Targeting of somatic hypermutation

Valerie H. Odegard and David G. Schatz‡*

Abstract | Somatic hypermutation (SHM) introduces mutations in the variable region of immunoglobulin genes at a rate of ~10^{-3} mutations per base pair per cell division, which is 10^6-fold higher than the spontaneous mutation rate in somatic cells. To ensure genomic integrity, SHM needs to be targeted specifically to immunoglobulin genes. The rare mistargeting of SHM can result in mutations and translocations in oncogenes, and is thought to contribute to the development of B-cell malignancies. Despite years of intensive investigation, the mechanism of SHM targeting is still unclear. We review and attempt to reconcile the numerous and sometimes conflicting studies on the targeting of SHM to immunoglobulin loci, and highlight areas that hold promise for further investigation.

V(D)J recombination
Somatic rearrangement of variable (V), diversity (D) and joining (J) regions of the genes that encode antigen receptors, which leads to repertoire diversity of both T-cell and B-cell receptors.

Transition mutations
Base changes in DNA in which a pyrimidine (cytidine (C) or thymidine (T)) is replaced by another pyrimidine, or a purine (adenosine (A) or guanosine (G)) is replaced by another purine.

Transversion mutations
Base changes in DNA in which a pyrimidine (cytidine (C) or thymidine (T)) is replaced by a purine (adenosine (A) or guanosine (G)), or a purine is replaced by a pyrimidine.

**VaxInnate Corporation, 300 George Street, Suite 311, New Haven, Connecticut 06511, USA.*
‡Yale Medical School, Section of Immunobiology, 300 Cedar Street, Box 208011, New Haven, Connecticut 06520-8011, USA.
Correspondence to D.G.S.
e-mail: david.schatz@yale.edu
doi:10.1038/nri1896

B cells express a diverse array of clonotypic cell-surface antigen receptors known as immunoglobulins, which are heterotetramers composed of two heavy chains and two light chains. These polypeptides are encoded by the three immunoglobulin loci, the heavy chain (IgH), the κ-light chain (Igκ) and the λ-light chain (Igλ) loci. Each of these three loci contains a variable region, which contributes to the antigen-binding domain of the immunoglobulin molecule, and a constant (C) region. The variable region of the heavy chain is generated from variable (V), diversity (D) and joining (J) gene segments, whereas the variable regions of the light chains are generated from V and J gene segments. There are multiple C regions encoded in the IgH locus, each with a specialized effector function. Immunoglobulin genes are assembled early in B-cell development by a genetic recombination programme known as V(D)J recombination. Subsequently, B cells enter the periphery where they can come into contact with their cognate antigens in the presence of antigen-specific T cells. A fraction of the B cells activated in this manner initiate the formation of oligoclonal germinal centres, in which they undergo two distinct immunoglobulin gene diversification processes, class switch recombination (CSR) and somatic hypermutation (SHM). CSR involves recombination between switch (S) regions to alter the C region of the IgH and therefore the effector function of the immunoglobulin molecule. By contrast, SHM introduces non-templated point mutations in the variable region of rearranged immunoglobulin heavy and light chain genes. SHM underlies the process of affinity maturation, which results in the preferential outgrowth of B cells expressing an immunoglobulin that has high affinity for its cognate antigen.

In recent years, substantial progress has been made in identifying the factors involved in SHM and in defining the molecular pathways that result in the introduction of mutations. By contrast, remarkably little is understood about the mechanisms that target SHM specifically to immunoglobulin genes, although considerable efforts have been made to address this issue. In this Review, we focus on the mechanisms by which SHM is targeted to immunoglobulin genes and discuss ideas that have been considered and discarded, experiments that have provided tantalizing leads, areas of considerable uncertainty and the opportunities offered by the recent advances in the enzymology of the reaction.

Features of SHM

The mutations introduced by SHM are predominately point mutations, although insertions and deletions are occasionally observed[1]. Transition mutations occur about twice as frequently as transversion mutations and a high proportion of mutations arise in the hotspot motif DGYW (where D denotes adenosine (A), guanosine (G) or thymidine (T); Y denotes cytidine (C) or T; and W denotes A or T) or its reverse complement WRCH (where R denotes A or G; and H denotes T, C or A), showing that SHM is influenced by the primary sequence of the DNA[2].

The role of transcription. Transcription underpins numerous key molecular features of SHM[3]. Eliminating transcription through an immunoglobulin locus results in a loss of SHM[4]. In a pre-B-cell line that supports SHM, the mutation rate of an immunoglobulin gene is proportional to the rate of transcription through that locus[5]. Mutations are confined to a 1–2 kilobase (kb) region in rearranged immunoglobulin genes. The variable-region

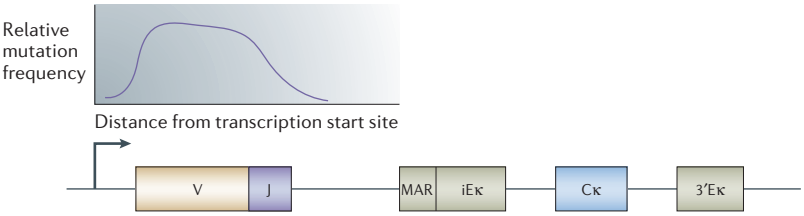

Figure 1 | **Gene-specific targeting of somatic hypermutation.** The somatic hypermutation machinery is targeted to a ~1.5 kilobase (kb) region in the three immunoglobulin loci. In this figure, the graph above the rearranged variable (V) and joining (J) gene segments that form the variable region of Igκ depicts the mutation domain in the κ-light chain (Igκ) locus. 3′Eκ, Igκ 3′ enhancer; Cκ, Igκ constant; iEκ, Igκ intronic enhancer; MAR, matrix attachment region.

promoter defines the 5′ boundary of this hypermutation domain with mutations beginning ~150 base pairs (bp) downstream of the transcription start site. The 3′ boundary is not as well defined, but the hypermutation domain includes the rearranged variable region and a portion of the 3′ flanking intronic region, whereas the C region is protected from mutation (FIG. 1). Mutation frequency is not constant throughout the mutated region and decreases exponentially with increasing distance from the transcription start site[6] (FIG. 1). Interestingly, duplication of a variable-region promoter immediately upstream of the Igκ C (Cκ) region in a transgene led to a low level of mutation in the Cκ region[7]. However, it remains to be determined whether a similar manipulation in the endogenous Igκ locus would lead to mutations in the Cκ region.

An influential model to explain the link between SHM and transcription proposes that a mutator factor is loaded onto RNA polymerase II at a variable-region promoter, is brought into the protein-encoding portion of the gene by the elongating transcription complex and is deposited onto the DNA, at which point it triggers the introduction of a mutation[7]. Transcriptional pausing or simple dissociation could release the factor from the elongation complex, perhaps preferentially at hotspot motifs. It has been suggested that the mutator has a constant probability of dissociating from the polymerase at each elongation step and that this results in the observed exponential decrease in mutation as the polymerase moves towards the 3′ end of the gene[6]. The relative lack of mutations in the first ~150 bp of the gene has been suggested to arise from a failure of the mutator to associate with the transcription complex until it has progressed past the promoter and entered the elongation phase[8,9].

Activation-induced cytidine deaminase. Activation-induced cytidine deaminase (AID) is required for SHM and CSR[10,11]. AID is also required for immunoglobulin gene conversion (GCV)[12,13], which is a homologous-recombination-based process used by various animals, including chickens and rabbits, to diversify immunoglobulin genes in B cells. Expression of AID is generally confined to germinal-centre B cells, but overexpression of AID in fibroblast cell lines and bacteria gives rise to a high frequency of SHM-like mutations in multiple, highly expressed genes[14–16]. Furthermore, constitutive AID expression in mice results in the accumulation in T cells of mutations in the genes encoding the T-cell receptor and MYC, and a high incidence of T-cell lymphomas[14]. These results raise the possibility that B-cell-specific cofactors or activities are needed to enforce immunoglobulin substrate specificity and perhaps to ensure a high rate of mutation. Mutation of the amino-terminal region of AID impairs SHM but leaves CSR intact, whereas the opposite is observed with certain carboxy-terminal-domain mutations of AID[17–19]. These studies are consistent with the possibility that cofactors needed for SHM and CSR interact with the amino- and carboxy-terminal domains of the protein, respectively. The carboxy-terminal region of AID also functions as a nuclear export signal and probably has an important role in regulating the amount of AID in the nucleus[20–22]. Recent evidence indicates that phosphorylation of AID by cyclic-AMP-dependent protein kinase (PKA) in activated B cells regulates the interaction of AID with replication protein A (RPA), which indicates another potential mechanism for control of AID activity that could be missing in non-lymphoid cells engineered to constitutively express AID[23–25].

The role of AID in SHM, CSR and GCV has been the subject of intensive investigation in recent years. Given its homology with the cytidine deaminase APOBEC1 (REF. 26), AID was initially thought to be an RNA-editing enzyme that modifies an mRNA target(s) to produce a new protein required for these reactions[10]. There is currently no direct experimental evidence to support the RNA-editing model. Instead, a growing body of evidence strongly indicates that AID acts directly on DNA, converting C to uridine (U) in immunoglobulin V and S regions.

Recombinant AID deaminates single-stranded DNA (ssDNA), but not RNA or double-stranded DNA (dsDNA), and has a preference for deaminating C nucleotides in SHM hotspot motifs[8,27–33]. Transcription, potentially with assistance from template supercoiling[34] or the ssDNA-binding protein RPA[23], is thought to have a key role in generating the ssDNA necessary for AID to act on dsDNA targets. An attractive model is that the small regions of ssDNA generated transiently by transcription are bound by RPA, which in turn directs AID activity to the variable region[23].

