

Biological Function of Activation-induced Cytidine Deaminase (AID)

Ritu Kumar^{1*}, Lauren J. DiMenna^{2*}, Jayanta Chaudhuri², Todd Evans¹

Activation-induced Cytidine Deaminase (AID) is an essential regulator of B cell diversification, but its full range of action has until recently been an enigma. Based on homology, it was originally proposed to be an RNA-editing enzyme, but so far, no RNA substrates are known. Rather, it functions by deaminating cytidine, and in this manner, coupled with base-excision repair or mismatch repair machinery, it is a natural mutator. This allows it to play a central role in adaptive immunity, whereby it initiates the processes of class switch



Prof. Jayanta Chaudhuri



Prof. Todd R. Evans

recombination and somatic hypermutation to help generate a diverse and high-affinity repertoire of immunoglobulin isotypes. More recently, it has been appreciated that methylated cytidine, already known as a key epigenetic mark on DNA controlling gene expression, can also be a target for AID modification. Coupled with repair machinery, this can facilitate the active removal of methylated DNA. This activity can impact the process of cellular reprogramming, including transition of a somatic cell to pluripotency, which requires major reshuffling of epigenetic memory. Thus, seemingly disparate roles for AID in controlling immune diversity and epigenetic memory have a common mechanistic basis. However, the very activity that is so useful for B cell diversity and cellular reprogramming is dangerous for the integrity of the genome. Thus, AID expression and activity is tightly regulated, and deregulation is associated with diseases including cancer. Here, we review the range of AID functions with a focus on its mechanisms of action and regulation. Major questions remain to be answered concerning how and when AID is targeted to specific loci and how this impacts development and disease. (*Biomed J* 2014;37:269-283)

Key words: B cells, demethylation, epigenetics

Seemingly disparate functions for AID, controlling B cell diversification and DNA methylation patterns

Class switch recombination and somatic hypermutation

B cells are critical components of the mammalian immune response as they generate antibodies or immu-

noglobulins (Igs) against a seemingly infinite number of antigens. Antibodies comprise two identical heavy (IgH) and light (IgL) chains held together by disulfide bonds, with the amino-terminus “variable” region providing antigen-binding specificity and the less-divergent carboxy-terminal “constant” region of IgH required for effector function. The initial repertoire of B (and T) cells is generated by V(D) J recombination which assembles the variable (V) region

*These authors contributed equally.

From the ¹Department of Surgery, Weill Cornell Medical College, New York, NY, USA; ²Immunology Program, Sloan-Kettering Institute, Memorial Sloan-Kettering Cancer Center, New York, NY, USA

Received: Dec. 03, 2013; Accepted: Feb. 24, 2014

Correspondence to: Prof. Jayanta Chaudhuri, Immunology Program, Sloan-Kettering Institute, Memorial Sloan-Kettering Cancer Center, New York, NY, USA. 1275 York Ave, Box 477, New York, NY 10021, USA. Tel: 212-639-5562, Fax: 212-794-4019, E-mail: chaudhuj@mskcc.org
Correspondence to: Prof. Todd Evans, Department of Surgery, Weill Cornell Medical College, New York, NY, USA. 1300 York Ave., LC-708, New York, NY 10065, USA. Tel: 212-746-9485, Fax: 212-746-8753, E-mail: tre2003@med.cornell.edu

DOI: 10.4103/2319-4170.128734

exons from IgH and IgL, with the diversity (D, IgH only) and joining (J) segments. Following antigenic stimulation in secondary lymphoid organs such as the spleen and lymph nodes, mature B cells undergo two additional diversification reactions: Somatic hypermutation (SHM) and class switch recombination (CSR). Through SHM, point mutations, and less frequently deletions and insertions, are generated in the DNA encoding variable regions at a very high rate (from 10^{-2} to 10^{-3} per base pair per generation), which can increase affinity of the antibody for its cognate antigen.^[1-3] Mutations occur throughout the variable region exons, but with the highest frequency at the RGYW/WRCY (R = A/G, Y = C/T, W = A/T) “hot spot” motif.^[4] CSR, on the other hand, occurs exclusively at the IgH locus and exchanges the default C μ constant region exons for one of the downstream IgH C-region (C $_h$) exons (C γ , C ϵ , or C α). Thus, instead of expressing IgM, the B cell now expresses a secondary isotype (IgG, IgE, or IgA, respectively), providing a new set of effector functions without altering the affinity for antigen.^[5,6]

The mouse IgH locus includes eight constant C $_h$ genes, each (except C δ) preceded by G:C rich, repetitive DNA elements, termed switch (S) regions, that serve as recombination targets [Figure 1A]. CSR involves recombination between two S regions, thereby deleting the intervening sequence and juxtaposing a new C $_h$ gene downstream of the IgH variable region exon. Each C $_h$ gene comprises individual transcription units, with a cytokine-inducible promoter driving transcription through an intervening (I)-exon, the intronic S region, and the C $_h$ exons.^[7] The primary transcript is spliced and polyadenylated, but the mature transcript does not code for any proteins and is referred to as a sterile or germline transcript.^[8,9] However, transcription plays a major mechanistic role in CSR as different combinations of cytokines drive transcription through different C $_h$ genes and promote CSR to that isotype.^[10] In addition to the requirement of S regions and transcription, CSR is absolutely dependent on the activity of the DNA cytidine deaminase [activation-induced cytidine deaminase (AID)].^[11]

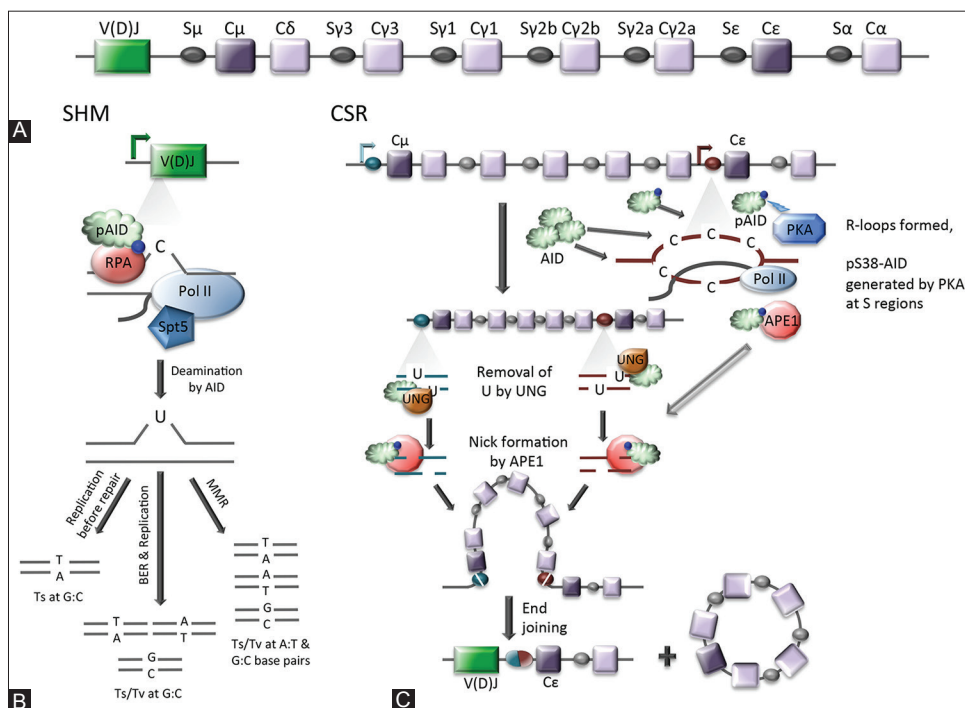


Figure 1: Overview of somatic hypermutation (SHM) and class switch recombination (CSR). (A) The mouse Ig heavy chain is composed of the variable region, which provides antigen specificity, and a series of eight constant region genes (C, purple squares), which determine the functional characteristics of the antibody molecule. Each of the C genes, except C δ , is preceded by a switch (S) region (black circles) which is vital to CSR. (B) SHM begins with transcription through the variable region (illustrated with a green arrow). A transcription bubble is created and stabilized through Spt5-induced Pol II stalling and the recruitment of RPA to AID phosphorylated at S38. AID deaminates the exposed cytosine residues (C) generating a uridine (U). If replication occurs before the U can be replaced, a transition mutation (Ts) will occur from the original G:C to T:A. If BER removes the U prior to replication, Ts or Tv (transversion) mutations will result from the G:C. Finally, if the U is removed by the MMR pathway, both Ts and Tv mutations will occur at G:C and A:T base pairs. (C) CSR is initiated by transcription through the S regions (shown in teal and red, respectively), which leads to the formation of stable R-loops. PKA and AID are recruited to the switch regions, where PKA will phosphorylate AID at S38 (represented by the lightning bolt and blue circle on the AID molecule). AID deaminates the C residues to U, which in turn is removed by UNG. Phosphorylated AID (pAID) recruits APE1, which generates a nick in the DNA. When this occurs on both strands, a DSB is generated. The broken switch regions are ligated together by one of the end-joining pathways and the intervening DNA is looped out.

