nature immunology

Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes

Arito Yamane[1,5], Wolfgang Resch[1,5], Nan Kuo[1], Stefan Kuchen[1], Zhiyu Li[1], Hong-wei Sun[2], Davide F Robbiani[3], Kevin McBride[3], Michel C Nussenzweig[3] & Rafael Casellas[1,4]

© 2011 Nature America, Inc. All rights reserved.

The cytidine deaminase AID hypermutates immunoglobulin genes but can also target oncogenes, leading to tumorigenesis. The extent of AID's promiscuity and its predilection for immunoglobulin genes are unknown. We report here that AID interacted broadly with promoter-proximal sequences associated with stalled polymerases and chromatin-activating marks. In contrast, genomic occupancy of replication protein A (RPA), an AID cofactor, was restricted to immunoglobulin genes. The recruitment of RPA to the immunoglobulin loci was facilitated by phosphorylation of AID at Ser38 and Thr140. We propose that stalled polymerases recruit AID, thereby resulting in low frequencies of hypermutation across the B cell genome. Efficient hypermutation and switch recombination required AID phosphorylation and correlated with recruitment of RPA. Our findings provide a rationale for the oncogenic role of AID in B cell malignancy.

Somatic hypermutation (SHM) introduces non-templated point mutations in genes encoding immunoglobulin variable (V) domains at a frequency of about one mutation per kilobase (kb) per cell per generation. In germinal centers, this activity generates antibody variants that are selected on the basis of their affinity for antigen. In addition to altering antibodies via hypermutation, the germinal center response also shapes the effector function of antibody genes through class-switch recombination (CSR). This reaction replaces the μ-chain immunoglobulin heavy-chain locus constant (C) region (*Igh-C*) for one of a set of downstream *Igh-C* exons[1,2].

AID initiates both CSR and SHM[3,4] by deaminating cytidine residues in single-stranded DNA (ssDNA), which is exposed during immunoglobulin transcription by RNA polymerase II (PolII)[5,6]. The resulting U:G mismatches are processed by base excision–repair and mismatch-repair pathways that produce mutations or double-strand DNA breaks, which are obligate intermediates for CSR[5,7]. During SHM, AID seems to act mainly in a region 1.5 kb downstream of the transcription start sites (TSSs) of genes encoding immunoglobulin V domains[8], whereas CSR-related double-strand DNA breaks map to switch (S) regions 1–12 kb in length that precede the participating *Igh-C* exons.

AID can also deaminate non-immunoglobulin genes, including *CD79A*, *CD79B*, *MYC*, *RHOH*, *PIM1*, *PAX5*, *BCL6* and *MIR142* and their mouse homologs[9–13]. Mutations occur at a lower frequency at such off-target sites than at immunoglobulin genes, in part because the initial lesions are repaired with high accuracy by physiological base excision–repair and mismatch-repair pathways[12]. Nevertheless, lesions in off-target sites for AID can cause deleterious mutations and large-scale chromosomal abnormalities that result in B cell lymphomas[14]. AID activity has also been linked to blast crisis progression in chronic myeloid leukemia[15], prostate malignancies[16] and gastric tumors[17].

SHM outside the immunoglobulin loci may not be entirely pathological. AID is expressed in pluripotent tissues and has been shown to deaminate 5-methylcytosines *in vitro*[18], which has led to the suggestion that it might mediate DNA demethylation in vertebrates[19–21]. Promiscuous AID activity, therefore, may promote developmental reprogramming in the embryo and potentially in activated B lymphocytes.

Several AID cofactors have been identified so far, including the ssDNA-binding protein RPA and protein kinase Ar1α[22–26]. RPA is of particular interest because of its established role in DNA recombination and repair[27]. In biochemical assays, RPA promotes the deamination of transcribed substrates by AID by stabilizing its interaction with ssDNA[22], which suggests that the role of RPA in SHM and CSR is to provide access of AID to target DNA. *In vivo*, the formation of RPA-AID complexes is facilitated by phosphorylation of AID at Ser38 by protein kinase Ar1α[23,26,28], and AID, RPA and protein kinase Ar1α all associate with sites of switch recombination, as determined by chromatin immunoprecipitation (ChIP)[24,29].

Despite the importance of AID in shaping the antibody response and in promoting malignancy, there is little understanding of how immunoglobulin genes are preferentially hypermutated, the extent

[1]Genomics & Immunity, National Institute of Arthritis, Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland, USA. [2]Biodata Mining and Discovery, National Institute of Arthritis, Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland, USA. [3]Laboratory of Molecular Immunology, Howard Hughes Medical Institute, The Rockefeller University, New York, New York, USA. [4]Center of Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA. [5]These authors contributed equally to this work. Correspondence should be addressed to M.C.N. (nussen@mail.rockefeller.edu) or R.C. (casellar@mail.nih.gov).

Received 2 August; accepted 2 November; published online 28 November 2010; doi:10.1038/ni.1964

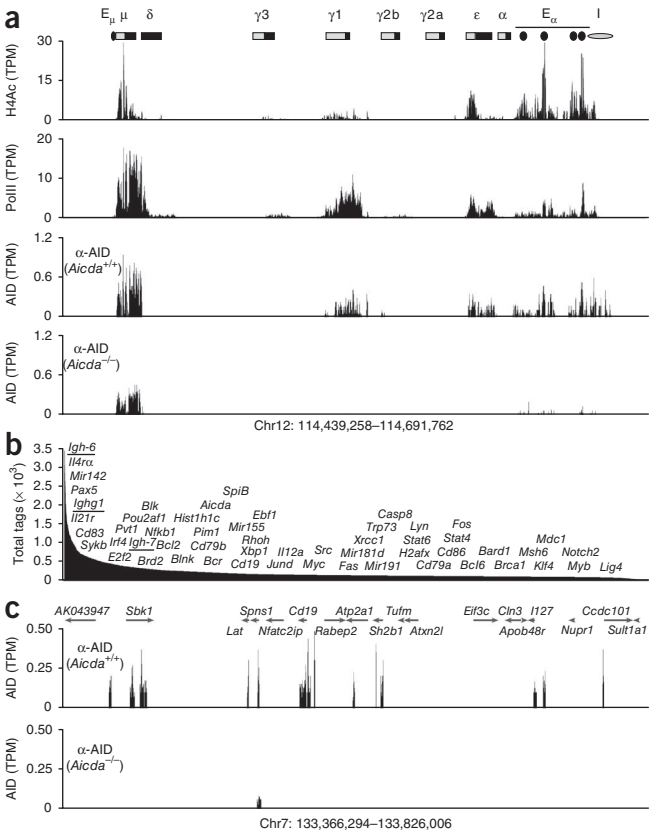

Figure 1 Extent of AID recruitment in activated B cells. (**a**) ChIP-seq analysis of the *Igh* locus (chromosome 12 (Chr12)) in B cells stimulated with LPS and IL-4, showing acetylation of histone H4 (H4Ac), as well as recruitment of PolII and AID; results are presented as sequence tags per million sequences (TPM) in 100-bp windows. Bottom row, specificity of AID immunoprecipitation, assessed in *Aicda*$^{-/-}$ cells with an AID-specific antibody (α-AID)[22]. Above, *Igh* locus, showing the S domains (light gray boxes) and C domains (black boxes) as well as the 5′ μ-chain enhancer (E_μ), 3′ α-chain enhancer (E_α) and insulator (I). (**b**) Total sequence tags of AID-recruiting genes (n = 5,910), showing the approximate positions of a subset of AID targets, including immunoglobulin genes (underlined). (**c**) Recruitment of AID at a domain of chromosome 7 ~500 kb in length containing sequence encoding the B cell–specific CD19 antigen, assessed in *Aicda*$^{+/+}$ and *Aicda*$^{-/-}$ (control) cells. Data are representative of four independent experiments.

of AID off-target activity or how AID finds its ssDNA substrate near TSSs. To address these issues, we have defined the genome-wide association of AID and RPA in the context of the activated B cell epigenome, transcriptome and PolII. We found enrichment for AID across the genome at pausing sites for PolII, with the greatest abundance at the immunoglobulin μ-chain gene (*Igh-6*). Thus, our data support and extend the observation that Spt5, a factor required for polymerase stalling and CSR, is required for the association of AID with the transcribing holoenzyme[30]. In contrast, however, RPA associated mainly with the immunoglobulin locus, and this interaction was dependent on phosphorylation AID at Ser38 and Thr140. We propose that interaction of AID with Spt5-PolII complexes results in deamination of DNA across the genome in a manner that is proportional to the amount of AID recruited. However, optimal SHM and CSR may also require the recruitment of additional AID cofactors.

