A licensing step links AID to transcription elongation for mutagenesis in B cells

Stephen P. Methot1,2,9, Ludivine C. Litzler1,3, Poorani Ganesh Subramani1,2, Anil K. Eranki1, Heather Fifield4, Anne-Marie Patenaude1,10, Julian C. Gilmore1, Gabriel E. Santiago5, Halil Bagci1,6, Jean-François Côté1,6,7, Mani Larijani4, Ramiro E. Verdun5,8 & Javier M. Di Noia1,2,3,7

Activation-induced deaminase (AID) mutates the immunoglobulin (Ig) genes to initiate somatic hypermutation (SHM) and class switch recombination (CSR) in B cells, thus underpinning antibody responses. AID mutates a few hundred other loci, but most AID-occupied genes are spared. The mechanisms underlying productive deamination versus non-productive AID targeting are unclear. Here we show that three clustered arginine residues define a functional AID domain required for SHM, CSR, and off-target activity in B cells without affecting AID deaminase activity or Escherichia coli mutagenesis. Both wt AID and mutants with single amino acid replacements in this domain broadly associate with Spt5 and chromatin and occupy the promoter of AID target genes. However, mutant AID fails to occupy the corresponding gene bodies and loses association with transcription elongation factors. Thus AID mutagenic activity is determined not by locus occupancy but by a licensing mechanism, which couples AID to transcription elongation.
The enzyme activation-induced deaminase (AICDA, referred to as AID, encoded by the Aicda gene) initiates genetic modifications at the immunoglobulin (Ig) genes in activated B cells1,2. AID catalyses the deamination of deoxycytidine to deoxyuridine on single-stranded DNA (ssDNA)3. This change is mutagenic, but further processing of the deoxyuridines by DNA repair enzymes underpins somatic hypermutation (SHM) and class switch recombination (CSR), which are indispensable for efficient antibody responses4–6. As deleterious side effects of SHM and CSR, AID can mutate and induce DNA damage outside the Ig loci, in many cases triggering chromosomal translocations4.

DNA repair pathways limit off-target mutations and DNA damage by AID5–7. Nevertheless, several additional layers of regulation are necessary to control AID oncogenic and cytotoxic activity8. Regulation of AID protein levels and nuclear access restrains both on- and off-target activities, but it is unclear whether they contribute to target specificity1. The preferential targeting of AID to the Ig genes and how AID mutates a small number of additional genomic loci while sparing most others is an area of active research9,10. The Ig loci possess an intrinsic ability to attract AID activity10, conferred in part by specialized cis-acting elements that combine transcriptional enhancers with multiple transcription factor-binding sites and can ectopically function to target SHM11,12. Similar elements have not been identified in AID off-targets, but these loci share with the Ig the characteristic of showing convergent transcription and being associated with strong super-enhancers13–15. Nonetheless, many highly transcribed genes have similar characteristics but are not mutated, so an additional layer of regulation must exist. The identity of the trans-acting factors targeting AID to the Ig loci is also elusive, though non-coding RNA and transcription factors likely have a function4. Genome-wide studies have identified a few factors that correlate with AID occupancy and mutagenic activity, such as RNA polymerase II (RNAPII), its associated factor Spt5 (Supt5h) and the RNA processing exosome16–18. Again, these factors function at a much larger number of loci than are mutated by AID and fail to explain AID’s specificity on their own.

Thus there is a three-tier system of AID targeting, with the Ig loci being targeted much more frequently than any AID off-targets but the latter restricted to a few hundred sites. Beyond specific examples of loci occupied but not mutated by AID19, the analysis of AID occupancy by chromatin immunoprecipitation (ChIP)—sequencing has suggested its association with ~6000 genes in B cells, while AID-induced damage is limited to some 300 loci7,13,14,20,21. This begs the question of why most sites bound by AID are spared from its activity.

Here we report a new functional domain of AID that is dispensable for enzymatic activity but necessary for on- and off-target biological activity in B cells. Systematic analysis of the function and interactome of AID variants with mutations in this arginine-rich (RR) domain reveals that they have a defect specifically in their association with the gene body of physiological and collateral target sites, explaining their failure to mutate. Our results uncover a licensing mechanism that most likely couples AID to transcription elongation, which can explain why occupancy is not sufficient to predict AID activity and suggest a new model for productive AID targeting. Our data also suggest that limiting nuclear levels of AID are important to enforce this licensing mechanism.

**Results**

Three arginines in AID α6 define a new functional domain. In previous structure–function analyses, we used a set of chimeric proteins in which contiguous regions of AID were replaced by their homologous region from APOBEC2 (A2)22–24. Only one of these, AID-A2#5, could mutate the *Escherichia coli* genome (Supplementary Fig. 1a, b). AID-A2#5 replaces a large C-terminal portion of AID, starting from the loop preceding alpha-helix 6 (a6) and eliminating the C-terminal E5 domain, which is necessary for CSR25. However, not only did adding back E5 not rescue CSR but this chimera also lacked IgV SHM activity when used to complement *Aicda*-/- B cells (Supplementary Fig. 1c, d). A smaller chimera, replacing only the a6 of AID with that of A2 (AID-A2 a6) had measurable activity in *E. coli* but not in B cells (Supplementary Fig. 1a–d). The functional defect of AID-A2 a6 could not be explained by differences in protein abundance or nuclear access (Supplementary Fig. 1b–e). These results suggested that the AID a6 contained residues required for SHM and CSR but dispensable to mutate *E. coli*.

We sought to identify single amino acid substitutions within AID a6 that could separate its ability to mutate *E. coli* from its biological activity in B cells. Comparing a three-dimensional molecular model of AID26 to the A2 structure27 showed several residue and charge differences in *E. coli* between these paralogues (Fig. 1a). To obtain AID variants with minimal structural alterations that could recapitulate the phenotype of the chimeras, we independently mutated several of these AID residues to the corresponding A2 residue. Three of these recapitulated the results obtained with the chimeras. AID R171Y, R174E and R178D mutated *E. coli* with the same efficiency as AID but were inactive for SHM and CSR (Fig. 1b–d). In contrast, adjacent mutations AID R177A and S173E maintained all three activities (Fig. 1a–d). Notably, Arg 171, 174 and 178 are conserved in AID from most jawed vertebrates but not in the APOBECs (Supplementary Fig. 1f) and form a contiguous AID surface (Fig. 1a)28,29. The natural AID variant R174S found in some immunoodeficient HIGM2 patients30 conserves DNA binding and processivity but has substantially reduced catalytic activity31. Furthermore, protein arginine residues are common contact points with nucleic acids32. Nonetheless, the R-mutants show similar DNA-binding affinity and deaminate ssDNA within a bubble substrate in vitro with similar specific activities to wild-type (wt) AID (Fig. 1e, f).

To test the role of the positive charge contributed by these Arg residues, we substituted each of them for lysine (Supplementary Fig. 2). AID R171K had reduced *E. coli* mutation activity and a proportional decrease in SHM and CSR, indicating a structural contribution to catalysis that Tyr can provide but Lys cannot. AID R174K mutated *E. coli* with 50% efficiency compared to AID and showed a proportional decrease in SHM but lacked CSR, indicating not only a structural contribution for R174 but also a specific role in CSR. AID R178K was indistinguishable from AID for all activities, indicating that, in this case, the charge is sufficient for biological function.

We conclude that AID arginines 171, 174 and 178 not only contribute in charge but also structurally to creating a functional domain necessary for SHM and CSR. Since the substitutions to the corresponding A2 residues consistently retained enzymatic activity, we used those mutants for dissecting the role of this functional domain (hereafter, the RR domain).

AID R-mutants enter the nucleus but lack off-target activity. Nuclear access of AID is restricted, with ~10% of AID being nuclear in homeostasis as a consequence of several mechanisms regulating AID nuclear residency and protein stability33. To exclude that the functional defect of the R-mutants was due to defective nuclear access, we analysed their subcellular localization. We used AID-deficient CH12 B cells, in which wt AID reconstituted CSR activity but none of the R-mutants did, despite similar expression levels (Supplementary Fig. 3a–c). Akin to wt
AID, the R-mutants were cytoplasmic in steady state and accumulated in the nucleus after inhibiting nuclear export with leptomycin B (LMB) and/or AID cytoplasmic retention with didemnin B (Did B) (Fig. 2a). Probing for AID by western blot (WB) in cytoplasmic and nuclear extracts from untreated cells showed that nuclear levels for all AID variants were similar to wt AID, even for R174E that showed reduced cytoplasmic levels compared to wt AID (Fig. 2b). AID can be trapped inside the cell by western blotting in cytoplasmic and nuclear extracts from untreated cells using an anti-AID antibody. The results show that the nuclear levels of all AID variants were similar to wt AID, even for R174E that showed reduced cytoplasmic levels compared to wt AID (Fig. 2b).
nucleus upon reforming of the nuclear envelope after mitosis. This was maintained by the R-mutants (Fig. 2c and Supplementary Fig. 3d). Thus the mechanisms regulating AID nuclear access are functional in the R-mutants.

