChIP-seq revealed 5514 genes occupied in WT cells, of which 396 were shared with AID-deficient samples, and were also regions that were hypersensitive to sonication, (e.g., the *Igh* locus). Given the preference of AID for accessible chromatin, this study distinguished specific AID peaks from the background in AID-deficient B cells by considering this hypersensitivity effect for normalizing the signal [23,82,112].

While independent confirmation of these data is desirable, a large number of occupied loci is consistent with the broadly distributed chromatin-bound endogenous AID observed in the CH12F3 mouse B cell line by immunofluorescence, and by the biochemical abundance of chromatin-bound AID [23,83]. A stepwise process of AID targeting might harmonize these observations with the much smaller number of genes that AID has been found to mutate [79–81,102,113] (Figure I). Nuclear AID could broadly associate to the chromatin through ribonucleoproteins [72], as suggested by RNA dependency [23]. Interaction with paused RNAPII–SPT5 could allow accumulation around the TSS, consistent with the CHIP-seq profile of endogenous AID in B cells [82]; this would then enable its access to selected gene bodies based on the mechanism discussed in Figure 3 in the main text. This model is analogous to how transcription or DNA repair factors find their target sequences by scanning chromatin, and provides a rationale for the broad association of AID with chromatin and its presence around the TSS of unmutated genes.

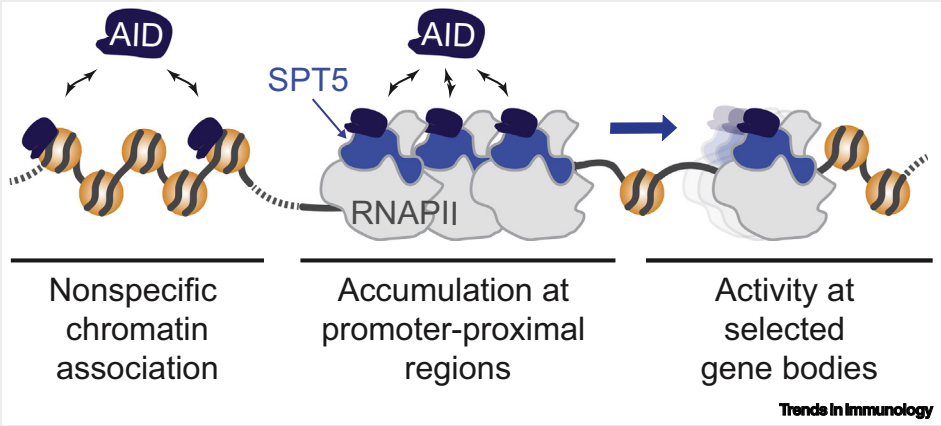

Figure I. Stepwise Model for Productive Deamination of Selected AID Off-Targets. While AID may be able to associate to chromatin in different ways, it could accumulate at paused promoter regions of many genes by virtue of its interaction with SPT5, but access only a subset of those genes bodies, for instance by a licensing mechanism and/or the abundance of ssDNA substrate. Abbreviations: AID, Activation-induced Cytidine Deaminase; RNAPII, RNA polymerase II.

CSR and SHM [84,85]. Notably, pS38 seems dispensable for AID activity on *Myc* and increasing pS38 increases *Igh–Myc* translocations without affecting CSR [86]. This suggests that pS38, and potentially other post-translational modifications, can provide context-dependent regulation of AID activity at different loci.

The *Ig* and off-target genes share some features that suggest similarities between the underlying mutation mechanism. One basic common requirement for AID targeting is transcription [72]. AID associates with the RNA polymerase II (RNAPII) complex and with multiple transcription, splicing, and RNA processing factors, as well as with RNA–DNA hybrid structures arising from transcription,

Box 3. Identification of AID Off-Targets

Three strategies have been used to identify AID off-target genes. First, sequencing of select genes revealed a mutation pattern resembling SHM in human B cell lymphoma samples expressing AID, notably diffuse large B-cell lymphoma (DLBCL) [114]. The identity of these relatively few genes can be biased by selection of mutations advantageous for the tumor cells and, accordingly, include many proto-oncogenes or tumor suppressors. Second, mid- and high-throughput sequencing analyses have been undertaken in normal, activated, or germinal center B cells, in mouse models. These efforts still sampled a small proportion of the genome, selecting genes based on defined criteria [79,80]. These approaches have also revealed that some AID-deaminated genes are subsequently repaired by UNG-BER and/or MMR with high efficiency [79,80]. Third, a series of studies have relied on methods that detect and/or trap AID-induced DSB [102,113,115]. These require off-target mutations to be converted to DSB, likely biasing the identification towards genes with the highest deamination load and/or displaying features that facilitate DSB. The latter could include high transcription and/or AID occupancy, enrichment in the WRCY motifs preferred by AID, the formation of secondary structures, or a combination thereof. For instance, G4 DNA may be preferred by AID and may amplify its activity by inducing AID oligomerization (Box 1) [18].