RESULTS

AID ChIP-seq in activated B cells

Although AID seems to target many non-immunoglobulin genes[9,10,12,31], the extent of its off-target activity is not known. To define the genome-wide occupancy of AID in B cells, we stimulated AID-sufficient (*Aicda*$^{+/+}$) B cells and AID-deficient (*Aicda*$^{-/-}$) control B cells with lipopolysaccharide (LPS) and interleukin 4 (IL-4) and analyzed the cells by ChIP coupled with deep sequencing (ChIP-seq) using antibody to AID (anti-AID). These culture conditions induce AID expression and CSR to IgG1 (*Ighg1*) or IgE (*Igh-7*). As expected, AID was immunoprecipitated from the *Igh* μ-chain (*Igh-6*), γ1-chain (*Ighg1*) and ε-chain (*Igh-7*) (**Fig. 1a**), which are recombining, actively transcribed (as determined by the association of PolII with chromatin) and acetylated at histone H4 (**Fig. 1a**). Unexpectedly, outside the *Igh* locus, AID was also associated with as many as 5,910 genes (or 12,200 islands; **Fig. 1b** and **Supplementary Table 1**). Analogous to published ChIP-seq studies[32], 396 genes (6.7%) also showed background signals in *Aicda*$^{-/-}$ cells (**Supplementary Fig. 1a**), probably because of the presence of sonication-hypersensitive chromatin in these sites (such as *Spns1* (**Fig. 1c**) or *Igh-6* (**Fig. 1a**); **Supplementary Text**). Nevertheless, with few exceptions, AID islands were distinguishable from background noise ($P < 0.0005$ (Wilcoxon rank-sum test); **Supplementary Fig. 1b**), and saturation studies confirmed the specificity of the AID islands observed in *Aicda*$^{+/+}$ cells (**Supplementary Fig. 2a**).

Ranking of genes associated with AID based on the absolute number of sequence tags positioned *Igh-6* (encoding IgM) as the gene with the greatest recruitment of AID (**Fig. 1b**). This was followed by other previously characterized AID targets, including *Pax5*, *Il21r*, *Cd83*, *Sykb*, *Pou2af1*, *Pim1*, *Cd79a*, *Cd79b*, *Aicda*, *Ebf1*, *Myc* and *H2afx*, among others (complete list, **Supplementary Table 1**). Notably, *Mir142*, a *Myc* translocation partner in mouse and human B cell tumors[14], was among the 15 genes with the greatest recruitment of AID (**Supplementary Table 1**). AID was also immunoprecipitated from *Bcr* and *SpiB* (**Fig. 1b**), presumed AID targets associated with chronic myelogenous leukemia[15] and a subset of diffuse large B cell lymphomas[33], respectively. Among the previously unknown targets identified, we found genes encoding microRNAs expressed in activated lymphocytes[34] (*Mir155*, *Mir181d* and *Mir21*), transcription factors involved in cell differentiation (*Stat6*, *Xbp1* and *Nfkb1*) and DNA-repair proteins (*Brca1*, *Mdc1* and *Lig4*; **Fig. 1b** and **Supplementary Table 1**). We conclude that AID is recruited to a large number of genes in activated B cells.

AID recruitment indicates hypermutation

To confirm the ChIP-seq results for AID reported above, we obtained B cells from mice transgenic for expression of AID under the control of a promoter-enhancer cassette from the immunoglobulin κ-chain gene (*Igk*) and deficient in uracil DNA glycosylase (*Igk*-AID *Ung*$^{-/-}$ mice)[9], then stimulated the cells with LPS and IL-4 and measured hypermutation. On the basis of results from a sample of nearly two dozen genes, these B cell cultures have been shown to achieve hypermutation frequencies similar to those measured in germinal centers and to maintain target selection specificity[9,30]. In agreement with the published data, mutation frequencies at five documented AID targets (*Pim1*, *Myc*, *Pax5*, *Cd83* and *H2afx*) in *Igk*-AID *Ung*$^{-/-}$ cells were similar to those reported before for *Ung*$^{-/-}$ Peyer's patch B cells also deficient in mutS homolog 2 (*Msh2*$^{-/-}$)[12] (**Supplementary Table 2**). In comparison, *Igk*-AID *Ung*$^{-/-}$ cultures had four times more hypermutation and more transition mutations than did *Igk*-AID *Ung*$^{+/+}$ cultures (**Supplementary Fig. 3**).

To examine additional predicted AID target genes, we amplified ~750 base pairs (bp) of DNA downstream of TSSs of genes associated with AID (n = 13) or not associated with AID (n = 11) and analyzed

© 2011 Nature America, Inc. All rights reserved.

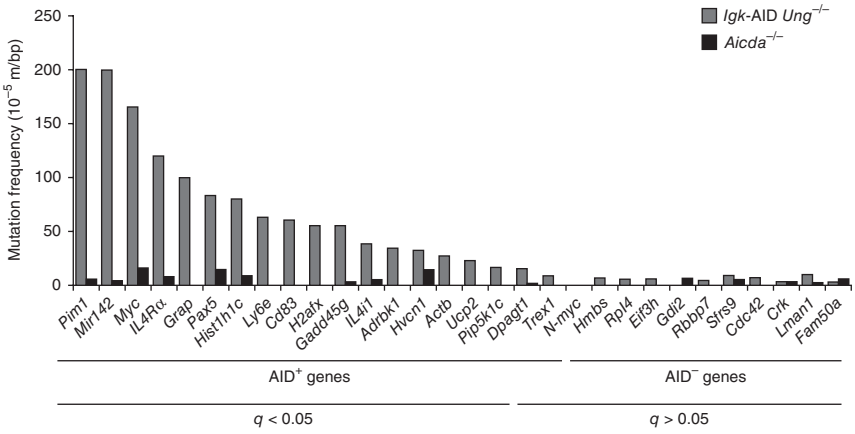

Figure 2 AID targets are somatically hypermutated. Mutation frequencies of genes that recruited AID (AID$^+$; $n = 9$) or did not recruit AID (AID$^-$; $n = 11$) in activated *Igk*-AID *Ung*$^{-/-}$ B cells and *Aicda*$^{-/-}$ B cells. Bottom, adjusted *P* values (*q*), calculated to control for false-discovery rate (Fisher's exact test). Data are representative of one experiment.

the sequences for point mutations. We analyzed a total of 1,496,058 bp. Most of the genes that recruited AID had significantly more mutations in *Igk*-AID *Ung*$^{-/-}$ B cells than in the background control *Aicda*$^{-/-}$ B cells (**Fig. 2** and **Supplementary Table 2**). Among the genes with the most mutations (beyond the known targets of AID) were *Il4ra*, *Grap*, *Hist1h1c*, *Ly6e*, *Gadd45g* and *Il4i1* (**Fig. 2** and **Supplementary Table 2**). Notably, more than 86% of all mutations were C:G to T:A transitions and showed a strong bias toward RGYW-WRCY hot spots (**Supplementary Table 2**), which confirmed that they were indeed the result of AID deamination activity in the absence of *Ung*. In contrast, the mutation frequency of genes associated with AID that were highly transcribed was not distinguishable from background ($q > 0.05$; **Fig. 2** and **Supplementary Table 2**). These results support the view that ChIP-seq for AID can be used to reliably predict the presence of AID activity.