Methylation provides a key epigenetic mark

DNA methylation in mammals appears restricted to the covalent addition of a methyl group to the 5-position of cytosine [5-methylcytosine (5mC)].^[12,13] The impact of 5mC in controlling mammalian gene regulation has led it to be considered the fifth base of DNA.^[14] While the existence of 5mC was identified even before DNA was recognized as the genetic material,^[15] the importance was underappreciated until a targeted mutation of the gene encoding an enzyme responsible for methylating cytosine was found to cause embryonic lethality.^[16] Evidence that DNA methylation regulates gene expression came from observations that silencing of the beta-globin locus is associated with methylation.^[17] The 5mC mark has since been shown to affect many cellular processes including genomic imprinting, X chromosome inactivation, preservation of chromosomal stability, genome defense, paramutation, tissue-specific gene regulation, cancer, and aging.^[18,19] The common mechanism is the capacity for methylation to modify gene function “epigenetically” without changing the coding sequence (in contrast to SHM or CSR). In mammals, DNA methylation occurs almost exclusively in the symmetric CpG context and is estimated to occur at ~70-80% of CpG dinucleotides throughout the genome.^[20] A small amount of non-CpG methylation is described in embryonic stem cells (ESCs).^[18,21] Promoter methylation is largely associated with transcriptional repression, whereas intragenic methylation correlates with transcriptional activity.^[22] DNA methylation patterns are dynamic, and variably methylated CpGs are found more often in gene bodies and intergenic regions, compared with promoters and upstream regulatory regions.^[23]

In vertebrates, DNA methylation marks are established and maintained by a family of DNA methyltransferases (DNMTs),^[24] which catalyze the transfer of a methyl group using S-adenosyl methionine (SAM) as the donor.^[24] In mouse embryos, the pattern of methylation inherited from the parental germline is rapidly lost after fertilization,^[25] and thus, the DNA of a pre-implantation blastocyst is relatively hypomethylated. After implantation, the embryo undergoes a wave of *de novo* methylation that establishes a new embryonic methylation pattern^[26-28] through the activity of *de novo* methyltransferases DNMT3a and DNMT3b.^[29] During post-implantation development, a second wave of DNA demethylation occurs in primordial germ cells (PGCs),^[30] to erase DNA methylation imprints established in the previous generation, followed by DNA methylation through DNMT3a and a non-catalytic paralog, DNMT3L.^[31] Once established, global DNA methylation must be stably maintained to preserve cellular identity. The maintenance methyltransferase DNMT1 restores hemimethylated DNA generated during DNA replication to the fully methylated

state.^[26] Until recently, a major gap in understanding control of methylation was the absence of known enzymes that actively remove methyl groups. This was bridged with the discovery that AID is one of several enzymes that can catalyze modification of 5mC, leading indirectly to removal of methylation.

Biological functions of AID

Discovery of AID

AID was discovered as a gene up-regulated in a cell line undergoing CSR.^[32] Mice with targeted deletion of AID have a complete block in CSR and SHM, while humans with hypomorphic AID mutations have severe defects in CSR and SHM, indicating that AID is an essential component of both secondary immune diversification reactions.^[11,33] AID is a single-strand (ss)-DNA-specific DNA cytidine deaminase converting deoxycytidines (dCs) into deoxyuridines (dUs).^[34-38] While it has been proposed based on its homology to the RNA-editing cytidine deaminase apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1 (APOBEC-1) that AID could deaminate RNA, no RNA substrate for AID has been identified.^[32]

AID and SHM

The DNA deamination activity of AID explains the observed mutation spectrum at variable region genes during SHM.^[39,40] In a simplified model, transcription through the variable region exons promotes the recruitment of AID to transcription bubbles, facilitated by the AID co-factor replication protein A (RPA). It has been proposed that RPA stabilizes ssDNA at transcription bubbles, thus allowing AID-mediated deamination at the variable region.^[41] If replication occurs prior to dU:dG repair, transition mutations are introduced at dC:dG base pairs. Alternatively, removal of dU by uracil DNA glycosylase (UNG) before replication creates an abasic site, which leads to the introduction of both transition and transversion mutations when followed by replication involving error-prone DNA polymerases. When the dU:dG mismatch is recognized by mismatch repair (MMR) proteins, coupled with gap-filling by error-prone DNA polymerases, the result can be point mutations at A:T residues and/or short insertions and deletions throughout the variable region. Indeed, consistent with this model, mutations in UNG lead to a substantial shift in dC:dG mutations toward transitions, while MMR mutations lead to a significant focusing of mutations in favor of dC:dG bases with a sharp decline in mutations at dA:dT bases [Figure 1B].

AID and CSR

The current model for CSR requires transcription through the G:C-rich S regions to promote formation of DNA: RNA hybrid structures (R-loops), thereby exposing

stretches of ssDNA on the non-template strand, providing a substrate for AID.^[42,43] The dUs introduced by AID are removed by the base-excision repair (BER) enzyme UNG to generate an abasic site, which is then converted into an ssDNA break by the apurinic/apyrimidinic endonuclease 1 (APE1).^[44] Note that both APE1 and APE2 may induce the ssDNA breaks and it is not yet clear whether APE2 or APE1 is more important.^[45,46] Creation of a similar nick nearby on the opposite strand yields a staggered double-strand break (DSB). In addition to the BER enzymes, components of the MMR pathway can also process a dU: dG mismatch, converting it into staggered DNA breaks.^[47] While the R-loop model satisfactorily explains AID activity on the non-template strand, the mechanism by which the template strand is deaminated is not fully elucidated. Activities of the RNA exosome complex combined with negative supercoiling upstream of elongating RNA polymerase II (Pol II) exposes ssDNA, which has been proposed to facilitate template strand deamination [Figure 1C].^[48,49] DSBs between two distinct S regions are synapsed and then ligated by end-joining through the activities of the classical non-homologous end-joining (C-NHEJ) and the poorly characterized alternative-NHEJ

pathways.^[6,50-53] Consistent with this model, mutations in UNG, APE1, MMR proteins Msh2, Msh6, and ExoI, and NHEJ components lead to impaired CSR.^[46,54,55]

Keeping AID in line

While the Ig genes are the primary targets of AID in an activated B cell, AID has the ability to target numerous non-Ig genes with a lower frequency.^[56-61] Because of the potential for inducing devastating genomic alterations, AID is tightly controlled at several levels, from transcriptional regulation and cellular localization to post-translational modifications and targeted recruitment to the Ig loci [Figure 2]. The *Aicda* locus (encoding AID) contains binding sites for multiple transcription factors (TFs) that regulate expression. For example, located within intron 1 is a region that binds E2F and c-MYB for repression in the resting state.^[62] Upon stimulation, several TFs including nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB), signal transducer and activator of transcription 6 (STAT6), and CCAAT-enhancer-binding protein (C/EBP) bind the enhancer region upstream of the transcriptional start site, while the activators PAX5 and

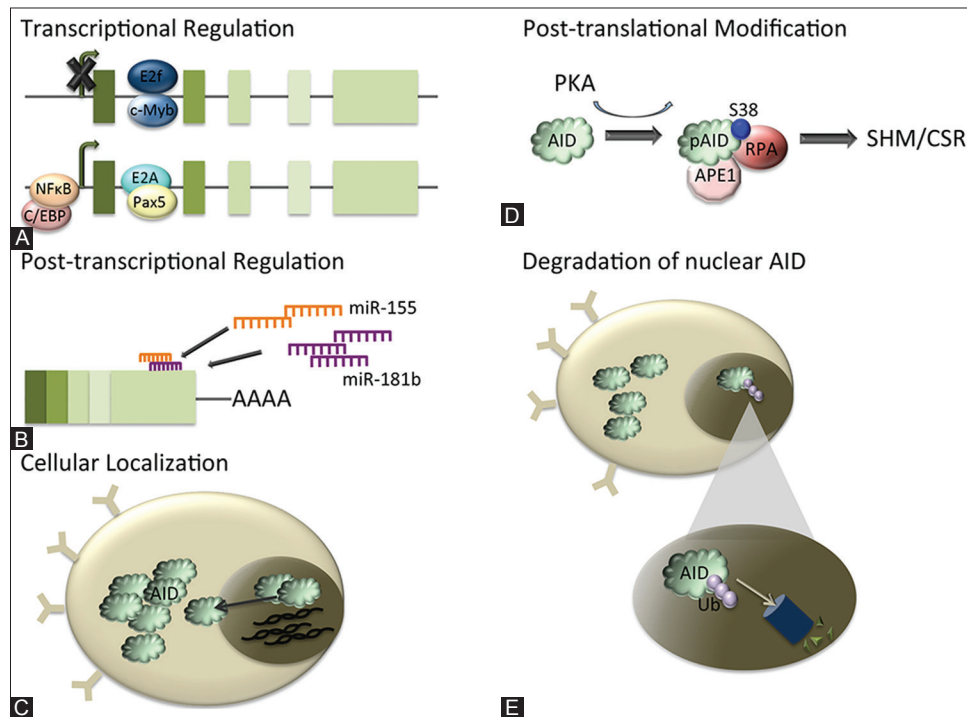


Figure 2: AID is regulated at multiple levels. (A) AID transcription is controlled through the binding of transcription factors (TFs) to enhancer or repressor regions. For example, when E2F and c-MYB are bound to the repressive region in intron 1, AID is not transcribed. Upon activation, other TFs bind to various regulatory regions to enhance transcription (green arrows). (B) miR-155 and -181b bind the 3'UTR of AID mRNA to regulate its expression. (C) AID is actively shuttled out of the nucleus and accumulates in the cytoplasm, thus preventing access to DNA. (D) PKA is recruited to the Ig locus upon antigenic stimulation, where it phosphorylates AID at S38, allowing AID to interact with RPA and APE1, which are vital for stabilization of the substrate and generating nicks in the DNA, respectively. Thus, this post-translational modification affects where AID activity can have the most impact. Note: Other phosphorylation sites have been identified, but S38 is currently the only one with a clearly defined mechanism of action. (E) Nuclear AID undergoes ubiquitination (Ub, violet spheres) and proteasomal degradation (blue cylinder), thus limiting the time it remains in contact with the DNA.