The overlap between the lists of AID off-target genes determined by each method is only partial, demonstrating that each approach has a different bias, which must be considered in AID targeting models. The implementation of unsupervised analysis of mutational signatures in large cancer sequence datasets has expanded the potential off-targets of AID in human samples, but this represents correlative data [116,117]. It would be possible to use a similar unbiased approach to obtain the full repertoire of AID off-targets in normal B cells by comparing germinal center B cells obtained from AID-proficient and -deficient mice, perhaps in combination with UNG and/or MMR deficiencies.

which have all been implicated in AID targeting [9,72]. As these are all ubiquitous factors, none by themselves can explain productive AID targeting; thus, they must be integrated into a comprehensive picture that in conjunction with chromatin architecture, could account for AID specificity.

Chromatin Architecture

Chromatin architecture has emerged as an important determinant of genome-wide mutagenic AID targeting. AID-susceptible loci are associated with highly accessible chromatin and significantly correlate with histone marks typical of active enhancers (i.e., H3K27Ac and H3K4me1) and transcription elongation (i.e., H3K36me3 and H3K79me2) [87]. AID activity is enriched at promoter-proximal exons and introns, but can also be detected at enhancers that form regulatory clusters with other promoters and enhancers through long-range interactions [81]. Most AID-susceptible clusters regulated by these **superenhancers** are highly interconnected and show greater accessibility than other regions (Figure 3A, B). However, the proportion of AID off-targets regulated by superenhancers ranges between 40 and 70% [80,81,88]. The discrepancy may originate from the different techniques used to detect AID off-target activity (Box 3).

An overlapping of superenhancers with gene bodies of susceptible loci can define the site of mutagenesis, at least in part, by producing antisense transcription that clashes with transcription from the gene promoter. This convergent transcription can stall the RNAPII, causing early termination, and/or recruit RNA processing factors to favor AID activity by exposing ssDNA on both strands [72,89] (Figure 3C). In transcription-linked *in vitro* deamination assays, the transcript would normally protect the template strand from deamination, as the nontemplate strand is usually mutated. However, purified AID can mutate both strands of an *Ig* V region when transcribed in the presence of SPT4 and SPT5, a heterodimeric complex that regulates RNAPII processivity [90]. This suggests that AID has an intrinsic ability to mutate both strands under those conditions, but it is likely that in the more complex context of chromatin, some target genes require convergent transcription to expose both DNA strands to AID [88,89].

Chromatin organization into topologically associated domains (TADs) can contribute to AID activity by bringing together distant DNA regions, as exemplified by chromatin extrusion to ensure