As the R-mutants still accessed the nucleus, we asked whether their functional defect was upstream or downstream from deamination. To do so, we measured AID-induced mutations at the Sμ region, the major AID target at the IgH locus during CSR. In reconstituted Aicda−/− Ung−/− B cells, AID R171Y and R178D had substantially lower mutation frequency than wt AID (Fig. 3a), proportionally to the SHM and CSR defects observed (Fig. 1c, d).

We next asked whether this apparent Ig-targeting defect of the R-mutants extended to other loci. As a proxy assay for genome-wide mutagenesis, we measured the decrease in fitness that is associated with AID off-target DNA damage. AID expression was sufficient to compromise cell fitness in competitive cultures of CH12 cells, and this effect was enhanced by using a low dose of Did B, which increases the proportion of AID in the nucleus and boosts off-target activity. Cells expressing the R-mutants showed no cell fitness defect, just like AID-deficient cells or those expressing catalytically inactive AID E58A (Fig. 3b). To confirm these results, we introduced the R-mutants into the hyperactive AID7.3 variant, which bears three point mutations (outside a6) that increase enzymatic activity three-fold, leading to proportionally higher SHM, CSR, chromosomal translocations and DNA damage in B cells. The AID7.3 R-mutants maintained hyperactivity in E. coli (Fig. 3c), yet they were still severely deficient for CSR, SHM and cytotoxicity in CH12 and DT40 B cells (Fig. 3d–f). We conclude that the R-mutants enter and accumulate in the nucleus similarly to wt AID but have globally reduced mutagenic activity in B cells.

The AID RR domain is dispensable for chromatin association. The nuclear fraction of AID is difficult to visualize in whole cells because of the signal coming from cytoplasmic AID. To test whether the chromatin association of the R-mutants was different from wt AID, we combined a nuclear wash technique with confocal microscopy. We first validated this in situ fractionation method on endogenous AID using CH12 cells, where eliminating the cytoplasmic signal allowed specific detection of AID in the nucleus (Fig. 4a). We then compared the R-mutants to wt AID in transduced AID-deficient CH12 cells (Fig. 4b). Washing away cytoplasm and nucleoplasm, as shown by the loss of the cytoplasmic AID and cell-wide green fluorescent protein (GFP) signals, revealed that AID E58A and the R-mutants showed the same chromatin association as wt AID (Fig. 4b).

To confirm the association of the R-mutants with chromatin, we used a biochemical fractionation protocol that uses an incomplete DNA digestion by micrococcal nuclease (MNase) followed by sequential extractions with increasing salt concentrations (Fig. 4c). As expected, the majority of AID was cytoplasmic, and the lack of cytoplasmic contamination in the isolated nuclei was confirmed by the ggc2-dehydrogenase (Gapdh) partition (Fig. 4d). Nuclear fractionation showed some Spt5 but no RNAPII in the MNase fraction. All RNAPII and most Spt5 were found in the low and high salt fractions, representing loosely and tightly held transcription complexes, respectively (Fig. 4d). The latter fraction also contained most of the chromatin, judging from nucleosome content (Fig. 4d). Notably, ~60% of the nuclear AID was found in the 600 mM NaCl extract (Fig. 4d, e), indicating tight chromatin association, with little AID in the MNase or 150 mM NaCl fractions. The rest of AID was present in a remaining pellet that was largely devoid of RNAPII, Spt5 or chromatin but contained...
Lamin B, thus defining a second pool of nuclear AID that is either part of non-soluble nuclear complexes not directly associated with transcription factors or precipitates during extraction. In either case, the data indicate at least two distinct pools of AID associated with chromatin/nuclear matrix, independently of catalytic activity and biological function. Importantly, the R-mutants had similar distribution profile and proportions as wt AID in all fractions (Fig. 4d, e), showing that they are normally associated with the chromatin in homeostasis. We conclude that the R-mutants dissociate the chromatin interaction from the biological activity of AID, suggesting that these are mechanistically distinct steps.

**AID–chromatin association needs Spt5 but not transcription.**
We investigated further the association of AID and the R-mutants to chromatin. Spt5 is important for AID activity in B cells and correlates with AID occupancy genome wide, but it is not known whether it is necessary or sufficient for AID–chromatin interaction. Spt5 knockdown in CH12 cells not only reduced CSR but also overall chromatin association of both endogenous and transduced AID (Supplementary Fig. 4a, b and Fig. 5a). Interestingly, the R-mutants were similarly dependent on Spt5 for chromatin association (Fig. 5a). We directly assessed whether an accumulation of Spt5 at the chromatin was sufficient to induce local AID recruitment by using U2OS cells that contain a genomic Lac operon array (LacO), which is recognized by the Lac repressor (LacR) (Fig. 5b). In this system, a mCherry-LacR-Spt5 fusion recruited AID-GFP, as well as the R-mutants, to the LacO locus (Fig. 5b). As controls, the mCherry-LacR fusion alone did not recruit AID-GFP, and APOBEC1-GFP was not recruited by mCherry-LacR-Spt5.

The association of AID to RNAPII depends on Spt5, but whether transcription or RNAPII itself are involved in retaining AID at the chromatin is unknown. To test this, we treated CH12 cells with actinomycin D (Act D), a DNA intercalating compound that disrupts transcription elongation. The eukaryotic RNA polymerases show unequal sensitivity to Act D, RNAPII>RNAPI>RNA-PIII. Act D at doses that inhibit RNAPI (0.04 μM) or RNAPII and II (0.4 μM) slightly reduced chromatin-bound RNAPII and caused some AID redistribution to distinct nuclear sites but did not significantly change the amount of AID associated with the chromatin (Fig. 5c). Interestingly, at 4 μM Act D, both RNAPII and AID were depleted from the chromatin (Fig. 5c). A time course at 2 μM Act D showed that RNAPII and AID were concomitantly depleted from the chromatin (Fig. 5d), suggesting that retention of AID at the chromatin might be dependent on RNAPII. The chromatin association of the R-mutants was similarly sensitive to Act D (Fig. 5e).

Surprisingly, AID was partially resistant to chromatin disruption by DNase treatment, which depletes nuclear Spt5 but leaves the nuclear envelope (evidenced by Lamin B staining) intact (Fig. 5f and Supplementary Fig. 4c). In contrast, nuclear AID was largely depleted after RNase treatment, despite chromatin-associated Spt5 being resistant (Fig. 5f).

We conclude that the broad association of AID to chromatin requires RNA and the transcription machinery but not transcriptional activity, with Spt5 accumulation sufficient to promote the recruitment of AID to chromatin, although this...
The RR domain is dispensable for AID association with tightly held nuclear complexes. a Representative confocal microscopic images of CH12 B cells either fixed directly (whole cell) or after nuclear washing (0.5% Triton +200 mM NaCl). Isolated nuclei were analysed by IF to detect endogenous AID and Spt5 and DNA stained by Dapi. Cells were stimulated (+CIT) to induce AID expression or not (−CIT). Representative of three experiments. b (Left) Representative confocal microscopic images of GFP, AID (detected by IF) and DNA (Dapi) in reconstituted AID-deficient CH12 B cells, imaged as in a. AID and GFP expression were linked via IRES. (Right) Nuclear AID signal was calculated relative to whole-cell intensity for each variant and normalized to the wt AID value. Plotted are means (bars) from 3 independent experiments (dots) with 20–77 cells per condition per experiment. a, b Magnification 630×. Scale bar, 10 μm. Laser power and/or gain were increased for imaging after nuclear wash (see Methods). c Scheme for biochemical fractionation of B cells. Fractions analysed are indicated in brackets. d Representative WBs on the indicated fractions from reconstituted AID-deficient CH12 B cells. Antibodies against AID, RNAPII, Spt5, Lamin B and Gapdh were used as indicated. Representative agarose gel with purified DNA stained using ethidium bromide and Ponceau-S staining of total protein are also shown. For gel source data, see supplementary Fig. 7. e Quantification of AID signal in each lane from d, normalized to the respective cytoplasmic AID. Means ± s.d. from three independent experiments.