AID preferentially targets genes in open chromatin

Transcription is required for AID activity *in vitro* and *in vivo*[1,5,6,35]. However, the relationship between transcription and AID targeting has been difficult to define because only a small number of genes have been assayed and this has been done in a limited way. To address this issue comprehensively, we used mRNA deep sequencing and ChIP-seq to compare AID, transcription and epigenetic marks on a genome-wide basis.

The genes that recruited AID ($n = 5{,}910$) had a median mRNA abundance 40 times greater than that of genes that did not recruit AID ($n = 12{,}775$; **Fig. 3a** and **Supplementary Table 1**). Consistent with that result, we detected PolII and trimethylated histone H3 Lys4 (H3K4me3), which are associated with gene activation[36,37], in essentially all (>95%) genes associated with AID, compared with <50% of genes not associated with AID (**Fig. 3b**). Conversely, genes that recruited AID showed depletion of the polycomb group–inhibitory mark H3K27me3 (<12%), whereas it was associated with nearly 40% of those that did not recruit AID (**Fig. 3b**). These features were particularly evident at the *Myc* locus and *Mycn* locus (encoding N-Myc). *Myc* is expressed and preferentially translocated to *Igh* in mature B cells, whereas *Mycn* is not. Consistent with those observations, H3K4me3 marks, the presence of PolII and mRNA synthesis correlated with substantial association of AID at the *Myc* locus (**Fig. 3c**), whereas *Mycn*, which is not expressed in activated B cells, showed silencing by H3K27me3, had little H3K4me3 and lacked

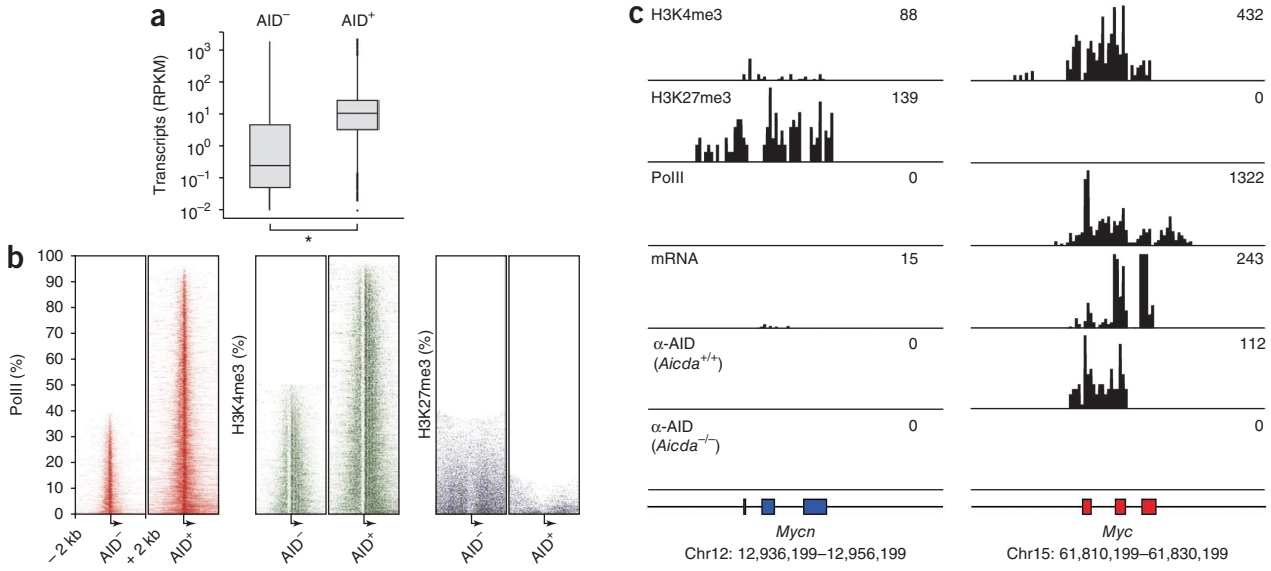

Figure 3 AID recruitment is biased toward actively transcribed genes associated with an open chromatin configuration. (**a**) Deep-sequencing analysis of mRNA for genes that recruited or did not recruit AID. Transcript abundance is presented as total mRNA sequences per gene length (in kb) per million aligned reads (RPKM), including the median (small horizontal lines), quartiles (vertical lines) and outliers (bolding). *$P < 0.0005$ (Wilcoxon rank-sum test). (**b**) Heat maps of ChIP-seq results for PolII, H3K4me3 and H3K27me3 at all (100%) genes that recruited or did not recruit AID, rank-ordered from least (top) to most (bottom) in terms of their association with PolII or chromatin modifications and centered on the TSS (arrow) of each gene ± 2 kb. $P < 0.0005$ (χ^2 test). (**c**) Aligned 'read' density for H3K4me3, H3K27me3, PolII, mRNA and AID at *Mycn* and *Myc* loci; numbers in plots (top right) indicate total sequenced tags aligned in each genomic domain. Data are representative of one experiment.

© 2011 Nature America, Inc. All rights reserved.

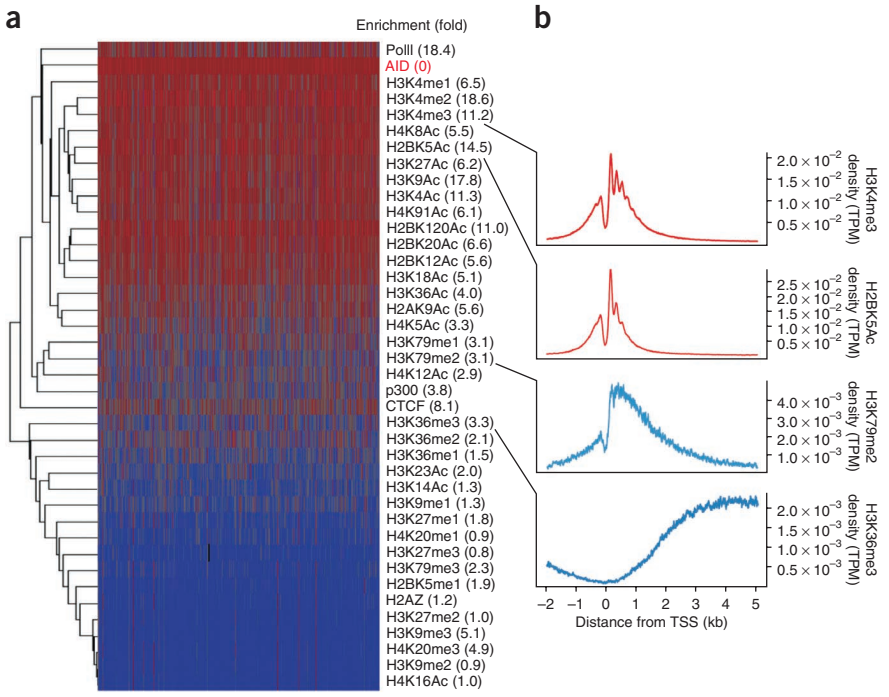

Figure 4 Epigenetic signature of AID recruitment. (**a**) Hierarchical clustering of AID, PolII, CTCF, p300 and 36 chromatin modifications on AID islands, presented as $\log_2$-transformed data and normalized Euclidean distances: red, high density; blue, low density. Numbers in parenthesis (right) indicate enrichment distance for each variable relative to that of AID. (**b**) Composite (metagene) profiles of ChIP-seq for H3K4me3, H2BK5Ac, H3K79me2 and H3K36me3, presented as tags per million aligned sequences per gene per nucleotide (density) versus position relative to the TSS (–2 kb to +5 kb). Data are representative of one experiment.

AID (**Fig. 3c**). Furthermore, we did not detect hypermutation at *Mycn*, whereas *Myc* intron 1 carried about 1×10^3 mutations per bp (**Supplementary Table 2**). Notably, the association of AID with *Myc* coincided with the mapped location of most AID-induced DNA breaks and canonical translocation breakpoints (at *Myc* exon 1–intron 1). Thus, we conclude that the interaction of AID with genes across the genome is biased toward genes associated with an open chromatin configuration. This feature provides a rationale for the greater incidence of *Myc* translocations than *Mycn* translocations in B lymphoid tumors.