After AID initiates SHM by the deamination of C nucleotides, the resulting U•G mismatch has several possible fates that could lead to mutation[35] (FIG. 2). If the mismatch is not repaired before the onset of DNA replication, DNA polymerases will insert an A nucleotide opposite the U nucleotide creating C>T and G>A transition mutations. If, however, the U nucleotide is removed by uracil-DNA glycosylase (UNG), an abasic site is created, replication of which should give rise to both transition and transversion mutations. In addition to activating UNG-dependent base-excision repair (BER), a U•G mismatch recruits the mismatch repair (MMR) machinery[36], which is thought to create mutations at A•T near the initiating U•G lesion, probably through an error-prone patch repair process (for review see REF. 35).

Hotspot motif
A short DNA motif (DGYW or WRCH; where D denotes adenosine (A), guanosine (G) or thymidine (T); Y denotes cytidine (C) or T; W denotes A or T; R denotes A or G; and H denotes T, C or A) in which mutations are preferentially inserted during somatic hypermutation.

Elongating transcription complex
A complex comprising the RNA polymerase and its associated proteins that is formed during the elongation phase of transcription.

Nuclear export signal
A highly conserved, leucine-rich sequence that facilitates protein trafficking from the nucleus to the cytoplasm.

Template supercoiling
Compaction of DNA by twisting and folding.

Abasic site
A site created by the loss of a purine or pyrimidine from DNA.

Base-excision repair
A DNA-repair pathway that removes single bases from DNA, such as uridine nucleotides arising by deamination of cytidine. Repair is initiated by a DNA glycosylase that is specialized for a particular class of damage.

Mismatch repair
A repair pathway that recognizes and corrects mismatched base pairs (typically those that arise from errors of chromosomal DNA replication).

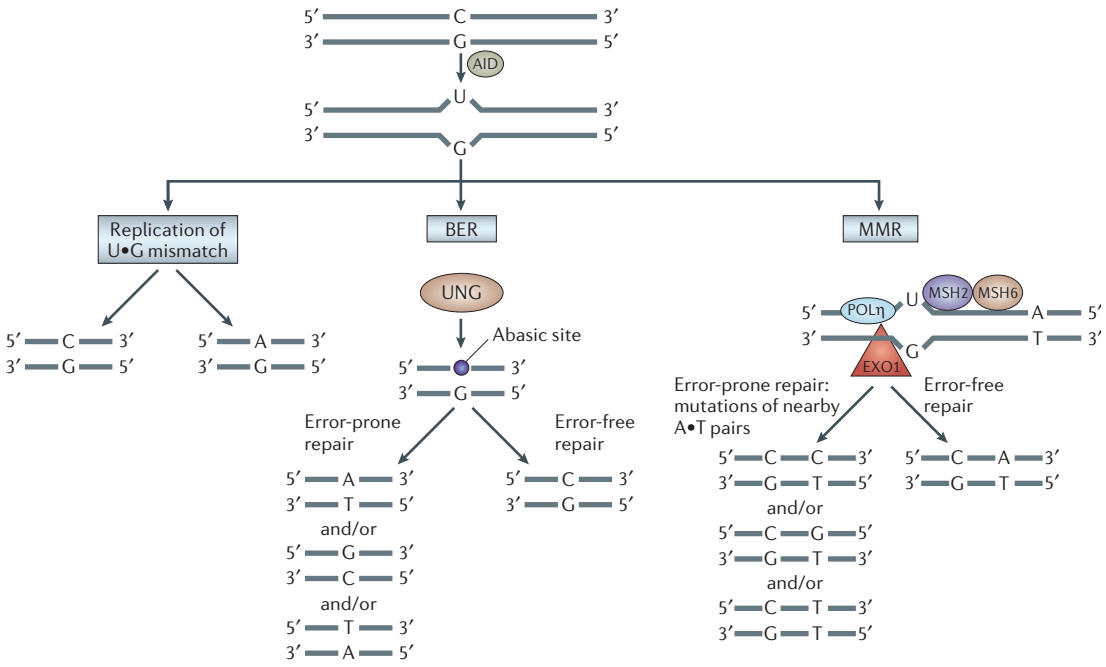

Figure 2 | **Activation-induced-cytidine-deaminase-dependent lesion repair.** Experimental evidence supports a model in which activation-induced cytidine deaminase (AID) initiates somatic hypermutation (SHM) by the deamination of cytidine (C) nucleotides on single-stranded DNA. As first proposed by Neuberger and colleagues, the resulting uridine (U)•guanosine (G) mismatch can be repaired by one of three pathways. For a mutation to be fixed at the site of deamination, error-free DNA repair must be 'perturbed' to become error-prone. If the mismatch is carried unrepaired to replication, DNA polymerases will insert an adenosine (A) opposite the U nucleotide, ultimately creating C>T and G>A transition mutations. In base-excision repair (BER), replication over an abasic site created by uracil-DNA glycosylase (UNG) will give rise to both transition and transversion mutations. A U•G mismatch also recruits the mismatch repair (MMR) machinery, which is thought to create mutations at A•T base pairs near the initiating U•G lesion. EXO1, exonuclease 1; MSH, homologue of *Escherichia coli* MutS; POLη, polymerase (DNA directed)-η.

It is important to note that U nucleotides are frequently incorporated in DNA during normal cell growth (independently of AID activity) by deamination of C nucleotides and by the misincorporation of U nucleotides during replication[37]. These U nucleotides are efficiently and accurately repaired, predominately by BER pathways that depend on the activity of uracil-DNA glycosylases. During SHM, however, the uracil lesion is repaired in an error-prone manner. Why repair should be error-prone in this context but error-free in others is not understood. It has been suggested that either the repair processes are saturated during SHM or error-free repair pathways are somehow perturbed specifically in the vicinity of immunoglobulin genes[38]. One mechanism by which error-free repair might be disrupted is indicated by the recent finding that the MRE11 (meiotic recombination 11 homologue)–RAD50 complex is able to cleave DNA at abasic sites, generating a 3′ end that cannot be extended by DNA polymerases[39]. Overexpression of the Nijmegen breakage syndrome 1 (NBS1) protein, which is part of the MRE11–RAD50–NBS1 (MRN) complex, increased SHM in a human B-cell line and GCV in a chicken B-cell line, consistent with the idea that this complex can participate in the processing of AID-generated lesions[39].

Given its putative function as a DNA mutator, it is surprising that AID is predominately located in the cytoplasm of B cells[20,22,40]. Immunohistochemistry indicates that the localization of AID to the nucleus of B cells occurs most frequently in regions of the germinal centre in which SHM occurs[41], which provides evidence that nuclear AID correlates with SHM. Accumulating evidence indicates that the regulation of subcellular localization has an important role in the regulation of AID activity[42]. One study indicated that AID contains a nuclear localization signal[20] and three groups identified a nuclear export signal near the carboxy terminus of AID[20–22], which indicates that AID can shuttle between the nucleus and cytoplasm of cells. Consistent with this, AID accumulates in the nucleus when nuclear export is inhibited[17,21–22] and a recent study showed that the induction of DNA lesions results in the retention of AID in the nucleus[21]. Nuclear retention of AID alone, however, does not result in an increase in the mutation frequency of physiological targets, indicating that further cofactors need to be imported along with AID to support SHM[17,21,22].

DNA lesions in SHM. It is generally accepted that AID-dependent DNA lesions are crucial for SHM and CSR, with the exact nature of these lesions and the pathways by which they are generated being areas of active

