B Cell Super-Enhancers and Regulatory Clusters Recruit AID Tumorigenic Activity

Jason Qian,[1,14] Qiao Wang,[2,14] Marei Dose,[1,14,*] Nathanael Pruett,[1,14] Kyong-Rim Kieffer-Kwon,[1,14] Wolfgang Resch,[1] Genqing Liang,[1] Zhonghui Tang,[3] Ewy Mathé,[1] Christopher Benner,[13] Wendy Dubois,[5] Steevenson Nelson,[1] Laura Vian,[1] Thiago Y. Oliveira,[2] Mila Jankovic,[2] Ofir Hakim,[6] Anna Gazumyan,[2] Rushad Pavri,[7] Parirokh Awasthi,[8] Bin Song,[9] Geng Liu,[9] Longyun Chen,[9] Shida Zhu,[9] Lionel Feigenbaum,[8] Louis Staudt,[10] Cornelis Murre,[4] Yijun Ruan,[3] Davide F. Robbiani,[2] Qiang Pan-Hammarström,[11,15] Michel C. Nussenzweig,[2,12,15] and Rafael Casellas[1,5,15,*]

[1]Genomics and Immunity, NIAMS, NIH, Bethesda, MD 20892, USA
[2]Laboratory of Molecular Immunology, The Rockefeller University, New York, NY 10065, USA
[3]Department of Genetic and Development Biology, Jackson Laboratory for Genomic Medicine, University of Connecticut, 400 Farmington, CT 06030, USA
[4]Division of Biological Sciences, Department of Molecular Biology, University of California, San Diego, La Jolla, CA 92093, USA
[5]Center of Cancer Research, NCI, NIH, Bethesda, MD 20892, USA
[6]Bar-Ilan University, Ramat-Gan 5290002, Israel
[7]Institute of Molecular Pathology (IMP), Vienna BioCenter, Doktor Bohr Gasse 7, Vienna 1030, Austria
[8]Science Applications International Corporation/Frederick, NCI-Frederick Cancer Research and Development Center, Frederick, MD 21702, USA
[9]Beijing Genomics Institute, Shenzhen, Shenzhen 518083, China
[10]Metabolism Branch, NCI, NIH, Bethesda, MD 20892, USA
[11]Department of Laboratory Medicine, Karolinska Institutet, Karolinska University Hospital, Huddinge, 14186 Stockholm, Sweden
[12]HHMI, The Rockefeller University, New York, NY 10065, USA
[13]The Integrative Genomics and Bioinformatics Core, Salk Institute for Biological Studies, La Jolla, CA 92037, USA
[14]Co-first author
[15]Co-senior author
*Correspondence: marei.dose@nih.gov (M.D.), rafael.casellas@nih.gov (R.C.)
http://dx.doi.org/10.1016/j.cell.2014.11.013

SUMMARY

The antibody gene mutator activation-induced cytidine deaminase (AID) promiscuously damages oncogenes, leading to chromosomal translocations and tumorigenesis. Why nonimmunoglobulin loci are susceptible to AID activity is unknown. Here, we study AID-mediated lesions in the context of nuclear architecture and the B cell regulome. We show that AID targets are not randomly distributed across the genome but are predominantly grouped within super-enhancers and regulatory clusters. Unexpectedly, in these domains, AID deaminates active promoters and eRNA$^+$ enhancers interconnected in some instances over megabases of linear chromatin. Using genome editing, we demonstrate that 3D-linked targets cooperate to recruit AID-mediated breaks. Furthermore, a comparison of hypermutation in mouse B cells, AID-induced kataegis in human lymphomas, and translocations in MEFs reveals that AID damages different genes in different cell types. Yet, in all cases, the targets are predominantly associated with topological complex, highly transcribed super-enhancers, demonstrating that these compartments are key mediators of AID recruitment.

INTRODUCTION

Although humans produce roughly equal numbers of B and T lymphocytes, up to 95% of lymphomas in the Western world are of B cell origin (Küppers, 2005). This overrepresentation originates in large part from misrepair of DNA lesions introduced by activation-induced cytidine deaminase (AID), a B cell-specific cytidine deaminase that initiates class switch recombination (CSR) and somatic hypermutation (SHM) of immunoglobulin (*Ig*) genes (Alt et al., 2013). Although AID preferentially targets *Ig* heavy and light chain loci, it also mutates and produces DNA breaks in non-*Ig* genes (Hakim et al., 2012; Liu et al., 2008; Robbiani et al., 2008). Among these off targets, a substantial number are oncogenes directly implicated in B cell lymphomagenesis, including *BCL6*, *Myc*, *MIR142*, *CD95*, *Pax5*, and *BCL7* (Chiarle et al., 2011; Hakim et al., 2012; Hasham et al., 2010; Kato et al., 2012; Klein et al., 2011; Müschen et al., 2000; Pasqualucci et al., 1998; Robbiani et al., 2009; Shen et al., 1998; Tsai et al., 2008). Recurrent DNA damage at these loci leads to oncogenic mutations and chromosomal translocations that activate proto-oncogenes by juxtaposing them to potent *Ig* enhancers (Nussenzweig and Nussenzweig, 2010). Accordingly, genetic ablation of AID markedly impairs the formation of *Ig*-translocations and the onset of B cell tumor development in mice (Kovalchuk et al., 2007, 2012; Ramiro et al., 2004; Robbiani et al., 2008; Takizawa et al., 2008).

Transcription facilitates AID targeting to *Ig* genes by at least three related mechanisms. First, *Ig* enhancers are required for

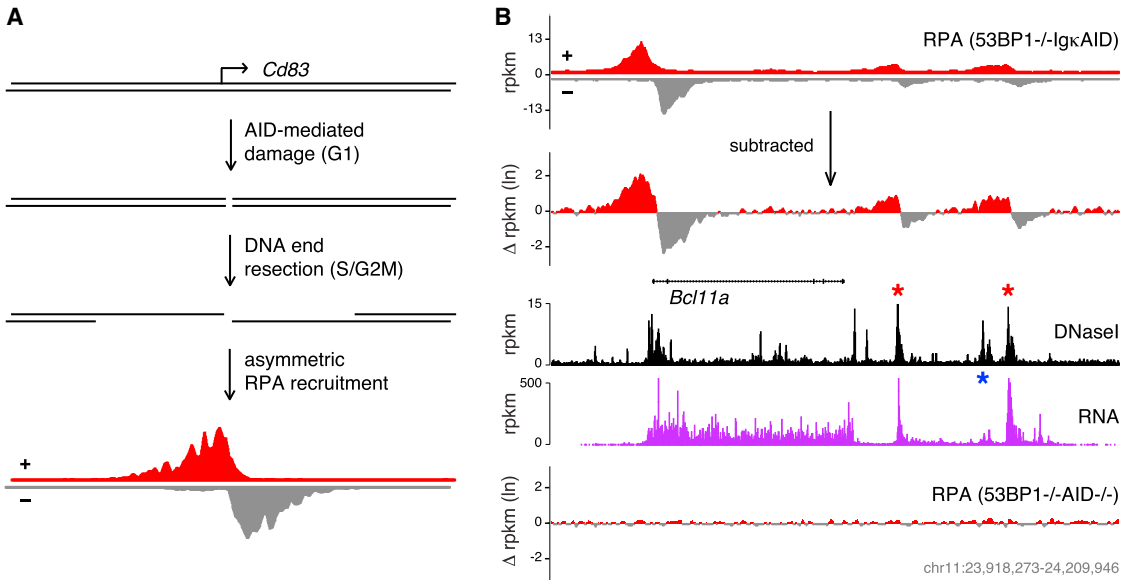


Figure 1. AID Damages Enhancer DNA

(A) Strategy to reveal AID-mediated breaks. In 53BP1$^{-/-}$ cells DNA lesions at AID off-targets (e.g., *Cd83*) in G1 are resected in S and G2M by HR repair nucleases, leading to asymmetric RPA binding that can be detected by ChIP-Seq.

(B) The visualization of RPA-Seq was improved by plotting the difference in ChIP signals between + and – strands. An algorithm was developed to efficiently detect asymmetric RPA occupancy. The new approach reveals two additional AID targets at the *Bcl11a* locus that overlap with enhancer elements (highlighted with red asterisks). The nontargeted enhancer is marked with a blue asterisk. DNaseI, RNA (GRO-seq) (Chiarle et al., 2011), and RPA control (53BP1$^{-/-}$AID$^{-/-}$) tracks are provided.

See also Figure S1 and Table S1A.

hypermutation and recombination of both variable (V) domains and switch (S) DNA repeats that precede antibody gene constant (C) regions (Buerstedde et al., 2014). Second, transcription of S repeats leads to substantial RNA PolII pausing (Rajagopal et al., 2009; Wang et al., 2009), and Spt5, a PolII pausing factor, enables hypermutation and recombination by associating with AID (Pavri et al., 2010). Third, the RNA degrading exosome complex displaces nascent S transcripts thereby rendering both DNA strands accessible to deamination (Basu et al., 2011). Whether these or additional mechanisms are responsible for promiscuous AID activity at non-*Ig* loci is unknown.