AID recruitment follows PolII distribution

Despite the general correlation between mRNA accumulation and AID, our findings also indicated that transcription itself cannot be used to predict genome-wide recruitment of AID. Instead, a considerable fraction of genes with no detectable AID were abundantly transcribed (**Fig. 3a**, and **Supplementary Table 1**), including most of the genes assayed above that did not recruit AID (**Fig. 2** and **Supplementary Table 2**). In addition, there was a wide variation in mRNA accumulation among genes targeted by similar amounts of AID (**Supplementary Table 1**). To clarify the precise nature of the interaction of AID with genes across the B cell genome, we determined the overlap between AID islands and histone modifications (n = 36), the insulator protein CTCF, the enhancer-binding acetyltransferase p300 and PolII in activated B cells (650,005 islands analyzed in total). For each variable, we calculated both its normalized Euclidean distance (**Fig. 4a**, dendrogram, and **Supplementary Table 3**) and its enrichment (**Fig. 4a**, right margin) relative to those in a random background model. These analyses confirmed the overall link between AID and active chromatin. For example, in addition to H3K27me3 (**Fig. 3b**), the inhibitory marks H4K20me1, H3K9me2 and H3K27me2 were rarely associated with AID islands (0.77 ≤ enrichment (fold) ≤ 1.01, **Fig. 4a** and **Supplementary Table 3**). Furthermore, most histone acetylation modifications, which as a group promote transcriptional activation, were closely associated with AID (**Fig. 4a**). Notably, hierarchical clustering of the distance matrix also indicated a preference of AID for promoter-proximal sequences (**Supplementary Table 3**). Methylation and acetylation marks that typically demarcate promoters of active genes (H3K4me1-H3K4me2-H3K4me3, H2BK5Ac, H3K27Ac and H3K9Ac) overlapped physically better with AID (6.17 ≤ enrichment (fold) ≤ 18.6) than did those showing enrichment at transcribed regions (H3K36me3, H3K79me2, H3K14Ac, H4K5Ac and H4K12Ac (1.26 ≤ enrichment (fold) ≤ 3.35); **Fig. 4b** and **Supplementary Table 3**).

To confirm that AID has high occupancy of promoter regions, we generated composite profiles of AID and PolII around TSSs (from –2 kb to +5 kb relative to the TSS). On the basis of PolII-binding activity[38], we classified AID-recruiting genes as either stalled (stalling index > 3 for n = 4,756 genes (80%)) or elongating (1 ≤ stalling index ≤ 3 for n = 352 genes (6%)). The stalling index (also known as the traveling or pausing index) is a measure of the ratio of the density of PolII at the promoter (±1 kb relative to the TSS) versus its density in the gene body. Rather than indicating the presence or absence of transcription, the stalling index reflects the dynamics of PolII assembly and promoter clearance[38]. In stalled genes, for example, the rate of promoter clearance is lower than that of holoenzyme assembly; as a result, PolII disproportionately accumulates (pauses) at promoter-proximal sequences. In contrast, in elongating genes, the rates of promoter assembly and clearance are more equivalent and thus PolII is more easily detected in gene bodies by ChIP-seq analysis[38]. We found that AID density closely matched the overall PolII profiles. In elongating genes, the density of AID and PolII peaked at the TSS and decreased thereafter to background density 4–5 kb downstream (**Fig. 5a**). Two illustrative examples of this correlation were *Mir142* and *Cd79b*, for which AID density quantitatively mirrored, nearly peak by peak, the PolII-association profiles at the gene body and transcription termination sites (**Fig. 5b**). In stalled genes, both PolII and AID were immunoprecipitated mainly from a 2-kb area centered on the TSS that included the basal promoter (**Fig. 5c**). *Lyn* and *Atm* were both in this promoter-stalled category (**Fig. 5d**). On the basis of these findings we conclude that genome-wide AID occupancy mirrors PolII density. Furthermore, the overall bias of AID toward promoter-proximal sequences is explained by the fact that, as in embryonic stem cells[38], most genes in activated B lymphocytes accumulate paused polymerases at promoter areas. These conclusions were further supported by the observation that the overall AID-recruitment profiles correlated better with PolII phosphorylated at Ser5 than with PolII phosphorylated at Ser2 (**Supplementary Fig. 4**). Of note, the former has been shown to associate mainly with transcriptional initiation and pausing, whereas sites of elongation and transcriptional termination show enrichment for the latter[39].

© 2011 Nature America, Inc. All rights reserved.

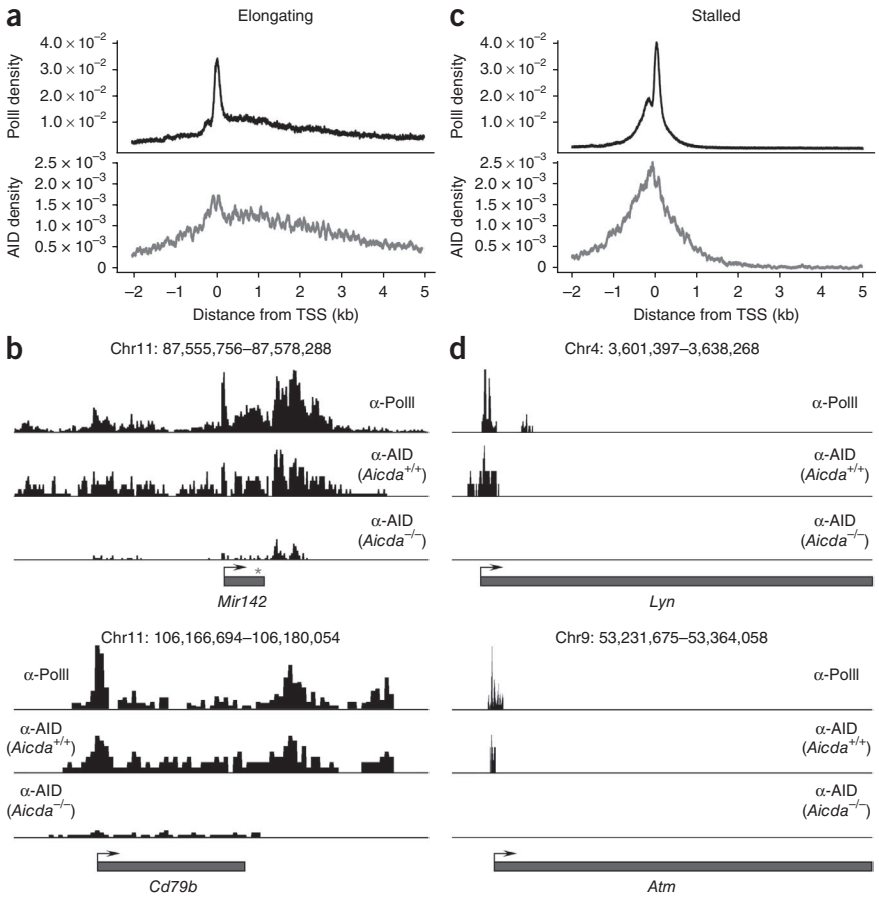

Figure 5 Genome-wide correlation between AID and PolII occupancy. (**a**) Composite profiles of the density of PolII and AID at elongating genes ($n = 352$) that recruit AID (presented as in **Fig. 4b**). (**b**) Quantitative correlation of PolII and AID at the *Mir142* and *Cd79b* loci. Bottom row, microsequencing background in *Aicda*$^{-/-}$ cells. (**c**) Composite profiles of the density of PolII and AID at stalled genes ($n = 4,756$) that recruit AID (presented as in **Fig. 4b**). (**d**) Quantitative correlation of PolII and AID at the *Lyn* and *Atm* loci (stalled genes). Data are representative of one experiment.

In addition to interacting with gene domains, a fraction of AID islands (34%) were intergenic, with three fourths of these localizing together with PolII, p300 and/or CTCF (**Supplementary Fig. 5a**). A particularly good example of this was the 3′ E_α enhancer, where both PolII and AID distinctly immunoprecipitated from each of the enhancer elements (**Fig. 1a**; other examples, **Supplementary Fig. 5b**). As a group, these sites may represent promoters of unknown genes, true intergenic AID-binding sites and/or regulatory elements at which AID is crosslinked as a result of long-range interactions with promoters.