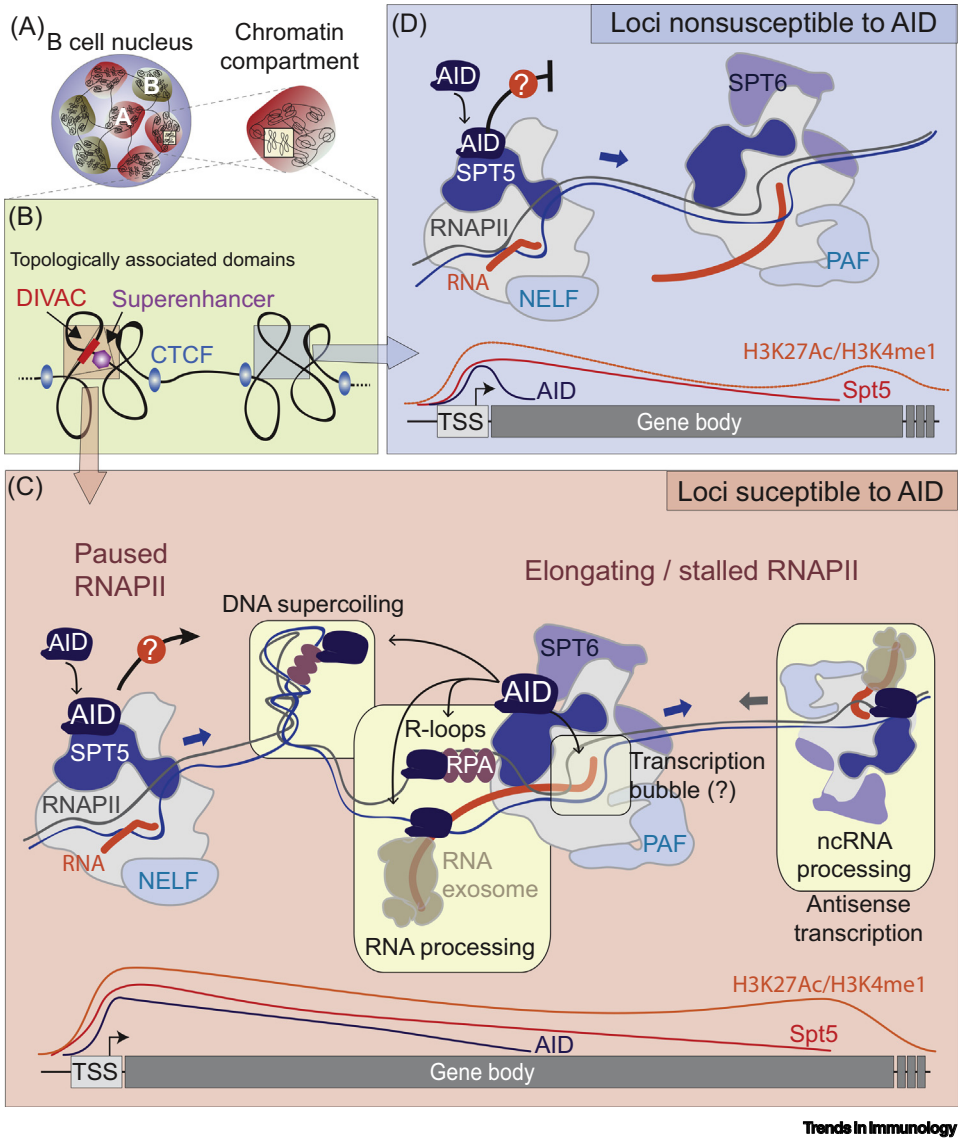

Figure 3. Multilayered Productive AID Targeting in B cells. (A) AID activity is restricted to type A chromatin compartments, which are accessible and transcriptionally active, versus transcriptionally inactive B compartments. (B) AID activity is also limited to some TADs of DNA, often affecting highly transcribed genes regulated by superenhancers, with clusters by long-range interactions between regulatory sequences. The *Ig* and some off-target genes additionally harbor *cis*-acting elements named DIVAC (diversification activators) that attract AID activity. (C, D) AID can be recruited to promoter or promoter-proximal paused RNAPII via SPT5. (C) AID accesses the gene body of susceptible loci through associations with the elongating RNAPII complex, via a putative licensing mechanism. At these loci, several mechanisms can provide ssDNA AID substrates (yellow boxes): DNA supercoiling, DNA secondary structures, and RNA processing of sense and antisense RNA. Transcriptional stalling, caused by nucleic acid secondary structures or by early termination due to antisense transcription, can also expose and favor AID jumping to ssDNA, possibly helped by the regulated association of AID to the ssDNA binding protein RPA. Stalling and SPT5 association may also allow time for AID to act on the small transcription bubble within the RNAPII. (D) At loci that are not susceptible to mutation, AID may fail its coupling with transcription elongation, and/or lack access to ssDNA substrates. Abbreviations: AID, Activation-induced Cytidine Deaminase; ncRNA, noncoding RNA; NELF, negative elongation factor (causes RNAPII pausing); PAF, RNAPII-associated factor complex; RNAPII, RNA polymerase II; ssDNA, single-stranded DNA; TAD, topologically associated domain; TSS, transcription start site.

CSR directionality [91]. Deletional CSR requires long-range interactions between the donor (Eμ) and acceptor (E_{acc}) S-region enhancers, and the *Igh* 3′regulatory region (3′RR) [92]. The loop extrusion mechanism envisages **cohesin**-loading *cis* elements in the *Igh*, drawing the 3′RR and Eμ together to form a CSR center. Transcription of an acceptor S region induced by cytokine signaling would expose E_{acc} cohesin-binding sites, driving chromatin extrusion for juxtaposing the E_{acc} with the CSR center, and thus arranging distal AID-induced DSBs in a direction favorable for deletional CSR. Additionally, long-range inter-TAD interactions promoted by noncoding RNA can contribute to AID function. Specifically, disrupting transcription by CRISPR/Cas9 on a TAD containing a long noncoding RNA that is 2.6 Mb away from the *Igh* is sufficient to halve the efficiency of CSR to IgA in mouse **Peyer's patch** B cells and in the CH12F3 B cell line [93].

Gene Deamination Susceptibility

Since neither transcription, AID occupancy, nor the presence of superenhancers is sufficient to predict mutation susceptibility of a gene, additional features must contribute to define susceptibility to AID. The presence of *cis* elements, quality of transcription, formation of DNA/RNA structures, and a licensing mechanism coupling AID to transcription elongation have been proposed [94] (reviewed later).

Cis Elements

Ig loci are privileged for SHM [95], implying that AID is preferentially targeted to and/or more active at those loci. The cause of this preference is still unclear but is independent of the *Ig* promoter. *Ig* loci have strong enhancers and are highly transcribed, and at least the *Igh* locus harbors a superenhancer [81]. *Ig* enhancers contain defined evolutionarily conserved elements, known as **DIVAC elements** (diversification activators), that are sufficient to attract SHM to a juxtaposed transcribed sequence [96]. DIVACs can affect mutability at the TAD level [97]. Thus, the insertion of a DIVAC into an otherwise nontargeted TAD can attract SHM, as shown in the **Ramos B cell** line [97]. In this study, adjacent TADs were spared, additionally suggesting a topologically delineated confinement of AID mutagenesis [97]. The same study found that some AID off-target genes have associated enhancers with DIVAC activity [97], which may explain why they are colaterally mutated by AID.