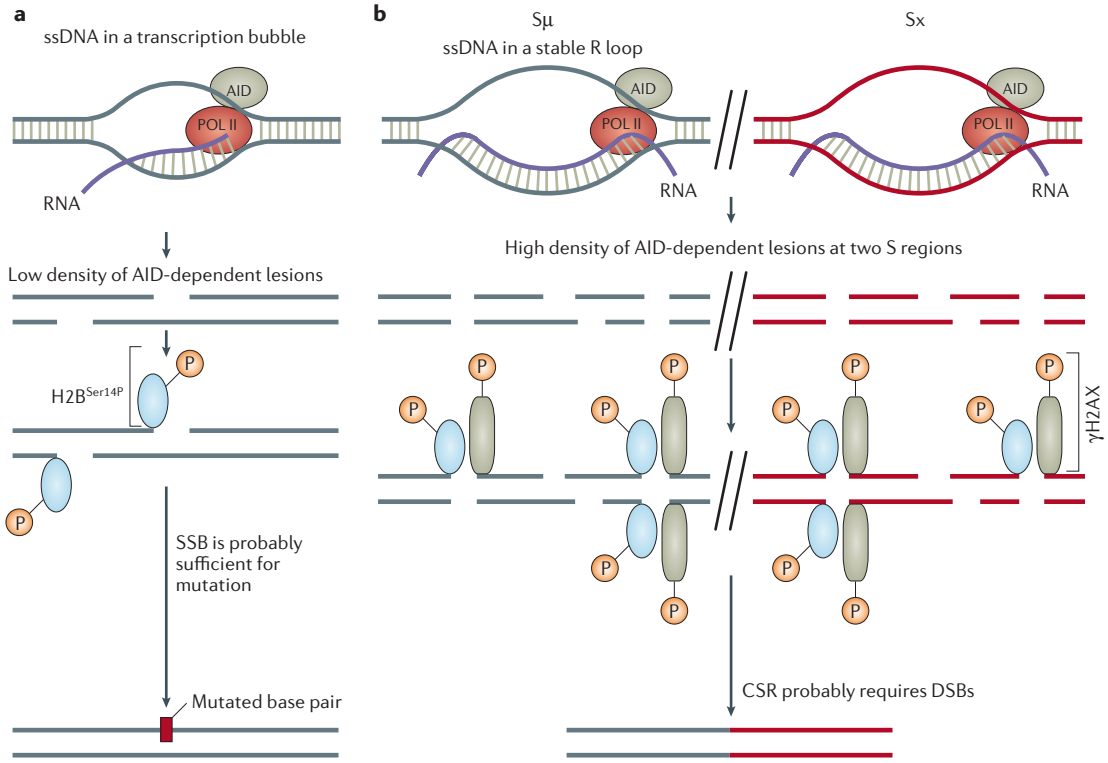

Figure 3 | **DNA lesions in somatic hypermutation and class switch recombination.** This model contrasts the introduction and repair of activation-induced cytidine deaminase (AID) dependent lesions in somatic hypermutation (SHM) (**a**) and class switch recombination (CSR) (**b**). In SHM, the transcription bubble exposes single-stranded DNA (ssDNA) to AID. In CSR, the repetitive nature of the transcribed switch (S) regions gives rise to R loops. Evidence indicates that further protein factors (such as replication protein A and cyclic-AMP-dependent protein kinase) are involved in the recruitment of AID to exposed ssDNA (not shown in this figure). One model proposes that the density of AID-dependent DNA lesions in variable and S regions differ, leading to the recruitment of distinct sets of repair factors that ultimately resolve AID-dependent DNA lesion in different ways. DSB, double-strand breaks; γH2AX, phosphorylated histone 2A family member X; $H2B^{Ser14P}$, phosphorylated histone 2B; POL II, polymerase II; SSB, single-strand breaks.

investigation (FIG. 3). CSR involves DNA double-strand breaks (DSBs) (reviewed in REF. 43). Initial studies found that blunt-ended DSBs in rearranged variable regions undergo SHM in constitutively hypermutating cell lines and primary germinal-centre B cells[44,45]. These blunt-ended DSBs, however, were found to arise independently of AID activity[46,47] and to accumulate in the variable regions of immunoglobulin genes in various cell types, which include fibroblasts and T cells[48]. The asymmetrical nature of these lesions indicates that they might be generated during DNA replication when the polymerase encounters a single-strand nick or gap[49]. By contrast, DSBs with resected ends were reported to be found only in variable regions undergoing SHM and to be AID dependent, raising the possibility that resected DSBs are intermediates in SHM[48]. Furthermore, DNA single-strand breaks (SSBs) were detected in Igλ variable regions that were mutating in germinal-centre B cells[50] (see below).

The carboxy-terminal tail of histone 2A family member X (H2AX) is rapidly phosphorylated (γH2AX) in response to DSBs in DNA[51]. DSBs and γH2AX are both associated with genomic regions undergoing V(D)J recombination or CSR, and H2AX is required for efficient CSR (reviewed in REF. 43). By contrast, H2AX is not required for SHM[52] and γH2AX is not detectable at Igλ loci undergoing SHM[53], although it has been detected at the IgH variable region of a hypermutating B-cell line[54]. These results, together with various other genetic and biochemical findings[42,50,55,56], strongly support a DNA-SSB model for SHM. It seems probable, however, that DSBs can be generated as a by-product of SHM because the reaction has been linked to chromosomal translocations (see below) and generates a low level of deletion and insertion mutations[1]. In addition to replication-induced DSBs, SSBs in close proximity on the two DNA strands could explain a subset of the SHM-associated DSBs detected in the studies mentioned earlier. Such a mechanism has been invoked to explain the DSBs observed in CSR[57]. To explain the lack of detectable γH2AX associated with SHM at the Igλ locus, it is appealing to think that SHM generates a lower density of SSBs, and therefore fewer DSBs, than CSR, which thereby keeps γH2AX concentrations below detectable limits. Alternatively, it is possible that the DSBs generated during SHM do not recruit γH2AX, supporting the idea that the DNA lesions introduced during SHM and CSR are repaired by different pathways[53].

Chromosomal translocation
An aberration of chromosome structure in which a portion of one chromosome is broken off and becomes attached to another.

SHM is occasionally mistargeted and contributes to lymphomagenesis. It was originally thought that SHM occurred only in rearranged variable regions, but an increasing number of studies using primary human B cells indicate that this is not the case. Mutations have been detected in B-cell lymphoma 6 (*BCL6*, which is a proto-oncogene) and *CD95* (which is a tumour suppressor gene) in normal human germinal-centre and memory, but not naive, B cells[58–60]. Aberrant SHM has also been suggested to mutate the proto-oncogenes *PIM1*, *MYC*, *RHOH* (RAS homologue gene-family member H) and *PAX5* (paired box gene/protein 5), contributing to the development of diffuse large B-cell lymphomas, which are tumours derived from germinal-centre B cells[61]. It should be noted, however, that aberrant SHM is not observed in all subtypes of lymphomas that are of germinal-centre or post-germinal-centre B-cell origin[61]. More recently, two genes that are not oncogenic, *B29* (also known as *CD79B*) and *PSMB5* (proteasome (prosome, macropain) subunit, β type, 5; also known as *MB1*), which encode B-cell-receptor-signalling proteins, were found to be substrates for SHM[62]. Very recently, *Bcl6* was reported to be hypermutated in Peyer's patch B cells from mice[63], which contrasts with several earlier reports claiming that the mouse gene is not a SHM target[3,64,65]. Indeed, we also found a substantial level of mutation in *Bcl6* in Peyer's patch B cells from mice (M. Liu and D.G.S., unpublished observations). The mutation frequency of non-immunoglobulin genes is ~50–100 fold lower than that of hypermutated variable regions[58–60,62], but the basis for this difference is not known. The acquired mutations resemble those introduced by SHM in rearranged variable regions as they are predominately transition mutations and are biased to DGYW hotspots. Furthermore, with a few exceptions, the mutation distribution in non-immunoglobulin genes is similar to that in hypermutated variable regions[58–60,62].

Mistargeted SHM can contribute to lymphomagenesis by mutating regulatory and coding sequences in proto-oncogenes and tumour suppressor genes, and by promoting chromosomal translocations. Mutations in *BCL6* are confined to a putative regulatory region which, if mutated, could dysregulate *BCL6* expression[58]. Although most mutations in *CD95* occur in the first 2 kb of the gene, a small proportion arise in exon 9, which encodes the death domain of the protein[60]. Mutations in exon 9 could confer resistance to CD95-induced apoptosis, which would allow tumour cells to escape CD95-mediated elimination. As noted earlier, mutations originating from SHM are detected in multiple proto-oncogenes in diffuse large B-cell lymphomas, which provides further evidence that SHM contributes to lymphomagenesis[58]. In addition, these proto-oncogenes are susceptible to chromosomal translocations in the regions where SHM-derived mutations are located[61]. AID activity, in the context of CSR, has been linked to genome instability and chromosomal translocations involving S regions[66–68]. Similarly, the genome instability associated with SHM is probably attributable to aberrant AID targeting and/or activity.

It is plausible that those genes that attract the SHM machinery contain a common targeting feature(s). Therefore, identifying structural and functional similarities between mutation targets could provide key insights to the mechanism(s) that underlies the specificity of the SHM reaction. Although the immunoglobulin and non-immunoglobulin genes that are subject to SHM do not have obvious primary sequence homologies (the DGYW motif, even though it is a hotspot for SHM, is found throughout the genome and therefore cannot be a primary sequence that targets SHM to immunoglobulin genes)[58,60–62], a detailed, comparative study of loci that undergo SHM has not been done.

Targeting SHM to immunoglobulin loci

The targeting of SHM can be thought of as occurring at two distinct levels. One, which we refer to as 'local' targeting, arises from the observation that SHM is confined to a 1–2 kb region downstream of the transcription start site in immunoglobulin loci. As summarized earlier, the coupling of a mutator factor (perhaps AID) to elongating RNA polymerase II provides a plausible, although as yet unproven, mechanism for local targeting. The second, which we term 'global' targeting, is the putative process that directs SHM specifically to immunoglobulin genes. Global targeting is still very much a mystery and there is no solid support for any hypothesis. The idea that there is some degree of global targeting is supported by the observation that not all transcribed genes in germinal-centre B cells undergo SHM[59,69](M. Liu and D.G.S., unpublished observations) and is reinforced by the idea that it would seem to be deleterious to genomic integrity for non-immunoglobulin genes to undergo SHM indiscriminately. The substrate specificity of AID is unlikely to explain the global targeting of SHM given the abundance of actively transcribed genes containing DGYW motifs that are not targets for SHM. One simple explanation would be that global targeting of SHM arises from a particularly high rate of transcription at immunoglobulin loci. Although this idea does not readily explain why other highly expressed genes would not be subject to SHM, it has not been rigorously tested. In the discussion that follows, we assume that factors other than (or in addition to) transcription rate are involved in the global targeting of SHM.

It seems probable that two aspects of SHM, AID-mediated deamination and error-prone DNA lesion repair, must be targeted to the variable regions of immunoglobulin loci. This could be accomplished by two independent mechanisms, one that recruits AID to immunoglobulin loci and another that ensures that repair occurs in an error-prone manner. Alternatively, a single mechanism that recruits AID specifically to immunoglobulin loci would suffice if AID then triggers error-prone repair and/or recruits error-prone repair factors (perhaps by virtue of its interaction with RPA, which interacts with error-prone repair factors)[23]. Global targeting might therefore be reduced to the question: how is AID targeted specifically to immunoglobulin loci? A recent study provides support for the idea that the C regions and extreme 5′ ends of immunoglobulin genes are not accessible to AID[9], indicating that local targeting might also be determined by the spatially restricted recruitment of AID.