Direct deamination of basal promoters by AID

In addition to producing conventional elongation of gene bodies, mammalian RNA polymerases also transcribe promoters in an orientation opposite that of the annotated gene[40]. Because of efficient polymerase stalling[38], the resulting antisense transcripts do not typically extend very far beyond the basal promoter[40]. In agreement with such studies, analysis of deep-sequenced RNA isolated from activated B cells[34] showed that PolII-stalling profiles at promoters of B cell genes coincided precisely with the TSSs of both sense and antisense RNA (**Fig. 6a**,**b**).

Although our observations indicated that SHM is associated with PolII stalling, published studies have indicated that the mutation track is unidirectional, occurring mainly 3′ of TSSs[8]. To determine whether divergent PolII stalling also renders basal promoters susceptible to SHM, we analyzed six AID-recruiting genes (*Pax5*, *Myc*, *Grap*, *Ighg1*, *Pim1* and *Il4ra*) for point mutations 5′ of their respective TSSs. Notably, except for *Il4ra* and *Grap*, all promoters had mutation frequencies above background in an AID-dependent manner (**Fig. 6c** and **Supplementary Table 4**). Mutation frequencies, which ranged between 1.7×10^{-4} and 4.7×10^{-4}, were equivalent to those measured at gene bodies of other AID targets, such as *Pax5* (7.0×10^{-4}) or *H2afx* (5.5×10^{-4}; **Supplementary Table 2** and **Supplementary Fig. 6**). Consistent with the PolII and small RNA profiles at basal promoters (**Fig. 6a**,**b**), promoter mutations were abundant near the TSS and became increasingly rare at upstream sequences (**Fig. 6c**). Notably, the overall mutation spectrum, consisting almost entirely (>94%) of transition mutations, was biased toward RGYW-WRCY hotspots, showing the distinctive footprint of AID (**Supplementary Table 4**). Finally, we also found hypermutation of the *Pax5* basal promoter in Peyer's patches from *Ung*$^{-/-}$*Msh2*$^{-/-}$ germinal centers (**Supplementary Table 4**), which demonstrated that hypermutation of this promoter also occurs as part of physiological immune responses. We conclude that AID can deaminate basal promoters in activated B cells in a manner consistent with divergent polymerase profiles. Together with the observation that PolII paused at immunoglobulin S core domains[41,42] (**Fig. 1a**), our results further support the idea that SHM is targeted across the genome to sites of polymerase stalling[30].

RPA is restricted to *Igh* loci

Co-occupancy of AID and PolII across the genome provides a mechanistic explanation for the widespread hypermutation observed in primary and transformed B lymphocytes. However, the disparity in the magnitude of the mutation load at immunoglobulin and off-target sites cannot be fully explained on the basis of AID-binding density. For example, whereas the association of AID with chromatin was ~1.5-fold greater at *Igh-6* than at *Mir142* (**Supplementary Table 1**), the mutation frequency at the 5′ end of the *Igh-6* S domain ($S\mu$) was 10 times that measured at *Mir142* (**Supplementary Table 2**). Furthermore, hypermutation at the $S\mu$ core is probably orders of magnitude higher[43].

Several lines of evidence indicate that the ssDNA-binding protein RPA may have a role in hypermutation. In biochemical assays, RPA enhances the deamination of transcribed DNA by AID[22,23], and RPA interacts in an AID-dependent manner with S regions in B cells undergoing CSR[24]. However, whether recruitment of RPA correlates with AID activity across the B cell genome or whether it is specific for immunoglobulin genes has not been fully explored. To determine the genomic occupancy by RPA in AID-expressing cells, we activated B cells with LPS and IL-4 and assayed the cells by ChIP-seq with

© 2011 Nature America, Inc. All rights reserved.

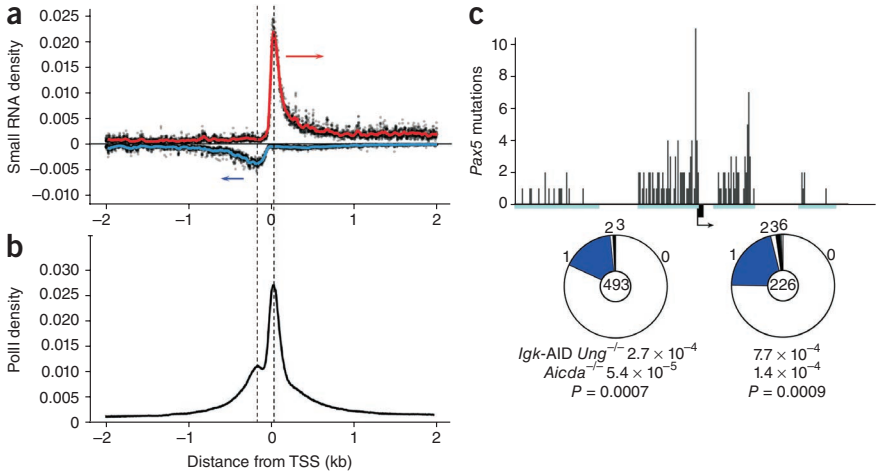

Figure 6 AID hypermutates basal promoters. (**a**) Alignment of sequences from small RNA cDNA libraries ($n = 14.7 \times 10^6$) relative to gene TSS (± 2 kb) in activated B cells. Arrows indicate sense of transcription. (**b**) PolII density at all genes in B cells activated with LPS and IL-4. Dashed lines indicate the coincidence of the two stalling peaks with the TSSs of both sense and antisense RNA in **a**. (**c**) AID-mediated hypermutation of the *Pax5* promoter-proximal area in *Igk*-AID *Ung*$^{-/-}$ B cells (top) detected from four PCR products (cyan rectangles); black box indicates *Pax5* exon 1, and arrow indicates TSS and direction of transcription. Middle, mutation frequency: segment size indicates the proportion of sequences with the number of mutations noted along circle margin; center, total sequences. Bottom, mutation frequency (calculated as total mutations per total base pairs sequenced) and *P* value (relative to *Aicda*$^{-/-}$ control). Data are representative of one experiment.

RPA-specific antibodies. Two RPA islands were reproducibly immunoprecipitated from *Igh*; one spanned the entire *Igh-6* μ-chain and one centered on the γ1-chain (**Fig. 7a**, top row). In contrast, only a very small number of sequence 'reads' aligned near Sμ in the absence of AID[24] (**Fig. 7a**, top and second rows). Notably, *Igk*-AID B cells, which have higher AID expression[9], had RPA signals 2.5 times more abundant than those measured in *Aicda*$^{+/+}$ cells (784.5 versus 318.6 tags per million sequences; **Fig. 7a**, top and third rows). In cells undergoing CSR to the γ3-chain (cells treated with LPS plus anti-δ-dextran), as expected, RPA was associated with both gene segments encoding the μ-chain and γ3-chain (data not shown). Thus, RPA is recruited to *Igh* in a manner that is dependent on and directly proportional to AID expression.

AID is phosphorylated at Ser38 and Thr140 in activated B cells[23,25,26,44]. Substitution of either residue does not affect AID deamination activity but interferes with SHM and CSR *in vivo*[44,45].

Furthermore, phosphorylation of Ser38 is required for the association of AID with RPA both *in vitro*[23] and *in vivo*[24]. To assess the role of AID phosphorylation in genome-wide RPA recruitment, we did ChIP-seq for RPA with mutant B cells in which AID Ser38 or AID Thr140 is replaced with an alanine residue[44]. The association of RPA with chromatin was much less (about one third as many sequence reads) in both AID-mutant samples than in *Aicda*$^{+/+}$ samples (**Fig. 7a**, fourth and fifth rows versus top row). These results indicate that phosphorylation of AID contributes to the interaction of RPA with *Igh*.