The mechanism of DIVAC activity is unclear. Evidence in the DT40 and Ramos B cell lines suggests that recruitment of different transcription factors to DIVACs enhances SHM without an effect on transcription [96,97]. Nevertheless, their relative contribution remains to be dissected. DIVACs are transcribed, producing **enhancer RNA (eRNA)** from both strands [97]. Bidirectional transcription, either convergent (as discussed above) or **divergent transcription**, and eRNA processing by the **RNA exosome** [89] can contribute to the AID targeting activity of the DIVAC (Figure 3C,D). Although eRNA has not been shown to be a universal feature of AID targets, it is possible that some intragenic AID off-target sites are due to bidirectional transcription driven by overlapping intragenic enhancers [88,89].

Transcription Quality and Nucleic Acid Structures

RNAPII and its processivity factor SPT5 are required for AID activity and colocalize with AID genome-wide [23,98]. SPT5 mediates the AID–RNAPII interaction and is sufficient to recruit AID to chromatin [23,98]. The function of SPT5 is to initially cause promoter-proximal pausing of RNAPII [99], and is thus enriched immediately downstream from the transcription start site (TSS). AID shows a similar genome-wide pattern by ChIP-seq in primary B cells, consistent with its accumulation at SPT5-rich sites (Box 2) [82,98]. During transcription elongation, SPT5 remains associated with and acts as a positive factor for elongating RNAPII; thus, both accumulate inside gene bodies upon **transcriptional stalling**, a phenomenon considered

key for AID activity [90,98,100,101]. Accordingly, stalled RNAPII and SPT5 are found within gene bodies at the *Ig* V and S regions, as well as at some AID off-targets [72,100,102], which could recruit AID to stalled transcription [98,100]. However, SPT5 is still abundant at the *Ig* constant regions and at many genes that are occupied by AID, but spared from mutation [80,81,88,98]. Since AID is an inefficient enzyme *in vitro*, RNAPII stalling via SPT5 or RNA–DNA structures might not only expose ssDNA but also provide time for AID to work [15,90] (Figure 3C). DNA–RNA R loops or G4 DNA formed during transcription can stall RNAPII [103], and in fact, are preferred AID substrates *in vitro* [20]; this might in turn contribute to AID susceptibility at certain loci. Another targeting mechanism has been proposed in which G4-RNA formed in the intron of the S-region transcript binds AID and brings it to the cDNA [104,105] (Box 1). However, while R loops and G4 can form in the S region and some AID off-targets such as *Myc* [19,20,89], their presence at the *Ig* V and at other AID off-targets, is unlikely and uncertain, respectively; therefore, other mechanisms must be at play for SHM [20] (Box 1). Transcriptional stalling may cause premature termination, another potential source of ssDNA demonstrated at the *Ig* V [101]. In addition, transcription-induced DNA supercoiling also provides ssDNA substrate to AID [106].

Coupling to Transcription Elongation

The identification of catalytically active AID mutants that can occupy the *Ig* promoter, but not the gene body, and fail to enable SHM or CSR, has suggested a model in which AID is recruited by SPT5 to promoter-proximal-paused RNAPII; in this model, AID gains access to the gene body by traveling with transcription elongation [23]. Supporting the link between AID activity and transcription elongation, factors that interact with RNAPII after the pause release, namely SPT5, RNAPII-associated factor (PAF) and the histone chaperones SPT6 and FAcilitates Chromatin Transcription (FACT), are important for SHM and/or CSR, and biochemically copurify with AID [72,107,108]. Whether histone chaperones facilitate AID activity at the *Ig* by their role in transcription or by defining histone modifications, is unclear. Of note, the RNA exosome has been linked to elongating RNAPII via SPT6 in *Drosophila* cells [109] which may facilitate AID activity. A more precise understanding of how the transition from transcriptional pausing/stalling to elongation has an impact on AID activity is required. A speculative model integrating several observations would have AID accessing the gene body with elongating RNAPII, but only being able to mutate genes in which the transcriptional and genomic contexts cause simultaneous RNAPII stalling and ssDNA exposure (Figure 3C and Outstanding Questions). In addition, the dissociation of AID from *Ig* loci to prevent mutation of constant regions must be somehow regulated – a step that has received little attention.

Concluding Remarks

Research carried out over the past two decades has provided a fundamental understanding of B cell-mediated adaptive immunity by uncovering the biological functions of AID, including its biochemical and structural properties, as well as the DNA repair pathways downstream of AID, and the regulatory mechanisms that facilitate productive AID targeting at *Ig* loci. Although there is no evidence of a specific AID targeting factor, the implication of ubiquitous transcription-associated factors, and the verification that several factors contribute to, but are not essential for SHM or CSR, suggests a multilayered AID targeting mechanism in which chromatin architecture, transcription quality, and AID regulation all contribute to define whether a locus is mutated or not, and to what extent. This is a fundamental mechanism to further understand, as it has direct implications in affinity maturation, and might eventually be manipulated to improve the efficiency of immunizations. In addition, future endeavors should be directed towards developing a better understanding of the biochemical events involved in antibody diversification processes, including enzymology and mechanisms of action during SHM and CSR.