effect could be indirect and likely requires additional factors, such as RNA. The fact that chromatin-associated AID exists in at least two distinct fractions, only one of which contains Spt5 and RNAPII (Fig. 4d), implies that AID is dynamically associated with these fractions. Constant cycling through Spt5 would explain why Spt5 knockdown evicts AID from the chromatin. The RR domain is dispensable for this dynamic interaction, yet necessary for function, suggesting that it mediates productive targeting of AID.
**Fig. 5 Requirements for chromatin association of AID.** a, c–f Representative confocal microscopic images of GFP, immunofluorescence (for AID and Spt5 or RNAPII) and DNA staining (Dapi) in reconstituted AID-deficient CH12 B cells. Nuclear wash was carried out as in Fig. 4. Magnification 630×. Scale bar, 10 μm. a (Left) CH12 cells were transduced with either a shRNA control or against Spt5 as indicated. (Right) Box plots (25–75 percentiles with median and 10–90 percentiles whiskers of AID and Spt5 signal from individual whole cells or isolated nuclei, normalized to the median of wt AID. One of the two independent experiments is plotted (21–66 cells per condition). Differences were evaluated by unpaired, two tailed t-test (ns, not significant, *P < 0.05, ****P < 0.0001). b (Left) Schematic of the lacO/LacR system. (Middle) Representative confocal microscopic images of GFP or mCherry and their overlay, including DNA (Dapi). Small boxes highlight typical mCherry foci. (Right) Quantification of GFP signal at each mCherry focus, normalized to non-focus nuclear GFP signal (dots). The mean for each construct (bars) is shown. Compilation of two independent experiments. Magnification 400×. Scale bar, 10 μm. c As in a, but cells were treated for 60 min with DMSO (0) or the indicated dose of Actinomycin D (Act D) prior to plating and washing. d As in a, but cells were treated with DMSO (t = 0) or for various times with 2 μM Act D prior to plating and washing. e As in a but cells were treated with 2 μM Act D for 120 min. c–e Average AID and RNAPII signal for individual nuclei (dots), with bars indicating population median, normalized to the median of untreated cells (dotted line), with 17–32 cells per condition from 1 experiment. f As in a but nuclei were incubated at 37 °C with PBS control, RNase or DNase during nuclear wash. Images are representative of 42–78 cells per condition and 2 independent experiments. IF immunofluorescence, A.U. arbitrary units.
**Modular nature of the RR domain.** To obtain mechanistic insight into the defect of the AID R-mutants, we asked whether a wt AID RR domain could rescue their function by fusing either wt AID or the R-mutants to the catalytically inactive AID E58A mutant. The control AID-AID E58A fusion protein had reduced mutagenic activity in *E. coli*, compared to the AID monomer, but still produced substantial CSR (Fig. 6a). The analogous R-mutant fusions had the same activity as AID-AID E58A not only in *E. coli* but also for CSR (Fig. 6a). In this experiment, catalytic activity is derived from the R-mutant AID, while AID E58A provides the RR domain function. This was confirmed by further mutating the RR domain of AID-E58A in these fusions, which did not affect activity in *E. coli* but eliminated CSR activity (Fig. 6a). This demonstrates that the two functions are modular and suggests

---

**Fig. 6** Protein–protein interactions of AID and R-mutants. (a) (Left) schemes of AID, AID E58A and fusions thereof indicating additional mutations. (Middle) Relative ability to produce rifampicin resistant (Rif-R) *E. coli* colonies is shown by means (bars) of median values (dots) from 2 to 4 independent experiments (5 clones/experiment), normalized to the median of AID from each experiment. (Right) Relative CSR capacity of the same proteins expressed −/− mouse B cells. Means (bars) from 4 to 10 independent mice (dots) from 2 to 5 independent experiments are shown, normalized to AID.

(b) MA plot showing the fold change in spectral counts (s.c.) as a function of average s.c. of all hits for wt AID versus R-mutants. Red and blue dots represent significant differences identified according to negative binomial distribution by the DESeq2 software. (c) Representative WB of the AID-BirA* fusions (anti-GFP) and biotinylated proteins from whole-cell lysates of AID-defiltered CH12 B cells reconstituted with AID variant-BirA* fusions. Ratios of Spt6 to AID signal normalized to AID-BirA* are indicated. Representative of two independent experiments.

(d) Log2 fold change of Hits (n) in selected known AID interacting partners. Circle size indicates relative abundance normalized to the variant with the most s.c. Circle colour represent actual s.c. Scales are shown.

(e) MA plot showing the fold change in spectral counts (s.c.) as a function of average s.c. of all hits for wt AID versus R-mutants. Red and blue dots represent significant differences identified according to negative binomial distribution by the DESeq2 software.

(f) Broad functional categories of BioID associations found reduced in the R-mutants compared to wt AID. (Right) Circle plots for hits in the indicated categories. Anti-GFP pull down from whole-cell lysates of AID-deficient CH12 B cells reconstituted with AID variant-GFP fusions. Sp6, Sp5, RNAPII and AID were detected by WB from immunoprecipitates or lysates. Representative of two independent experiments. Streptavidin pull down of biotinylated proteins from whole-cell lysates of AID-deficient CH12 B cells reconstituted with AID variant-BirA* fusions after 24 h pulse of biotin. Sp6 and AID were detected by WB from either pull downs or lysates. Ratios of Sp6 to AID signal normalized to AID-BirA* are indicated. Representative of two independent experiments.
that the R-mutants are still intrinsically capable of deaminating the Ig locus.

**AID R-mutants lose specific interactions in vivo.** Our results suggest that the RR domain may mediate an interaction necessary for targeting AID activity. To identify proteins that associate with AID in a manner that depends on the intact RR domain, we compared the interactome of AID and two R-mutants in live B cells. We used BioID, a proximity-based biotin labelling technique in which a bait is fused to the promiscuous BirA* biotin ligation that can label the protein environment of the bait in an ~10 nm radius. We generated an AID-BirA* fusion, which was active for CSR in transduced Aicda−/− B cells, and its R171Y or R178D derivatives (Fig. 6b). Addition of biotin to the cultures led to biotinylation of proteins in cells expressing BirA* fusions (Fig. 6c), proportionally to AID-BirA* expression levels (Supplementary Fig. 5a, b). The vast majority of AID interactions were conserved in the R-mutants, including many validated AID interactors (Fig. 6d, e and Supplementary Fig. 5f), further confirming the structural integrity of the mutants. Using four different statistical methods, we identified 54 proteins that showed reduced interaction with both R-mutants compared to AID, 29 of these were identified by at least two methods (Fig. 6d, Supplementary Fig. 5c–e, and Supplementary Table 1, see Methods). Functional annotation of the BioID hits showed nine chromatin-associated factors, including the histone chaperones Spt6 (Supp6) and Nap114 as well as the P-TEFb component Cdk9, which are functionally linked to transcription elongation (Fig. 6f). We also found seven factors linked to co-transcriptional RNA processing: splicing and mRNA transport factors, as well as Dicer which can bind to dsRNA resulting from convergent transcription induced by R-loops. Of these, only Spt6 has been previously shown to interact with AID by co-immunoprecipitation (co-IP). Interestingly, standard pull down from cell extracts showed that GFP-tagged R-mutants still co-IP Spt6 (Fig. 6g). However, streptavidin pull down of proteins biotinylated in live CH12 B cells expressing AID- or R-mutants-BirA* confirmed the lack of interaction with the R-mutants in live cells (Fig. 6h). These data together strongly suggest that the BioID result reflects the loss of the functional interaction between the R-mutants and Spt6, rather than their physical inability to form a complex with Spt6. We conclude that the defect in the R-mutants lies after chromatin association but prior to the transcriptional step in which Spt6 is recruited.