E2A bind the repressive region, effectively de-repressing AID expression.^[62-64]

The stability of AID mRNA is regulated by two micro-RNAs: miR-155 and miR-181b. While miR-181b is down-regulated upon B cell activation, miR-155 expression is increased, and they likely play complementary roles in restricting AID expression to a finite window of time.^[65,66] Mutation of the AID 3'UTR target sites for either miRNA leads to increased CSR and mutations in non-Ig genes. As an added layer of protection, AID protein is actively shuttled out of the nucleus via its C-terminal nuclear export sequence (NES). Loss of the NES region affects not only export but also protein stability and CSR.^[67] This hints at yet another regulatory mechanism: Preventing aberrant AID activity through proteosomal degradation of nuclear AID following poly-ubiquitination.^[68] Alternatively, nuclear AID may be degraded in a ubiquitination-independent manner through interaction with Regulatory (REG)- γ .^[69]

In B cells, AID is phosphorylated at multiple residues. Phosphorylation at serine 38 (S38) is carried out by the cAMP-dependent protein kinase A (PKA).^[70] The pS38-AID interacts with RPA, which stabilizes ssDNA substrates of AID during SHM.^[42,70] Additionally, RPA recruited to S regions by pS38-AID probably protects DNA ends from resection prior to end-joining during the completion phase of CSR.^[71] Recently, pS38-AID was also shown to interact with APE1, suggesting that AID phosphorylation promotes formation of DSBs at S regions during CSR.^[72] Additionally, ATM, a major component of the DNA damage response pathway, is required for AID phosphorylation at S38 (likely through activation of PKA) and interaction with APE1. Strikingly, AID phosphorylation at S38 is dependent on DSBs, demonstrating that DSB formation at S regions is both dependent on and required for AID phosphorylation.^[72] Taken together, a positive feedback loop amplifies DNA breaks at S regions through a phosphorylation- and ATM-dependent interaction of AID with APE1, which serves as a paradigm for concerted regulation of a generally toxic reaction where the same molecule, ATM, participates in both the generation and repair of DSBs.

While AID has the potential to bind to a large number of transcribed genes in B cells, S regions and Ig variable regions serve as major AID targets. Thus, following all the intricate regulation of its expression and localization, there still exist mechanisms to actively recruit AID to the Ig loci. Several AID-interacting proteins may mediate locus specificity. The germinal center associated nuclear protein (GANP) is induced in germinal center B cells, binds variable region transcripts, and has been reported to recruit AID to V genes during SHM.^[73] The 14-3-3 proteins, with ability to bind DNA with altered conformations (e.g. cruciform DNA), bind to AID and to AGCT (a subset of RGYW sequences), and are implicated as scaffold proteins that recruit AID to S regions.^[74] AID also interacts with SPT5, which binds to stalled Pol II.^[75] Depletion

of SPT5 markedly reduces recruitment of AID to Ig and non-Ig sequences. Chromatin immunoprecipitation (ChIP)-sequence analyses showed that co-occupancy of AID and SPT5 on stalled Pol II sites is predictive of AID-dependent mutations at the corresponding DNA sequence. Finally, AID can interact with the RNA-binding splicing regulator polypyrimidine tract binding protein 2 (PTBP2).^[76] Depletion of PTBP2 significantly impairs association of AID with S regions, with a concomitant defect in CSR. The mechanism by which PTBP2 targets AID to the Ig locus during CSR remains unknown.

Promiscuity leads to disease

The potential for AID to generate lesions outside the Ig loci has been implicated in the generation of Ig-partnered chromosomal translocations prevalent in many B cell lymphomas and leukemias.^[77,78] For example, c-MYC-IGH translocations are prominent in Burkitt's lymphomas and mouse plasmacytomas,^[79,80] whereas BCL2-IGH translocations are found in follicular and diffuse large cell lymphomas.^[81,82] The requirement of AID in generating DSBs at both the Ig and the non-Ig loci has been directly demonstrated.^[83] When mapped, AID-dependent DSBs were found to include non-transcribed regions, frequently within repeated sequence elements.^[84] However, how these translocations are related to lymphomagenesis is not entirely clear. While these translocations may be a major component of transformation,^[85] they can also be found in healthy individuals.^[81,82] In some cancers, including acute lymphoblastic leukemia and chronic myeloid leukemia, AID serves as a source of secondary mutations, enhancing tumor progression.^[86,87] Thus, in a cell where AID is actively expressed to enhance a protective immune response, failure to confine AID activity to the Ig locus can turn a protective response into a potentially fatal cancer.

Due to its role in targeted mutation of the Ig locus, the fact that aberrant AID activity is involved in B cell-based cancers is not particularly startling. Perhaps more surprising, AID activity was discovered in several types of solid tumors. A subset of lung tumors expresses AID, with the expression concordant with the number of mutations in p53.^[88] Likewise, some patients with early-stage oral squamous cell carcinoma have AID expressed specifically in the epithelial cells associated with the disease.^[89] Insight comes from transgenic mice that ubiquitously express AID (AID-Tg).^[59,90] While the majority of AID-Tg mice succumb to lymphomas, others develop hepatocellular carcinoma (HCC), lung cancer, or gastric tumors. Furthermore, each organ-specific tumor displays a different profile of mutations, e.g. c-Myc is highly mutated in lung tumors but not in tumors of stomach or liver, while k-Ras is most highly mutated in gastric tumors but not in tumors of lung or liver. The most highly mutated genes are relevant to the development of respective cancers in humans and are highly transcribed in that particular organ type.

This does not explain why AID would be expressed in these tissues in human cancers. In HCC and colorectal cancer, chronic inflammation leads to the secretion of cytokines capable of inducing AID expression.^[91,92] Eventually this results in p53 mutations, which can be found in cancerous and non-cancerous tissues from these patients. For the HCC patients, AID expression is significantly increased in individuals with chronic hepatitis (CH) or liver cirrhosis (LC), two patient groups at high risk for developing HCC. In this case, transforming growth factor beta (TGFβ), which is expressed in CH and LC, induces AID expression in primary human hepatocytes. Similarly, a subset of colorectal cancers has been linked to chronic inflammatory bowel diseases which are known for expression of tumor necrosis factor alpha (TNFα) and Th2 immune responses. In colon epithelial samples, TNFα, interleukin (IL)-4, and IL-13 influence AID expression leading to p53 mutations. Other solid tumors linked to inflammation-based AID activity include cholangiocarcinoma, Barrett’s esophageal adenocarcinoma, and gastric cancer [Figure 3].

Putting on another hat: AID can facilitate activate DNA demethylation

DNA demethylation is necessary to activate specific genes or for resetting the epigenetic state, and can occur pas-

sively by DNA replication in the absence of DNMT1.^[26,28,93] However, demethylation of DNA without replication is observed in the zygotic paternal genome^[94-96] and in PGCs,^[30,97] which suggests alternate mechanisms. Furthermore, demethylation of the promoter–enhancer region of the IL-2 gene is replication-independent in T lymphocytes.^[98] The demethylation of DNA injected into zebrafish embryos occurs in a replication-independent manner,^[99] and quantification of methylation levels in zebra-fish suggests active demethylation at fertilization.^[100] In plants, bifunctional DNA glycosylases excise the 5mC base and then cleave the DNA backbone at the abasic site. In vertebrates, DNA glycosylases may have a role, but only after 5mC is oxidized and/or deaminated. In both cases, the resultant single-nucleotide gap is subsequently filled with an unmodified cytidine through the BER pathway. It was suggested that vertebrate orthologs of *Escherichia coli* AlkB demethylases (dioxygenases) might operate through direct methyl group removal;^[101] however, none have direct 5mC demethylation activity. It was also considered that methyl-binding domain protein 2b (Mbd2b) might directly remove the methyl group, leaving cytosine and methanol as products.^[102] However, as Mbd2b lacks an enzymatic domain, this mechanism remains controversial. Another potential pathway in vertebrates could involve two demonstrated glycosylases: Thymine DNA glycosylase (TDG) or methyl-binding domain protein 4 (Mbd4). However, both TDG and Mbd4 have very weak 5mC base

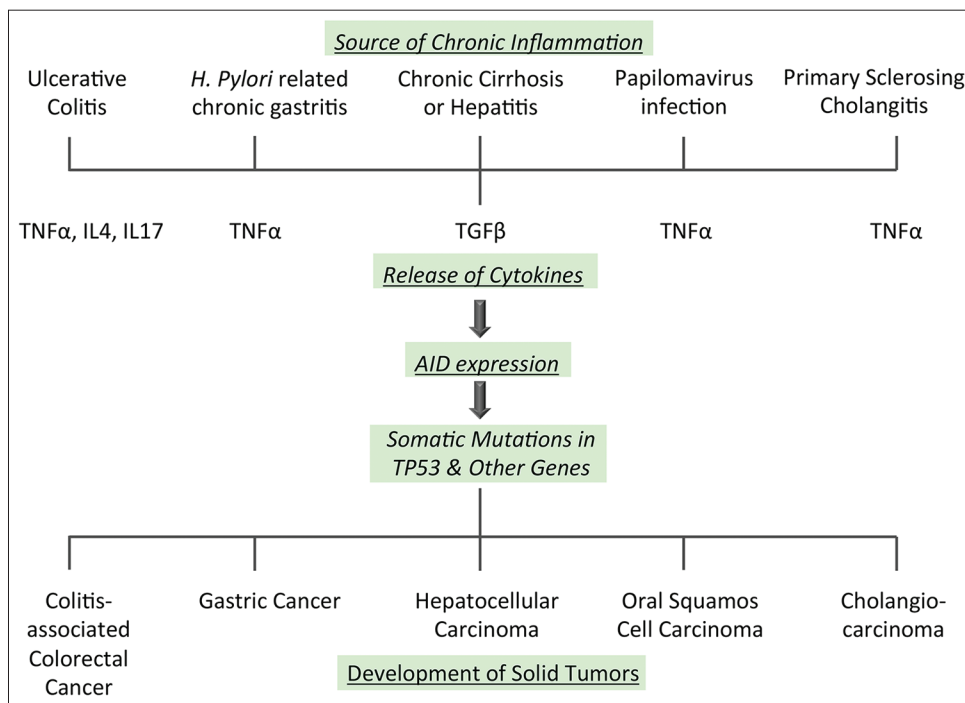


Figure 3: Chronic inflammation leads to aberrant AID expression and subsequent mutations that promote tumorigenesis in solid organ systems. Several chronic diseases have been implicated in cancer development. Only those cancers known to express AID in a subset of patients and which have had that expression linked to chronic inflammation are shown here. The common theme for all of these cancers is that chronic inflammation generates a setting of cytokine signaling, generally TNF , which in turn induces AID expression in the epithelial cells. AID then deaminates different genes, including the oncogene p53, leading to mutations that promote tumorigenesis.