Death domain
A protein–protein interaction domain found in many proteins that are involved in signalling and apoptosis.

Enhancer
A control element in DNA to which regulatory proteins bind and influence the rate of gene transcription. Enhancers function in an orientation- and position-independent manner (that is, they can function either upstream or downstream of the associated gene, or in an intron).

Matrix attachment region
DNA sites that are linked to the nuclear matrix. They are thought to have roles in gene regulation and higher order chromosome organization.

Cis-*acting DNA elements*. In other regulated DNA processes, such as transcription and V(D)J recombination, *cis*-acting elements function as platforms on which transcription factors assemble[70]. These assemblies of protein factors have various functions, most notably mediating enhancer and promoter interactions, and mediating the recruitment of chromatin-modifying enzymes and the transcription machinery. It is appealing to think that a unique assembly of protein factors, for example on immunoglobulin enhancers, could recruit the SHM machinery specifically to immunoglobulin loci (FIG. 4; TABLE 1). In this case, the mistargeting of SHM to oncogenes could indicate that these genes have regulatory elements that are structurally or functionally similar to the targeting elements in immunoglobulin loci.

The finding that Igκ transgenes that are designed to resemble a rearranged Igκ locus can be substrates for SHM[71,72] opened the way for analysis of the *cis*-acting elements that recruit the mutation machinery. Initial studies using a 16 kb Igκ transgene, which contained endogenous Igκ sequences beginning at the Igκ promoter and continuing 9 kb past Cκ, indicated that it contained at least a minimum set of *cis*-acting elements that are required for targeting SHM to the Igκ variable region[71,72]. Attempts to create similar systems for the Igλ[73,74] and IgH[75–78] loci have not been as successful, indicating that further sequences from these loci are needed for efficient hypermutation, perhaps to modulate transgene accessibility and transcription, or to function as direct targets of the SHM machinery.

It was proposed that the primary targets of SHM, rearranged variable regions, contain a targeting motif for the hypermutation machinery. However, when the rearranged V and J gene segments of Igκ (VJκ) were replaced by heterologous DNA sequences in the context of an Igκ transgene, SHM was not affected[79]. Furthermore, unlike CSR, SHM does not seem to target specialized RNA or DNA secondary structures, but instead the mutability of a region is determined by the collection of hotspot motifs in the sequence[80]. These studies indicate that the variable region itself is not necessary for recruiting SHM, at least in the context of a transgene.

The correlation between transcription and SHM indicates that transcriptional regulatory elements could be involved in targeting SHM to immunoglobulin loci. Transgenic studies have shown that heterologous RNA polymerase II promoters can support SHM[76,81], but whether this is true in an endogenous context and applies universally to active RNA polymerase II promoters remains to be determined. Two enhancers, one in the intron between Jκ5 and Cκ (known as the intronic enhancer, iEκ) and one 9 kb 3′ of Cκ (known as the 3′ enhancer, 3′Eκ), have been identified and extensively analysed[82,83]. In the context of an Igκ transgene, deletion of iEκ and its flanking matrix attachment region (MAR) abolished SHM but expression of the transgene was maintained[81], indicating that iEκ has a role in SHM that is distinct from its role in facilitating transcription. Deletion of 3′Eκ in the context of a similar transgene markedly reduced SHM of the Igκ variable region but also resulted in a dramatic reduction in transcription of the transgene[81]. As other studies have indicated that a poorly expressed transgene does not mutate[5,84], it was not possible to determine whether 3′Eκ has a direct role in targeting or merely maintains the required level of transcription. Together, these studies indicate that both iEκ and 3′Eκ are required for efficient SHM of an Igκ transgene, although subsequent experiments from the same group showed that they are not sufficient for transgene mutation[81,85].

Attempts to define putative targeting elements in iEκ and 3′Eκ were largely uninformative because any observable decrease in SHM was accompanied by a decrease in transgene expression[84]. Furthermore, Igκ-transgene expression and mutability can be heavily influenced by the site at which the transgene integrates[84,86], which hinders the interpretation of transgenic studies. In addition, Igκ transgenes, such

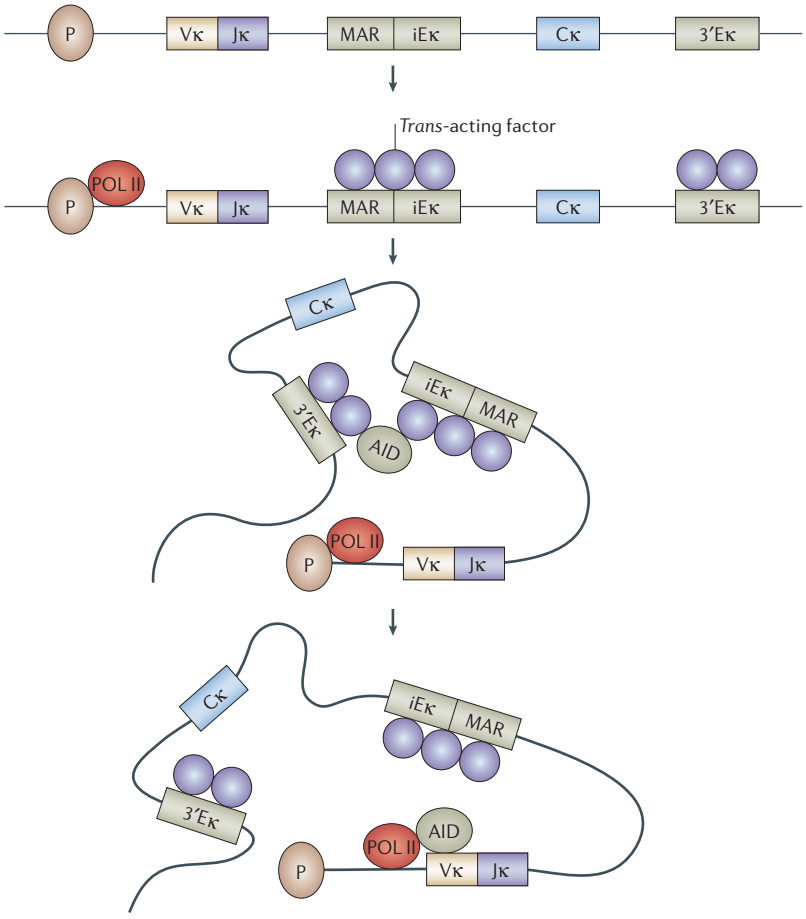

Figure 4 | **Recruitment of activation-induced cytidine deaminase by *trans*-acting factors.** One model predicts that a collection of *trans*-acting factors bound to immunoglobulin enhancers form a specific structural platform that recruits activation-induced cytidine deaminase (AID) to the immunoglobulin locus (depicted here is the κ-light-chain (Igκ) locus). Although not shown in this figure, protein factors bound to the immunoglobulin promoter (P) could also be involved in this process. AID is then loaded onto the RNA polymerase, which carries AID to the rearranged variable (V) and joining (J) gene segments. 3′Eκ, Igκ 3′ enhancer; Cκ, Igκ constant; iEκ, Igκ intronic enhancer; MAR, matrix attachment region; POL II, polymerase II.

as Igλ and IgH transgenes, often mutate at far lower frequencies than those observed for endogenous immunoglobulin variable regions[87]. This observation calls into question whether conventional transgenes provide a reliable reflection of the mechanisms that target SHM to endogenous immunoglobulin loci. Therefore, despite initially encouraging findings and much effort, the study of transgenes has yet to identify a single well-defined *cis*-acting element for targeting SHM.

Given the variable nature of transgenes, it is important to determine the contribution of endogenous immunoglobulin enhancers to SHM targeting. Analysis of SHM in mice lacking 3′Eκ led to two conclusions that conflicted with those of the Igκ-transgene studies: first, 3′Eκ is not essential for Igκ transcription[88]; and second, 3′Eκ is not required for antigen-driven SHM[89]. More recently, Xu and colleagues[90] have shown that the endogenous iEκ is not required for SHM of Igκ, providing further evidence that immunoglobulin transgenes do not accurately reflect the role of *cis*-acting elements in SHM. Deletion of the endogenous IgH intronic enhancer (Eμ) and its flanking MARs in a constitutively mutating cell line did not reduce the frequency of SHM of the IgH locus[91]. In mice, SHM of IgH was not markedly affected by deletion of either the core region of Eμ[91,92] or the 3′ enhancer of the endogenous IgH locus[93]. Taken together, these results raise important questions about the role of enhancers in the global targeting of SHM, although the possibility remains that immunoglobulin enhancer elements, and perhaps other *cis*-acting elements, have redundant functions in targeting SHM. The recent discovery of a third Igκ enhancer located at the extreme 3′ end of the locus indicates that the Igκ transgenes used so far do not contain the complete set of *cis*-acting regulatory elements that are present in the endogenous Igκ locus[94,95]. The absence of such elements in Igκ transgenes could explain the striking effects observed in the absence of one Igκ enhancer[81]. The discrepancies between studies of transgenic and endogenous immunoglobulin loci indicate that conventional transgenes might not provide accurate insight to the molecular mechanism of targeting. Therefore, results from early transgenic studies should be complemented with further studies on endogenous immunoglobulin loci to ultimately determine the role of *cis*-acting elements in targeting SHM.