Outstanding Questions

What is the precise mechanism leading to mutagenic repair of AID-induced uracil residues within Ig loci?

Does the Shieldin complex in CSR have a similar mechanism of action as Shelterin-mediated telomere end processing?

Is the enzymology for the SHM and CSR processes completely understood?

How much each of the proposed AID targeting mechanisms contributes to SHM, CSR, and off-target mutagenesis? Which of these mechanisms are in fact linked and which are independent of each other?

What exact combination of genomic features and factors promotes AID activity after occupancy?

What is the mechanistic role of G4 and R-loop structures in AID targeting to *Ig* and other genes?

Is there a coordination between the transcriptional landscape of susceptible genes and AID post-translational modifications or licensing?

Acknowledgments

Some primary literature was not cited due to reference number limitations. Work in the laboratory of J.D.N. is supported by grants from the Canadian Institutes of Health Research (PJ-155944 and PJ-166028) and The Cancer Research Society. J.D.N is a Merit scholar from the Fonds de recherche de Quebec – Santé. Work in the laboratory of A.M. is supported by grants from the Canadian Institutes of Health Research (PJT-153307 and PJT-156330).

References

1. Schatz, D.G. *et al.* (1989) The V(D)J recombination activating gene, RAG-1. *Cell* 59, 1035–1048
2. Peters, A. and Storb, U. (1996) Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity* 4, 57–65
3. Rogozin, I.B. and Diaz, M. (2004) Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. *J. Immunol.* 172, 3382–3384
4. Muramatsu, M. *et al.* (1999) 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.* 274, 18470–18476
5. Revy, P. *et al.* (2000) Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* 102, 565–575
6. Muramatsu, M. *et al.* (2000) Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 102, 553–563
7. Arakawa, H. *et al.* (2002) Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. *Science* 295, 1301–1306
8. Morimoto, R. *et al.* (2020) Cytidine deaminase 2 is required for VLRB antibody gene assembly in lampreys. *Sci. Immunol.* 5
9. Methot, S.P. and Di Noia, J.M. (2017) Molecular mechanisms of somatic hypermutation and class switch recombination. *Adv. Immunol.* 133, 37–87
10. Pham, P. *et al.* (2003) Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature* 424, 103–107
11. Larijani, M. and Martin, A. (2012) The biochemistry of activation-induced deaminase and its physiological functions. *Semin. Immunol.* 24, 255–263
12. Rogozin, I.B. and Kolchanov, N.A. (1992) Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochim. Biophys. Acta* 1171, 11–18
13. Larijani, M. *et al.* (2007) AID associates with single-stranded DNA with high affinity and a long complex half-life in a sequence-independent manner. *Mol. Cell. Biol.* 27, 20–30
14. Pham, P. *et al.* (2011) Analysis of a single-stranded DNA-scanning process in which activation-induced deoxycytidine deaminase (AID) deaminates C to U haphazardly and inefficiently to ensure mutational diversity. *J. Biol. Chem.* 286, 24931–24942
15. King, J.J. *et al.* (2015) Catalytic pocket inaccessibility of activation-induced cytidine deaminase is a safeguard against excessive mutagenic activity. *Structure* 23, 615–627
16. Senavirathne, G. *et al.* (2015) Activation-induced deoxycytidine deaminase (AID) co-transcriptional scanning at single-molecule resolution. *Nat. Commun.* 6, 10209
17. Pham, P. *et al.* (2016) Structural analysis of the activation-induced deoxycytidine deaminase required in immunoglobulin diversification. *DNA Repair (Amst)* 43, 48–56
18. Qiao, Q. *et al.* (2017) AID recognizes structured DNA for class switch recombination. *Mol. Cell* 67, 361–373.e4
19. Duquette, M.L. *et al.* (2005) AID binds to transcription-induced structures in c-MYC that map to regions associated with translocation and hypermutation. *Oncogene* 24, 5791–5798
20. Pavri, R. (2017) R loops in the regulation of antibody gene diversification. *Genes (Basel)* 8, 154
21. Larijani, M. and Martin, A. (2007) Single-stranded DNA structure and positional context of the target cytidine determine the enzymatic efficiency of AID. *Mol. Cell. Biol.* 27, 8038–8048
22. Bohn, J.A. *et al.* (2017) APOBEC3H structure reveals an unusual mechanism of interaction with duplex RNA. *Nat. Commun.* 8, 1021
23. Methot, S.P. *et al.* (2018) A licensing step links AID to transcription elongation for mutagenesis in B cells. *Nat. Commun.* 9, 1248
24. Krokan, H.E. and Bjoras, M. (2013) Base excision repair. *Cold Spring Harb. Perspect. Biol.* 5, a012583
25. Di Noia, J. and Neuberger, M.S. (2002) Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* 419, 43–48
26. Di Noia, J.M. *et al.* (2006) SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair. *EMBO J.* 25, 585–595
27. Krijger, P.H. *et al.* (2013) Rev1 is essential in generating G to C transversions downstream of the Ung2 pathway but not the Msh2+Ung2 hybrid pathway. *Eur. J. Immunol.* 43, 2765–2770
28. Kano, C. *et al.* (2012) Analysis of mice deficient in both REV1 catalytic activity and POLH reveals an unexpected role for POLH in the generation of C to G and G to C transversions during Ig gene hypermutation. *Int. Immunol.* 24, 169–174
29. Rada, C. *et al.* (2002) Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr. Biol.* 12, 1748–1755
30. Jiricny, J. (2006) The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol.* 7, 335–346
31. Chahwan, R. *et al.* (2012) AIDing antibody diversity by error-prone mismatch repair. *Semin. Immunol.* 24, 293–300
32. Cascalho, M. *et al.* (1998) Mismatch repair co-opted by hypermutation. *Science* 279, 1207–1210
33. Rada, C. *et al.* (1998) Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity* 9, 135–141
34. Wiesendanger, M. *et al.* (2000) Somatic hypermutation in MutS homologue (MSH)3-, MSH6-, and MSH3/MSH6-deficient mice reveals a role for the MSH2-MSH6 heterodimer in modulating the base substitution pattern. *J. Exp. Med.* 191, 579–584
35. Bardwell, P.D. *et al.* (2004) Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice. *Nat. Immunol.* 5, 224–229
36. Kim, N. *et al.* (1999) Different mismatch repair deficiencies all have the same effects on somatic hypermutation: intact primary mechanism accompanied by secondary modifications. *J. Exp. Med.* 190, 21–30
37. Zeng, X. *et al.* (2001) DNA polymerase eta is an A-T mutator in somatic hypermutation of immunoglobulin variable genes. *Nat. Immunol.* 2, 537–541
38. Green, B. *et al.* (2011) The mismatch repair pathway functions normally at a non-AID target in germinal center B cells. *Blood* 118, 3013–3018
39. Frieder, D. *et al.* (2009) The concerted action of Msh2 and UNG stimulates somatic hypermutation at A. T base pairs. *Mol. Cell Biol.* 29, 5148–5157
40. Krijger, P.H. *et al.* (2009) Dependence of nucleotide substitutions on Ung2, Msh2, and PCNA-Ub during somatic hypermutation. *J. Exp. Med.* 206, 2603–2611
41. Girelli Zubani, G. *et al.* (2017) Pms2 and uracil-DNA glycosylases act jointly in the mismatch repair pathway to generate Ig gene mutations at A-T base pairs. *J. Exp. Med.* 214, 1169–1180
42. Thientosapol, E.S. *et al.* (2017) Proximity to AGCT sequences dictates MMR-independent versus MMR-dependent mechanisms for AID-induced mutation via UNG2. *Nucleic Acids Res.* 45, 3146–3157
43. Dingler, F.A. *et al.* (2014) Uracil excision by endogenous SMUG1 glycosylase promotes efficient Ig class switching and impacts on A:T substitutions during somatic mutation. *Eur. J. Immunol.* 44, 1925–1935