excision activity, relative to their activity on thymine.^[103,104]

As described above, deamination of cytidine by AID leads to removal of the modified base, with subsequent restoration or mutation through BRE or MMR machinery. Acting upon 5mC, this could provide an indirect method to remove the methyl group. Progress in understanding active DNA demethylation comes from recent studies that focused on such mechanisms initiated by two unrelated protein families.^[105] One family comprises the ten eleven translocation (TET) proteins, which initiate DNA demethylation by oxidizing 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), or 5-carboxylcytosine (5caC) isoforms that are then replaced with cytidine by the BER machinery. The second family includes enzymes in the cytidine deaminase family, including AID and APOBEC-1.^[106] These two pathways (TET and AID) are thought to be the major initiators of DNA demethylation [Figure 4], although other

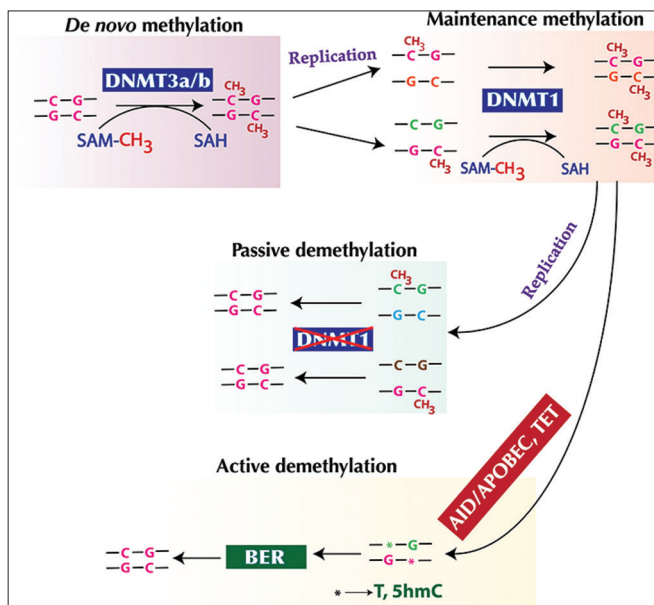


Figure 4: DNA methylation and demethylation. De novo DNA methylation involves addition of a methyl group to the cytosine of unmethylated DNA catalyzed by the enzymes DNA methyltransferase (DNMT) 3a and 3b. Pre-established DNA methylation is maintained by another family member, DNMT1, with a preference for hemimethylated CpG sites, such as those generated by DNA replication. DNMTs utilize S-adenosyl-L-methionine (SAM) as a methyl group donor. DNA demethylation, which is removal of the methyl group, can be achieved by either of two mechanisms – passive or active. Passive DNA demethylation involves DNA replication in the absence of DNMT1. Active demethylation (given the absence of demethylases) is a two-step process. The first step is the modification of 5mC by enzymes such as from the AID/APOBEC or ten eleven translocation (TET) families of proteins. AID/APOBEC enzymes deaminate 5-methylcytosine (5mC) to thymine (T), whereas TET enzymes hydroxylate 5mC to 5-hydroxymethylcytosine (5hmC). Intermediates generated after cytosine modification can then be replaced by cytosine by DNA base-excision repair (BER) machinery.

mechanisms are still plausible. For example, the DNMTs are capable of functioning as calcium and redox-state dependent DNA methylases, at least *in vitro*.^[107]

Active DNA methylation by AID

Evidence coupling DNA deamination with BER came from studies in zebrafish embryos.^[108] Demethylation of injected DNA fragments, as well as the whole genome, occurs in a time-dependent manner coinciding with the up-regulation of AID/APOBEC genes. Loss-of-function experiments revealed that AID/APOBEC also reduces steady-state methylation during normal development. Knockdown of AID results in loss of neurons in 1-day embryos, due at least in part to increased CpG methylation at the neurod2 promoter.^[108]

Lineage tracing during mouse embryogenesis using an AID: CRE knock-in allele clearly marked non-lymphoid cells during early development.^[109] High expression levels of AID are found in cells undergoing epigenetic reprogramming, including oocytes, with moderate levels in embryonic germ (EG) cells, ESCs, and E12.5 PGCs isolated from the genital ridge.^[106] Indeed, when DNA methylation is compared between wild-type and AID-deficient mice, AID-deficient PGCs are substantially more methylated, and this difference occurs throughout the genome, with introns, intergenic regions, and transposons being relatively more methylated than exons. Furthermore, AID deficiency interferes with genome-wide erasure of germline DNA methylation patterns. DNA from AID-null PGCs is hypermethylated compared to PGCs with wild-type AID, whether considering promoters, differentially methylated regions (DMRs), including imprinted genes that are demethylated during PGC development (*Dazl*, *H19*, *Lit1*), or retrotransposons.^[110] In addition to this loss-of-function evidence, ectopic expression of AID in human fibroblast cells (WI-38) results in expression of melanoma antigen-1 (MAGE-1) and cancer/testis antigen-1 (CAGE-1), associated with CpG demethylation in those promoters.^[111]

Mechanisms of active DNA demethylation by AID

Based on the zebrafish model, it was proposed that removal of 5mC by AID is a two-step process.^[108] Initial deamination of 5mC to thymine by AID results in a G:T mismatch, followed by the replacement of mutagenic T with C using the BER system [Figure 5A]. This two-step process is invoked because over-expression of AID/APOBEC enzymes alone does not alter methylation; coupling of AID deaminase activity with the thymine glycosylase MBD4 results in demethylation of the zebrafish genome. On the other hand, overexpression of MBD4 alone has little effect on DNA methylation. MBD4 contains both methylated-DNA binding and glycosylase domains and preferentially binds to 5mC:T mismatches.^[112] Moreover,

a DNA damage response gene, *gadd45a*, cooperates with AID and MBD4 to enhance demethylation, most likely by bridging the deaminase and glycosylase enzymes, as suggested by co-immunoprecipitation analysis.^[108] Consistent with this model, DNA hypermethylation occurs in mice lacking TDG. Knockdown of TDG with shRNA blocks the reactivation of an *in vitro* methylated Oct4 gene in EG cells, and co-immunoprecipitation (co-IP) experiments confirmed that TDG also forms a complex with AID and growth arrest and DNA-damage-inducible protein 45a (GADD45a).^[113]

Other studies suggest that AID might not directly deaminate 5mC during DNA demethylation.^[114,115] An alternate mechanism would have AID deaminate 5hmC to 5-hydroxymethyl uracil (5hmU), followed by base correction through BER^[105] [Figure 5B]. 5hmC can be generated by TET enzymes, a group of Fe (II)/2-oxoglutarate-dependent dioxygenases recently identified as 5mC hydroxylases.^[116,117] AID over-expression significantly decreases 5hmC levels induced by TET1 in HEK293 cells, while adeno-associated virus (AAV)-mediated over-expression of AID significantly decreases the endogenous levels of 5hmC in the adult mouse dentate gyrus.^[118] Mammalian TDG exhibits robust activity, comparable to single-strand selective monofunctional uracil DNA glycosylase (SMUG1,^[105] a TDG-related glycosylase), on a 5hmU:G mismatch in double-stranded DNA, the expected product of deamination following hydroxylation of 5mC.^[113] In contrast, an *in vitro* study demonstrated that AID enzymatic activity is inversely proportional to the electron cloud size of C5-cytosine, such that H > F > methyl >> hydroxymethyl, indicating AID is unlikely to participate in a 5hmC removal reaction.^[119] However, AID alters the subcellular localization of TET family members, such that

nuclear TET1 is translocated to the cytoplasm by AID and the translocated TET1 co-localizes with AID.^[116] TET1 is required to remove methylation on imprinted genes of PGCs, such that embryos derived from knockout null fathers (paternal knockout) display a “maternalized” pattern of gene expression from the paternal alleles.^[120] Clearly, further studies are required to understand the relationship of TET-AID and the impact of enzyme subcellular localization on DNA demethylation.

Reprogramming to pluripotency requires removal of epigenetic memory

Reprogramming refers to a change in cell identity, and involves not only changes in gene expression but also the erasure of epigenetic marks including DNA methylation. The role of reprogramming during normal development is controversial, but it can clearly be achieved by manipulation, including the reversal of a somatic cell back to a pluripotent state. This has been demonstrated by injection of a somatic cell nucleus into an enucleated egg [somatic cell nuclear transfer (SCNT)],^[121,122] by fusion of a pluripotent cell with a somatic cell,^[123] or most famously, by expression of four TFs.^[124] Reprogramming can also be achieved using small molecule compounds that target the “pluripotency network.”^[125] Generation of induced pluripotent stem cells (iPSCs) represents a major step forward for the field of regenerative medicine, since it allows the creation of a pluripotent cell, capable of forming any cell type, from any differentiated cell. Generation of iPSCs from a patient provides a pathway for disease modeling “in a dish” or even generating transplantable cells that would be recognized as “self,” largely avoiding the problem of immune rejection.