Here, we examine promiscuous AID activity and its relationship to chromosome folding and the B cell regulome. We find that AID-mediated lesions occur predominantly within B cell super-enhancers and regulatory clusters. Furthermore, we show that the structural and transcriptional features of these domains help explain AID tumorigenic activity in the B cell compartment of mice and humans.

RESULTS

AID Damages Enhancer DNA

To study AID off-targeting activity, we made use of replication protein A chromatin immunoprecipitation (RPA-ChIP) that labels DNA breaks in the 53BP1$^{-/-}$ background (Hakim et al., 2012). B cells isolated from these mice are defective for nonhomologous end joining (NHEJ), and AID-mediated lesions that are induced in G1 are aberrantly processed in S and G2M by homologous recombination (Yamane et al., 2013). As a result, DNA-ends are resected leading to asymmetrical accumulation of RPA and Rad51 around DNA breaks and these proteins can be detected by chromatin immunoprecipitation (Figure 1A)

To improve the sensitivity of the assay, we developed an algorithm that detects asymmetric RPA recruitment with high precision, and the difference in ChIP signals between upper (+) and lower (–) DNA strands was plotted on a log scale (Figure 1B). The new approach revealed 92 additional genomic sites associated with RPA in 53BP1$^{-/-}$IgκAID B cells (236 total targets; Table S1A available online). Conversely, we detected a single RPA asymmetric peak in 53BP1$^{-/-}$AID$^{-/-}$ cells (not shown). At the *Bcl11a* locus, for instance, we found two additional sites downstream of the promoter (120 and 180 kb away) that display asymmetric RPA accumulation in the presence of AID but not in its absence (Figure 1B). Notably, a fraction of the peaks (33, or 14%) did not overlap with TSSs but were associated with DNaseI hypersensitive sites corresponding to B cell enhancers (red asterisks in Figure 1B) (Kieffer-Kwon et al., 2013). Consistent with this interpretation, AID targets distal from TSSs displayed the epigenetic signature of active enhancers: H2AZ^{low}H3K4me3^{low}H3K4me1^{high} (Kouzine et al., 2013; not shown). Thus, in addition to promoter proximal sequences, AID damages enhancer DNA.

Nuclear Compartmentalization of AID Activity

AID activity is confined to the interphase nucleus (Petersen et al., 2001), where the genome is partitioned into a hierarchy of

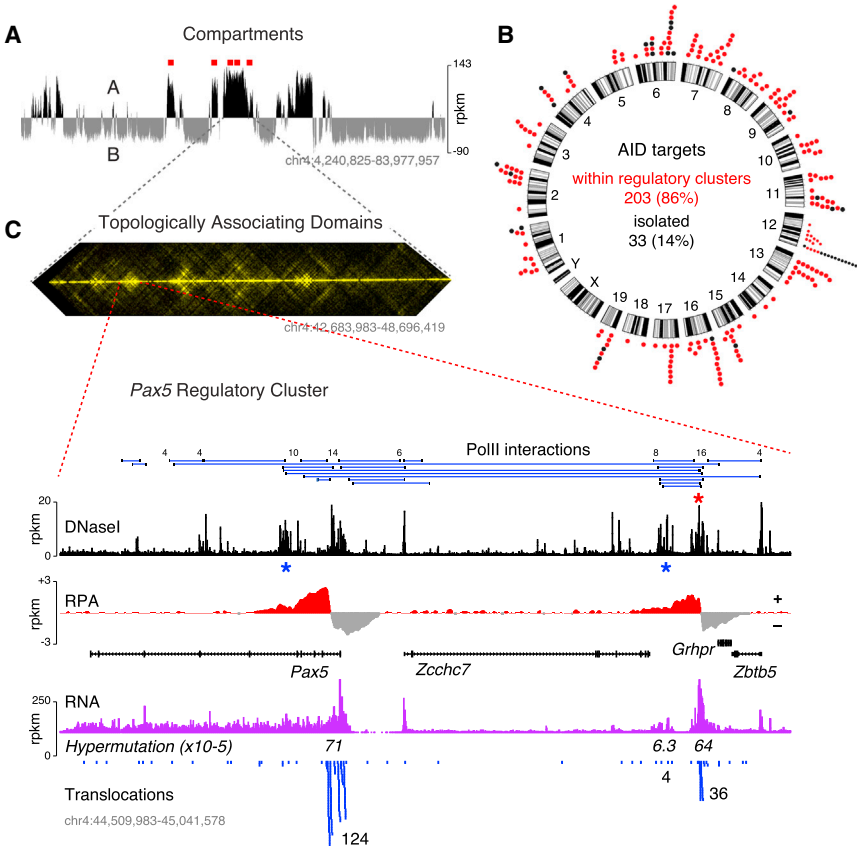


Figure 2. Tethering and Compartmentalization of AID Targets in the Mouse Genome

(A) AID targets are largely found within A compartments (black upper track) as defined by Hi-C. Red dots identify the location of damaged loci within the genomic domain. The Hi-C data was obtained from pro-B cells. All other experiments involving mouse B cells in the manuscript were done with activated B cells.

(B) Circos plot shows the genome-wide distribution of AID targets that are either tethered within regulatory clusters (red dots) or isolated (black dots).

(C) Upper: heat map of *cis*-interaction frequencies revealing TADs within the domain chr4:42,683,983-48,696,419. Lower: *Pax5* gene regulatory cluster, as defined by PolII long-range interactions. The targeted promoter is associated with nondamaged (blue asterisks) and damaged (red asterisk) enhancers. DNaseI hypersensitivity, RNA, hypermutation, and chromosomal translocations (TC-Seq) are also shown. The number of interactions is provided above the ChIA-PET links.

See also Figure S1.

structures, including A-B compartments, topologically associating domains (TADs), and clusters of interactive gene regulatory elements (Gibcus and Dekker, 2013). The finding that both promoters and enhancers undergo AID-mediated damage suggests that AID targets might also be clustered in the B cell nucleus. In support of this idea, nearly half of all targets (110 of 236) were located within ~90 kb of each other, a distance that is markedly different from a random model (~4 Mb, Figure S1A). Prompted by these observations, we analyzed the distribution of RPA+ sites in the context of genome folding, as defined by chromosome conformation capture (3C) techniques.

Hi-C maps from pro-B cells (Lin et al., 2012) revealed that 96% of AID targets (233 of 236) are located within A compartments (Table S1A; Figure 2A). These compartments are generally gene-rich, DNaseI-hypersensitive, and transcriptionally active (Lieberman-Aiden et al., 2009), features that agree well with AID's preference for transcribed chromatin.

In eukaryotes, TADs divide A-B compartments into nuclear subdomains containing clusters of multiple regulatory elements tethered by long-range interactions (Gibcus and Dekker, 2013; Li et al., 2012). To examine the distribution of AID targets vis-à-vis this architecture we made use of a PolII ChIA-PET map from activated B cells (Kieffer-Kwon et al., 2013). This technique combines PolII ChIP with 3C technology to define the promoter-enhancer interactome. Remarkably, while 47% of active promoters in B lymphocytes are not anchored in regulatory clusters, (Table S1A), 86% of AID targets were preferentially tethered to neighboring promoters and enhancers within regulatory clusters ($p < 10^{-15}$, Figure 2B and Experimental Procedures). In some cases, these clusters connected multiple AID targets. For instance, at the *Pax5* locus the targeted promoter was linked by long-range interactions with three enhancer domains, one of which (~250 kb away) was also damaged by AID (Figure 2C). Likewise, the targeted *Ly6a*, *Ly6e*, and *Rohema* promoters in chromosome 15 formed a topological cluster spanning ~100 kb (Figure S1B). Importantly, the vast majority of AID targets (84%) were tethered to regulatory elements within the same TADs (e.g., *Pax5* cluster, Figure 2C), consistent with the notion that these domains restrict chromatin mobility (Gibcus and Dekker, 2013). A notable exception was the histone H1 gene family, where AID targets from two noncontiguous compartments physically associated over 2.1 Mb (Figure S1C). We conclude that AID preferentially damages promoters and enhancers tethered by long-range interactions within gene regulatory clusters.

AID Targeting Is Largely Confined to B Cell Super-Enhancers

Super-enhancers (SEs) or stretch enhancers were recently identified as a special subset of regulatory elements (Hnisz et al., 2013; Lovén et al., 2013; Parker et al., 2013; Whyte et al., 2013). They represent exceptionally large enhancer domains primarily associated with highly transcribed genes controlling cell identity. Because of the known correlation between transcription and AID activity, we asked whether regulatory clusters targeted by AID might represent SE domains. To this end, we used H3K27Ac and a published algorithm (Whyte et al., 2013) to catalog SEs in stimulated B cells. Consistent with the high degree of activation in the presence of LPS+IL-4, we uncovered 1,003 SEs in cultured B cells (Figure S2A). By comparison, 13% of 86

human tissues surveyed displayed >1,000 SEs (Hnisz et al., 2013). In agreement with such studies, activated B cell SEs spanned DNA regions an order of magnitude greater than conventional enhancers, and they were densely occupied by the Mediator complex (Figure S2B).