**AID R-mutants fail to occupy the target genes body.** The R-mutants were able to interact with RNAPII (Fig. 6c, g), yet did not mutate B cells. As Spt6 is recruited to promoter-proximal paused RNAPII, and Spt6 only to active transcription, we hypothesized that the R-mutants might fail to progress from paused to elongating RNAPII. We therefore compared the occupancy of wt AID and the R-mutants near the transcription start site (TSS) versus the gene body at the physiological and one prominent off-targets of AID. We used an antibody against the E5 domain of AID to perform ChIP from Aicda−/− mouse B cells complemented with either wt AID or the R-mutants and stimulated for CSR to IgG1. We first analysed the Sµ-region, the major AID target at the IgH locus during CSR, comparing AID occupancy at the region around the TSS where paused RNAPII is expected, to the downstream Sµ region (Fig. 7a). ChIP showed that wt AID was present in all amplicons but was highest at the Sµ region (Fig. 7a), where elongating RNAPII was previously shown to stall. This profile was not an artefact of AID overexpression, as endogenous AID had a similar distribution (Fig. 7b). Spt6 was also present in all amplicons (Fig. 7b), as expected for a highly transcribed gene. In contrast, the R-mutants were present at the TSS but depleted from the Sµ gene body (Fig. 7a), and we obtained the same result at the Sy1 region (Fig. 7b). Moreover, we were able to detect low but reproducible ChIP signals for AID at the IL4Ra locus, a known AID off-target, where the R-mutants were specifically depleted within the gene body but showed normal occupancy at the promoter, compared to wt AID (Fig. 7d). As a control, no AID or R-mutants were detected at either the promoter or the gene body of Gapdh (Fig. 7e), which is not occupied by endogenous AID. To test whether the same defect could underlie the inability of the R-mutants to do SHM, we analysed AID occupancy at the IgV region in complemented DT40 Aicda−/− ΔΨVΔVA B cells. Again, the R-mutants were equally recruited to the IgV promoter but substantially depleted from the gene body, compared to wt AID (Fig. 7f). As an additional control in all ChIPs, we used AID E58A, which showed similar occupancy as wt AID at all at the regions tested, indicating that the simple lack of deamination ability could not explain the observed occupancy profile. All AID variants showed similar expression levels in B cells (Fig. 7g).

We conclude that the R-mutants can occupy the promoter-proximal regions of all physiological and at least one off AID targets, yet fail to occupy the corresponding gene bodies. Thus, the lack of association between the R-mutants and Spt6 most likely reflects the uncoupling of the AID R-mutants from elongating and/or stalled RNAPII, which suggests the existence of a licensing step for productive targeting of AID.

**Constitutively nuclear AID bypasses licensing.** The RR domain was dispensable for mutagenesis in E. coli, which also depends on transcription, thus suggesting that the licensing step can be bypassed. We therefore asked whether a constitutively nuclear AID might rescue the defect of the R-mutants by mimicking the E. coli situation, where proximity to the genome is not regulated. We introduced the R-mutants into AIDΔE5, a nuclear AID variant, that results in >3-fold higher steady-state levels of AID in mouse B cells complemented with E. coli AIDΔE5 ΔΨVΔVA (Fig. 8b). In contrast to their effect on full-length AID, mutations R174E and R178D did not prevent the detrimental effect on cell fitness or ability to generate H2AX foci of AIDΔE5 in B cells (Fig. 8c, d). Moreover, when expressed in Aicda−/− DT40 B cells, AIDΔE5 R174E and AIDΔE5 R178D produced similar SHM levels than AIDΔE5 at the IgV, as judged by the IgM-loss assay as well as by mutation frequency and profile (Fig. 8e, f and Supplementary Fig. 6). While AIDΔE5 is deficient for CSR, it can target and deaminate the Sµ. AIDΔE5 R174E deaminated the Sµ with the same frequency and profile as AIDΔE5 in complemented Aicda−/− Ung primary B cells, with R171Y and R178D producing less but still substantial mutations (Fig. 8g). Interestingly, AIDΔE5 R171Y consistently produced less SHM, DNA damage and Sµ mutations than AIDΔE5 and the other two mutants, despite equal expression (Fig. 8h), suggesting an additional role of this residue for mutagenesis in B cells. Nonetheless, the results with the other two mutants indicate that at least R174 and R178 are dispensable for deaminating the physiological AID targets when the proposed licensing step is bypassed, which can be achieved when nuclear AID levels are not limiting. We conclude that any possible deamination defect in the specific R-mutants used here are insufficient to explain their biological defects and that limiting AID nuclear levels helps enforcing licensing.

**Discussion** The genes that are mutated by AID in B cells share a transcriptional landscape associated with super-enhancers and convergent...
transcription. This demonstrates that AID targeting is based on the local chromatin architecture, rather than the fixed features of the target genes. The mechanism underlying the productive targeting of AID remains unclear, as it cannot be explained by differential recruitment of AID given that AID occupancy is insufficient for mutation. Here we uncover an obligatory licensing step for AID that provides insight into this fundamental question.

We show that specific substitutions of either Arg residues 174 or 178, and to a lesser extent Arg 171, dissociate the ability of AID to associate with chromatin and occupy promoter-proximal regions from its ability to induce mutations in the downstream genes in B cells. It was recently proposed that these Arg in AID α6 form part of an assistant patch required for deaminating structured DNA, especially G4 DNA that forms at the S-regions. Several lines of evidence indicate that the functional defect of our R-mutants cannot be explained by a potential defect in deaminating structured DNA: (1) None of the R-mutants affect transcription; (2) Catalytically inactive AID E58A can occupy the target regions; (3) When licensing is bypassed using the AIDΔE5 variant, R174E has no effect on, and R178D and R171Y only reduce by 2–3-fold, the ability to mutate the Sμ (Fig. 8g), a much less dramatic effect than these mutations have on CSR activity. In fact, AID R174E, which has virtually no CSR activity (Fig. 1d and ref. 50), retains at least 50% enzymatic activity on G4 DNA; (4) the IgV does not form G4 DNA and AIDΔE5 R174E and AIDΔE5 R178D show normal SHM (Fig. 8f); (5) These nuclear AID mutants also produce normal levels of off-target DNA damage in B cells (Fig. 8d). On the other hand, AIDΔE5 R171Y is less efficient in deaminating the Sμ and even more impaired for SHM and DNA damage, indicating an additional defect compared to R174E and R178D. This defect could be biochemical but remains to be studied. Thus our results do not dispute the involvement of the α6 Arg residues in assisting deamination of G4 DNA but provide good evidence that they have an additional function, which we propose is AID licensing. AID and the R-mutants broadly associate with chromatin. We show that Spt5 accumulation can promote AID recruitment to chromatin, fitting the observation that AID accumulates at

![Fig. 7 AID R-mutants are uncoupled from transcription elongation.](image)

a) ChIP qPCR results using anti-AID antibody on extracts from reconstituted Aicda−/− mouse B cells activated with LPS and IL-4. Means ± s.e.m. from two independent experiments are plotted. A scheme of each genomic region analysed is shown at the top of each panel with the qPCR amplicons indicated. b) ChIP qPCR for endogenous AID and Spt6 from wt mouse B cells treated and analysed as in a, including amplicons in the non-transcribed Sμ region. Means ± s.e.m. from three independent experiments are plotted. c) (Bottom) ChIP qPCR for AID from reconstituted DT40 Aicda−/− ΔΨV γ B cells. Means ± s.e.m. from three independent experiments are plotted. An additional amplicon within Gapdh was used as a negative control. g) WB of cell extract from reconstituted Aicda−/− mouse B cells, probed with antibodies recognizing GFP and AID. For gel source data, see supplementary Fig. 7.
genomic Spt5 peaks\(^{17,48}\), although this could be indirect and other factors might share this capacity. Several factors seem to contribute to the association of AID with chromatin. The physical presence of RNAPII may be required, but transcription itself is dispensable. Our data indicate that Spt5 is required to maintain chromatin-wide association of AID, as Spt5 depletion evicts AID from chromatin. On the other hand, RNase treatment also evicts AID from the chromatin but does not evict Spt5. Moreover, a large proportion of chromatin-associated AID is in a fraction devoid of RNAPII or Spt5, as shown by salt fractionation. One possible explanation for these observations is that the broad distribution of AID at the chromatin reflects dynamic associations or shuttling between Spt5 complexes and at least another distinct RNA-dependent complex. We propose that dynamic association with chromatin and accumulation at Spt5-rich regions allow AID to sample multiple loci. Similar diffusion and accumulation behaviour has been described for transcription factors searching for their target DNA sequences\(^ {52}\). Unlike transcription factors, AID does not recognize a specific DNA sequence\(^ {2}\), and the licensing step we report would provide the extra level of
regulation to recognize deamination targets after associating with Sp15-rich promoters.

The current model for AID targeting posits that AID mutates when RNAPII stalls, but this does not explain how AID associates with stalled polymerase, beyond invoking a role for Sp15. Like wt AID, the R-mutants associate with Sp15 and are recruited to the promoter-proximal region of various target sites (Fig. 7) but fail to act, thereby uncovering an additional level of regulation. Our results are consistent with AID being recruited to the TSS and progressing with transcription elongation until the polymerase becomes stalled, rather than being directly recruited to stalled polymerases. Abnormally high nuclear AID levels can bypass this licensing step, possibly by direct recruitment to stalled polymerases, resulting in a highly mutagenic enzyme. This would provide a rationale for the abundance of mechanisms that restrict AID nuclear access, which would serve to enforce licensing.