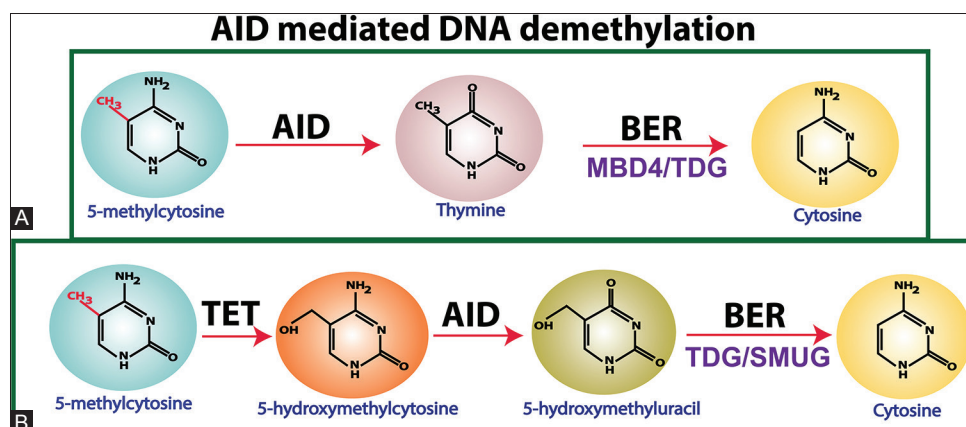


Figure 5: DNA demethylation induced by activation-induced cytidine deaminases. (A) AID can induce DNA demethylation by catalyzing deamination of 5-methylcytosine (5mC) to thymine, after which the thymine is replaced by cytosine by methyl-CpG binding domain protein 4 (MBD4) DNA glycosylase (shown in zebrafish) or thymine-DNA (TDG) glycosylase (shown in mammals). (B) An alternative mechanism has been proposed where ten eleven translocation (TET) family enzymes hydroxylate 5mC to 5-hydroxymethylcytosine (5hmC), and AID then catalyzes deamination of 5-hydroxymethylcytosine to 5-hydroxymethyluracil, which is replaced with cytosine by base-excision repair glycosylases such as TDG and single-strand selective monofunctional uracil DNA glycosylase 1 (SMUG1).

Reprogramming requires erasure of epigenetic memory of the starting somatic cell, and failure to do so results in incomplete transition to pluripotency. Most cloned embryos obtained through SCNT lack sufficient DNA demethylation resulting in aberrant epigenetic reprogramming,^[96] and in a high proportion of embryos, the distribution of methylated DNA is abnormal.^[126] Cloned embryos resemble donor cells in their overall genomic methylation status, which is very different from that of primary blastocysts produced *in vitro* or *in vivo*.^[127,128] Reprogramming, including using TFs, is considered an inefficient process,^[129-131] and one of the reasons that it fails in most somatic cells is due to retention of epigenetic memory from the cell of origin.^[132-134] Incomplete DNA demethylation is considered a key barrier toward achieving ground-state pluripotency, and treatment with DNMT inhibitors improves the overall efficiency of the reprogramming process.^[135] Demethylation of promoters for key pluripotency genes such as *Nanog* and *Oct4* may be a rate-limiting step.^[135,136] Thus, understanding the process of DNA demethylation may help eliminate the roadblocks of reprogramming.

AID-dependent demethylation in reprogramming induced by cell fusion

The first report implicating AID in reprogramming studied pluripotency induced in human fibroblasts by fusion with mouse ESCs^[105] [Figure 6A]. In that study, the onset of expression for two key human pluripotency genes, *OCT4* and *NANOG*, was shown as early as 1 day after fusion, correlating with the progressive replication-independent loss of methylation at the *OCT4* and *NANOG* promoters. Indeed, knockdown of mouse and human AID using small interfering RNAs (siRNAs) blocked induction of *OCT4* and *NANOG* in the heterokaryons, including inhibition of CpG demethylation in the human promoters. Furthermore, ChIP analysis revealed significant binding of AID to heavily methylated regions of human *OCT4* and *NANOG* promoters, but not at the promoter regions of *Oct4* and *Nanog* in mouse ESCs, despite higher levels of AID protein in ESCs. However, interpretation of these results was questioned by subsequent studies showing that AID is not expressed during fusion of rat fibroblasts with mouse ESCs, and that ectopic expression of AID does not enhance the reprogramming process or demethylation of *Oct4*.^[137]

AID-dependent demethylation during induction of pluripotency by TFs

Because its role in reprogramming was unclear, we devised a definitive test to see whether AID promotes pluri-

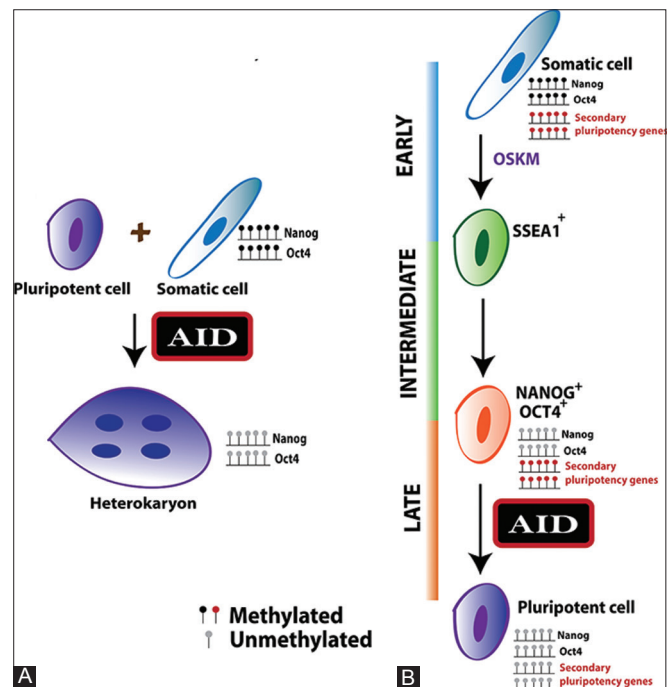


Figure 6: AID facilitates reprogramming by inducing the demethylation of pluripotency genes. (A) AID induces pluripotency in heterokaryons generated by fusion of a pluripotent stem cell with a somatic cell, facilitating DNA demethylation of *NANOG* and *OCT4* genes. (B) AID stabilizes pluripotency achieved by reprogramming through expression of four TFs (*OCT4*, *SOX2*, *KLF4*, and *cMYC*) by removing epigenetic memory of secondary pluripotency genes through its DNA demethylation activity at relatively late stages of reprogramming.

potency during reprogramming.^[138] For this purpose, we used AID-knockout fibroblasts, rather than relying on knockdown approaches that may be inefficient or suffer from off-target effects. Surprisingly, when these knockout cells were transduced to express reprogramming TFs (*OCT4*, *SOX2*, *KLF4*, and *cMYC*, or *OSKM*),^[124] they initiated reprogramming faster, compared to wild-type cells from the same strain. The AID-null cells generated more and larger initial iPSC-like colonies. However, these colonies were not stable and 3-4 weeks following initiation of reprogramming, they failed to maintain expression of pluripotency markers and they differentiated. A genome-wide molecular analysis of methylation revealed the cause. The genome of reprogramming cells that lack AID remains hypermethylated, and many of these hypermethylated genes fail to be properly up-regulated, including a set of known pluripotency-associated genes, such as *Apobec-1*, *Gdf3*, *Rex1*, *Zfp296*, *Cbx7*, *Dnmt1*, *Dnmt3L*, and *Tet3*. The expression profiles of the differentiating AID-null “iPSC-like” cells cluster most closely to fibroblasts, the source somatic cell type. When individual iPSC-like clones were picked and passaged, again most of the AID-null clones differentiated. However, some of the clones stabilized for pluripotency. Those that did expressed normal levels of pluripotency genes, were

normalized for methylation patterns, and could differentiate to form cells representing all three germ layers. Therefore, AID is not essential for reprogramming, but it is a key factor that facilitates the process [Figure 6B]. One potential issue for interpreting any studies using the original AID knockout^[11] is that the *Nanog* gene resides next to AID, and the knockout mice are not congenic with controls, since they retain this chromosome from the original ESC line used for derivation (also encoding the glycosylase MBD4).^[139] Thus, while it is possible that pluripotency-related genes located on chromosome 6 could be differentially expressed based on the origin of the chromosome (CBA vs. C57BL/6), inefficient reprogramming was recapitulated in a pure C57BL/6 AID^{-/-} mouse, suggesting that AID deficiency is the primary cause underlying this phenotype.^[138] In the absence of AID, reprogramming initiates well, but must eventually be compensated by another mechanism, perhaps involving passive removal through passaging. It will be interesting to see if iPSCs generated in the absence of AID have a lighter mutation load, given the capacity of AID to function as a natural mutator. If so, iPSCs generated (albeit inefficiently) without AID might be safer to use in future regenerative strategies. Furthermore, the early iPSCs that retain epigenetic memory in the absence of AID might provide a more efficient source of progenitors for directed differentiation back to a mature differentiated cell of the same lineage.