44. Wang, Q. *et al.* (2017) The cell cycle restricts activation-induced cytidine deaminase activity to early G1. *J. Exp. Med.* 214, 49–58
45. Xu, Z. *et al.* (2010) 14-3-3 adaptor proteins recruit AID to 5'-AGCT-3'-rich switch regions for class switch recombination. *Nat. Struct. Mol. Biol.* 17, 1124–1135
46. Rada, C. *et al.* (2004) Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol. Cell* 16, 163–171
47. Yan, C.T. *et al.* (2007) IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature* 449, 478–482
48. Cortizas, E.M. *et al.* (2013) Alternative end-joining and classical nonhomologous end-joining pathways repair different types of double-strand breaks during class-switch recombination. *J. Immunol.* 191, 5751–5763
49. Ling, A.K. *et al.* (2018) Double-stranded DNA break polarity skews repair pathway choice during intrachromosomal and interchromosomal recombination. *Proc. Natl. Acad. Sci. U. S. A.* 115, 2800–2805
50. Petersen, S. *et al.* (2001) AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. *Nature* 414, 660–665
51. Fradet-Turcotte, A. *et al.* (2013) 53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark. *Nature* 499, 50–54
52. Bunting, S.F. *et al.* (2010) 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell* 141, 243–254
53. Di Virgilio, M. *et al.* (2013) Rif1 prevents resection of DNA breaks and promotes immunoglobulin class switching. *Science* 339, 711–715
54. Chapman, J.R. *et al.* (2013) RIF1 is essential for 53BP1-dependent nonhomologous end joining and suppression of DNA double-strand break resection. *Mol. Cell* 49, 858–871
55. Zimmermann, M. *et al.* (2013) 53BP1 regulates DSB repair using Rif1 to control 5' end resection. *Science* 339, 700–704
56. Escribano-Diaz, C. *et al.* (2013) A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. *Mol. Cell* 49, 872–883
57. Noordermeer, S.M. *et al.* (2018) The shieldin complex mediates 53BP1-dependent DNA repair. *Nature* 560, 117–121
58. Dev, H. *et al.* (2018) Shieldin complex promotes DNA end-joining and counters homologous recombination in BRCA1-null cells. *Nat. Cell Biol.* 20, 954–965
59. Gupta, R. *et al.* (2018) DNA repair network analysis reveals shieldin as a key regulator of NHEJ and PARP inhibitor sensitivity. *Cell* 173, 972–988.e23
60. Barazas, M. *et al.* (2018) The CST complex mediates end protection at double-strand breaks and promotes PARP inhibitor sensitivity in BRCA1-deficient cells. *Cell Rep.* 23, 2107–2118
61. Ghezraoui, H. *et al.* (2018) 53BP1 cooperation with the REV7-shieldin complex underpins DNA structure-specific NHEJ. *Nature* 560, 122–127
62. Findlay, S. *et al.* (2018) SHLD2/FAM35A co-operates with REV7 to coordinate DNA double-strand break repair pathway choice. *EMBO J.* 37
63. Setiaputra, D. and Durocher, D. (2019) Shieldin - the protector of DNA ends. *EMBO Rep.* 20
64. Grawunder, U. *et al.* (1997) Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* 388, 492–495
65. Lee-Theilen, M. *et al.* (2011) CtIP promotes microhomology-mediated alternative end joining during class-switch recombination. *Nat. Struct. Mol. Biol.* 18, 75–79
66. Bothmer, A. *et al.* (2013) Mechanism of DNA resection during intrachromosomal recombination and immunoglobulin class switching. *J. Exp. Med.* 210, 115–123
67. Lu, G. *et al.* (2016) Ligase I and ligase III mediate the DNA double-strand break ligation in alternative end-joining. *Proc. Natl. Acad. Sci. U. S. A.* 113, 1256–1260
68. Masani, S. *et al.* (2016) Redundant function of DNA ligase 1 and 3 in alternative end-joining during immunoglobulin class switch recombination. *Proc. Natl. Acad. Sci. U. S. A.* 113, 1261–1266