In conclusion, our data are consistent with a model (Fig. 8) in which AID samples the genome by dynamic and broad chromatin association. The low concentration of AID in the nucleus would favor its accumulation at promoter-proximally paused RNAPII, probably via Sp15 but likely involving additional factors. Only those loci within a permissive transcriptional landscape would contain the factors that mediate the coupling of AID to transcription elongation, licensing AID to deaminate the downstream region, as shown for the IgV and Sµ regions. The R-mutants fail to interact with Sp15 in live cells but have not lost biochemical interaction ability with this factor, suggesting that licensing takes place before their encounter during transcription, rather than Sp15 being directly involved. Elucidating the molecular mechanism of licensing will require considerable work but it is likely that it takes place within super-enhancers. Our result at Ibra suggests that licensing permits mutation of AID off-targets (which will need to be confirmed genome-wide) but is probably more active at the Ig loci and absent from those loci that are not mutated by AID. We speculate that the nuclear deaminase APOBEC3B and other AID paralogues with antiviral activity have lost the RR domain and thereby licensing to prevent mutating the self-genome during transcription.

Methods

Animals. Aicda−/− mice (a gift from Dr. T. Honjo, Kyoto University, Japan) and Aicda+/−/Ung−/− mice (Ung−/− mice were a gift of Dr. H. Krokan, Norwegian University of Science and Technology, Norway) in C57BL/6J background were bred at the specific pathogen-free facility of the Institut de Recherches Cliniques de Montreal. Male and female mice of 6 weeks to 8 months old were used as a source of B cells. Animals were euthanized in CO2 chamber. All animal work was approved by the IRCM animal protection committee (AUP 2015-10) in accordance with the guidelines of the Canadian Council for Animal Care.

DNA constructs. Retroviral vector pMXs human AID-, AIDΔE5- and AIDΔ7.3-ires-GFP, as well as mouse AID-GFP, have been described. Human AID fusions were assembled as BamHI-AID-EcoRI-Linker-HindIII-AID-ES8A-Xhol cassettes into pTrcHisA (ThermoFisher) or pMX-ires-GFP (Cell Biolabs). The linker encoded a flexible (SGGGG)3 peptide. Human SPT3 was PCR amplified and cloned into a gateway-compatible mCherry-LaCrep destination vector (a gift from Dr. D. DeRocher, University of Toronto, Canada). Mouse AID and Linker-BirA′ were PCR amplified using gateway-compatible primers and cloned into appropriate donor vectors to generate AID-BirA′ fusions into a homologous gateway-compatible pMX-ires-GFP bearing a gateway cassette cloned BamHI-EcoRI by using Multisite gateway technology (Invitrogen). AID variants were engineered using PCR amplification with ad hoc oligonucleotides or by quick-change site-directed mutagenesis using KOD1 DNA polymerase (Toyobo Inc.). For oligonucleotide sequences, see Supplementary Table 2.

Cell culture and transduction. CH12 cells (A kind gift from Dr T Hongo, Kyoto University, Japan) and primary B lymphocytes were cultured in RPMI 1640 media (Wiset) at 37 °C with 5% (vol/vol) CO2. U2OS Iaco cells (Obtained from Dr. A Arthwein, Lady Davis Institute, Montreal) were cultured in McCoy’s 5a media (Wiset). Media were supplemented with 10% foetal bovine serum (Wiset), 1% penicillin/streptomycin (Wiset) and 0.1 mM 2-mercaptopethanol (Biohop). DT40 Aicda−/− ΔΨV cells (A kind gift from Dr H Arakawa, IFOM, Italy) were cultured in RPMI 1640 supplemented with as above plus 1% chicken serum (Wiset). CH12 cells stably expressing short hairpin RNA against AID have been described. All cell lines were regularly tested for mycoplasma, and cell line identity was inferred from relevant functional assays. Naïve splenic B cells from Aicda−/− or Aicda+/−/Ung−/− mice were isolated from total spleencells depleted with anti-CD43 microbeads (Miltenyi Biotec, Cat.#130-049-801) in an autoMACS cell separator (Miltenyi Biotec). For retroviral complementation of DT40 or CH12 cells, VSV-G, MLY gag-pol and pMx vectors (1:1:4 ratio, 2.5 μg DNA total) were transduced into HEK293 cells using Trans-IT LT-1 (Minus Bio, Cat.# MIR 2905). HEK293 were chosen because of their high transfection efficiency and virions’ production confirmed in test infection assays. Retrovirus for primary cell transduction was produced using Plat-E ecdotropic packaging cells (A kind gift of Dr. T Kitamura, University of Tokyo, Japan) and transduced with pMx vectors. For infections, 1 mL of HEK293 or 1.5 mL Plat-E supernatant at 48 h post-transfection was used to infect 106 B cells in 24-well plates, in the presence of 8 μM L1 polybrene (Sigma, Hexadimethrine bromide Cat.# H9268), by spinning at 600 × g for 90 min at 32 °C. Medium was replaced 4 h later.

Reagents and antibodies. Stock aliquots: 50 μg mL−1 LMB in ethanol (LC Laboratories), 20 mM Dihemadin B in dimethyl sulphoxide (NCS 325319; provided by the Natural Products Branch, National Cancer Institute, Bethesda, MD) and 2 mM Actinomycin D in DMSO (Santa Cruz Biotechnology). Drugs were kept at −20 °C in the dark and blotted dried before each experiment. Antibodies used for immunofluorescence (IF): (1F): for human AID; rat monoclonal antibody (mAb) anti-AID (1:500, EK 5G9, Cell Signaling), for mouse AID; rat mAb anti-AID (1:250, mAb-2, Bioscience), rabbit anti-SPT3 (1:500, EPR51452(2), Abcam), rabbit anti-RNAPII (1:100, H-224, SC Biotechnology), goat anti-LaminB (1:500, M-20, SC Biotechnology), Rabbit anti-ɣH2AX (1:800, #2577, Cell Signaling). With the exception of anti-rat Dylight 550 (1:500, SA5-10027, ThermoFisher Scientific), all