Table 1 | **The role of immunoglobulin enhancers in somatic hypermutation**

System	Substrate	Role of enhancers	Reference
IgH locus enhancers			
Mice	IgH transgene	Eμ is not sufficient to target SHM, but transgenes that recombine with the IgH locus, and are therefore in close proximity to the complete set of IgH *cis*-regulatory regions, mutate	73
Mice	Igλ transgene	Eμ does not target SHM to Igλ transgene	77
Mice	IgH transgene	Eμ and the 3′ IgH enhancer do not target SHM	119
Mice	IgH transgene	The IgH 3′ enhancer is not sufficient for SHM	93
Mice	IgH transgene	Hypersensitive sites 3b and 4 of the IgH 3′ regulatory region are important for SHM	120
Mice	Endogenous IgH locus	Hypersensitive sites 3b and 4 of the IgH 3′ regulatory region are dispensible for SHM	121
Mice	Endogenous IgH locus	Eμ is not required for SHM of the endogenous IgH locus	92
Hybridomas	Endogenous IgH locus	Combined deletion of the Eμ and MAR of the IgH locus indicates that they are not required for SHM	91
Igκ locus enhancers			
Mice	Igκ transgene	iEκ and 3′Eκ are required for full SHM; iEκ most probably targets SHM and 3′Eκ is required for transcription	81
Mice	Igκ transgene	iEκ and 3′Eκ are not sufficient for targeting SHM to the Igκ transgene	85
Mice	Igκ transgene	MAR, iEκ, 3′Eκ and the 3′ flanking regions of 3′Eκ are required for SHM	84
Mice	Endogenous Igκ locus	The 3′Eκ enhancer is not required for SHM of the Igκ locus	89
Ramos cell line and mice	Non-immunoglobulin transgene	iEκ and 3′Eκ are sufficient for SHM	44
Mice	Endogenous Igκ locus	Neither 3′Eκ nor iEκ is required for SHM of the Igκ locus	90
Igλ locus enhancers			
Mice	Igλ transgene	The Igλ 2–4 enhancer, a described enhancer within the Igλ locus, supports low mutation frequency	73
Mice	Igλ transgene	The Igλ 2–4 enhancer, but not 3′Eκ, drives SHM of the Igλ transgene	74

3′Eκ, Igκ 3′ enhancer; Eμ, IgH intronic enhancer; iEκ, Igκ intronic enhancer; IgH, immunoglobulin heavy chain; Igκ, immunoglobulin κ-light-chain; Igλ immunoglobulin λ-light-chain; MAR, matrix attachment region; SHM, somatic hypermutation.

Trans-*acting factors.* Each immunoglobulin locus has at least two enhancer elements, which contain binding sites for several *trans*-acting factors, including nuclear factor-κB (NF-κB), octamer-binding transcription factor proteins, E12 and E47 (collectively known as E2A), transcription factor binding to IgH 3′Eμ (TFE3) and ETS family proteins[96]. In addition, most of these factors are expressed by germinal-centre B cells[97]. Targeted disruption of any of the genes encoding these factors typically has pleiotropic effects on transcription and early lymphocyte development, making it difficult to determine the role of a given factor in SHM by this method[97,98].

E2A is alternatively spliced to generate mRNAs that encode two ubiquitously expressed proteins, E12 and E47. These transcription factors bind a motif known as an E-box (CANNTG, where N denotes A, T, G or C) and have important roles in lymphocyte development[99]. E47 and E12 bind enhancer elements in all three immunoglobulin loci[96] and the regulatory region of *BCL6*, a gene that is mistargeted by SHM[59]. In addition, genes targeted for SHM in T-cell lymphomas that have been isolated from AID transgenic mice all contain E47-binding sites in their *cis*-regulatory regions[100]. Curiously, the accidental incorporation of two E-box motifs in the hypermutation domain of a transgene resulted in a substantial increase in mutation frequency that was not caused by an increase in transgene expression[87]. Whether the effects of E-box motifs located in the mutated region itself are similar to those caused by such motifs located in immunoglobulin enhancers or other *cis*-acting elements remains to be determined. Furthermore, the function of E-box motifs in endogenous immunoglobulin loci in SHM has not yet been addressed. Interestingly, a recent study finds that overexpression of E47 leads to increased GCV in chicken DT40 B cells, strengthening the connection between *E2A* gene products and the targeting of AID[101].

Conventional histone modifications. Each of the four core histones (H1–H4) contains a 20–35 amino-acid amino-terminal tail that extends from the surface of the nucleosome and contains sites for post-translational modifications. Acetylation of lysine residues in the histone tails of H3 and H4 is tightly correlated with transcriptional activation and chromatin accessibility. Acetylated lysine residues in a histone tail can decrease histone–DNA interactions and promote accessibility of the DNA for transcription. In addition, acetylated histone tails have been shown to regulate processes such as DNA replication, histone deposition and DNA repair by recruiting proteins that have an acetyl-lysine-binding domain, known as a bromodomain[102].

Histone acetylation has been associated with accessible gene segments and S regions during V(D)J recombination and CSR, respectively[70,103,104]. One study indicated that histone acetylation participates in the regulation of SHM as well[54]. Using the hypermutation-inducible BL2 cell line, IgH-variable-region chromatin was found to be hyperacetylated compared with the C region of IgM (Cμ) following activation of SHM. Treatment with trichostatin A (TSA), a histone-deacetylase inhibitor, caused increased acetylation of Cμ and a concomitant increase in SHM-generated mutations in this region. The authors concluded that histone acetylation is involved in targeting SHM specifically to the rearranged variable regions in immunoglobulin loci[54].

Our recent study of similar issues examined both IgH and immunoglobulin light chain loci in primary mouse B cells[53]. Histone acetylation, methylation of H3 at lysine residue 4 (H3–K4 methylation) and DNA methylation of the variable regions were found to be similar in naive and germinal-centre B cells. Interestingly, although the IgH variable region was found to be hyperacetylated compared with Cμ in germinal-centre B cells, no differences in histone acetylation and H3–K4 methylation were observed between the Igλ variable and C regions. The distinct acetylation patterns observed at the IgH and Igλ loci might be explained by the different sizes of these two loci. It is well established that, in many genes, levels of histone acetylation are highest near transcription start sites with a gradual decline over the length of the gene[105,106]. Therefore, one hypothesis is that the low level of histone acetylation seen at Cμ might reflect its greater distance from a transcription start site compared with the Igλ C (Cλ) region. Overall, our recent study strongly indicates that, *in vivo*, dynamic alterations of histone acetylation and H3–K4 methylation do not accompany the activation of SHM and that these chromatin modifications do not, by themselves, confer targeting of SHM specifically to variable regions in immunoglobulin loci[53]. Because similar patterns of histone modification were seen in naive and germinal-centre B cells, these modifications might represent a precondition for SHM that is established at, or before, the mature B-cell stage[53]. It should be noted that total H3 and H4 acetylation were measured, potentially obscuring the modulation of acetylation at specific lysine residues in response to the activation of SHM. Furthermore, because only a small sample of known chromatin modifications have been analysed, it is possible that a specific histone code at immunoglobulin variable regions, comprised of modifications not assessed as yet, targets SHM.

Phosphorylated histone H2B. Phosphorylation of histone H2B on serine 14 (denoted $H2B^{Ser14P}$) was first observed in the context of apoptosis[107,108] and subsequently in response to DSBs induced by γ-irradiation[109]. More recently, $H2B^{Ser14P}$ was shown to correlate temporally and spatially with both SHM and CSR[53]. Phosphorylation of H2B during SHM and CSR is AID dependent and associated with the recruitment of mammalian sterile kinase 1 (MST1)[53], which is the kinase responsible for H2B phosphorylation during apoptosis[107,108]. The role of $H2B^{Ser14P}$ in SHM and CSR has not been determined, but it is conceivable that this modification has a role in the recruitment or retention of AID or components of the error-prone repair pathways that resolve AID-dependent lesions. Previous studies indicate that $H2B^{Ser14P}$ can self-aggregate[107,108], raising the possibility that $H2B^{Ser14P}$ has a role in tethering broken DNA ends together in response to a lesion. Self-aggregation of $H2B^{Ser14P}$ could also result in the formation of a specific chromatin structure that

Octamer-binding transcription factor proteins
A family of transcription factors that recognize and bind a conserved eight nucleotide motif found in the regulatory regions of many genes, including immunoglobulin genes.

DT40 cells
A chicken B-cell line that undergoes high rates of homologous recombination and immunoglobulin gene conversion.

BL2 cell line
A human B-cell lymphoma cell line that can be stimulated to undergo SHM.

H3–K4 methylation
The addition of methyl groups to lysine residue 4 on histone H3.

functions as a docking site for lesion-repair factors (such as the MRN complex or error-prone polymerases), or perhaps the phosphorylated histone functions as such a site. In yeast, phosphorylation of histone H2A at sites of DNA damage helps to recruit components of the DNA repair machinery[110,111], and it is attractive to think that $H2B^{Ser14P}$ has a similar role in SHM and CSR.