Conclusion: AID as both a friend and foe in B cells and reprogramming

AID is a critical regulator of our immune system, but it is a dangerous enzyme. Abnormal AID expression has been implicated in the development of autoimmunity, the hallmark of which is the production of autoantibodies of high affinity for host tissues. The best-studied case is systemic lupus erythematosus (SLE), in which the pathogenic antibodies are hypermutated and generally switched to the IgG isotype.^[140-142] In mouse models of SLE, AID expression is elevated and directly linked to the production of the multireactive pathogenic autoantibodies that characterize the human disease.^[143] Increased AID expression has also been found in patients with Sjögren's syndrome and rheumatoid arthritis, and is linked to autoantibody production.^[144,145] Further evidence comes from a study of experimental autoimmune encephalomyelitis, in which AID^{-/-} mice are resistant to disease development and only produce low-affinity IgM antibodies.^[146] AID-induced mutations have also been shown to play an important role in activating self-reactive T cells, at least in the MRL/lpr model of SLE, but a similar role in other autoimmune diseases has not been discounted.^[147]

Such studies leave the impression that for AID, less is better, in terms of preventing disease. However, lack of AID activity may also cause autoimmunity. Patients with

mutations in AID have a disease known as hyper-IgM syndrome (HIGM), characterized by increased serum IgM and low-affinity antibodies, as would be expected, but a subset of these patients also develops autoimmunity.^[33,148,149] Furthermore, HIGM patients who do not exhibit an autoimmune disease have autoreactive IgM antibodies in their sera. Corroborating data from AID^{-/-} mice shows antibodies directed at the gastric tissues in older mice correlating with severe gastritis.^[150] It has also been shown that mouse AID^{-/-} B cells have a more diverse repertoire of germline-encoded antigen receptors than their wild-type counterparts, and the cells are more activated even in the absence of a directed immune response.^[151] It has been proposed that AID may play an important role in the development of B cell tolerance to self, although more work is needed to rule out the immunodeficiency resulting from lack of AID as the true cause.^[152]

The capacity for AID to remove epigenetic memory may also have relevance for disease beyond the immune system. The AID^{-/-} mouse strain has small litters and we showed that generation of ESCs from knockout blastomeres is relatively inefficient, suggesting functions during early embryonic development.^[138] In a breast cancer cell model, AID expression is enhanced by inflammatory cytokines, leading to AID-dependent changes in transcriptional regulators of the epithelial-mesenchymal transition (EMT) and increased invasive behavior.^[153] Activation of EMT is an important step contributing to tumor metastasis and this may add to the potential for AID to be oncogenic. Clearly, AID is a fascinating enzyme with broad capacity to regulate development, the immune system, and human health. Major questions remain for how this essential but dangerous protein is regulated and targeted to specific loci, and how deregulation of its function contributes to disease mechanisms.

REFERENCES

1. McKean D, Huppi K, Bell M, Staudt L, Gerhard W, Weigert M. Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. *Proc Natl Acad Sci USA* 1984;81:3180-4.
2. Rudikoff S, Pawlita M, Pumphrey J, Heller M. Somatic diversification of immunoglobulins. *Proc Natl Acad Sci USA* 1984;81:2162-6.
3. Sablitzky F, Wildner G, Rajewsky K. Somatic mutation and clonal expansion of B cells in an antigen-driven immune response. *EMBO J* 1985;4:345-50.
4. Rogozin IB, Diaz M. 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* 2004;172:3382-4.
5. Honjo T, Kinoshita K, Muramatsu M. Molecular mechanism of class switch recombination: Linkage with somatic hypermutation. *Annu Rev Immunol* 2002;20:165-96.

6. Stavnezer J, Guikema JE, Schrader CE. Mechanism and regulation of class switch recombination. *Annu Rev Immunol* 2008;26:261-92.
7. Lennon GG, Perry RP. C mu-containing transcripts initiate heterogeneously within the IgH enhancer region and contain a novel 5'-nontranslatable exon. *Nature* 1985;318:475-8.
8. Alt FW, Rosenberg N, Enea V, Siden E, Baltimore D. Multiple immunoglobulin heavy-chain gene transcripts in Abelson murine leukemia virus-transformed lymphoid cell lines. *Mol Cell Biol* 1982;2:386-400.
9. Stavnezer-Nordgren J, Sirlin S. Specificity of immunoglobulin heavy chain switch correlates with activity of germline heavy chain genes prior to switching. *EMBO J* 1986;5:95-102.
10. Snapper CM, Finkelman FD, Paul WE. Regulation of IgG1 and IgE production by interleukin 4. *Immunol Rev* 1988;102:51-75.
11. 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;102:553-63.
12. Ratel D, Ravanat JL, Berger F, Wion D. N6-methyladenine: The other methylated base of DNA. *Bioessays* 2006;28:309-15.
13. Hattman S. DNA-[adenine] methylation in lower eukaryotes. *Biochemistry (Mosc)* 2005;70:550-8.
14. Ambinder RF, Robertson KD, Tao Q. DNA methylation and the Epstein-Barr virus. *Semin Cancer Biol* 1999;9:369-75.
15. Johnson TB, Cowgill RD. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus. *J Am Chem Soc* 1925;47:2838-44.
16. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 1992;69:915-26.
17. McGhee JD, Ginder GD. Specific DNA methylation sites in the vicinity of the chicken beta-globin genes. *Nature* 1979;280:419-20.
18. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev* 2002;16:6-21.
19. Bender J. DNA methylation and epigenetics. *Annu Rev Plant Biol* 2004;55:41-68.
20. Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, *et al.* Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res* 1982;10:2709-21.
21. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 2009;462:315-22.
22. Ball MP, Li JB, Gao Y, Lee JH, LeProust EM, Park IH, *et al.* Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* 2009;27:361-8.
23. Consortium EP, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57-74.
24. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 2010;11:204-20.
25. Kafri T, Ariel M, Brandeis M, Shemer R, Urven L, McCarrey J, *et al.* Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes Dev* 1992;6:705-14.
26. Reik W, Dean W, Walter J. Epigenetic reprogramming in mammalian development. *Science* 2001;293:1089-93.
27. Li E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet* 2002;3:662-73.
28. Jaenisch R, Bird A. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nat Genet* 2003;33 Suppl: 245-54.
29. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 1999;99:247-57.
30. Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W, *et al.* Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* 2002;117:15-23.
31. Chen ZX, Riggs AD. DNA methylation and demethylation in mammals. *The J Biol Chem* 2011;286:18347-53.
32. Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO, *et al.* 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;274:18470-6.
33. Revy P, Muto T, Levy Y, Geissmann F, Plebani A, Sanal O, *et al.* Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* 2000;102:565-75.
34. Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E, Alt FW. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* 2003;422:726-30.
35. Dickerson SK, Market E, Besmer E, Papavasiliou FN. AID mediates hypermutation by deaminating single stranded DNA. *J Exp Med* 2003;197:1291-6.
36. Nabel CS, Lee JW, Wang LC, Kohli RM. Nucleic acid determinants for selective deamination of DNA over RNA by activation-induced deaminase. *Proc Natl Acad Sci USA* 2013;110:14225-30.
37. Pham P, Bransteitter R, Petruska J, Goodman MF. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature* 2003;424:103-7.
38. 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;4:452-6.
39. Gearhart PJ, Bogenhagen DF. Clusters of point mutations are found exclusively around rearranged antibody variable genes. *Proc Natl Acad Sci USA* 1983;80:3439-43.
40. Lebecque SG, Gearhart PJ. Boundaries of somatic mutation in rearranged immunoglobulin genes: 5' boundary is near the promoter, and 3' boundary is approximately 1 kb from V (D) J gene. *J Exp Med* 1990;172:1717-27.
41. Chaudhuri J, Alt FW. Class-switch recombination: Interplay of transcription, DNA deamination and DNA repair. *Nat Rev Immunol* 2004;4:541-52.
42. Chaudhuri J, Khuong C, Alt FW. Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature* 2004;430:992-8.
43. Yu K, Lieber MR. Nucleic acid structures and enzymes in the