69. Mohni, K.N. *et al.* (2019) HMCES maintains genome integrity by shielding abasic sites in single-strand DNA. *Cell* 176, 144–153.e13
70. Shukla, V. *et al.* (2020) HMCES functions in the alternative end-joining pathway of the DNA DSB repair during class switch recombination in B cells. *Mol. Cell* 77, 1154
71. Halabelian, L. *et al.* (2019) Structural basis of HMCES interactions with abasic DNA and multivalent substrate recognition. *Nat. Struct. Mol. Biol.* 26, 607–612
72. Casellas, R. *et al.* (2016) Mutations, kataegis and translocations in B cells: understanding AID promiscuous activity. *Nat. Rev. Immunol.* 16, 164–176
73. Zan, H. and Casali, P. (2013) Regulation of Aicda expression and AID activity. *Autoimmunity* 46, 83–101
74. Methot, S.P. *et al.* (2015) Consecutive interactions with HSP90 and eEF1A underlie a functional maturation and storage pathway of AID in the cytoplasm. *J. Exp. Med.* 212, 581–596
75. Le, Q. and Maizels, N. (2019) Activation-induced deaminase (AID) localizes to the nucleus in brief pulses. *PLoS Genet.* 15, e1007968
76. Patenaude, A.-M. *et al.* (2009) Active nuclear import and cytoplasmic retention of activation-induced deaminase. *Nat. Struct. Mol. Biol.* 16, 517–527
77. Hu, Y. *et al.* (2013) A combined nuclear and nucleolar localization motif in activation-induced cytidine deaminase (AID) controls immunoglobulin class switching. *J. Mol. Biol.* 425, 424–443
78. Ito, S. *et al.* (2004) Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1975–1980
79. Liu, M. *et al.* (2008) Two levels of protection for the B cell genome during somatic hypermutation. *Nature* 451, 841–845
80. Alvarez-Prado, A.F. *et al.* (2018) A broad atlas of somatic hypermutation allows prediction of activation-induced deaminase targets. *J. Exp. Med.* 215, 761–771
81. Qian, J. *et al.* (2014) B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell* 159, 1524–1537
82. Yamane, A. *et al.* (2011) Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat. Immunol.* 12, 62–69
83. McBride, K.M. *et al.* (2006) Regulation of hypermutation by activation-induced cytidine deaminase phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* 103, 8798–8803
84. Basu, U. *et al.* (2005) The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature* 438, 508–511
85. Chaudhuri, J. *et al.* (2004) Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature* 430, 992–998
86. Mu, Y. *et al.* (2017) Phosphorylation promotes activation-induced cytidine deaminase activity at the Myc oncogene. *J. Exp. Med.* 214, 3543–3552
87. Wang, Q. *et al.* (2014) Epigenetic targeting of activation-induced cytidine deaminase. *Proc. Natl. Acad. Sci. U. S. A.* 111, 18667–18672
88. Meng, F.-L. *et al.* (2014) Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* 159, 1538–1548
89. Pefanis, E. *et al.* (2014) Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature* 514, 389–393
90. Pham, P. *et al.* (2019) AID–RNA polymerase II transcription-dependent deamination of IgV DNA. *Nucleic Acids Res.* 1–15
91. Zhang, X. *et al.* (2019) Fundamental roles of chromatin loop extrusion in antibody class switching. *Nature* 575, 385–389
92. Wuerffel, R. *et al.* (2007) S-S synapsis during class switch recombination is promoted by distantly located transcriptional elements and activation-induced deaminase. *Immunity* 27, 711–722
93. Rothschild, G. *et al.* (2020) Noncoding RNA transcription alters chromosomal topology to promote isotype-specific class switch recombination. *Sci. Immunol.* Published online February 7, 2020. https://doi.org/10.1126/sciimmunol.aay5864
94. Betz, A.G. *et al.* (1994) Elements regulating somatic hypermutation of an immunoglobulin kappa gene: critical role for the intron enhancer/matrix attachment region. *Cell* 77, 239–248