Future directions and possibilities

Genome-wide SHM? The discovery that multiple non-immunoglobulin genes are subject to hypermutation in normal B cells shows that SHM is not as rigorously targeted as was once believed[58–62]. Curiously, Wabl and colleagues found that artificial retroviral substrates lacking any immunoglobulin element mutate when integrated at many positions in the genome of hypermutating B-cell lines[112]. The data were consistent with a high rate of mutation at some insertion sites and a low rate at many other sites. Broadly similar results were recently reported using the same retroviral vectors in the hypermutating cell line Ramos[113]. The implication is, at least using this experimental system, that AID-mediated SHM can occur in many genomic locations and does not require immunoglobulin gene elements. Wabl and colleagues suggest that AID and other components of the SHM machinery operate as general, genome-wide mutators, and that targeting of SHM to immunoglobulin loci is much less specific than is widely believed. According to such a model, in the absence of a targeting mechanism one could imagine that immunoglobulin genes seem to be more highly mutated than the remainder of the genome as a result of several factors. First, germinal-centre B cells expressing mutated variable regions are positively selected by affinity maturation (although it is not clear how this explains the high levels of mutations seen in non-productively rearranged variable regions[114–116]). Second, immunoglobulin genes might be more highly transcribed and have a higher density of DGYW hotspots than many other genes. And third, only certain portions of the genome and certain genes support efficient targeting by SHM for as yet unknown reasons (perhaps related to chromatin structure), and not enough sequencing has been performed so far to identify many of the efficiently mutated non-immunoglobulin genes. Before the genome-wide mutation model can become widely accepted, it will be necessary to rule out possible artefacts of the experimental system used in these studies[112,113], including the use of transformed cell lines and retroviral delivery of the mutation substrate. Extensive sequencing of actively transcribed genes in germinal-centre B cells, especially introns where mutations could accumulate without deleterious effects on the encoded protein, will help to explain the extent to which SHM is targeted to immunoglobulin genes.

Cis-acting elements in the targeting of SHM. To conclusively define the role of immunoglobulin *cis*-acting elements in SHM, it will be important to focus on studies of endogenous immunoglobulin loci. The targeted insertion of a rearranged VJκ that is driven by a constitutive promoter in the context of an enhancer-less Igκ allele might help to clarify the role of immunoglobulin enhancers in targeting SHM to the Igκ locus. It is possible that further *cis*-elements located in immunoglobulin loci and/or combinatorial effects of multiple immunoglobulin *cis*-elements regulate the targeting of SHM. Further targeted deletions and insertions could be used to address this point.

Chromatin structure. The finding that $H2B^{Ser14P}$ correlates with SHM (and CSR) raises the possibility that specific chromatin modification(s) has a role in targeting one or more steps of SHM to immunoglobulin loci. As conventional chromatin modifications are typically coordinately regulated throughout the genome, it is difficult to envision such modifications constituting a unique targeting signal. It is possible, however, that a specific combination of less common modifications provides a platform recognized by components of the hypermutation machinery.

SHM might be targeted by higher order chromatin organization. Subnuclear positioning has been suggested to regulate both transcription and V(D)J recombination[117], raising the possibility that this process is involved in SHM as well. It would be interesting to determine whether immunoglobulin loci cluster at a particular location in the nucleus of germinal-centre B cells. Such a location could be particularly permissive to SHM owing to a high relative concentration of components of the hypermutation machinery. If nuclear localization is important during SHM, it possible that non-immunoglobulin targets, such as *BCL6*, are located in close proximity to immunoglobulin loci and that this leads to the somatic mutations (and translocations) detected in these genes.

A recent study showed that the regulatory regions of genes located on different chromosomes can physically interact and that this interaction can regulate gene expression[118]. It is conceivable that such interactions could regulate further genetic programmes. In the context of SHM, the regulatory regions of immunoglobulin loci or accessory genes could associate in a way that mediates targeting by AID and/or error-prone repair factors.

AID, at least when purified from insect cells, is found to be associated with RNA, and degradation of the RNA is required for AID deaminase activity[29]. One might therefore speculate that in B cells, association of AID with RNA has a negative regulatory role, or conceivably that association with specific RNAs has a role in targeting AID to immunoglobulin genes.

The mechanism by which SHM is targeted to immunoglobulin variable regions remains largely unknown. Revisiting old ideas with improved experimental systems and testing new hypotheses that are based on recent advances should provide insight to this long-standing mystery. An understanding of the mechanisms targeting SHM to immunoglobulin genes should explain both how genomic integrity is maintained in germinal-centre B cells and why SHM is occasionally mistargeted to oncogenes.