- immunoglobulin class switch recombination mechanism. *DNA Repair (Amst)* 2003;2:1163-74.
44. Petersen-Mahrt SK, Harris RS, Neuberger MS. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature* 2002;418:99-103.
 45. Guikema JE, Linehan EK, Tsuchimoto D, Nakabeppu Y, Strauss PR, Stavnezer J, *et al.* APE1- and APE2-dependent DNA breaks in immunoglobulin class switch recombination. *J Exp Med* 2007;204:3017-26.
 46. Schrader CE, Guikema JE, Wu X, Stavnezer J. The roles of APE1, APE2, DNA polymerase beta and mismatch repair in creating S region DNA breaks during antibody class switch. *Philos Trans R Soc Lond B Biol Sci* 2009;364:645-52.
 47. 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;16:163-71.
 48. Basu U, Meng FL, Keim C, Grinstein V, Pefanis E, Eccleston J, *et al.* The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell* 2011;144:353-63.
 49. Shen HM, Storb U. Activation-induced cytidine deaminase (AID) can target both DNA strands when the DNA is supercoiled. *Proc Natl Acad Sci USA* 2004;101:12997-3002.
 50. Dinkelmann M, Spehalski E, Stoneham T, Buis J, Wu Y, Sekiguchi JM, *et al.* Multiple functions of MRN in end-joining pathways during isotype class switching. *Nat Struct Mol Biol* 2009;16:808-13.
 51. Lee-Theilen M, Matthews AJ, Kelly D, Zheng S, Chaudhuri J. CtIP promotes microhomology-mediated alternative end joining during class-switch recombination. *Nat Struct Mol Biol* 2011;18:75-9.
 52. Robert I, Dantzer F, Reina-San-Martin B. Parp1 facilitates alternative NHEJ, whereas Parp2 suppresses IgH/c-myc translocations during immunoglobulin class switch recombination. *J Exp Med* 2009;206:1047-56.
 53. Xie A, Kwok A, Scully R. Role of mammalian Mre11 in classical and alternative nonhomologous end joining. *Nat Struct Mol Biol* 2009;16:814-8.
 54. Di Noia J, Neuberger MS. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* 2002;419:43-8.
 55. Imai K, Slupphaug G, Lee WI, Revy P, Nonoyama S, Catalan N, *et al.* Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nat Immunol* 2003;4:1023-8.
 56. Dorsett Y, Robbiani DF, Jankovic M, Reina-San-Martin B, Eisenreich TR, Nussenzweig MC. A role for AID in chromosome translocations between c-myc and the IgH variable region. *J Exp Med* 2007;204:2225-32.
 57. Liu M, Duke JL, Richter DJ, Vinuesa CG, Goodnow CC, Kleinstein SH, *et al.* Two levels of protection for the B cell genome during somatic hypermutation. *Nature* 2008;451:841-5.
 58. Pasqualucci L, Migliazza A, Fracchiolla N, William C, Neri A, Baldini L, *et al.* BCL-6 mutations in normal germinal center B cells: Evidence of somatic hypermutation acting outside Ig loci. *Proc Natl Acad Sci USA* 1998;95:11816-21.
 59. Robbiani DF, Bunting S, Feldhahn N, Bothmer A, Camps J, Deroubaix S, *et al.* (2009) AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. *Mol Cell* 1998;36:631-41.
 60. 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;280:1750-2.
 61. Yamane A, Resch W, Kuo N, Kuchen S, Li Z, Sun HW, *et al.* Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat Immunol* 2011;12:62-9.
 62. Tran TH, Nakata M, Suzuki K, Begum NA, Shinkura R, Fagarasan S, *et al.* B cell-specific and stimulation-responsive enhancers derepress Aicda by overcoming the effects of silencers. *Nat Immunol* 2010;11:148-54.
 63. Maul RW, Gearhart PJ. AID and somatic hypermutation. *Adv Immunol* 2010;105:159-91.
 64. Lee-Theilen M, Chaudhuri J. Walking the AID tightrope. *Nat Immunol* 2010;11:107-9.
 65. de Yebenes VG, Belver L, Pisano DG, Gonzalez S, Villasante A, Croce C, *et al.* miR-181b negatively regulates activation-induced cytidine deaminase in B cells. *J Exp Med* 2008;205:2199-206.
 66. Teng G, Hakimpour P, Landgraf P, Rice A, Tuschl T, Casellas R, *et al.* MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. *Immunity* 2008;28:621-9.
 67. 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;106:6736-41.
 68. Aoufouchi S, Faili A, Zober C, D'Orlando O, Weller S, Weill JC, *et al.* Proteasomal degradation restricts the nuclear lifespan of AID. *J Exp Med* 2008;205:1357-68.
 69. Uchimura Y, Barton LF, Rada C, Neuberger MS. REG-gamma associates with and modulates the abundance of nuclear activation-induced deaminase. *J Exp Med* 2011;208:2385-91.
 70. Cheng HL, Vuong BQ, Basu U, Franklin A, Schwer B, Astarita J, *et al.* Integrity of the AID serine-38 phosphorylation site is critical for class switch recombination and somatic hypermutation in mice. *Proc Natl Acad Sci USA* 2009;106:2717-22.
 71. Vuong BQ, Lee M, Kabir S, Irimia C, Macchiarulo S, McKnight GS, *et al.* Specific recruitment of protein kinase A to the immunoglobulin locus regulates class-switch recombination. *Nat Immunol* 2009;10:420-6.
 72. Vuong BQ, Herrick-Reynolds K, Vaidyanathan B, Pucella JN, Ucher AJ, Donghia NM, *et al.* A DNA break- and phosphorylation-dependent positive feedback loop promotes immunoglobulin class-switch recombination. *Nat Immunol* 2013;14:1183-9.
 73. Maeda K, Singh SK, Eda K, Kitabatake M, Pham P, Goodman MF, *et al.* GANP-mediated recruitment of activation-induced cytidine deaminase to cell nuclei and to immunoglobulin variable region DNA. *J Biol Chem* 2010;285:23945-53.
 74. Xu Z, Fulop Z, Wu G, Pone EJ, Zhang J, Mai T, *et al.* 14-3-3 adaptor proteins recruit AID to 5'-AGCT-3'-rich switch regions for class switch recombination. *Nat Struct Mol Biol* 2010;17:1124-35.
 75. Pavri R, Gazumyan A, Jankovic M, Di Virgilio M, Klein I, Ansarah-Sobrinho C, *et al.* Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* 2010;143:122-33.
 76. Nowak U, Matthews AJ, Zheng S, Chaudhuri J. The splicing regulator PTBP2 interacts with the cytidine deaminase AID and promotes binding of AID to switch-region DNA. *Nat Immunol* 2011;12:160-6.

77. Nishida K, Tamura A, Nakazawa N, Ueda Y, Abe T, Matsuda F, *et al.* The Ig heavy chain gene is frequently involved in chromosomal translocations in multiple myeloma and plasma cell leukemia as detected by *in situ* hybridization. *Blood* 1997;90:526-34.
78. Robbiani DF, Nussenzweig MC. Chromosome translocation, B cell lymphoma, and activation-induced cytidine deaminase. *Annu Rev Pathol* 2013;8:79-103.
79. Dalla-Favera R, Bregni M, Erikson J, Patterson D, Gallo RC, Croce CM. Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. *Proc Natl Acad Sci U S A* 1982;79:7824-7.
80. Taub R, Kirsch I, Morton C, Lenoir G, Swan D, Tronick S, *et al.* Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. *Proc Natl Acad Sci USA* 1982;79:7837-41.
81. Janz S, Potter M, Rabkin CS. Lymphoma- and leukemia-associated chromosomal translocations in healthy individuals. *Genes Chrom Cancer* 2003;36:211-23.
82. Limpens J, de Jong D, van Krieken JH, Price CG, Young BD, van Ommen GJ, *et al.* Bcl-2/JH rearrangements in benign lymphoid tissues with follicular hyperplasia. *Oncogene* 1991;6:2271-6.
83. Robbiani DF, Bothmer A, Callen E, Reina-San-Martin B, Dorsett Y, Difilippantonio S, *et al.* AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. *Cell* 2008;135:1028-38.
84. Staszewski O, Baker RE, Ucher AJ, Martier R, Stavnezer J, Guikema JE. Activation-induced cytidine deaminase induces reproducible DNA breaks at many non-Ig Loci in activated B cells. *Mol Cell* 2011;41:232-42.
85. Adams JM, Harris AW, Pinkert CA, Corcoran LM, Alexander WS, Cory S, *et al.* The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. *Nature* 1985;318:533-8.
86. Feldhahn N, Henke N, Melchior K, Duy C, Soh BN, Klein F, *et al.* Activation-induced cytidine deaminase acts as a mutator in BCR-ABL1-transformed acute lymphoblastic leukemia cells. *J Exp Med* 2007;204:1157-66.
87. Klemm L, Duy C, Iacobucci I, Kuchen S, von Levetzow G, Feldhahn N, *et al.* The B cell mutator AID promotes B lymphoid blast crisis and drug resistance in chronic myeloid leukemia. *Cancer Cell* 2009;16:232-45.
88. Shinmura K, Igarashi H, Goto M, Tao H, Yamada H, Matsuura S, *et al.* Aberrant expression and mutation-inducing activity of AID in human lung cancer. *Ann Surg Oncol* 2011;18:2084-92.
89. Nakanishi Y, Kondo S, Wakisaka N, Tsuji A, Endo K, Muroso S, *et al.* Role of activation-induced cytidine deaminase in the development of oral squamous cell carcinoma. *PloS One* 2013;8:e62066.
90. Okazaki IM, Hiari H, Kakazu N, Yamada S, Muramatsu M, Kinoshita K, *et al.* Constitutive expression of AID leads to tumorigenesis. *J Exp Med* 2003;197:1173-81.
91. Endo Y, Marusawa H, Kou T, Nakase H, Fujii S, Fujimori T, *et al.* Activation-induced cytidine deaminase links between inflammation and the development of colitis-associated colorectal cancers. *Gastroenterology* 2008;135:889-98, 898.e1-3.
92. Kou T, Marusawa H, Kinoshita K, Endo Y, Okazaki IM, Ueda Y, *et al.* Expression of activation-induced cytidine deaminase in human hepatocytes during hepatocarcinogenesis. *Int J Cancer* 2007;120:469-76.
93. Surani MA. Reprogramming of genome function through epigenetic inheritance. *Nature* 2001;414:122-8.
94. Oswald J, Engemann S, Lane N, Mayer W, Olek A, Fundele R, *et al.* Active demethylation of the paternal genome in the mouse zygote. *Curr Biol* 2000;10:475-8.
95. Mayer W, Niveleau A, Walter J, Fundele R, Haaf T. Demethylation of the zygotic paternal genome. *Nature* 2000;403:501-2.
96. Dean W, Santos F, Stojkovic M, Zakhartchenko V, Walter J, Wolf E, *et al.* Conservation of methylation reprogramming in mammalian development: Aberrant reprogramming in cloned embryos. *Proc Natl Acad Med USA* 2001;98:13734-8.
97. Lee J, Inoue K, Ono R, Ogonuki N, Kohda T, Kaneko-Ishino T, *et al.* Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. *Development* 2002;129:1807-17.
98. Bruniquel D, Schwartz RH. Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. *Nature Immun* 2003;4:235-40.
99. Collas P. Modulation of plasmid DNA methylation and expression in zebrafish embryos. *Nucl Acids Res* 1998;26:4454-61.
100. Mhanni AA, McGowan RA. Global changes in genomic methylation levels during early development of the zebrafish embryo. *Dev Genes Evol* 2004;214:412-7.
101. Sedgwick B. Repairing DNA-methylation damage. *Nat Rev Mol Cell Biol* 2004;5:148-57.
102. Bhattacharya SK, Ramchandani S, Cervoni N, Szyf M. A mammalian protein with specific demethylase activity for mCpG DNA. *Nature* 1999;397:579-83.
103. Cortazar D, Kunz C, Saito Y, Steinacher R, Schar P. The enigmatic thymine DNA glycosylase. *DNA Repair (Amst)* 2007;6:489-504.
104. Zhu B, Zheng Y, Angliker H, Schwarz S, Thiry S, Siegmann M, *et al.* 5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence. *Nucl Acids Res* 2000;28:4157-65.
105. Bhutani N, Burns DM, Blau HM. DNA demethylation dynamics. *Cell* 2011;146:866-72.
106. Morgan HD, Dean W, Coker HA, Reik W, Petersen-Mahrt SK. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: Implications for epigenetic reprogramming. *J Biol Chem* 2004;279:52353-60.
107. Chen CC, Wang KY, Shen CK. DNA 5-methylcytosine demethylation activities of the mammalian DNA methyltransferases. *J Biol Chem* 2013;288:9084-91.
108. Rai K, Huggins IJ, James SR, Karpf AR, Jones DA, Cairns BR. DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* 2008;135:1201-12.
109. Rommel PC, Bosque D, Gitlin AD, Croft GF, Heintz N, Casellas R, *et al.* Fate mapping for activation-induced cytidine deaminase (AID) marks non-lymphoid cells during mouse development. *PloS One* 2013;8:e69208.
110. Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, *et al.* Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* 2010;463:1101-5.
111. Isobe T, Song SN, Tiwari P, Ito H, Yamaguchi Y, Yoshizaki K. Activation-induced cytidine deaminase auto-activates and triggers aberrant gene expression. *FEBS Lett* 2013;587:2487-92.