At all three *Ig* loci, AID-mediated damage occurred within SE domains interconnected by long-range interactions (Figures 3A and S2C). Remarkably, 76% (179 of 236) of all AID targets were linked to SEs, a significant enrichment over what is expected by chance ($p < 1 \times 10^{-15}$, see Experimental Procedures). As an example, both the *Aicda*- and *Apobec1*-targeted genes are interconnected within the same SE (Figure 3C). Thus, AID on- and off-targeting activity occurs primarily within SE domains.

A key characteristic of SEs is that they are largely cell-type specific. Consistent with this, more than 50% of AID-targeted SEs were only present in stimulated B cells when compared to 18 primary mouse cells and tissues (Figure S2D). The analysis included SEs from developing pro-B cells (Whyte et al., 2013), which only displayed 32% overlap with activated counterparts (Figure S2D). Hence, most AID-mediated damage occurs within SEs acquired during development.

Approximately 80% (824 of 1,003) of B cell SEs did not harbor AID-mediated damage (Figure 3B). Notably, SEs containing AID targets could be distinguished from nontargeted ones in that they were more accessible (higher H3K27Ac, $p = 1 \times 10^{-25}$, Figure 3D), larger in size ($p = 3 - 10^{-9}$, Figure 3E), and their associated promoters were transcribed at higher levels ($p = 4 \times 10^{-10}$, Figure 3F). In addition, the extent of 3D connectivity was significantly higher at targeted SEs ($p = 3 \times 10^{-17}$, Figure 3G). We conclude that AID targets are preferentially associated with SEs displaying a high degree of accessibility, transcription, and structural complexity.

Functional Attributes of AID-Targeted Regulatory Elements

Within SEs, genes undergoing AID-mediated damage are linked to both targeted and nontargeted elements. For instance, of 11 enhancers associated with *Myc*, only two showed asymmetric RPA occupancy (Figure 4A). To characterize features that might distinguish these two enhancer groups, we measured hypersensitivity to DNaseI but found no significant differences ($p = 0.9$, Figure 4B). Conversely, targeted enhancers were consistently transcribed, as determined by GRO-Seq analysis ($p = 3 \times 10^{-5}$, Figure 4C). For instance, of the two enhancers upstream of *Pax5*, only the one displaying high levels of eRNA synthesis was associated with RPA, chromosomal translocations, and somatic hypermutation (Figure 2C). Additional examples at the *Bcl11a* locus are provided in Figure 1B. Similarly, the RPA+ *Myc* enhancers at the mid-point of *Pvt1* were transcribed at higher levels compared to those lacking RPA (Figure 4A). Of note, *Igκ* translocations involving this particular *Myc* enhancer cluster are selected during plasmacytomagenesis (Huppi et al., 2011).

Consistent with eRNA synthesis, PolII and PolII long-range interactions were significantly higher at enhancers associated with AID-mediated lesions ($p = 2 \times 10^{-6}$, Figure 4D and not shown). The PolII stalling factor Spt5, implicated in AID recruitment (Pavri et al., 2010), was also enriched in RPA+ enhancers ($p = 6 \times 10^{-4}$, Figure 4E). Importantly, these features were particularly prominent at hypermutated *Igh* Eμ and *Igκ* Ei enhancers, whereas they were consistently low at the nontargeted *Igλ* E3-1 and E3-1 s enhancers (Figures 4C–4E; Table S1B). Conversely, no differences were found in the recruitment of CTCF, a factor involved in nuclear architecture ($p = 0.03$, Figure 4F). A separate analysis showed that these same features distinguished AID-targeted from nontargeted promoters (Figure S3A). Thus, AID preferentially deaminates transcriptionally active promoters and enhancers that engage in frequent long-range interactions.

Interacting Targets within SEs Cooperate to Recruit AID Activity

The clustering of AID targets in the mouse genome suggests that they may cooperate or synergize to recruit AID to SE domains. To directly test this idea we asked whether a nontargeted, but otherwise highly transcribed promoter could recruit hypermutation when linked to a damaged gene cluster. To this end, we inserted the ubiquitin-C (*Ubc*) gene promoter from chromosome 5 in lieu of the *Il4ra* promoter in chromosome 7 to generate Il4ra$^{u/u}$ mice (Figure S3B). In activated B cells, *Il4ra* and flanking *Nsmce1* and *Il21r* overlap with SEs and interact extensively creating a multiple-promoter gene cluster (Figure 5A). In the presence of AID, all three genes undergo DNA double-strand breaks (Figure 5A), whereas no damage is detected at *Ubc* (Figure S3C).

Fluorocytometric analysis of Il4ra$^{u/u}$ and Il4ra$^{+/+}$ B cells showed comparable levels of cell surface Il4ra receptor (Figure S3D). Consistent with this result, knockin B cells proliferated normally and underwent wild-type levels of γ1 recombination (Figures S3E and S3F). Importantly, H3K27Ac and RNA-Seq showed little or no differences in SE location or expression of *Nsmce1*, *Il4ra*, or *Il21r* between the two cell types (Figures 5B, 5C, and S4A). To measure chromatin contacts at the knockin allele we applied an improved version of 4C-Seq that characterizes local architecture at high resolution (van de Werken et al., 2012). The analysis showed that the knocked-in *Ubc* promoter associates with flanking *Nsmce1* and *Il21r* genes at wild-type frequencies (Figure S4B). Similar results were obtained when using the *Il21r* promoter as bait (Figure S4C). Thus, neither transcription nor the architecture of the *Nsmce1-Il4ra-Il21r* locus appeared disrupted following promoter replacement.

To directly assess AID activity we bred the Il4ra^{u} allele into the Ung$^{-/-}$IgκAID background, which enables measurement of hypermutation in *ex-vivo* cultures (Hakim et al., 2012). Il4ra$^{u/u}$Ung$^{-/-}$IgκAID and Il4ra$^{+/+}$Ung$^{-/-}$IgκAID B cells were stimulated for 7 days and mutations downstream of *Ubc* were assessed at chromosomes 5 (native configuration) and 7 (knockin alleles). Consistent with the lack of DNA breaks at *Ubc* in chromosome 5 (Figure S3C), biological triplicates revealed background mutation at this site, comparable to the average PCR error rate measured in AID$^{-/-}$ cells (SHM(f) = 13.6×10^{-5} versus 8.7×10^{-5}; Figure 5D; Table S1B). Notably, in Il4ra$^{u/u}$ cells *Ubc* displayed a significant increase in mutation frequency in chromosome 7 compared to its native site (SHM(f) = 59.2×10^{-5}, fold change = 4.3, $p = 0.0005$, Figure 5D). This mutation frequency was nearly that of *Il4ra* in wild-type cells (80.5×10^{-5}, Figure 5D). *Mir142*, *Pim1*, and *Myc*, which are not directly associated with the *Il4ra* locus, showed no significant changes in