1. Goossens, T., Klein, U. & Kuppers, R. Frequent occurrence of deletions and duplications during somatic hypermutation: implications for oncogene translocations and heavy chain disease. *Proc. Natl Acad. Sci. USA* **95**, 2463–2468 (1998).
2. Rogozin, I. B. & 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.* **172**, 3382–3384 (2004).
3. Storb, U. *et al. Cis*-acting sequences that affect somatic hypermutation of Ig genes. *Immunol. Rev.* **162**, 153–160 (1998).
4. Fukita, Y., Jacobs, H. & Rajewsky, K. Somatic hypermutation in the heavy chain locus correlates with transcription. *Immunity* **9**, 105–114 (1998).
5. Bachl, J., Carlson, C., Gray-Schopfer, V., Dessing, M. & Olsson, C. Increased transcription levels induce higher mutation rates in a hypermutating cell line. *J. Immunol.* **166**, 5051–5057 (2001).
6. Rada, C. & Milstein, C. The intrinsic hypermutability of antibody heavy and light chain genes decays exponentially. *EMBO J.* **20**, 4570–4576 (2001).
7. Peters, A. & Storb, U. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity* **4**, 57–65 (1996).
Duplication of the Igκ promoter immediately upstream of the Cκ region targeted SHM to this region in the context of an Igκ transgene. In light of this finding, a transcription-coupled model for SHM was first proposed.
8. Shen, H. M., Ratnam, S. & Storb, U. Targeting of the activation-induced cytosine deaminase is strongly influenced by the sequence and structure of the targeted DNA. *Mol. Cell. Biol.* **25**, 10815–10821 (2005).
9. Longerich, S., Tanaka, A., Bozek, G., Nicolae, D. & Storb, U. The very 5′ end and the constant region of Ig genes are spared from somatic mutation because AID does not access these regions. *J. Exp. Med.* **202**, 1443–1454 (2005).
10. Muramatsu, M. *et al.* Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553–563 (2000).
11. Revy, P. *et al.* Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the hyper-IgM syndrome (HIGM2). *Cell* **102**, 565–575 (2000).
12. Arakawa, H., Hauschild, J. & Buerstedde, J. M. Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. *Science* **295**, 1301–1306 (2002).
13. Harris, R. S., Sale, J. E., Petersen-Mahrt, S. K. & Neuberger, M. S. AID is essential for immunoglobulin V gene conversion in a cultured B cell line. *Curr. Biol.* **12**, 435–438 (2002).
14. Okazaki, I. M. *et al.* Constitutive expression of AID leads to tumorigenesis. *J. Exp. Med.* **197**, 1173–1181 (2003).
Transgenic mice designed to express AID ubiquitously develop T-cell lymphomas that have mutations in the genes encoding the T-cell receptor and MYC. Therefore, dysregulated expression of AID can lead to malignancy *in vivo*.
15. Petersen-Mahrt, S. K., Harris, R. S. & Neuberger, M. S. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature* **418**, 99–103 (2002).
16. Yoshikawa, K. *et al.* AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. *Science* **296**, 2033–2036 (2002).
17. Shinkura, R. *et al.* Separate domains of AID are required for somatic hypermutation and class-switch recombination. *Nature Immunol.* **5**, 707–712 (2004).
18. Ta, V. T. *et al.* AID mutant analyses indicate requirement for class-switch-specific cofactors. *Nature Immunol.* **4**, 843–848 (2003).
19. Barreto, V., Reina-San-Martin, B., Ramiro, A. R., McBride, K. M. & Nussenzweig, M. C. C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. *Mol. Cell* **12**, 501–508 (2003).
20. Ito, S. *et al.* Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc. Natl Acad. Sci. USA* **101**, 1975–1980 (2004).
21. Brar, S. S., Watson, M. & Diaz, M. Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks. *J. Biol. Chem.* **279**, 26395–26401 (2004).
22. McBride, K. M., Barreto, V., Ramiro, A. R., Stavropoulos, P. & Nussenzweig, M. C. Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. *J. Exp. Med.* **199**, 1235–1244 (2004).
23. Chaudhuri, J., Khuong, C. & Alt, F. W. Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature* **430**, 992–998 (2004).
24. Pasqualucci, L., Kitaura, Y., Gu, H. & Dalla-Favera, R. PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells. *Proc. Natl Acad. Sci. USA* **103**, 395–400 (2006).
25. Basu, U. *et al.* The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature* **438**, 508–511 (2005).
26. Muramatsu, M. *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.* **274**, 18470–18476 (1999).
27. Chaudhuri, J. *et al.* Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* **422**, 726–730 (2003).
This elegant study shows that transcription can target AID activity to dsDNA by generating the ssDNA substrate preferred by this enzyme, providing a mechanistic link between transcription and SHM.
28. Pham, P., Bransteitter, R., Petruska, J. & Goodman, M. F. Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature* **424**, 103–107 (2003).
This group shows that AID targets WRC motifs in ssDNA, introduces multiple mutations in one molecule of DNA (and thereby seems to function processively), and prefers to act on the non-transcribed strand of DNA that is exposed during transcription. Together, these findings provide a biochemical basis for several key features of SHM.
29. Bransteitter, R., Pham, P., Scharff, M. D. & Goodman, M. F. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl Acad. Sci. USA* **100**, 4102–4107 (2003).
30. Dickerson, S. K., Market, E., Besmer, E. & Papavasiliou, F. N. AID mediates hypermutation by deaminating single stranded DNA. *J. Exp. Med.* **197**, 1291–1296 (2003).
31. Yu, K., Huang, F. T. & Lieber, M. R. DNA substrate length and surrounding sequence affect the activation-induced deaminase activity at cytidine. *J. Biol. Chem.* **279**, 6496–6500 (2004).
32. Ramiro, A. R., Stavropoulos, P., Jankovic, M. & Nussenzweig, M. C. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nature Immunol.* **4**, 452–456 (2003).
33. Sohail, A., Klapacz, J., Samaranayake, M., Ullah, A. & Bhagwat, A. S. Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. *Nucleic Acids Res.* **31**, 2990–2994 (2003).
34. Shen, H. M. & Storb, U. Activation-induced cytidine deaminase (AID) can target both DNA strands when the DNA is supercoiled. *Proc. Natl Acad. Sci. USA* **101**, 12997–13002 (2004).
35. Neuberger, M. S., Harris, R. S., Di Noia, J. & Petersen-Mahrt, S. K. Immunity through DNA deamination. *Trends Biochem. Sci.* **28**, 305–312 (2003).
36. Wilson, T. M. *et al.* MSH2–MSH6 stimulates DNA polymerase η, suggesting a role for A:T mutations in antibody genes. *J. Exp. Med.* **201**, 637–645 (2005).
37. Krokan, H. E., Drablos, F. & Slupphaug, G. Uracil in DNA--occurrence, consequences and repair. *Oncogene* **21**, 8935–8948 (2002).
38. Reynaud, C. A., Aoufouchi, S., Faili, A. & Weill, J. C. What role for AID: mutator, or assembler of the immunoglobulin mutasome? *Nature Immunol.* **4**, 631–638 (2003).
39. Larson, E. D., Cummings, W. J., Bednarski, D. W. & Maizels, N. MRE11/RAD50 cleaves DNA in the AID/UNG-dependent pathway of immunoglobulin gene diversification. *Mol. Cell* **20**, 367–375 (2005).
40. Rada, C., Jarvis, J. M. & Milstein, C. AID-GFP chimeric protein increases hypermutation of Ig genes with no evidence of nuclear localization. *Proc. Natl Acad. Sci. USA* **99**, 7003–7008 (2002).
41. Cattoretti, G. *et al.* Nuclear and cytoplasmic AID in extrafollicular and germinal center B cells. *Blood* **107**, 3967–3975 (2006).
42. Longerich, S., Basu, U., Alt, F. & Storb, U. AID in somatic hypermutation and class switch recombination. *Curr. Opin. Immunol.* **18**, 164–174 (2006).
43. Chaudhuri, J. & Alt, F. W. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nature Rev. Immunol.* **4**, 541–552 (2004).
44. Papavasiliou, F. N. & Schatz, D. G. Cell-cycle-regulated DNA double-stranded breaks in somatic hypermutation of immunoglobulin genes. *Nature* **408**, 216–221 (2000).
45. Bross, L. *et al.* DNA double-strand breaks in immunoglobulin genes undergoing somatic hypermutation. *Immunity* **13**, 589–597 (2000).
46. Papavasiliou, F. N. & Schatz, D. G. The activation-induced deaminase functions in a postcleavage step of the somatic hypermutation process. *J. Exp. Med.* **195**, 1193–1198 (2002).
47. Bross, L., Muramatsu, M., Kinoshita, K., Honjo, T. & Jacobs, H. DNA double-strand breaks: prior to but not sufficient in targeting hypermutation. *J. Exp. Med.* **195**, 1187–1192 (2002).
48. Zan, H., Wu, X., Komori, A., Holloman, W. K. & Casali, P. AID-dependent generation of resected double-strand DNA breaks and recruitment of Rad52/Rad51 in somatic hypermutation. *Immunity* **18**, 727–738 (2003).
49. Haber, J. E. Hypermutation: give us a break. *Nature Immunol.* **2**, 902–903 (2001).
50. Kong, Q. & Maizels, N. DNA breaks in hypermutating immunoglobulin genes: evidence for a break-and-repair pathway of somatic hypermutation. *Genetics* **158**, 369–378 (2001).
51. Redon, C. *et al.* Histone H2A variants H2AX and H2AZ. *Curr. Opin. Genet. Dev.* **12**, 162–169 (2002).
52. Reina-San-Martin, B. *et al.* H2AX is required for recombination between immunoglobulin switch regions but not for intra-switch region recombination or somatic hypermutation. *J. Exp. Med.* **197**, 1767–1778 (2003).
53. Odegard, V. H., Kim, S. T., Anderson, S. M., Shlomchik, M. J. & Schatz, D. G. Histone modifications associated with somatic hypermutation. *Immunity* **23**, 101–110 (2005).
54. Woo, C. J., Martin, A. & Scharff, M. D. Induction of somatic hypermutation is associated with modifications in immunoglobulin variable region chromatin. *Immunity* **19**, 479–489 (2003).
55. Rada, C., Di Noia, J. M. & Neuberger, M. S. 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 (2004).
56. Di Noia, J. & Neuberger, M. S. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* **419**, 43–48 (2002).
57. Schrader, C. E., Linehan, E. K., Mochegova, S. N., Woodland, R. T. & Stavnezer, J. Inducible DNA breaks in Ig S regions are dependent on AID and UNG. *J. Exp. Med.* **202**, 561–568 (2005).
58. Pasqualucci, 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* **95**, 11816–11821 (1998).
59. Shen, H. M., 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* **280**, 1750–1752 (1998).
This paper marks the first time that a gene other than the immunoglobulin genes was shown to be a target of SHM in normal B cells.
60. Muschen, M. *et al.* Somatic mutation of the *CD95* gene in human B cells as a side-effect of the germinal center reaction. *J. Exp. Med.* **192**, 1833–1840 (2000).
61. Pasqualucci, L. *et al.* Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* **412**, 341–346 (2001).
Multiple proto-oncogenes were found to contain SHM-like mutations in diffuse large B-cell lymphomas, pinpointing a possible origin for genomic instability during certain types of B-cell tumorigenesis.
62. Gordon, M. S., Kanegai, C. M., Doerr, J. R. & Wall, R. Somatic hypermutation of the B cell receptor genes *B29* (*Igβ*, *CD79b*) and *mb1* (*Igα*, *CD79a*). *Proc. Natl Acad. Sci. USA* **100**, 4126–4131 (2003).
63. Muto, T. *et al.* Negative regulation of activation-induced cytidine deaminase in B cells. *Proc. Natl Acad. Sci. USA* **103**, 2752–2757 (2006).