In contrast to the ChIP-seq results for AID, which identified nearly 6,000 genomic targets of AID (>12,000 islands), we identified only a small number of islands outside *Igh* by ChIP-seq for RPA (~200 per sample; **Supplementary Table 5** and **Supplementary Fig. 7**). We detected most of these islands in only one of the four independent ChIP-seq experiments with *Aicda*$^{+/+}$ cells (including *Igk*-AID), and those that we reproducibly detected in all samples (16 in total) were also present in the absence of AID (data not shown). Notably, a substantial fraction of non-immunoglobulin islands localized together with background noise obtained with *Aicda*$^{-/-}$ lymphocytes after ChIP-seq for AID (for example, the *Mir142* locus; **Figs. 5b** and **7b**). Our failure to detect RPA-specific

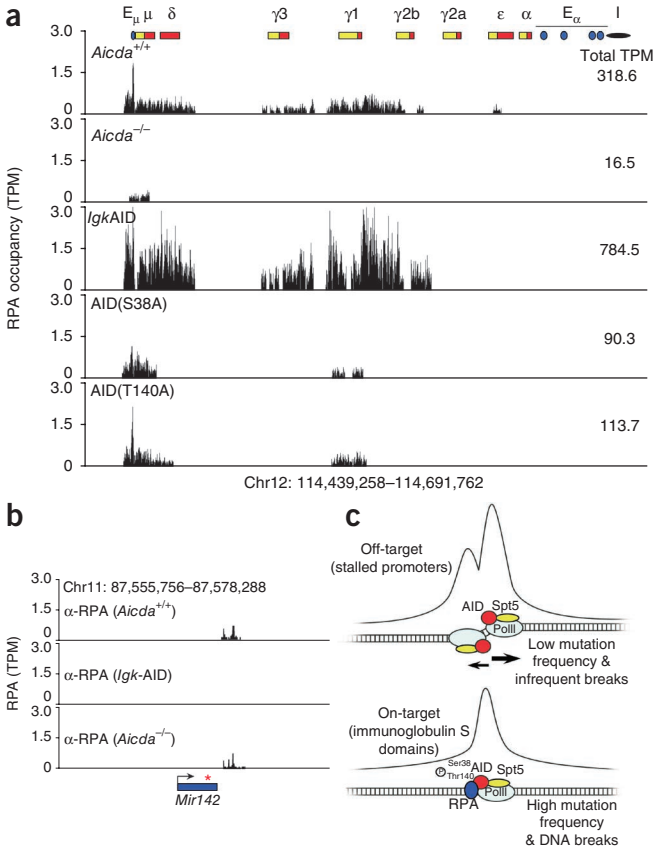

Figure 7 Recruitment of RPA to on-target sites of AID. (**a**) RPA occupancy at the immunoglobulin gene locus in B cells (genotype, top left) stimulated for 72 h with LPS and IL-4. AID(S38A) or AID(T140A), replacement of AID Ser38 or AID Thr140, respectively, with alanine. Numbers in plots (top right) indicate total sequence tags per million for the entire locus. (**b**) Background signals at the *Mir142* locus after ChIP-seq analysis of RPA. Red asterisk indicates position of the microRNA in the microRNA transcript (arrow). (**c**) AID recruitment and activity in B lymphocytes. Top: AID accumulates predominantly at PolII pausing sites across the genome by interacting with the stalling factor Spt5. At proximal promoter areas, RPA is either absent or fails to accumulate in sufficient amounts to fully engage AID activity. This results in little hypermutation both upstream and downstream of TSSs. With few exceptions, hypermutation at these sites is efficiently erased by base excision–repair and mismatch-repair activity. Bottom: at on-target immunoglobulin sites, RPA is recruited as a result of site-specific phosphorylation (P) of AID at Ser38 and Thr140. This activity enhances DNA deamination and the formation of double-stranded DNA breaks, particularly in the core of immunoglobulin S domains, where PolII frequently stalls.

© 2011 Nature America, Inc. All rights reserved.

targets outside *Igh* was not due to lack of sequence tag saturation, because a combined analysis of *Aicda*$^{+/+}$ samples ($n = 3$) and *Igk*-AID samples ($n = 2$) analyzed by ChIP-seq for RPA (44,764,916 total tags) did not show substantial association of RPA with AID-recruiting genes (**Supplementary Fig. 8** and **Supplementary Table 6**). We conclude that in contrast to genomic occupancy by AID itself, genomic occupancy by its cofactor RPA is restricted to immunoglobulin genes (**Fig. 7c**).

DISCUSSION

By means of deep sequencing, we have defined the global occupancy by AID and its cofactor RPA in the B cell genome. To better characterize the data, we further annotated the B cell genome by comprehensively mapping 36 epigenetic marks, the mRNA transcriptome, PolII, p300 and CTCF binding. We found that the association of AID with genes across the genome was widespread and correlated with activating chromatin marks. Among those we found acetylation of H3, H4 and H2B, as well as all three methylated forms of H3K4. By extension, genes with an inhibitory chromatin configuration (such as H3K27me3 and H3K20me3) failed to recruit AID or hypermutate. One example of this was *Mycn*, which is epigenetically and transcriptionally silent, does not bind AID and is rarely involved in chromosomal translocations in the mature B cell compartment[46,47].

Our data have established a tight correlation between AID and PolII, as predicted by the coimmunoprecipitation of AID and PolII from *ex vivo*–activated B cells[29]. The pausing factor Spt5 is required for the linkage of AID to the transcription apparatus[30]. In agreement with those data, we have now shown that AID associated mainly with paused polymerases at promoter-proximal sequences across the genome. Published observations have suggested a link between AID activity and pausing of gene transcription. For example, stalling of PolII at tandem repeats of the immunoglobulin S domain has been associated with AID activity during CSR[41,42]. One interpretation of those results was that pausing of the transcription machinery might promote DNA deamination by facilitating the interaction of AID with ssDNA substrates[41,42]. Consistent with that hypothesis, we found that AID occupancy and hypermutation coincided with stalling of divergent polymerases upstream of TSSs, a feature not previously appreciated. In addition, deep-sequencing studies have shown tight correlation between genome-wide recruitment of AID and sites of ssDNA (A.Y. and R.C., unpublished data). On the basis of these findings, we postulate that Spt5-stalled polymerases recruit AID across the genome, thus explaining the degree of AID's promiscuity in B cells.

The large number of AID targets explains the broad genomic instability observed in primary and premalignant cells after sustained or aberrant AID expression[14]. The data also underscore the decisive role of base-excision repair and mismatch repair in safeguarding the genome from promiscuous SHM. A pertinent example is the *Myc* proto-oncogene, which accumulates substantial hypermutation in *Ung*$^{-/-}$ or *Ung*$^{-/-}$*Msh2*$^{-/-}$ cells but is fully protected in wild-type germinal center or activated B cells[9,12], as shown here. Despite efficient repair, however, the *Myc* locus often participates in large-scale chromosomal alterations and translocations that are dependent on AID[14,46,48,49]. Thus, high-fidelity repair at off-target sites is not sufficient to prevent (and perhaps even promotes) DNA breaks, chromosomal translocations and B cell malignancy. In this context, we have shown that AID occupancy at *Myc* coincided precisely with mapped sites for canonical translocation breakpoints. We anticipate that the AID ChIP-seq data will help identify new tumor-inducing targets of AID.

The broad recruitment of AID in the B cell genome raises the question of whether AID has additional functions beyond diversification of immunoglobulin genes. Studies have linked cytidine deamination and AID to the elusive mechanism of DNA demethylation. In zebrafish, AID and Apobec2 seem to be required for the demethylation of exogenous DNA[21], and AID deficiency is reported to result in genome-wide hypermethylation of mouse primordial germ cells[20]. The reprogramming of mouse-human heterokaryons, which requires the demethylation of promoters of genes encoding key transcription factors, is also facilitated by AID[19]. On the basis of those observations, it has been proposed that high AID expression in germinal center B cells might also engage in active demethylation of the B cell genome[19,20]. Our observation that AID deaminated basal promoters would be consistent with that hypothesis.