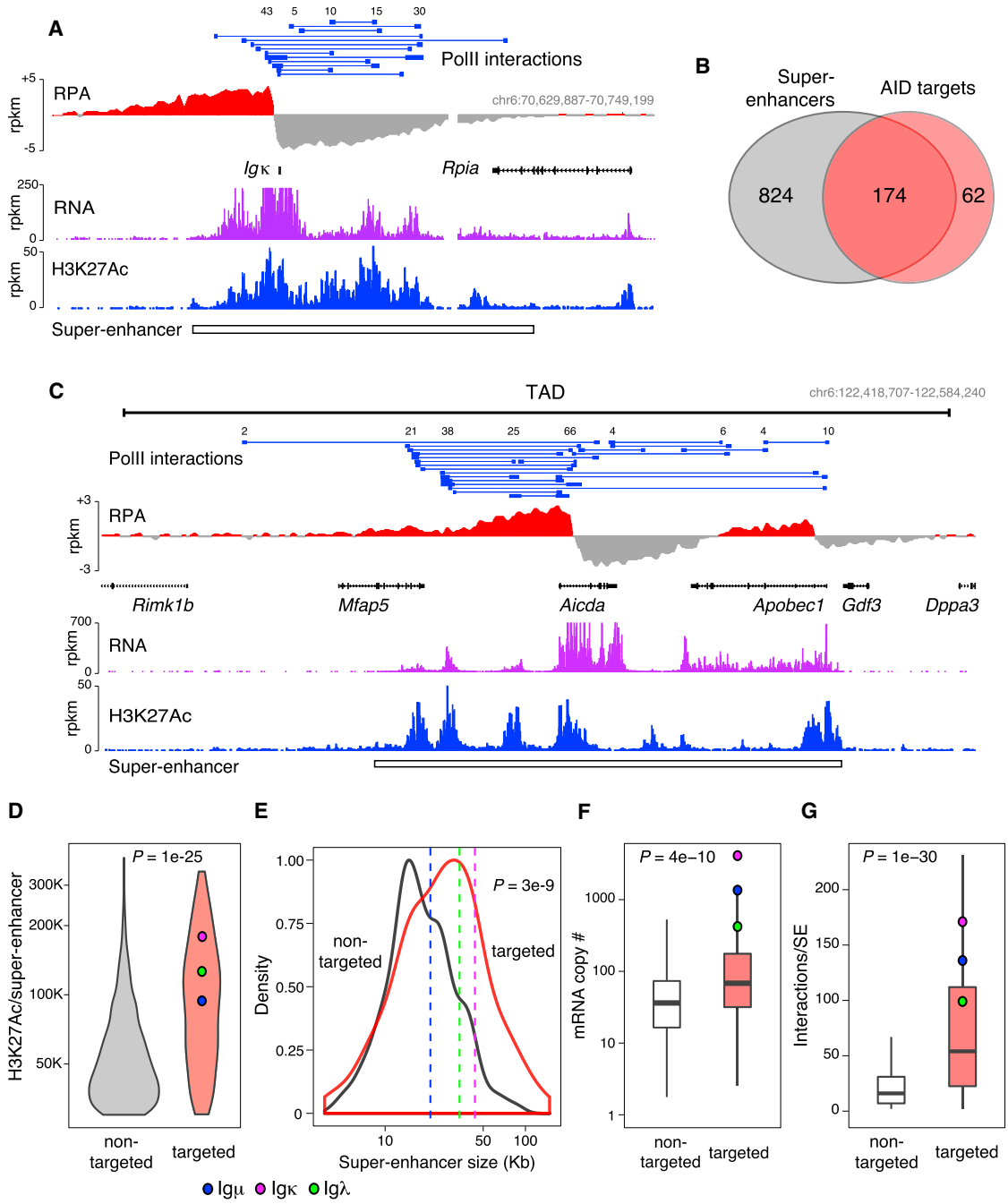


Figure 3. AID-Targeted Regulatory Clusters Are Predominantly Associated with B Cell SEs

(A) AID activity at the *Igκ* locus occurs within a 65 kb SE domain displaying long-range chromatin interactions. PolII interactions, RPA, RNA, and H3K27Ac profiles are provided.

(B) Venn diagram showing the fraction of AID targets associated with B cell SEs.

(C) Example of AID off-targeted SEs at the *Aicda-Apobec1* TAD in chromosome 6.

(D) H3K27Ac signal at targeted and nontargeted SEs. Igμ (blue, chr12:114640978-114669901), Igκ (magenta, chr6:70659188-70724456), and Igλ (green, chr.16:19002804-19067747) SEs are highlighted.

(E) Size distribution of total constituent enhancers in targeted (red line) or nontargeted (black line) SEs.

(F and G) Box plots showing the absolute expression or PolII-mediated connections at targeted (red) and nontargeted (open) SEs. Data are represented as the mean ± SEM.

See also Figure S2.

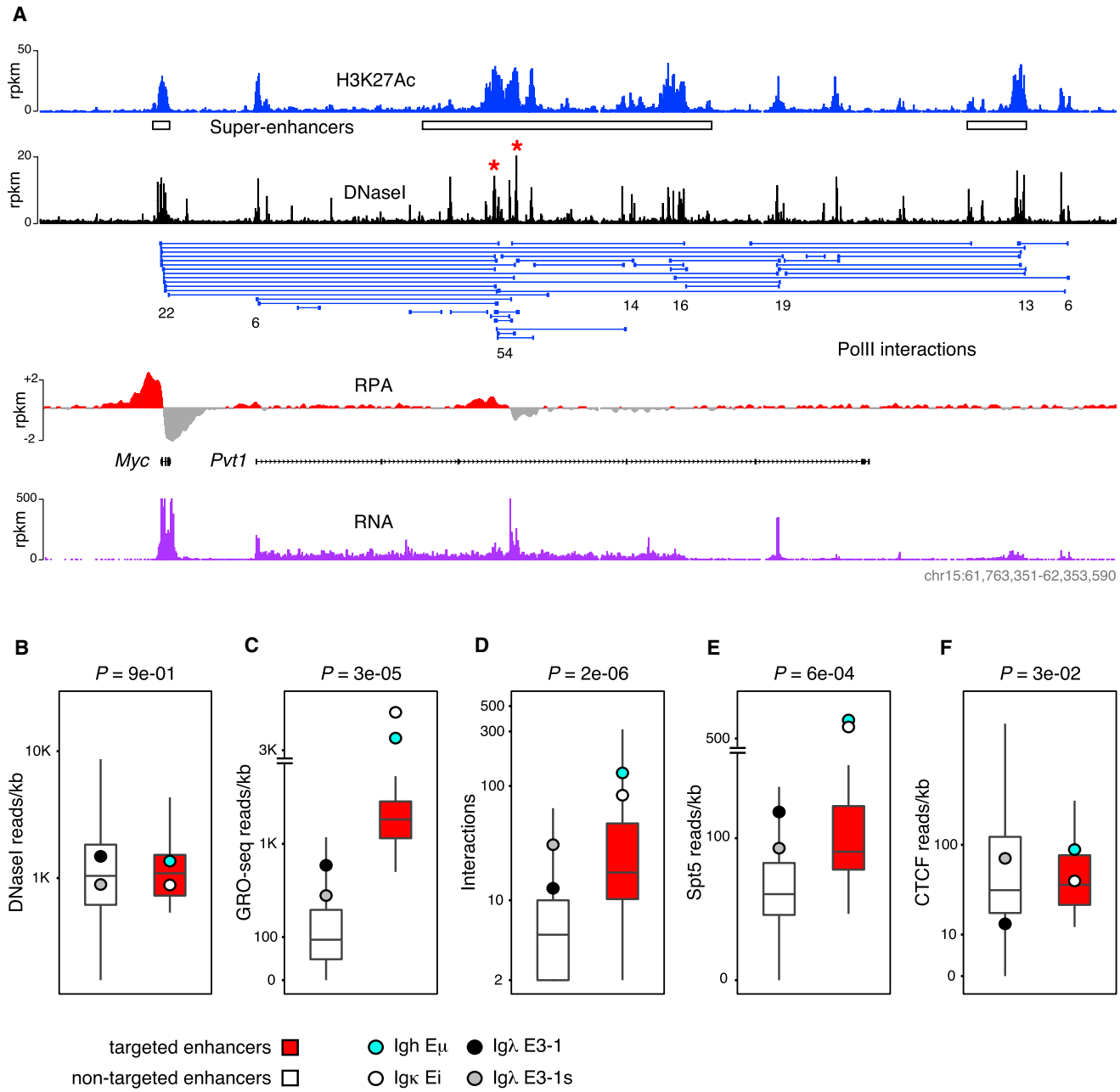


Figure 4. Defining Features of Targeted Enhancers

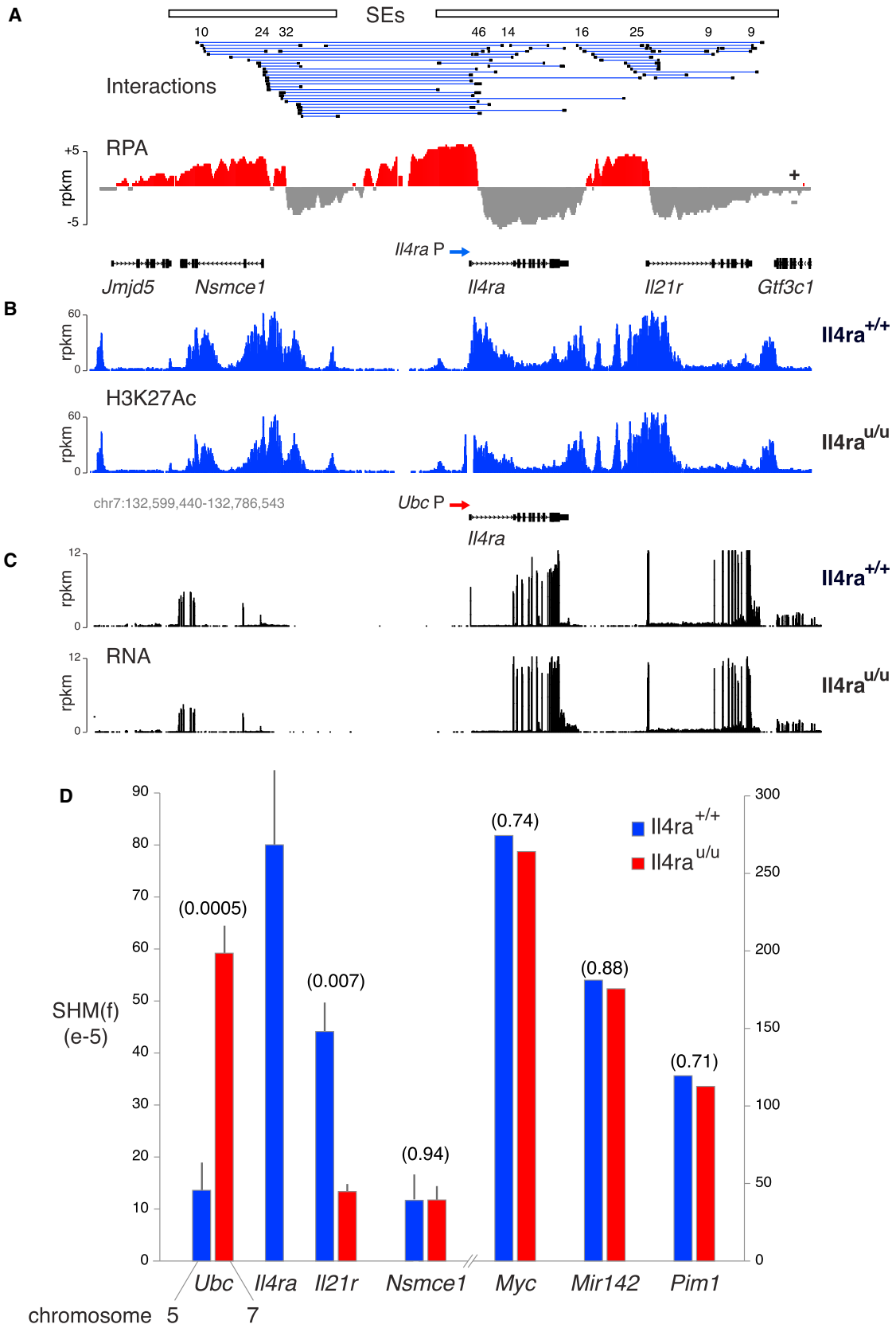
(A) *Myc* locus showing the distribution of SEs (H3K27Ac-Seq), enhancers (DNaseI-Seq), PolII long-range interactions (ChIA-PET), AID-mediated damage (RPA-Seq), and RNA synthesis (GRO-Seq). AID-targeted enhancers are denoted with red asterisks.