112. Hendrich B, Hardeland U, Ng HH, Jiricny J, Bird A. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* 1999;401:301-4.
113. Cortellino S, Xu J, Sannai M, Moore R, Caretti E, Cigliano A, *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* 2011;146:67-79.
114. Larijani M, Frieder D, Sonbuchner TM, Bransteitter R, Goodman MF, Bouhassira EE, *et al.* Methylation protects cytidines from AID-mediated deamination. *Mol Immunol* 2005;42:599-604.
115. 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: 4102-7.
116. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 2010;466:1129-33.
117. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 2009;324:930-5.
118. Guo JU, Su Y, Zhong C, Ming GL, Song H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* 2011;145:423-34.
119. Rangam G, Schmitz KM, Cobb AJ, Petersen-Mahrt SK. AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud. *PLoS One* 2012;7:e43279.
120. Yamaguchi S, Shen L, Liu Y, Sandler D, Zhang Y. Role of Tet1 in erasure of genomic imprinting. *Nature* 2013;504:460-4.
121. Briggs R, King TJ. Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs. *Proc Natl Acad Sci USA* 1952;38:455-63.
122. Gurdon JB, Elsdale TR, Fischberg M. Sexually mature individuals of *Xenopus laevis* from the transplantation of single somatic nuclei. *Nature* 1958;182:64-5.
123. Cowan CA, Atienza J, Melton DA, Eggan K. Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. *Science* 2005;309:1369-73.
124. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006;126:663-76.
125. Hou P, Li Y, Zhang X, Liu C, Guan J, Li H, *et al.* Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. *Science* 2013;341:651-4.
126. Beaujean N, Taylor J, Gardner J, Wilmut I, Meehan R, Young L. Effect of limited DNA methylation reprogramming in the normal sheep embryo on somatic cell nuclear transfer. *Biol Reprod* 2004;71:185-93.
127. Bourc'his D, Le Bourhis D, Patin D, Niveleau A, Comizzoli P, Renard JP, *et al.* Delayed and incomplete reprogramming of chromosome methylation patterns in bovine cloned embryos. *Curr Biol* 2001;11:1542-6.
128. Kang YK, Koo DB, Park JS, Choi YH, Chung AS, Lee KK, *et al.* Aberrant methylation of donor genome in cloned bovine embryos. *Nat Genet* 2001;28:173-7.
129. Marion RM, Strati K, Li H, Murga M, Blanco R, Ortega S, *et al.* A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature* 2009;460:1149-53.
130. Li H, Collado M, Villasante A, Strati K, Ortega S, Canamero M, *et al.* The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature* 2009;460:1136-9.
131. Eminli S, Foudi A, Stadtfeld M, Maherali N, Ahfeldt T, Mostoslavsky G, *et al.* Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. *Nat Genet* 2009;41:968-76.
132. Ghosh Z, Wilson KD, Wu Y, Hu S, Quertermous T, Wu JC. Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells. *PLoS One* 2010;5:e8975.
133. Marchetto MC, Yeo GW, Kainohana O, Marsala M, Gage FH, Muotri AR. Transcriptional signature and memory retention of human-induced pluripotent stem cells. *PLoS One* 2009;4:e7076.
134. Bar-Nur O, Russ HA, Efrat S, Benvenisty N. Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells. *Cell Stem Cell* 2011;9:17-23.
135. Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* 2008;454:49-55.
136. Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, *et al.* A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* 2012;151:1617-32.
137. Foshay KM, Looney TJ, Chari S, Mao FF, Lee JH, Zhang L, *et al.* Embryonic stem cells induce pluripotency in somatic cell fusion through biphasic reprogramming. *Mol Cell* 2012;46:159-70.
138. Kumar R, DiMenna L, Schrode N, Liu TC, Franck P, Munoz-Descalzo S, *et al.* AID stabilizes stem-cell phenotype by removing epigenetic memory of pluripotency genes. *Nature* 2013;500:89-92.
139. Hogenbirk MA, Heideman MR, Velds A, van den Berk PC, Kerkhoven RM, van Steensel B, *et al.* Differential programming of B cells in AID deficient mice. *PLoS One* 2013;8:e69815.
140. Jiang C, Foley J, Clayton N, Kissling G, Jokinen M, Herbert R, *et al.* Abrogation of lupus nephritis in activation-induced deaminase-deficient MRL/lpr mice. *J Immunol* 2007;178:7422-31.
141. van Es JH, Gmelig Meyling FH, van de Akker WR, Aanstoot H, Derksen RH, Logtenberg T. Somatic mutations in the variable regions of a human IgG anti-double-stranded DNA autoantibody suggest a role for antigen in the induction of systemic lupus erythematosus. *J Exp Med* 1991;173:461-70.
142. Zan H, Zhang J, Ardeshtna S, Xu Z, Park SR, Casali P. Lupus-prone MRL/fas^{lpr}/lpr mice display increased AID expression and extensive DNA lesions, comprising deletions and insertions, in the immunoglobulin locus: Concurrent upregulation of somatic hypermutation and class switch DNA recombination. *Autoimmunity* 2009;42:89-103.
143. Hsu HC, Wu Y, Yang P, Wu Q, Job G, Chen J, *et al.* Overexpression of activation-induced cytidine deaminase in B cells is associated with production of highly pathogenic autoantibodies. *J Immunol* 2007;178:5357-65.
144. Bombardieri M, Barone F, Humby F, Kelly S, McGurk M, Morgan P, *et al.* Activation-induced cytidine deaminase expression in follicular dendritic cell networks and interfollicular large B cells supports

- functionality of ectopic lymphoid neogenesis in autoimmune sialoadenitis and MALT lymphoma in Sjogren's syndrome. *J Immunol* 2007;179:4929-38.
145. Humby F, Bombardieri M, Manzo A, Kelly S, Blades MC, Kirkham B, *et al.* Ectopic lymphoid structures support ongoing production of class-switched autoantibodies in rheumatoid synovium. *PLoS Med* 2009;6:e1.
146. Sun Y, Peng I, Senger K, Hamidzadeh K, Reichelt M, Baca M, *et al.* Critical role of activation induced cytidine deaminase in experimental autoimmune encephalomyelitis. *Autoimmunity* 2013;46:157-67.
147. Jiang C, Zhao ML, Waters KM, Diaz M. Activation-induced deaminase contributes to the antibody-independent role of B cells in the development of autoimmunity. *Autoimmunity* 2012;45:440-8.
148. Durandy A, Cantaert T, Kracker S, Meffre E. Potential roles of activation-induced cytidine deaminase in promotion or prevention of autoimmunity in humans. *Autoimmunity* 2013;46:148-56.
149. Quartier P, Bustamante J, Sanal O, Plebani A, Debre M, Deville A, *et al.* Clinical, immunologic and genetic analysis of 29 patients with autosomal recessive hyper-IgM syndrome due to Activation-Induced Cytidine Deaminase deficiency. *Clin Immunol* 2004;110:22-9.
150. Hase K, Takahashi D, Ebisawa M, Kawano S, Itoh K, Ohno H. Activation-induced cytidine deaminase deficiency causes organ-specific autoimmune disease. *PLoS One* 2008;3:e3033.
151. Fagarasan S, Muramatsu M, Suzuki K, Nagaoka H, Hiai H, Honjo T. Critical roles of activation-induced cytidine deaminase in the homeostasis of gut flora. *Science* 2002;298:1424-7.
152. Kuraoka M, Holl TM, Liao D, Womble M, Cain DW, Reynolds AE, *et al.* Activation-induced cytidine deaminase mediates central tolerance in B cells. *Proc Natl Acad Sci USA* 2011;108:11560-5.
153. Munoz DP, Lee EL, Takayama S, Coppe JP, Heo SJ, Boffelli D, *et al.* Activation-induced cytidine deaminase (AID) is necessary for the epithelial-mesenchymal transition in mammary epithelial cells. *Proc Natl Acad Sci USA* 2013;110: E2977-86.