A prominent feature of our results is that whereas AID was highly promiscuous, its cofactor RPA seemed to be specific for immunoglobulin genes under the conditions tested. However, our data do not exclude the possibility that RPA is recruited to any given off-target site in only a fraction of the cells being assayed. A signal present in only some cells and in different genomic locations in subpopulations of cells would not be detected above background. Heterogeneity among dividing cells is probably the reason we did not detect interaction of RPA with the DNA-replication machinery in our nonsynchronized cultures. As expected from published work[24], the localization of RPA to *Igh* requires phosphorylation of AID. The disparity between the genome-wide recruitment of AID and RPA is also consistent with the idea that RPA may function as an amplifier of AID activity on the immunoglobulin locus[22,23]. In biochemical assays, RPA seems to stabilize ssDNA displaced by the transcribing holoenzyme[22,23,50], thus providing AID with a window of opportunity to initiate cytidine deamination. Given the established role of RPA in DNA repair[27], an additional, not mutually exclusive possibility is that RPA functions downstream of the initial DNA damage, for example, by stabilizing ssDNA exposed during the repair phase of an AID lesion. In conclusion, we have identified here the broad range of genes targeted by AID, and the epigenetic and PolII stalling signature associated with targeting. We have demonstrated that AID recruitment alone is insufficient to explain the difference in mutator activity at immunoglobulin and off-target genes.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/natureimmunology/.

Accession codes. GEO: ChIP-seq data for PolII and AID, GSE24178; ChIP-seq data for mRNA, GSE21630.

Note: Supplementary information is available on the Nature Immunology website.

ACKNOWLEDGMENTS

We thank D. Schatz for comments on the manuscript; J. Chaudhuri (Memorial Sloan-Kettering Cancer Center) and F. Alt (Harvard University) for antibodies to AID; J. Simone for cell sorting; G. Gutierrez for technical assistance with the genome analyzer; and C. Ansarah-Sobrinho and S. Nelson for help with sequencing. Supported by the National Institutes of Health (Intramural Research Program of the National Institute of Arthritis and Musculoskeletal and Skin Diseases; and AI037526 to M.C.N.) and the Howard Hughes Medical Institute (M.C.N.).

AUTHOR CONTRIBUTIONS

A.Y. did deep sequencing, cloning and conventional sequencing experiments; W.R. and H.-w.S. analyzed data; N.K. contributed data; Z.L. maintained the mouse colonies and cultured cells; D.F.R. contributed the *Igk*-AID mice; M.C.N. made suggestions for experiments and reviewed and wrote sections of the manuscript; R.C. designed the experiments and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

© 2011 Nature America, Inc. All rights reserved.

Published online at http://www.nature.com/natureimmunology/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Stavnezer, J., Guikema, J.E. & Schrader, C.E. Mechanism and regulation of class switch recombination. *Annu. Rev. Immunol.* **26**, 261–292 (2008).
2. Honjo, T., Kinoshita, K. & Muramatsu, M. Molecular mechanism of class switch recombination: linkage with somatic hypermutation. *Annu. Rev. Immunol.* **20**, 165–196 (2002).
3. 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).
4. 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).
5. Di Noia, J.M. & Neuberger, M.S. Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* **76**, 1–22 (2007).
6. Peled, J.U. *et al.* The biochemistry of somatic hypermutation. *Annu. Rev. Immunol.* **26**, 481–511 (2008).
7. Delker, R.K., Fugmann, S.D. & Papavasiliou, F.N. A coming-of-age story: activation-induced cytidine deaminase turns 10. *Nat. Immunol.* **10**, 1147–1153 (2009).
8. Peters, A. & Storb, U. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity* **4**, 57–65 (1996).
9. Robbiani, D.F. *et al.* AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. *Mol. Cell* **36**, 631–641 (2009).
10. 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).
11. 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).
12. Liu, M. *et al.* Two levels of protection for the B cell genome during somatic hypermutation. *Nature* **451**, 841–845 (2008).
13. 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).
14. Nussenzweig, A. & Nussenzweig, M.C. Origin of chromosomal translocations in lymphoid cancer. *Cell* **141**, 27–38 (2010).
15. Klemm, L. *et al.* The B cell mutator AID promotes B lymphoid blast crisis and drug resistance in chronic myeloid leukemia. *Cancer Cell* **16**, 232–245 (2009).
16. Lin, C. *et al.* Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell* **139**, 1069–1083 (2009).
17. Matsumoto, Y. *et al. Helicobacter pylori* infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. *Nat. Med.* **13**, 470–476 (2007).
18. Morgan, H.D., Dean, W., Coker, H.A., Reik, W. & Petersen-Mahrt, S.K. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J. Biol. Chem.* **279**, 52353–52360 (2004).
19. Bhutani, N. *et al.* Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* (2010).
20. Popp, C. *et al.* Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101–1105 (2010).
21. Rai, K. *et al.* DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* **135**, 1201–1212 (2008).
22. 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).
23. Basu, U. *et al.* The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature* **438**, 508–511 (2005).
24. Vuong, B.Q. *et al.* Specific recruitment of protein kinase A to the immunoglobulin locus regulates class-switch recombination. *Nat. Immunol.* **10**, 420–426 (2009).
25. McBride, K.M. *et al.* Regulation of hypermutation by activation-induced cytidine deaminase phosphorylation. *Proc. Natl. Acad. Sci. USA* **103**, 8798–8803 (2006).
26. 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).
27. Wold, M.S. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu. Rev. Biochem.* **66**, 61–92 (1997).
28. 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).
29. Nambu, Y. *et al.* Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science* **302**, 2137–2140 (2003).
30. Pavri, R. *et al.* Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* **143**, 122–133 (2010).
31. Pasqualucci, L. *et al.* Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* **412**, 341–346 (2001).
32. Ji, Y. *et al.* The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell* **141**, 419–431 (2010).
33. Lenz, G. *et al.* Aberrant immunoglobulin class switch recombination and switch translocations in activated B cell-like diffuse large B cell lymphoma. *J. Exp. Med.* **204**, 633–643 (2007).
34. Kuchen, S. *et al.* Regulation of microRNA expression and abundance during lymphopoiesis. *Immunity* **32**, 828–839 (2010).
35. Storb, U. *et al.* Targeting of AID to immunoglobulin genes. *Adv. Exp. Med. Biol.* **596**, 83–91 (2007).
36. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
37. Mikkelsen, T.S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
38. Rahl, P.B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).
39. Fuda, N.J., Ardehali, M.B. & Lis, J.T. Defining mechanisms that regulate RNA polymerase II transcription *in vivo*. *Nature* **461**, 186–192 (2009).
40. Buratowski, S. Transcription. Gene expression–where to start? *Science* **322**, 1804–1805 (2008).
41. Wang, L., Wuerffel, R., Feldman, S., Khamlichi, A.A. & Kenter, A.L. S region sequence, RNA polymerase II, and histone modifications create chromatin accessibility during class switch recombination. *J. Exp. Med.* **206**, 1817–1830 (2009).
42. Rajagopal, D. *et al.* Immunoglobulin switch mu sequence causes RNA polymerase II accumulation and reduces dA hypermutation. *J. Exp. Med.* **206**, 1237–1244 (2009).
43. Xue, K., Rada, C. & Neuberger, M.S. The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in msh2–/– ung–/– mice. *J. Exp. Med.* **203**, 2085–2094 (2006).
44. McBride, K.M. *et al.* Regulation of class switch recombination and somatic mutation by AID phosphorylation. *J. Exp. Med.* **205**, 2585–2594 (2008).
45. Cheng, H.L. *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* **106**, 2717–2722 (2009).
46. Kovalchuk, A.L. *et al.* AID-deficient Bcl-xL transgenic mice develop delayed atypical plasma cell tumors with unusual Ig/Myc chromosomal rearrangements. *J. Exp. Med.* **204**, 2989–3001 (2007).
47. Malynn, B.A. *et al.* N-myc can functionally replace c-myc in murine development, cellular growth, and differentiation. *Genes Dev.* **14**, 1390–1399 (2000).
48. Takizawa, M. *et al.* AID expression levels determine the extent of cMyc oncogenic translocations and the incidence of B cell tumor development. *J. Exp. Med.* **205**, 1949–1957 (2008).
49. Robbiani, D.F. *et al.* Activation induced deaminase is required for the chromosomal translocations in c-myc that lead to c-myc/IgH translocations. *Cell* **135**, 1028–1038 (2008).
50. Chaudhuri, J. & Alt, F.W. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat. Rev. Immunol.* **4**, 541–552 (2004).