(B–F) Box plots comparing the extent of protein recruitment (B), DNaseI-Seq, eRNA synthesis (C), GRO-Seq, PolII interactions (D), PETs, Spt5 (E), and CTCF occupancy (F) at targeted (red boxes) and nontargeted (open boxes) enhancers. Data are represented as the mean ± SEM.

See also Figure S3.

hypermutation following gene targeting (fold change = 1.0–1.1, p > 0.7; Figure 5D; Table S1B). Hence, regulatory sequences at the *Nsmce1-Il4ra-Il21r* locus promote AID activity at *Ubc*.

To determine whether the *Il4ra* promoter also facilitates AID activity at flanking genes, we measured hypermutation at *Il21r* and *Nsmce1* in wild-type and knockin B cells. At *Nsmce1*, mutation could not be detected above background (Figure 5D). Conversely, at *Il21r*, where SHM(f) was 44.3 × 10^{-5} in wild-type cells, we observed a statistically significant decrease in Il4ra$^{u/u}$ (13.2 × 10^{-5}, p = 0.007, Figure 5D), indicating that replacement of the *Il4ra* promoter for *Ubc* has a negative effect on mutation of *Il21r* more than 50 kb downstream. We conclude



(legend on next page)

that both the *Il4ra* promoter and additional regulatory sequences at the *Nsmce1-Il4ra-Il21r* gene cluster enable off-targeting hypermutation by AID. The findings thus support a model where topologically linked elements within targeted SEs cooperate to recruit AID-mediated damage.

AID Targets in Human Lymphomas Overlap with Regulatory Clusters and SEs

Despite the known link between AID activity and human B cell tumor development (Klein and Dalla-Favera, 2008; Seifert et al., 2013), a comprehensive map of AID targets in the human genome is lacking. To directly address this question and to validate our findings in mouse B cells we mapped AID activity in the Ramos Burkitt's lymphoma line and in primary diffuse large B cell lymphoma (DLBCL). These tumors derive from germinal center or postgerminal center B cells and frequently display evidence of AID activity (Lossos et al., 2004; Pasqualucci et al., 2004; Sale and Neuberger, 1998). To efficiently detect hypermutation in Ramos we developed a deep-sequencing assay (SHM-Seq) by disrupting the mismatch repair gene *MSH2* by genome editing with a cassette expressing AID and Ugi, an inhibitor of the base excision repair factor Ung (Figure S5A). The resulting cell line is therefore both Ung- and Msh2-deficient, a combination that in mouse B cells leads to high levels of AID-mediated transition mutations at *Ig* and off-target loci (Hakim et al., 2012; Liu et al., 2008). Following 300 days of culture, the targeted cell line was single-cell sorted, individual clones were expanded, and DNA associated with H3K4me3, a histone mark that overlaps with AID activity (Yamane et al., 2011), was isolated and microsequenced (Figure S5A). Nontargeted and AID$^{-/-}$ Ramos cells were used as controls.

Analysis of 26 clones revealed 11,344 mutations relative to nontargeted and AID$^{-/-}$ controls. As expected, 92% of the substitutions were transitions. At *IGH* we detected 1,474 mutations (SHM(f) = 1.0×10^{-2}), mostly downstream of VDJ and Sμ promoters (Figure 6A; Table S1C). Likewise, the *IGH*-translocated *MYC* allele was highly mutated (SHM(f) = 5.0×10^{-3}, Figure 6B). The nontranslocated *MYC* allele was also targeted but at a frequency ~20-fold lower (SHM(f) = 2.2×10^{-4}, not shown). Other oncogenes often targeted in human lymphomas showed evidence of AID activity, including *MIR142*, *BCL6*, *BCL7A*, *MSH6*, and *ID3* (Table S1C). In total, 60 sites were hypermutated with high confidence, including four conventional enhancers (false discovery rate [FDR] < 10^{-16}, see Experimental Procedures).

Our mouse studies were performed with B cells overexpressing AID and in ex-vivo cultures, where SHM is limited. To map AID activity in unmanipulated cells we next performed whole-genome sequencing (~40× coverage) of ten DLBCL primary tumors isolated from lymph node biopsies. Somatic substitutions were defined by sequencing normal blood cells from the same patients. A total of 145,997 mutations were identified concomitant with deletions, insertions, amplifications, and chromosomal translocations. To classify AID hypermutation targets with high confidence we took advantage of the processive nature of AID deamination, which can generate clusters of transition mutations in individual clones (Lada et al., 2012; Taylor et al., 2013). These mutation showers or kataegis were recently uncovered by whole-genome sequencing of B cell and nonhematopoietic tumors (Alexandrov et al., 2013; Bolli et al., 2014; Chen et al., 2014; Nik-Zainal et al., 2012; Sakofsky et al., 2014). In the latter, particularly in breast tumors, kataegis was ascribed to processive deamination by the AID-related enzyme APOBEC3B (Alexandrov et al., 2013; Taylor et al., 2013).

We identified 105 kataegic sites in DLBCL associated with 30 genes (Table S1D). Four features implicated AID in the etiology of these mutation clusters. First, 82% of kataegis overlapped with transcribed promoter sequences, AID's preferred targeting domain (Figure S5B). This is in stark contrast to published non-B cell tumors (Alexandrov et al., 2013), where <6% of the kataegis were associated with TSSs ($p < 1 \times 10^{-10}$, Figure S5B). Second, also in contrast to other tumors, kataegis in DLBCL were recurrent, in that they always involved the *IG* loci and in most instances other mouse AID targets such as *PIM1*, *PAX5*, *RHOH*, *CIITA*, *MIR142*, *BCL6*, and the AID gene itself *AICDA* (Figure 6C; Table S1D). Third, 71% of the mutations were C > T transitions, consistent with the notion that kataegis results from DNA replication over cytidine deamination of resected DNA (Sakofsky et al., 2014; Taylor et al., 2013). Fourth, targeted cytidines bear the hallmark of AID activity (Taylor et al., 2013), i.e., they occur in a sequence context that recapitulates AID's preference for WRCY hotspots (Chi-square test $p < 1 \times 10^{-15}$, Figure 6D) (Rogozin and Kolchanov, 1992). Conversely, mutated Cs in breast tumors only differed from the genome average in that they were preceded mostly by a T (Figure 6D), consistent with the deamination profile of APOBEC3B (Alexandrov et al., 2013; Burns et al., 2013). These results thus support the proposal that kataegis in human lymphomas stem from AID activity.