Fig. 8 Nuclear exclusion is necessary to enforce AID licensing. a (Left) Representative confocal microscopic images of AID-GFP and AIDΔE5-GFP fusions in CH12 B cells. (Right) Nuclear GFP signal was measured as the overall GFP signal overlapping with a nuclear mask, generated using Dapi signal. Cells expressing each construct were fixed and imaged with identical settings in parallel. Individual cell values (dots) are plotted relative to the median of AID-GFP, from one experiment. Control shows untransfected cells signals. Differences were evaluated by unpaired, two tailed t-test. b Relative mutagenic activity of AIDΔE5 variants in E. coli, measured as the frequency of rifampicin-resistant (Rif-R) colony-forming units (cfu). Means (bars) of medians (dots) from 2 independent experiments (5 cultures/experiment), normalized to AID. c Effect of AID or AIDΔE5 variants-ires-GFP on the competitive growth of transduced AID-deficient CH12 B cells. Means GFP+/GFP− ratio ± s.e.m. over time from two independent experiments, each normalized to maximal value. d (Left) Representative confocal microscopic images of GFP and γH2AX immunofluorescence with DNA staining (Dapi) in Aicda+/− mouse B cells activated with LPS and IL-4 complemented with AID variants-ires-GFP vectors. (Right) The proportion of cells with ≥ 5 γH2AX foci per nucleus is plotted from 1 experiment, with n cells counted. d Magnification 630×. Scale bar, 10 μm. e SHM capacity of AIDΔE5 variants-ires-GFP measured by IgM-loss over time in complemented DT40 Aicda+/− ΔΨV cells. B cells. One of the two independent experiments is shown, f Number and frequency (mutations per base pair) of mutations scored at the IgV of DT40 Aicda+/− ΔΨV. B cells expressing AIDΔE5 variants obtained from GFP+ cells sorted at day 3 post-transduction (see a) from one experiment. Mutation load pie charts, with slices representing proportion of sequences with the indicated number of mutations and total sequences analysed indicated in the centre. g (Left) Mutation profiles and (right) mutation load, as in f scored at the Sµ region of Aicda+/− Une−/− mouse B cells transduced with AIDΔE5 variants. h WB of cell extracts from reconstituted mouse B cells probed with antibodies recognizing GFP and the N-terminus of AID. For gel source data, see supplementary Fig. 7. i Model for AID targeting (see Discussion)
other IF secondary Abs were AlexaFluor conjugates (1:500, invitrogen): anti-rat-IgG1, 400 nM (GFP), 543 nM (Alexa546 and DyLight550) and 633 nM (Alexa680), using 100 nM DMSO, 10 ng mL
0.1 mM EDTA, 1 mM DTT), and centrifuged at 20,000 RPM for 15 min. The nuclear wash protocol was adapted from ref. 39. Briefly, CH12 cells were plated on poly-L-lysine coverslips and washed 1× with CSK buffer (10 mM PIPES, 300 mM sucrose, 200 mM MgCl2, 1 mM EDTA and 1× fresh complete protease inhibitor (CPI, Roche)). Cells were then either fixed directly in formaldehyde (whole cell) or washed to remove cytoplasm and loosely held proteins. Washing was done by sequentially incubating the samples 1× in: CSK buffer for 1 min, CSK + 0.01% triton X-100 for 1 min, CSK + 0.5% triton X-100 for 20 min. For nuclease treatments, the last wash was 10 min on ice, then 10 min at 37 °C in CSK buffer containing 10 mg mL
10 mM Tris, pH 7.5, 2.5 μM MgCl2, 50 mM NaCl and 1 mM DTT) in a
0.5% triton X-100 (PBS-T), 0.01% Triton X-100 for 20 min. For nuclease treatments, the samples were incubated in: CSK buffer for 1 min, CSK + 0.01% triton X-100 for 1 min, CSK + 0.5% triton X-100 for 20 min. For nuclease treatments, the last wash was 10 min on ice, then 10 min at 37 °C in CSK buffer containing 10 mg mL
anti-rat-IgG1, 400 nM (GFP), 543 nM (Alexa546 and DyLight550) and 633 nM (Alexa680), using 100 nM DMSO, 10 ng mL
400 nM (GFP), 543 nM (Alexa546 and DyLight550) and 633 nM (Alexa680), using anti-IgM conjugated with R-phycoerythrin (clone M-1, SouthernBiotech). Gating strategy for CSR is shown in Supplementary Fig. 8a.
34080, Thermo Scientific.
577 complemented by retroviral transduction with pMX AicdaΔ
EcoRI fragments encoding the IgV region was amplified36. Oligonucleotide sequences are in Supplementary Table 2.
peptidase. SHM was measured by
Supplementary Table 2. Amplicons were cloned into pGEM-T easy (Promega), and individual clones were sequenced at Macrogen (Seoul, Korea). The distribution and frequency of mutations were calculated as described37. Sequences were aligned using Sequencher (Gene Codes Corp.) and trimmed to remove vector and primer sequences, and electropherograms were manually inspected to confirm bona fide mutations before compiling the total and per sequence mutation numbers.

other IF secondary Abs were AlexaFluor conjugates (1:500, invitrogen): anti-rat-IgG1, 400 nM (GFP), 543 nM (Alexa546 and DyLight550) and 633 nM (Alexa680), using 100 nM DMSO, 10 ng mL
0.1 mM EDTA, 1 mM DTT), and centrifuged at 20,000 RPM for 15 min. The nuclear wash protocol was adapted from ref. 39. Briefly, CH12 cells were plated on poly-L-lysine coverslips and washed 1× with CSK buffer (10 mM PIPES, 300 mM sucrose, 200 mM MgCl2, 1 mM EDTA and 1× fresh complete protease inhibitor (CPI, Roche)). Cells were then either fixed directly in formaldehyde (whole cell) or washed to remove cytoplasm and loosely held proteins. Washing was done by sequentially incubating the samples 1× in: CSK buffer for 1 min, CSK + 0.01% triton X-100 for 1 min, CSK + 0.5% triton X-100 for 20 min. For nuclease treatments, the last wash was 10 min on ice, then 10 min at 37 °C in CSK buffer containing 10 mg mL
10 mM Tris, pH 7.5, 2.5 μM MgCl2, 50 mM NaCl and 1 mM DTT) in a

34080, Thermo Scientific.

Competitive growth assays. The effect on cell fitness was analysed by using AID-deficient CH12 B cells complemented with AID variants-ires-GFP or empty vector (GFP) by retroviral transduction. Depending on the transduction efficiency, the population of infected cells was sometimes kept as is or mixed with untransduced (GFP−) cell. Cells were maintained in culture with normal medium monitoring the GFP+ to GFP− cell ratio over time by flow cytometry immediately after transduction. Gating strategy is shown in Supplementary Fig. 8d. In some experiments, cells 1× 106 at the Su+ region, we amplified a 1749 bp fragment starting at the TSS in experiments using full-length AID or 607 bp fragment overlapping the S-region in Supplementary Table 2. Amplicons were cloned into pGEM-T easy (Promega), and individual clones were sequenced at Macrogen (Seoul, Korea). The distribution and frequency of mutations were calculated as described37. Sequences were aligned using Sequencher (Gene Codes Corp.) and trimmed to remove vector and primer sequences, and electropherograms were manually inspected to confirm bona fide mutations before compiling the total and per sequence mutation numbers.

34080, Thermo Scientific.

AID shutting and nuclear wash protocol. Retrovirally complemented AID-deficient CH12 cells were treated for 2 h with DMSO, 10 ng mL
LMB (a CRM1 inhibitor), 100 nM Did B (an EEF1A inhibitor) or both drugs combined before harvesting. Cells were washed with phosphate-buffered saline (PBS), then plated on poly-L-lysine-coated coverslips and fixed in 3% (v/v) formaldehyde for 10 min and then washed 3× in PBS. The nuclear wash protocol was adapted from ref. 38. Briefly, CH12 cells were plated on poly-L-lysine coverslips and washed 1× with CSK buffer (10 mM PIPES, 300 mM sucrose, 200 mM NaCl, 5 mM MgCl2, 1 mM EDTA and 1× fresh complete protease inhibitor (CPI, Roche)). Cells were then either fixed directly in formaldehyde (whole cell) or washed to remove cytoplasm and loosely held proteins. Washing was done by sequentially incubating the samples 1× in: CSK buffer for 1 min, CSK + 0.01% triton X-100 for 1 min, CSK + 0.5% triton X-100 for 20 min. For nuclease treatments, the last wash was 10 min on ice, then 10 min at 37 °C in CSK buffer containing 10 mg mL
10 mM Tris, pH 7.5, 2.5 μM MgCl2, 50 mM NaCl and 1 mM DTT) in a

Microscopy. Images were acquired at room temperature using ZEN 2010 on a Zeiss LSM 700 confocal microscope with excitation lasers at 405 nM (Dapi), 488 nM (GFP), 543 nM (Alexa546 and Dylight550) and 633 nM (Alexa680), using either 405/1.3 or 633/1.4 oil immersion objectives, and collected with a Hamamatsu Photometrics CC:14000, 1 rat at Su was analysed in uncaged AicdaΔ
EcoRI fragments encoding the IgV region was amplified36. Oligonucleotide sequences are in Supplementary Table 2. Amplicons were cloned into pGEM-T easy (Promega), and individual clones were sequenced at Macrogen (Seoul, Korea). The distribution and frequency of mutations were calculated as described37. Sequences were aligned using Sequencher (Gene Codes Corp.) and trimmed to remove vector and primer sequences, and electropherograms were manually inspected to confirm bona fide mutations before compiling the total and per sequence mutation numbers.

34080, Thermo Scientific.