95. Yeap, L.-S. *et al.* (2015) Sequence-intrinsic mechanisms that target AID mutational outcomes on antibody genes. *Cell* 163, 1124–1137
96. Buerstedde, J.-M. *et al.* (2014) Targeting of somatic hypermutation by immunoglobulin enhancer and enhancer-like sequences. *PLoS Biol.* 12, e1001831
97. Senigl, F. *et al.* (2019) Topologically associated domains delineate susceptibility to somatic hypermutation. *Cell Rep.* 29, 3902–3915.e8
98. Pavri, R. *et al.* (2010) Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* 143, 122–133
99. Chen, F.X. *et al.* (2018) Born to run: control of transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* 19, 464–478
100. Maul, R.W. *et al.* (2014) Spt5 accumulation at variable genes distinguishes somatic hypermutation in germinal center B cells from *ex vivo*-activated cells. *J. Exp. Med.* 211, 2297–2306
101. Wang, X. *et al.* (2014) A source of the single-stranded DNA substrate for activation-induced deaminase during somatic hypermutation. *Nat. Commun.* 5, 4137
102. Klein, I.A. *et al.* (2011) Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* 147, 95–106
103. Skourti-Stathaki, K. and Proudfoot, N.J. (2014) A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev.* 28, 1384–1396
104. Zheng, S. *et al.* (2015) Non-coding RNA generated following lariat debranching mediates targeting of AID to DNA. *Cell* 161, 762–773
105. Ribeiro de Almeida, C. *et al.* (2018) RNA helicase DDX1 converts RNA G-quadruplex structures into R-loops to promote IgH class switch recombination. *Mol. Cell* 70, 650–662.e8
106. Parsa, J.-Y. *et al.* (2012) Negative supercoiling creates single-stranded patches of DNA that are substrates for AID-mediated mutagenesis. *PLoS Genet.* 8, e1002518
107. Begum, N.A. *et al.* (2012) The histone chaperone Spt6 is required for activation-induced cytidine deaminase target determination through H3K4me3 regulation. *J. Biol. Chem.* 287, 32415–32429
108. Okazaki, I.-m. *et al.* (2011) Histone chaperone Spt6 is required for class switch recombination but not somatic hypermutation. *Proc. Natl. Acad. Sci. U. S. A.* 108, 7920–7925
109. Andrulis, E.D. *et al.* (2002) The RNA processing exosome is linked to elongating RNA polymerase II in *Drosophila*. *Nature* 420, 837–841
110. Lorenz, M. *et al.* (1995) Switch transcripts in immunoglobulin class switching. *Science (New York, N.Y.)* 267, 1825–1828
111. Hogenbirk, M.A. *et al.* (2012) Reassessing genomic targeting of AID. *Nat. Immunol.* 13, 797–798 author reply 798-800
112. Yamane, A. *et al.* (2012) Reply to "Reassessing genomic targeting of AID". *Nat. Immunol.* 13, 798–800
113. Chiarle, R. *et al.* (2011) Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* 147, 107–119
114. Pasqualucci, L. *et al.* (2004) Expression of the AID protein in normal and neoplastic B cells. *Blood* 104, 3318–3325
115. Staszewski, O. *et al.* (2011) Activation-induced cytidine deaminase induces reproducible DNA breaks at many non-Ig Loci in activated B cells. *Mol. Cell* 41, 232–242
116. Alexandrov, L.B. *et al.* (2013) Signatures of mutational processes in human cancer. *Nature* 500, 415–421
117. Kasar, S. *et al.* (2015) Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. *Nat. Commun.* 6, 8866