We next characterized AID targets from Burkitt's and DLBCL tumors in the context of nuclear architecture and SEs. To this end we mapped PolII ChIA-PET and H3K27Ac in Ramos and used germinal center B cells isolated from human tonsils as substitutes for primary DLBCL (see Experimental Procedures). Consistent with the mouse results we found a strong overlap between hypermutated genes and SE domains (57%–70%,

Figure 5. Tethered Regulatory Elements Cooperate to Recruit AID Activity

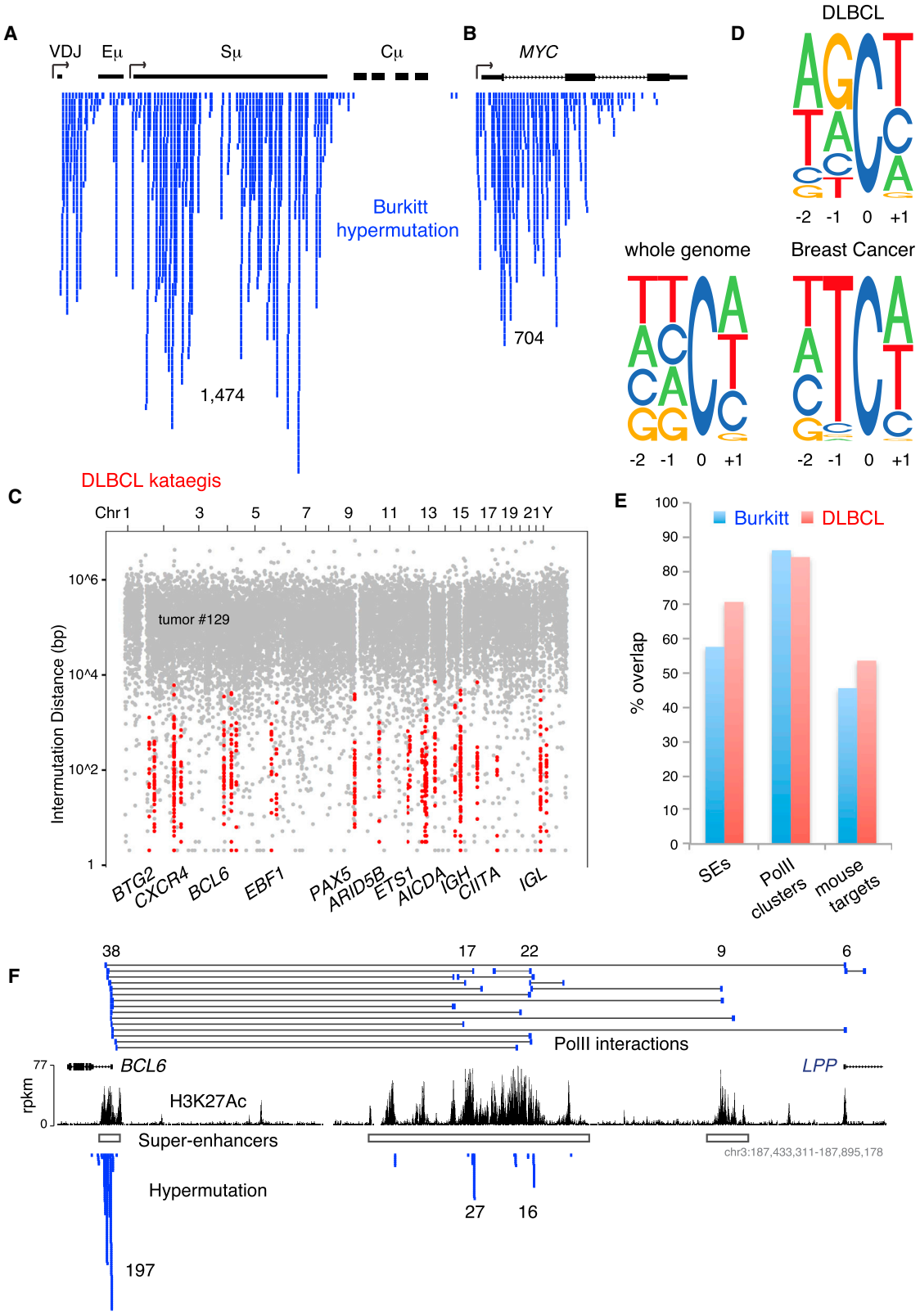
(A) *Il4ra, Il21r*, and *Nsmce1* form a promoter-gene cluster on mouse chromosome 7. Long-range interactions, DNA damage, and SEs are shown. *Il4ra*$^{u/u}$ knockin mice were created by replacing the *Il4ra* promoter (P, blue arrow) for that of *Ubc* (red arrow).

(B) H3K27Ac in wild-type and knockin mouse B cells.

(C) mRNA expression (plotted as rpkm values).

(D) Hypermutation frequency at *Ubc, Il4ra*, *Il21r*, *Nsmce1*, *Myc*, *Mir142*, and *Pim1* genes was measured in Il4ra$^{+/+}$ (blue bars) and Il4ra$^{u/u}$ (red bars) activated B cells. P values shown were calculated with Student's t test for triplicates experiments (*Ubc*, *Il21r*, *Nsmce1*) and Fisher's exact test (*Myc*, *Mir142*, *Pim1*) for single experiments. Hypermutation at *Ubc* was measured on chromosome 5 in Il4ra$^{+/+}$ (blue bar) and on chromosome 7 in Il4ra$^{u/u}$ (red bar). Data are represented as the mean ± SEM.

See also Figures S3 and S4 and Table S1B.



(legend on next page)

Figure 6E, $p < 1 \times 10^{-15}$, see Experimental Procedures). Furthermore, 83%–85% of the targets were anchored by PolII long-range interactions (Figure 6E). For instance, at the *BCL7A* gene regulatory cluster in Ramos both the promoter and upstream enhancers were hypermutated (Figure S5C). Another notable example was the *BCL6* promoter and a linked SE domain >250 kb upstream (Figure 6F). Importantly, while only the *BCL6* promoter was associated with kataegis in DLBCL, a survey of 26 genomes from other primary human lymphomas (Alexandrov et al., 2013) revealed the presence of kataegis at the upstream SE domain (Figure S5D). Altogether, the results demonstrate that both in mouse and human B cells AID mutates tethered regulatory elements associated with SEs and regulatory clusters.

AID Targets a Specific Microenvironment Rather Than a Determined Set of Genes

The kataegis and SHM-Seq analyses of B cell tumors revealed that 57%–85% of human AID targets overlap with SEs and regulatory clusters, whereas the overlap with mouse targets was only 45%–53% (Figure 6E). The strong inference is that rather than mutating a specific set of genes, AID targets topologically complex, highly transcribed domains. To directly test this idea we mapped AID-induced translocations in MEFs using TC-Seq (Klein et al., 2011). Primary $AID^{-/-}$ MEFs carrying I-SceI sites at *Myc* and *Igh* ($Myc^{I}Igh^{I}AID^{-/-}$) were transduced with I-SceI alone or I-SceI and AID. A total of 15,272 unique, mappable rearrangements to Myc^{I} were captured from 40 million $AID^{-/-}$ MEFs, and 28,265 from 40 million AID-expressing MEFs (2 libraries each, Table S1E). Similar to B cells (Klein et al., 2011), a large fraction (20%–44%) of the rearrangements in MEFs occurred in *cis* within a 250 kb window around I-SceI (Figure S6A). Furthermore, translocations were associated with genes more frequently than predicted by a random model (binomial test $p < 0.0001$, Table S1E). Using stringent criteria, we identified 29 and 43 AID-dependent translocation hotspots in MEFs and B cells, respectively (Table S1F). Remarkably, while the majority of these hotspots were genic (>84%), only three (11%) were shared between fibroblasts and lymphocytes (Figure 7A). This result indicates that the cell type alters the landscape of genomic rearrangements induced by AID.

Because the spatial organization of the genome is not random but compartmentalized (Lieberman-Aiden et al., 2009), it is possible that the cell type-restricted translocation to Myc^{I} results from differences in nuclear organization. However, 4C-Seq showed that the *Myc* interactome in fibroblasts and lymphocytes was highly similar (Pearson's $\rho = 0.88$, Figure S6B), consistent with the observation that nuclear interactions do not correlate with the frequency of AID-mediated translocations (Hakim et al., 2012).

To explore the contribution of transcription to cell type-specific targeting, we next measured RNA synthesis. We found that, in general, genes associated with translocation hotspots displayed higher transcription in the respective cell type (Figure S6C). For example, *Pax5* and *Cd83* were only targeted and expressed in B cells, while MEF-specific hotspots *Ctgf* and *Wisp1* were only transcribed in fibroblasts (Figure S6D). Furthermore, while *Myc* was frequently translocated to the *Igh* I-SceI site in MEFs, we failed to detect rearrangements to S domains, which in fibroblasts are transcriptionally silent (Figure S6E). To assess whether differential AID targeting was also associated with SE domains we analyzed publicly available H3K27Ac profiles. We found that, similar to results obtained with B cells, AID activity at hotspot genes in MEFs occurred largely within the context of SEs (71%, $p < 1 \times 10^{-10}$, Figure 7B and Experimental Procedures). Importantly, this correlation applied to genes that were expressed in both cell types but were targeted in only one of them, such as *Flnb* on chromosome 14 (Figure 7C) and *Pim1* on chromosome 17 (Figure S7A). Altogether, the findings demonstrate that whereas AID damages a different set of genes in MEFs and B cells, in both cell types the targets are preferentially associated with SEs domains.

DISCUSSION

Recurrent translocation to non-*Ig* loci in B cell cancers is due in part to DNA damage by AID (Chiarle et al., 2011; Hakim et al., 2012; Klein et al., 2011; Zhang et al., 2012). However, the genomic features responsible for recruiting DNA damage are unknown. Our studies of mouse B cells, human lymphomas, and MEFs reveal that a major unifying property of AID targets is that they are predominantly clustered within highly active SEs and regulatory clusters (Figure S7B). As discussed below, the functional and architectural properties of these domains help explain why their associated genes are susceptible to AID tumorigenic activity.