AID shutting and nuclear wash protocol. Retrovirally complemented AID-deficient CH12 cells were treated for 2 h with DMSO, 10 ng mL
LMB (a CRM1 inhibitor), 100 nM Did B (an EEF1A inhibitor) or both drugs combined before harvesting. Cells were washed with phosphate-buffered saline (PBS), then plated on poly-L-lysine-coated coverslips and fixed in 3% (v/v) formaldehyde for 10 min and then washed 3× in PBS. The nuclear wash protocol was adapted from ref. 38. Briefly, CH12 cells were plated on poly-L-lysine coverslips and washed 1× with CSK buffer (10 mM PIPES, 300 mM sucrose, 200 mM NaCl, 5 mM MgCl2, 1 mM EDTA and 1× fresh complete protease inhibitor (CPI, Roche)). Cells were then either fixed directly in formaldehyde (whole cell) or washed to remove cytoplasm and loosely held proteins. Washing was done by sequentially incubating the samples 1× in: CSK buffer for 1 min, CSK + 0.01% triton X-100 for 1 min, CSK + 0.5% triton X-100 for 20 min. For nuclease treatments, the last wash was 10 min on ice, then 10 min at 37 °C in CSK buffer containing 10 mg mL
10 mM Tris, pH 7.5, 2.5 μM MgCl2, 50 mM NaCl and 1 mM DTT) in a

placed at 37 °C for 5 min. CaCl₂ was added to 1 mM and DNA digested by adding Micrococcal nuclease (New England Biolabs) to 9.6 μL for 10 min. Digestion was stopped by adding EDTA to 2 mM final. An aliquot was saved as the total nuclear fraction. Nuclei were pelleted for 10 min at 10,000 × g, and the supernatant was saved as the MNase fraction. Nuclei were then re-suspended in 700 μL of 150 mM extraction buffer (10 mM Tris pH 7.4, 140 mM NaCl, 1 mM MgCl₂, 2 mM EGTA, 0.1 % Triton X-100, 1× CPI) and incubated for 2 h on a rocker at 4 °C. After centrifugation for 10 min at 500 × g, the supernatant was saved as the 150 mM fraction. Nuclei were then re-suspended in 700 μL of extraction buffer, with 590 mM NaCl and incubated overnight at 4 °C. After centrifugation for 10 min at 500 × g, the supernatant was saved as the 600 mM fraction. Nuclei were finally re-suspended in 100 μL Tris pH 7.4, 150 mM NaCl, 1 mM EDTA and kept as the pellet fraction. DNA was purified using a PCR Purification Kit (BioBasic, Cat.# RS614) and then run on an agarose gel in order to confirm efficient digestion and DNA extraction. Protein fractions were analysed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and WB.

Chromatin immunoprecipitation. For ChIP in primary B cells, naive Aicda/−/− mouse B cells were stimulated with 12 μg/mL LPS and 50 ng/mL IL-4 and retrovirally transduced 24 h later. Cells were harvested at 18–20 h postinfection when GFP+/+ cell proportions were 30–50%. For ChIP in DT40 Aicda/−/− cells, repeated retroviral transduction achieved ~75–90% GFP+ cells, and cells were expanded for 48 h. ChIP procedures have been described in detail. Briefly, cells fixed with 1% formaldehyde for 30 min, were lysed in RIPA buffer and sonicated for 30 s at 30% amplitude (3 × 10 s bursts with a 2 s interval). Protein A-Sepharose slurry before incubating overnight with 2–5 μg anti-AID antibody (328.8b, Active Motif) at 4 °C. In ChIPs from DT40 extract, Protein G magnetic microbeads were used and IP purified by mMacs columns (Miltenyi Biotech). DNA was purified and used as template in real-time PCR reactions containing 1× SYBR Green Mix (Applied Biosystems), 1/10 fraction ChIP-enriched DNA and 100 nM primers (see Supplementary Table 2 for primer sequences). Plates were read in an Applied Biosystems StepOnePlus instrument. Standard curves with different amounts of the input extracts were run in each plate for each individual amplification. Brieﬂy, reactions were run in 1× extraction buffer (10 mM Tris pH 7.4, 140 mM NaCl, 1 mM MgCl₂, 2 mM EGTA, 0.1 % Triton X-100, 1× CPI) and incubated for 2 h on a rocker at 4 °C. After centrifugation for 10 min at 500 × g, the supernatant was saved as the 600 mM fraction. Nuclei were finally re-suspended in 100 μL Tris pH 7.4, 150 mM NaCl, 1 mM EDTA and kept as the pellet fraction. Protein identifications were performed using Human RefSeq Version 57 and the iprophet pipeline62 integrated in ProHits63.

Statistical analyses. We used four different methods to identify wt AID–BioID interactions that were signiﬁcantly and consistently reduced in both R-mutants tested. Comparative results for the four methods are provided in Supplementary Table 1. All methods were implemented through ad-hoc scripts in R version 3.3.2 that are available upon request. Tables were handled in R using the data.table package. Method 1 used fold-enrichment (Fold), calculated using mean spectral counts (s.c.). First, s.c. were normalized in all samples to their corresponding Bira+ s.c. levels. Interactors with at least 2.5-fold enrichment in wt AID over the R mutants or vice versa were considered differential interactors. To eliminate interactions that were not AID-specific, only interactions with at least five-fold enrichment over APOBEC2 were considered. Method 2 calculated a Z-score using log-transformed fold-enrichment values (Normz). Method 3 calculated local Z-scores using a sliding window of 10% of the data points around the candidate in an ratio-intensity R-I plot (Maz)65. In methods 2 and 3, hits with high or local, respectively, Z-scores 2.2 (i.e., 2× the s.d. from the mean) were considered as differential interactors. Methods 1–3 assume normal distribution of the data. Method 4 used DESeq2 v.1.14.166, an R package that uses generalized linear models to fit a negative binomial distribution to the data (DispRes). The R-mutants were deﬁned as reference and default parameters from the package were used to carry out the analysis and calculate differential interactions for AID. Preys with multiple-testing adjusted p-values (Benjamini–Hochberg procedure) < 0.1 were considered as signiﬁcant. All ﬁgures were plotted using ggplot2 package in R. Dot plots were generated using the ProHit-Viz web tool67.

Where indicated, for pairwise comparisons of treatments or conditions, differences were evaluated by unpaired, two tailed t-test with alpha = 0.05 using Prism (GraphPad).

Code availability. R scripts used for statistical analyses of BioID data are available from the corresponding author upon reasonable request.

Data availability. Sequences of mouse Sµ and DT40 IgV regions have been deposited in GenBank as follows: For Sµ region in Fig. 8g, AIDA5 MG904385-MG904404, AIDA5 MG904427. AIDA5 MG904428-MG904449, AIDA5 MG904450–MG904469. For DT10 IgV region in Fig. 8f, AIDA5 MG904474–MG904484. AIDA5 MG904485–MG904492. AIDA5 MG904493–MG904502, AIDA5 MG904453–MG904557, AIDA5 MG904458–MG904592. For Sµ region in Fig. 3a, AID MG904593-MG904622, AID MG904623-MG904654. AID MG904655-MG904685. Mass spectrometric data have been deposited in MassIVE under ID MSV000081963. All other data generated or analyzed during this study are included in this published article and its supplementary information files.

Received: 11 August 2017 Accepted: 8 February 2018
Published online: 28 March 2018

References
1. Methot, S. P. & Di Noia, J. M. Molecular mechanisms of somatic hypermutation and cell switch recombination. Adv. Immunol. 133, 37–87 (2017).
2. Peled, J. U. et al. The biochemistry of somatic hypermutation. Annu. Rev. Immunol. 26, 481–511 (2008).
3. 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).
4. Casellas, R. et al. Mutations, kataegis and translocations in B cells: understanding AID promiscuous activity. Nat. Rev. Immunol. 16, 164–176 (2016).
5. Cortizas, E. M. et al. UNG protects B cells from AID-induced telomere loss. J. Exp. Med. 213, 2459–2472 (2016).
6. Hasham, M. G. et al. Widespread genomic breaks generated by activation-induced cytidine deaminase are prevented by homologous recombination. Nat. Immunol. 11, 820–826 (2010).
7. Liu, M. et al. Two levels of protection for the B cell genome during somatic hypermutation. Nature 451, 841–845 (2008).
8. Vaidyanathan, B., Yen, W.-F., Pucella, J. N. & Chaudhuri, J. AIDing chromatin and transcription-coupled orchestration of immunoglobulin class switch recombination. Front. Immunol. 5, 120 (2014).
9. Robbiani, D. F. & Nussenzweig, M. C. Chromosome translocation, B cell hypermutation, and activation-induced cytidine deaminase. Annu. Rev. Pathol. 8, 79–103 (2013).
10. Yeap, L.-S. et al. Sequence-intrinsic mechanisms that target AID mutational outcomes on antibody genes. Cell 163, 1124–1137 (2015).
11. Buerstedde, J.-M., Alikunia, J., Arakawa, H., McDonald, J. J. & Schatz, D. G. Targeting of somatic hypermutation by immunoglobulin enhancer and enhancer-like sequences. *Proc. Biol. 12*, 11001831 (2017).

12. Dunnick, W. A. et al. Switch recombination and somatic hypermutation are controlled by the heavy chain 3′ enhancer region. *J. Exp. Med. 206*, 2613–2623 (2009).

13. Qian, J. et al. B cell super-enhancers and regulatory clusters recruit AID transcription factors. *Cell 159*, 1524–1537 (2014).