© 2011 Nature America, Inc. All rights reserved.

ONLINE METHODS

ChIP-seq. Resting splenic B cells were isolated from 6- to 8-week-old wild-type C57BL6/J or *Aicda*$^{-/-}$ mice (backcrossed for 12 generations to C57BL6/J mice) with anti-CD43 Microbeads (anti-Ly48; Miltenyi Biotech) and were cultured for 60 h with LPS (50 μg/ml; Sigma) and IL-4 (5 ng/ml; Sigma). In total, 57 mice were used for ChIP-seq studies. For quality control, switch recombination to IgG1 was checked before deep-sequencing analysis (**Supplementary Fig. 9**). Chromatin was digested by incubation of collected cells with micrococcal nuclease (1 unit/ml; Sigma) for analysis of most histone modifications except H3K79me1, H3K79me2 and H3K79me3. For those last three modifications, as well as PolII, p300, CTCF and AID, cells were crosslinked for 10 min at 37 °C with 1% (vol/vol) formaldehyde, followed by quenching with 0.125 M glycine (final concentration). Crosslinked cell samples were then sonicated to obtain DNA fragments 200–500 bp in length. For each ChIP sample, sonicated (or nuclease-treated) chromatin (from 1×10^7 cells) was precleared with 40 μl protein A or protein G magnetic beads, followed by incubation overnight at 4 °C in radioimmunoprecipitation assay (RIPA) buffer (10 mM Tris, pH 7.6, 1 mM EDTA, 0.1% (wt/vol) SDS, 0.1% (wt/vol) sodium deoxycholate and 1% (vol/vol) Triton X-100) with 40 μl protein A or G magnetic beads (Invitrogen) and 5–10 μg antibody (**Supplementary Table 7**). Beads were washed twice with RIPA buffer alone, then twice with RIPA buffer plus 0.3 M NaCl, twice with LiCl buffer (0.25 M LiCl, 0.5% (vol/vol) Igepal CA630 and 0.5% (wt/vol) sodium deoxycholate), once with Tris-EDTA buffer (Tris 10 mM, pH 8.0, and 1 mM EDTA) plus 0.2% (vol/vol) Triton X-100, and once with Tris-EDTA buffer alone. ChIP DNA was extracted for 4 h at 65 °C in Tris-EDTA buffer with 0.3% (wt/vol) SDS and proteinase K (1 mg/ml). Samples were sequenced on a Genome Analyzer IIx after being prepared for microsequencing according to the manufacturer's protocol (Illumina). Each genome-wide analysis represents a single immunoprecipitation analyzed in one lane of the Ilumina slide. The number of immunoprecipitations is as follows: AID, four (including an *Aicda*$^{-/-}$ control); RPA, ten (details, **Supplementary Table 5**); PolII, four; H3K4me3, H3K4me27, H2BK5Ac, H3K27Ac, H3K9Ac, H2BK120Ac and CTCF, three; other samples, one. Although the Ab1 antibody to AID has been used for ChIP analysis before[22], its immunoprecipitation specificity was further tested by immunoblot analysis (**Supplementary Fig. 10**); all experiments used Ab1 except those in **Supplementary Figure 2**.

Hypermutation analysis. CD43$^-$ splenocytes from *Igk*-AID *Ung*$^{-/-}$ or *Aicda*$^{-/-}$ mice (at a density of 0.1×10^6 cells per ml) were cultured for 120 h with LPS and IL-4, plus anti-CD180 (anti-RP105; 0.5 μg/ml; RP/14; BD Pharmingen). Cells were passaged at 72 h at a dilution of 1–4. Genomic DNA (50 ng) was amplified for 30 cycles with Phusion DNA polymerase (New England Biolabs) and specific primers (**Supplementary Table 8**). For nested PCR, two 20-cycle amplifications were done. The amplicon was cloned with a Zero Blunt PCR kit (Invitrogen) and sequenced.

Mice. All animal experiments were done according to the National Institutes of Health guidelines for laboratory animals and were approved by the Scientific Committee of the National Institute of Arthritis, Musculoskeletal and Skin Diseases and National Cancer Institute Animal Facilities.

Sequencing-based mRNA analysis. The mRNA from activated B cells was sequenced according to the manufacturer's protocol for deep-sequencing analysis of mRNA (Illumina). A detailed version of this protocol has been published[34].

Image processing and alignment. Standard Illumina image-analysis and base-calling software was used for determination of short-read sequences and corresponding qualities as well as application of the default quality filters. Pipeline output was transformed to the fastq format and reads were aligned to the Build 37 assembly of the National Center for Biotechnology Information mouse genome data (July 2007; NCBI37/mm9) with Bowtie 0.9.9.3 software, including the following command line options: –best–all -m1 -n2 (reads that align uniquely in the best alignment stratum, allowing up to two mismatches).

ChIP-seq analysis. Peaks of enrichment, assessed relative to those of a random background model, were identified with SICER 1.03 software (spatial clustering approach for the identification of ChIP-enriched regions). For AID, the parameters used were as follows: window size, 100; gap size, 100; e, 0.000001; redundancy, 1; fragment size, 127 (*Aicda*$^{+/+}$) or 115 (*Aicda*$^{-/-}$). Estimates of fragment size were calculated from the data on the basis of a shifting algorithm that attempts to minimize a Shannon entropy measure. For other modifications, a more lenient 'e' value of 0.05 was used. Read counts were adjusted for library size and are presented as reads per million nonredundant aligned reads. Density plots presented here include only windows that were on enriched peak regions. A combination of custom Python and R software was used for downstream analysis of SICER output (for example, the assignment of individual reads to gene regions and calculation of cumulative summary plots at TSSs).

AID ChIP-seq saturation analysis. Multiple lanes in multiple flow cells were combined for *Aicda*$^{+/+}$ samples (eight lanes; 143,995,816 reads) and *Aicda*$^{-/-}$ samples (six lanes; 134,418,212 reads) for ChIP-seq with anti-AID. For saturation analysis, each data set was split into randomly generated subsets of approximately 6×10^6 reads, then increasing numbers of subsets were merged, and each set generated in this way was analyzed separately as described above.

Enrichment of ChIP-seq tags on AID islands. The total number of aligned reads of a ChIP-seq experiment was used to estimate the expected number of such reads that would be on AID islands by chance and was compared with the actual number of reads on the islands. The significance of the enrichment as well as the 99% confidence interval was determined with a two-tailed binomial test.

Hypermutation *q* values. The following procedure was used for the adjustment of *P* values to control for false-discovery rate. Fisher's exact test was used to test the significance of the difference between *Aicda*$^{+/+}$ and *Aicda*$^{-/-}$ samples in their mutation rate. To control for false discovery rate, the *P* values were further adjusted with the p.adjust function in R software.

ChIP-seq of small RNA (divergent transcription analysis). On the basis of previous ChIP-seq analysis of small RNA, tags that aligned uniquely with the genome and either mapped to protein-coding domains or could not be mapped to any annotated regions were retrieved from the read database for activated B cells (three experiments) and germinal center cells (one experiment). Up to four identical tags were retained and mapped to the RefSeq annotation. Approximately 1 million tags could be assigned to a known transcript, and 0.1 million could be assigned to a putative divergent transcript on the opposite strand of a known transcript up to –2 kb upstream of the TSS.

Hierarchical clustering. The Partek Genomics Suite version 6.5 was used for hierarchical clustering, with $\log_2$-transformed data and normalized Euclidean distance (average Euclidean distance as implemented in Partek). A distance matrix calculated from the densities of each modification and AID on the regions with significant AID tag densities was used for hierarchical clustering.

© 2011 Nature America, Inc. All rights reserved.

doi:10.1038/ni.1964