SEs represent a special subset of regulatory clusters, where chromatin accessibility and transcriptional activity are an order of magnitude higher than at other active sites (Parker et al., 2013; Whyte et al., 2013). Both accessibility and transcription have long been recognized as prerequisites to *Ig* gene deamination (Alt et al., 2013). Our experiments show that along with size and long-range interconnectivity, the presence of a SE can differentiate targeted from nontargeted regulatory elements. For instance, a model based on these combined features can

Figure 6. AID Targets in Human Lymphomas Are Associated with Long-Range Chromatin Interactions and SEs

(A and B) The SHM-Seq protocol detects AID-mediated hypermutation in Ramos B cells, including at the *IGH* (A) and *MYC* (B) loci.

(C) Rainfall plot displaying the distance between neighboring mutations across the genome of a DLBCL primary tumor (#129). Kataegic domains of clustered mutations are depicted with red dots. Some of the genes associated with kataegis are highlighted.

(D) Representation of sequence context at positions −2, −1, and +1 flanking mutated Cs in DLBCL or breast cancer kataegis. The average context of Cs in the entire human genome is also shown.

(E) Percent overlap between hypermutated genes from Ramos Burkitt's lymphoma (blue bars) or primary DLBCL (red bars) in SEs (left), PolII long-range interactions (middle), or mouse AID targets (right).

(F) AID hypermutation of the *BL6* regulatory cluster in Ramos cells. SEs, PolII long-range interactions, and hypermutation are provided.

See also Figure S5 and Tables S1C and S1D.

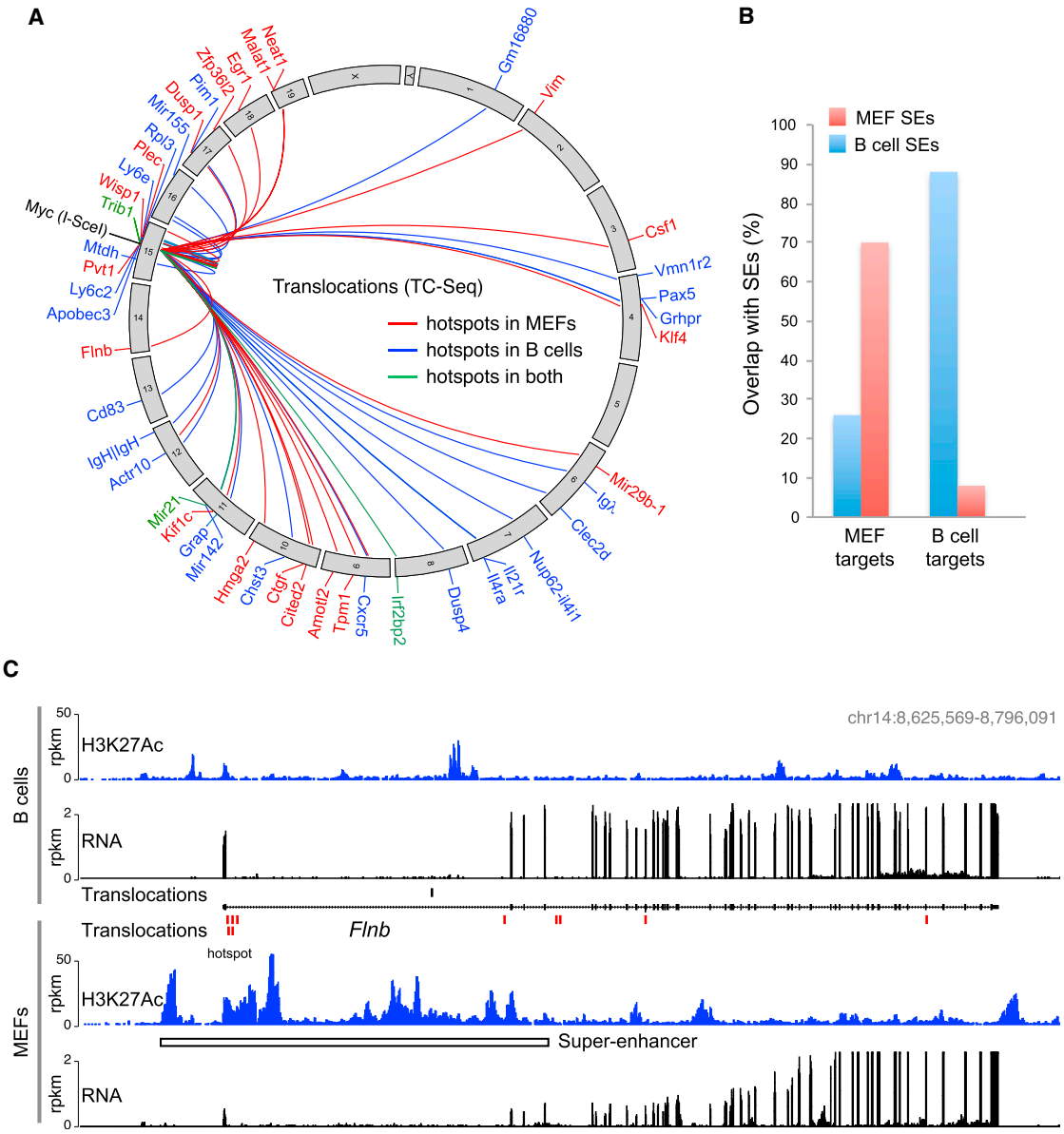


Figure 7. AID Damages Different Genes in Different Cell Types

(A) Circos diagram showing hotspots of AID-dependent chromosome translocations to *Myc*$^{I\text{-}SceI}$ in MEFs and B cells. Hotspots only present in B cells (blue lines), MEFs (red lines), or both cell types (green lines) are provided.

(B) Overlap of AID targets in MEFs (red bars) or B cells (blue bars) with SEs.

(C) *Myc* translocations to *Flnb* are primarily detected in MEFs (red bars), where the gene is associated with a SE domain. Conversely, a single translocation is detected in B cells (black bar).

See also Figures S6 and S7 and Tables S1E and S1F.

accurately predict 91% of mouse AID targets at a false discovery rate of 9% (Figure S7C; Experimental Procedures). The underlying assumption is that, as a group, these properties help create a nuclear microenvironment highly suitable to AID-mediated deamination. The fact that our data cannot predict AID targeting in its totality implies that additional parameters might also be at play. Specific transcription factors for instance have been shown to facilitate AID recruitment to *Ig* genes (Buerstedde et al., 2014). Small RNA processing by the Exosome complex is another example (Pefanis et al., 2014). Furthermore, in the accompanying paper, Alt and colleagues uncovers a strong correlation between convergent transcription and AID-mediated damage (Meng et al., 2014 in this issue of *Cell*).

Another unexpected finding is that within targeted SEs AID not only damages promoter proximal sequences but also cognate enhancers. These are invariably transcribed and more frequently

anchored by PolII long-range interactions. Both features likely render enhancer DNA accessible to cytidine deamination and double-strand break formation.

The link between AID activity and SEs sheds new light on the class of genes damaged in activated and germinal center B cells. Genome-wide maps of SHM, DNA breaks, and chromosomal translocations have consistently uncovered two sets of genes enriched among AID targets: oncogenes involved in proliferation and apoptosis (e.g., *Myc*, *Pim1*, *Jund*, *Bcl2*) and genes that feature prominently in B cell development and activation (*Pax5*, *Cd79b*, *Aicda*, *Irf8*, *Bach2*, *Nfκb*). Although AID's predilection for these gene groups has been unclear, they fit well with the observation that in all tissues examined so far, SEs largely control expression of cell identity genes as well as oncogenes that regulate cell cycle and differentiation. Examples of these are pluripotency genes in ES cells, genes critical for islet function in the pancreas, and *MYC* in multiple myeloma (Lovén et al., 2013; Parker et al., 2013; Whyte et al., 2013). By the same token, our TC-Seq analysis showed that targeted SEs in MEFs control expression of genes critical for fibroblast proliferation and maturation (e.g., *Ctgf*, *Wisp1*, *Amotl2*).

Another defining feature of SEs is that their constituent regulatory elements work in cooperation or synergistically to drive gene expression (Lovén et al., 2013). Our knockin experiments between the nontargeted *Ubc* promoter and the *Nsmce1-Il4ra-Il21r* targeted gene cluster provide compelling evidence that cooperation is also key to promiscuous AID-mediated damage. This feature helps explain why AID targets are clustered in the B cell genome. At the same time, it suggests that only networks of functionally cooperating elements can create the proper conditions for AID promiscuous activity. It is important to point out that these conditions are not exclusive to SE domains, but that they also typify highly interactive regulatory clusters not directly associated with SEs (e.g., H1 gene family). The *Ubc-Il4ra* experiment also provides a rationale to earlier observations showing that heterologous promoters not typically damaged in germinal centers (e.g., β-globin, B29, or PolI promoters) can recruit hypermutation when juxtaposed to *Ig* enhancers (Betz et al., 1994; Fukita et al., 1998; Tumas-Brundage and Manser, 1997). In both cases, AID exploits long-range interactions to act at a distance on nontargeted sequences.

In conclusion, rather than targeting a predetermined gene set, AID tumorigenic activity is focused on nuclear microenvironments that share a common set of architectural, transcriptional, and regulatory features.

EXPERIMENTAL PROCEDURES

Extended Experimental Procedures are provided in the Supplemental Information section.