14. Meng, F.-L. et al. Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell 159*, 1538–1548 (2014).

15. Pefanis, E. et al. Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature 514*, 389–393 (2014).

16. Nambu, Y. et al. Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science 302*, 2133–2140 (2003).

17. 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).

18. Basu, U. et al. The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell 144*, 353–363 (2011).

19. Matthews, A. J., Husain, S. & Chaudhuri, J. Binding of AID to DNA does not correlate with mutational activity. *J. Immunol. 193*, 252–257 (2014).

20. Hakim, O. et al. DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. *Nature 484*, 69–74 (2012).

21. Staszewski, O. et al. Activation-induced cytidine deaminase induces reproducible DNA breaks at many non-Ig Loci in activated B cells. *Mol. Cell 52*, 231–242 (2013).

22. Conticello, S. G. et al. Interaction between antibody-differentiation enzyme AID and spliceosome-associated factor CTNNBL1. *Mol. Cell 31*, 474–484 (2008).

23. Orthein, A. et al. Regulation of activation-induced deaminase stability and antibody gene diversification by Hsp90. *J. Exp. Med. 207*, 2751–2765 (2010).

24. Patenaude, A.-M. et al. Active nuclear import and cytoplasmic retention of AID in B lymphocytes. *Nat. Rev. Mol. Cell Biol. 15*, 167–177 (2015).

25. Barreto, V., Reina-San-Martin, B. R., 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).

26. Meito, S. P. et al. Consecutive interactions with Hsp90 and eIF1A underlie a functional maturation and storage pathway of AID in the cytoplasm. *J. Exp. Med. 212*, 581–596 (2015).

27. Prochnow, C., Branstetter, R., Klein, M. G., Goodman, M. F. & Chen, X. S. The APOBEC-2 crystal structure and functional implications for the deaminase AID. *Nature 445*, 447–451 (2007).

28. King, J. J. et al. Catalytic pocket inaccessibility of activation-induced cytidine deaminase is a safeguard against excessive mutagenic activity. *Structure 23*, 615–627 (2015).

29. Pham, P. et al. Structural analysis of the activation-induced deoxyuridine deaminase required in immunoglobulin diversification. *DNA Repair 43*, 48–56 (2016).

30. Quartier, P. et al. Clinical, immunologic and genetic analysis of 29 patients with autosomal recessive hyper-IgM syndrome due to activating deaminase deficiency. *Clin. Immunol. 110*, 22–29 (2004).

31. Mu, Y., Prochnow, C., Pham, P., Chen, X. S. & Goodman, M. F. A structural basis for the biochemical behavior of activation-induced deoxyuridine deaminase class-switch recombination-defective hyper-IgM-2 mutants. *J. Biol. Chem. 287*, 28007–28016 (2012).

32. Nadassy, K., Wodak, S., J. & Janin, J. Structural features of protein-nucleic acid recognition sites. *Biochemistry 38*, 1999–2017 (1999).

33. Orthein, A. & Di Noia, J. M. Activation induced deaminase: how much and where? *Semin. Immunol. 24*, 246–254 (2012).

34. Wang, Q. et al. The cell cycle restricts activation-induced cytidine deaminase activity to early G1. *J. Exp. Med. 214*, 49–58 (2017).

35. Xue, K., Rada, C. & Neuberger, M. S. The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in mice. *Nature 445*, 627 (2007).

36. Zahn, A. et al. Activation induced deaminase C-terminal domain links DNA transcription and ARN polymerase II. *Nat. Rev. Mol. Cell Biol. 14*, 167–177 (2013).

37. Skourti-Stathaki, K., Kamieniarz-Gdula, K. & Proudfoot, N. J. R-loops induce transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the non-template strand. *Nat. Immunol. 4*, 452–456 (2003).

38. Qiao, Q. et al. AID recognizes structured DNA for class switch recombination. *Cell 67*, 361–373.e4 (2017).

39. Puccella, J. N. & Chaudhuri, J. AID invited to the G4 Summit. *Mol. Cell 67*, 355–357 (2017).

40. Chen, J. et al. Single-molecule dynamics of enhancer-assembly in active promoters. *Nucleic Acids Res. 41*, 1285 (2014).

41. Mau, R. W. et al. Spi5 accumulation at variable genes distinguishes somatic hypermutation in germline center B cells from ex vivo-activated cells. *J. Exp. Med. 211*, 2297–2306 (2014).

42. Lackey, L. et al. APOBEC3B and AID have similar nuclear import mechanisms. *J. Mol. Biol. 419*, 301–314 (2012).

43. Nielen, H. et al. Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication. *Cell 5*, 1059–1065 (2000).

44. Orthein, A. et al. A mechanism for the suppression of homologous recombination in G1 cells. *Nature 528*, 422–426 (2015).

45. Arakawa, H., Sarbassak, H. & Buerstedde, J. M. Activation-induced cytidine deaminase initiates immunoglobulin gene conversion and hypermutation by a common intermediate. *PloS Biol. 2*, e179 (2004).

46. Cortizas, E. M. et al. Alternative end-joining and classical nonhomologous end-joining pathways repair different types of double-strand breaks during class-switch recombination. *J. Immunol. 191*, 5751–5763 (2013).

47. Monta, S., Kojima, T. & Kitamura, T. Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene Ther. 7*, 1063–1066 (2000).

48. Larjani, M. et al. AID associates with single-stranded DNA with high affinity and a long complex half-life in a sequence-independent manner. *Mol. Cell. Biol. 27*, 20–30 (2007).

49. Czajkowska, A. L. et al. Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions. *Sci. Signal. 6*, e115 (2013).

50. Kessner, D., Chambers, M., Burke, R., Agus, D. & Mallick, P. ProToxWizard: open source software for rapid proteomics tools development. *Bioinformatics 24*, 2534–2536 (2008).

51. Shiteyberg, D. et al. TheIdr: multi-level integrative analysis of shotgun proteome data improves peptide and protein identification rates and error estimates. *Mol. Cell Proteom. 10*, M111 007690 (2011).

52. Liu, G. et al. ProHit: integrated software for mass spectrometry-based interaction proteomics. *Nat. Biotechnol. 28*, 1015–1017 (2010).

53. Quackenbush, J. Microarray data normalization and transformation. *Nat. Genet. 25*, 250–255 (2000).

54. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol. 15*, 5 (2014).

55. Knight, J. D. et al. A web-tool for visualizing quantitative protein-protein interaction data. *Proteomics 15*, 1432–1436 (2015).

Acknowledgements

We thank Dr. N. Francis for critical reading and Dr. F. Robert, Dr. A. Orthein and Dr. A. Zahn for discussions. We thank Jean-Félix Côté for technical help and the assistance of D. Fillion with microscopy, E. Massicotte, J. Lord, J. Leconte and P. St-Onge with flow cytometry, S. Pavli with mass spectrometry and M. Cawthorne and E. Thivierge with animal care. We thank Dr. M. Neuberger, Dr. A. Orthein, Dr. D. Durocher, Dr. R. Pavli and Dr. T. Honjo for sharing reagents. This work was funded by grants from the Cancer Research Society (to J.M.D.N.) and Canadian institutes of Health research (MOP130535).
to J.M.D.N., MOP111132 to M.L.), CCSRI innovation grant 702145 (to M.L.), NIH R01GM121595 (to R.E.V.) and a Discovery grant from The Natural Sciences and Engineering Research Council of Canada (to J.-F.C.). S.P.M. and H.B. were supported by doctoral training awards from the Fonds de Recherche du Québec-Santé. S.P.M. and L.C. L. were supported by doctoral fellowships from the Cole Foundation. J.-F.C. holds the Transatl Chair in Breast Cancer Research and is supported by an FRQS Senior investigator career award. J.M.D.N. holds the Canada Research Chair in Genetic Diversity.

Author contributions
Conceptualization, S.P.M., L.C.L. and J.M.D.N.; methodology, S.P.M. and L.C.L.; formal analysis, P.G.S.; investigation, S.P.M., L.C.L., A.K.E., H.F., A.-M.P., J.C.G., H.B., G.E.S. and R.E.V.; data curation, P.G.S. and H.B.; writing—original draft, S.P.M. and J.M.D.N.; writing—reviewing and editing, S.P.M., L.C.L., P.G.S., H.B., J.-F.C., M.L., R.E.V. and J.M. D.N.; funding acquisition, J.M.D.N.; supervision, J.-F.C., M.L., R.E.V. and J.M.D.N.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-03387-6.

Competing interests: The authors declare no competing interests.