64. Hori, M., Qi, C. F., Torrey, T. A., Huppi, K. & Morse, H. C.. The *Bcl6* locus is not mutated in mouse B-cell lineage lymphomas. *Leuk. Res.* **26**, 739–743 (2002).
65. Neuberger, M. S. *et al.* Monitoring and interpreting the intrinsic features of somatic hypermutation. *Immunol. Rev.* **162**, 107–116 (1998).
66. Ramiro, A. R. *et al.* Role of genomic instability and p53 in AID-induced *c-myc–Igh* translocations. *Nature* **440**, 105–109 (2006).
67. Ramiro, A. R. *et al.* AID is required for *c-myc/IgH* chromosome translocations *in vivo*. *Cell* **118**, 431–438 (2004).
68. Franco, S. *et al.* H2AX prevents DNA breaks from progressing to chromosome breaks and translocations. *Mol. Cell* **21**, 201–214 (2006).
69. Shen, H. M., Michael, N., Kim, N. & Storb, U. The TATA binding protein, *c-Myc* and *survivin* genes are not somatically hypermutated, while *Ig* and *BCL6* genes are hypermutated in human memory B cells. *Int. Immunol.* **12**, 1085–1093 (2000).
70. Hesslein, D. G. & Schatz, D. G. Factors and forces controlling V(D)J recombination. *Adv. Immunol.* **78**, 169–232 (2001).
71. O'Brien, R. L., Brinster, R. L. & Storb, U. Somatic hypermutation of an immunoglobulin transgene in κ transgenic mice. *Nature* **326**, 405–409 (1987).
72. Sharpe, M. J., Milstein, C., Jarvis, J. M. & Neuberger, M. S. Somatic hypermutation of immunoglobulin κ may depend on sequences 3′ of Cκ and occurs on passenger transgenes. *EMBO J.* **10**, 2139–2145 (1991).
73. Klotz, E. L. & Storb, U. Somatic hypermutation of a λ 2 transgene under the control of the λ enhancer or the heavy chain intron enhancer. *J. Immunol.* **157**, 4458–4463 (1996).
74. Kong, Q., Zhao, L., Subbaiah, S. & Maizels, N. A λ 3′ enhancer drives active and untemplated somatic hypermutation of a λ 1 transgene. *J. Immunol.* **161**, 294–301 (1998).
75. Giusti, A. M. & Manser, T. Hypermutation is observed only in antibody H chain V region transgenes that have recombined with endogenous immunoglobulin H DNA: implications for the location of *cis*-acting elements required for somatic mutation. *J. Exp. Med.* **177**, 797–809 (1993).
76. Tumas-Brundage, K. & Manser, T. The transcriptional promoter regulates hypermutation of the antibody heavy chain locus. *J. Exp. Med.* **185**, 239–250 (1997).
77. Hengstschlager, M., Williams, M. & Maizels, N. A λ 1 transgene under the control of a heavy chain promoter and enhancer does not undergo somatic hypermutation. *Eur. J. Immunol.* **24**, 1649–1656 (1994).
78. Sohn, J., Gerstein, R. M., Hsieh, C. L., Lemer, M. & Selsing, E. Somatic hypermutation of an immunoglobulin μ heavy chain transgene. *J. Exp. Med.* **177**, 493–504 (1993).
79. Yelamos, J. *et al.* Targeting of non-Ig sequences in place of the V segment by somatic hypermutation. *Nature* **376**, 225–229 (1995).
This group shows that the rearranged variable region is not required for targeting SHM to an Igκ transgene. This study initiated the search for *cis*-acting elements that target SHM in transgenic mice.
80. Michael, N. *et al.* Effects of sequence and structure on the hypermutability of immunoglobulin genes. *Immunity* **16**, 123–134 (2002).
81. Betz, A. G. *et al.* Elements regulating somatic hypermutation of an immunoglobulin κ gene: critical role for the intron enhancer/matrix attachment region. Cell **77**, 239–248 (1994).
This pivotal study shows that the Igκ intronic enhancer is required for targeting SHM to an Igκ transgene and that this activity is separate from the enhancer's role in transcription. In addition, the Igκ 3′ enhancer is also required for SHM of the transgene, but this function is most probably due to this enhancer's role in stimulating transcription. Together, these data lead to a model in which the SHM machinery is specifically recruited by immunoglobulin enhancers to immunoglobulin loci.
82. Queen, C. & Stafford, J. Fine mapping of an immunoglobulin gene activator. *Mol. Cell. Biol.* **4**, 1042–1049 (1984).
83. Meyer, K. B. & Neuberger, M. S. The immunoglobulin κ locus contains a second, stronger B-cell-specific enhancer which is located downstream of the constant region. *EMBO J.* **8**, 1959–1964 (1989).
84. Goyenechea, B. *et al.* Cells strongly expressing Igκ transgenes show clonal recruitment of hypermutation: a role for both MAR and the enhancers. *EMBO J.* **16**, 3987–3994 (1997).
85. Klix, N. *et al.* Multiple sequences from downstream of the Jκ cluster can combine to recruit somatic hypermutation to a heterologous, upstream mutation domain. *Eur. J. Immunol.* **28**, 317–326 (1998).
86. Jolly, C. J. & Neuberger, M. S. Somatic hypermutation of immunoglobulin κ transgenes: association of mutability with demethylation. *Immunol. Cell. Biol.* **79**, 18–22 (2001).
87. Michael, N. *et al.* The E box motif CAGGTG enhances somatic hypermutation without enhancing transcription. *Immunity* **19**, 235–242 (2003).
88. Gorman, J. R. *et al.* The Igκ enhancer influences the ratio of Igκ versus Igλ B lymphocytes. *Immunity* **5**, 241–252 (1996).
89. van der Stoep, N., Gorman, J. R. & Alt, F. W. Reevaluation of 3′E_κ function in stage- and lineage-specific rearrangement and somatic hypermutation. *Immunity* **8**, 743–750 (1998).
Using mice lacking the Igκ 3′ enhancer, this study shows that the Igκ 3′ enhancer is not essential for transcription or mutation of the endogenous Igκ locus; these results are different from those obtained from Igκ transgenes.
90. Inlay, M. A. *et al.* Roles of the immunoglobulin κ light chain intronic and 3′ enhancers in *Igk* somatic hypermutation. *J. Immunol.* **177**, 1146–1151 (2006).
91. Ronai, D., Iglesias-Ussel, M. D., Fan, M., Shulman, M. J. & Scharff, M. D. Complex regulation of somatic hypermutation by *cis*-acting sequences in the endogenous *IgH* gene in hybridoma cells. *Proc. Natl Acad. Sci. USA* **102**, 11829–11834 (2005).
92. Perlot, T., Alt, F. W., Bassing, C. H., Suh, H. & Pinaud, E. Elucidation of *IgH* intronic enhancer functions via germ-line deletion. *Proc. Natl Acad. Sci. USA* **102**, 14362–14367 (2005).
93. Tumas-Brundage, K. M., Vora, K. A. & Manser, T. Evaluation of the role of the 3′α heavy chain enhancer [3′α E(hs1, 2)] in Vh gene somatic hypermutation. *Mol. Immunol.* **34**, 367–378 (1997).
94. Liu, Z. M. *et al.* Chromatin structural analyses of the mouse Igκ gene locus reveal new hypersensitive sites specifying a transcriptional silencer and enhancer. *J. Biol. Chem.* **277**, 32640–32649 (2002).
95. Schlissel, M. S. Regulation of activation and recombination of the murine *Igκ* locus. *Immunol. Rev.* **200**, 215–223 (2004).
96. Staudt, L. M. & Lenardo, M. J. Immunoglobulin gene transcription. *Annu. Rev. Immunol.* **9**, 373–398 (1991).
97. Henderson, A. & Calame, K. Transcriptional regulation during B cell development. *Annu. Rev. Immunol.* **16**, 163–200 (1998).
98. Henderson, A. J. & Calame, K. L. Lessons in transcriptional regulation learned from studies on immunoglobulin genes. *Crit. Rev. Eukaryot. Gene Expr.* **5**, 255–280 (1995).
99. Bain, G. & Murre, C. The role of E-proteins in B- and T-lymphocyte development. *Semin. Immunol.* **10**, 143–153 (1998).
100. Kotani, A. *et al.* A target selection of somatic hypermutations is regulated similarly between T and B cells upon activation-induced cytidine deaminase expression. *Proc. Nat. Acad. Sci. USA* **102**, 4506–4511 (2005).
101. Conlon, T. M. & Meyer, K. B. The chicken Ig light chain 3′-enhancer is essential for gene expression and regulates gene conversion via the transcription factor E2A. *Eur. J. Immunol.* **36**, 139–148 (2006).
102. Kurdistani, S. K. & Grunstein, M. Histone acetylation and deacetylation in yeast. *Nature Rev. Mol. Cell Biol.* **4**, 276–284 (2003).
103. Nambu, Y. *et al.* Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science* **302**, 2137–2140 (2003).
104. Li, Z., Luo, Z. & Scharff, M. D. Differential regulation of histone acetylation and generation of mutations in switch regions is associated with Ig class switching. *Proc. Natl Acad. Sci. USA* **101**, 15428–15433 (2004).
105. Kuo, M. H., vom Baur, E., Struhl, K. & Allis, C. D. Gcn4 activator targets Gcn5 histone acetyltransferase to specific promoters independently of transcription. *Mol. Cell* **6**, 1309–1320 (2000).
106. Liang, G. *et al.* Distinct localization of histone H3 acetylation and H3–K4 methylation to the transcription start sites in the human genome. *Proc. Natl Acad. Sci. USA* **101**, 7357–7362 (2004).
107. Cheung, W. L. *et al.* Apoptotic phosphorylation of histone H2B is mediated by mammalian sterile twenty kinase. *Cell* **113**, 507–517 (2003).
108. Ahn, S. H. *et al.* Sterile 20 kinase phosphorylates histone H2B at serine 10 during hydrogen peroxide-induced apoptosis in *S. cerevisiae*. *Cell* **120**, 25–36 (2005).
109. Fernandez-Capetillo, O., Allis, C. D. & Nussenzweig, A. Phosphorylation of histone H2B at DNA double-strand breaks. *J. Exp. Med.* **199**, 1671–1677 (2004).
110. Downs, J. A. *et al.* Binding of chromatin-modifying activities to phosphorylated histone H2A at DNA damage sites. *Mol. Cell* **16**, 979–990 (2004).
111. Downs, J. A., Lowndes, N. F. & Jackson, S. P. A role for *Saccharomyces cerevisiae* histone H2A in DNA repair. *Nature* **408**, 1001–1004 (2000).
112. Wang, C. L., Harper, R. A. & Wabl, M. Genome-wide somatic hypermutation. *Proc. Natl Acad. Sci. USA* **101**, 7352–7356 (2004).
113. Parsa, J. Y. *et al.* AID mutates a non-immunoglobulin transgene independent of chromosomal position. *Mol. Immunol.* 14 March 2006 (doi:10.1016/j.molimm.2006.02.003).
114. Dorner, T. *et al.* Analysis of the frequency and pattern of somatic mutations within nonproductively rearranged human variable heavy chain genes. *J. Immunol.* **158**, 2779–2789 (1997).
115. Delpy, L., Sirac, C., Le Morvan, C. & Cogne, M. Transcription-dependent somatic hypermutation occurs at similar levels on functional and nonfunctional rearranged IgH alleles. *J. Immunol.* **173**, 1842–1848 (2004).
116. Dunn-Walters, D. K., Dogan, A., Boursier, L., MacDonald, C. M. & Spencer, J. Base-specific sequences that bias somatic hypermutation deduced by analysis of out-of-frame human IgVH genes. *J. Immunol.* **160**, 2360–2364 (1998).
117. Kosak, S. T. *et al.* Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* **296**, 158–162 (2002).
118. Spilianakis, C. G., Lalioti, M. D., Town, T., Lee, G. R. & Flavell, R. A. Interchromosomal associations between alternatively expressed loci. *Nature* **435**, 637–645 (2005).
119. Johnston, J. M. *et al.* Analysis of hypermutation in immunoglobulin heavy chain passenger transgenes. *Eur. J. Immunol.* **26**, 1058–1062 (1996).
120. Terauchi, A. *et al.* A pivotal role for DNase I-sensitive regions 3b and/or 4 in the induction of somatic hypermutation of IgH genes. *J. Immunol.* **167**, 811–820 (2001).
121. Morvan, C. L. *et al.* The immunoglobulin heavy-chain locus *hs3b* and *hs4* 3′ enhancers are dispensable for VDJ assembly and somatic hypermutation. *Blood* **102**, 1421–1427 (2003).

Competing interests statement
The authors declare no competing financial interests.

DATABASES
The following terms in this article are linked online to:
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
AID | APOBEC1 | *BCL6* | *CD79B* | *CD95* | E2A | H2AX | IgH | Igκ | Igλ | NBS1 | *PAX5* | *PIM1* | RAD50
Access to this links box is available online.