4C-Seq

The 4C assay was performed as previously described van de Werken et al. (2012) with minor modifications. Ten million mouse B cells were crosslinked in 2% formaldehyde at 37°C for 10 min. The reaction was quenched by the addition of glycine (final concentration of 0.125 M). Cells were then washed with cold PBS and lysed (10 mM Tris-HCl, pH 8.0, 10 mM NaCl, 0.2% NP-40, 1× complete protease inhibitors [Roche]) at 4°C for 1 hr. Nuclei were incubated at 65°C for 30 min, 37°C for 30 min in 500 μl of restriction buffer (New England BioLabs DpnII buffer) containing 0.3% SDS. To sequester SDS, Triton X-100 was then added to a final concentration of 1.8%. DNA digestion was performed with 400 U of DpnII (New England Biolabs) at 37°C overnight. After heat inactivation (65°C for 30 min), the reaction was diluted to a final volume of 7 ml with ligation buffer containing 100 U T4 DNA Ligase (Roche) and incubated at 16°C overnight. Samples were then treated with 500 μg Proteinase K (Ambion) and incubated overnight at 65°C to reverse formaldehyde crosslinking. DNA was then purified by phenol extraction and ethanol precipitation. For circularization, the ligation junctions were digested with Csp6I (Fermentas) at 37°C overnight. After enzyme inactivation and phenol extraction, the DNA was religated in a 7 ml volume (1,000 U T4 DNA Ligase, Roche). Three micrograms of 4C library DNA was amplified with Expand Long template PCR System (Roche). Thermal cycle conditions were DNA denaturing for 2 min at 94°C, followed by 30 cycles of 15 s at 94°C, 1 min at 58°C, 3 min at 68°C, and a final step of 7 min at 68°C. Baits were amplified with inverse PCR primers as follows: *Il4ra* with DpnII: _4C 5′-TCAGGTAGTTCCATGGGATC-3′, Il4ra_Csp6i 5′-ATCTCTGCACCAGACATCAG-3′ and *Il21r* with IL21r_DpnII CCAGACCTACTTAGCAGATC, and IL21r_Csp6i: ACTTAGACACTGCTCAGCTG. 4C-amplified DNA was microsequenced with the Illumina platform.

ACCESSION NUMBERS

The two Gene Expression Omnibus (GEO) and one Sequence Read Archive (SRA) accession numbers for the deep-sequencing data reported in this paper are GSE62063, GSE61523, and SRP046243, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2014.11.013.

AUTHOR CONTRIBUTIONS

Q.P.-H., M.C.N., and R.C. designed experiments. J.Q., Q.W., N.P., and K.-R.K.-K. performed most key experiments. S.N., L.V., M.J., O.H., A.G., D.F.R., and R.P. performed experiments. G. Liang., P.A., W.D., and L.F. generated and maintained Il4ra mice. M.D., W.R., E.M., Z.T., T.Y.O., and C.B. performed bioinformatics. B.S., G. Liu., L.C., and S.Z. analyzed WGSs. L.S., C.M., and Y.R. provided expertise advice. M.C.N. and R.C. wrote the manuscript.

ACKNOWLEDGMENTS

We thank Daniel Hodson for isolating GC B cells, Ethan Tyler for designing Figure S7B, Gustavo Gutierrez for technical assistance with deep-sequencing, David Bosque and Tom Eisenreich for managing the mouse colonies, and Jim Simone, Klara Velinzon, and Yelena Shatalina for FACS sorting. This work was supported by the Intramural Research Programs of the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) and the National Cancer Institute (NCI), the Jackson Laboratory fund (JAX19020120), the European Research Council (ERC) (242551-ImmunoSwitch), and NIH grant AI037526 to M.C.N. M.C.N. is a Howard Hughes Medical Institute (HHMI) Investigator. Animal experiments were performed according to NIH and Rockefeller approved protocols. This study used the high-performance computational capabilities of NIH Helix Systems (http://helix.nih.gov).

Received: May 26, 2014
Revised: September 26, 2014
Accepted: October 30, 2014
Published: December 4, 2014

REFERENCES

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., et al.; Australian

Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MMML-Seq Consortium; ICGC PedBrain (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

Alt, F.W., Zhang, Y., Meng, F.L., Guo, C., and Schwer, B. (2013). Mechanisms of programmed DNA lesions and genomic instability in the immune system. Cell *152*, 417–429.

Basu, U., Meng, F.L., Keim, C., Grinstein, V., Pefanis, E., Eccleston, J., Zhang, T., Myers, D., Wasserman, C.R., Wesemann, D.R., et al. (2011). The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. Cell *144*, 353–363.

Betz, A.G., Milstein, C., González-Fernández, A., Pannell, R., Larson, T., and Neuberger, M.S. (1994). Elements regulating somatic hypermutation of an immunoglobulin kappa gene: critical role for the intron enhancer/matrix attachment region. Cell *77*, 239–248.

Bolli, N., Avet-Loiseau, H., Wedge, D.C., Van Loo, P., Alexandrov, L.B., Martincorena, I., Dawson, K.J., Iorio, F., Nik-Zainal, S., Bignell, G.R., et al. (2014). Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. Nat. Commun. *5*, 2997.

Buerstedde, J.M., Alinikula, J., Arakawa, H., McDonald, J.J., and Schatz, D.G. (2014). Targeting of somatic hypermutation by immunoglobulin enhancer and enhancer-like sequences. PLoS Biol. *12*, e1001831.

Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., Refsland, E.W., Kotandeniya, D., Tretyakova, N., Nikas, J.B., et al. (2013). APOBEC3B is an enzymatic source of mutation in breast cancer. Nature *494*, 366–370.

Chen, X., Bahrami, A., Pappo, A., Easton, J., Dalton, J., Hedlund, E., Ellison, D., Shurtleff, S., Wu, G., Wei, L., et al.; St. Jude Children's Research Hospital–Washington University Pediatric Cancer Genome Project (2014). Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. Cell Reports *7*, 104–112.

Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.J., Myers, D.R., Choi, V.W., Compagno, M., Malkin, D.J., et al. (2011). Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. Cell *147*, 107–119.

Fukita, Y., Jacobs, H., and Rajewsky, K. (1998). Somatic hypermutation in the heavy chain locus correlates with transcription. Immunity *9*, 105–114.

Gibcus, J.H., and Dekker, J. (2013). The hierarchy of the 3D genome. Mol. Cell *49*, 773–782.

Hakim, O., Resch, W., Yamane, A., Klein, I., Kieffer-Kwon, K.-R., Jankovic, M., Oliveira, T., Bothmer, A., Voss, T.C., Ansarah-Sobrinho, C., et al. (2012). DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. Nature *484*, 69–74.

Hasham, M.G., Donghia, N.M., Coffey, E., Maynard, J., Snow, K.J., Ames, J., Wilpan, R.Y., He, Y., King, B.L., and Mills, K.D. (2010). Widespread genomic breaks generated by activation-induced cytidine deaminase are prevented by homologous recombination. Nat. Immunol. *11*, 820–826.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Huppi, K., Pitt, J., Wahlberg, B., and Caplen, N.J. (2011). Genomic instability and mouse microRNAs. Toxicol. Mech. Methods *21*, 325–333.

Kato, L., Begum, N.A., Burroughs, A.M., Doi, T., Kawai, J., Daub, C.O., Kawaguchi, T., Matsuda, F., Hayashizaki, Y., and Honjo, T. (2012). Nonimmunoglobulin target loci of activation-induced cytidine deaminase (AID) share unique features with immunoglobulin genes. Proc. Natl. Acad. Sci. USA *109*, 2479–2484.

Kieffer-Kwon, K.-R., Tang, Z., Mathe, E., Qian, J., Sung, M.H., Li, G., Resch, W., Baek, S., Pruett, N., Grøntved, L., et al. (2013). Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. Cell *155*, 1507–1520.

Klein, U., and Dalla-Favera, R. (2008). Germinal centres: role in B-cell physiology and malignancy. Nat. Rev. Immunol. *8*, 22–33.

Klein, I.A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D.F., et al. (2011). Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. Cell *147*, 95–106.

Kouzine, F., Wojtowicz, D., Yamane, A., Resch, W., Kieffer-Kwon, K.-R., Bandle, R., Nelson, S., Nakahashi, H., Awasthi, P., Feigenbaum, L., et al. (2013). Global regulation of promoter melting in naive lymphocytes. Cell *153*, 988–999.

Kovalchuk, A.L., duBois, W., Mushinski, E., McNeil, N.E., Hirt, C., Qi, C.F., Li, Z., Janz, S., Honjo, T., Muramatsu, M., et al. (2007). AID-deficient Bcl-xL transgenic mice develop delayed atypical plasma cell tumors with unusual Ig/Myc chromosomal rearrangements. J. Exp. Med. *204*, 2989–3001.

Kovalchuk, A.L., Ansarah-Sobrinho, C., Hakim, O., Resch, W., Tolarová, H., Dubois, W., Yamane, A., Takizawa, M., Klein, I., Hager, G.L., et al. (2012). Mouse model of endemic Burkitt translocations reveals the long-range boundaries of Ig-mediated oncogene deregulation. Proc. Natl. Acad. Sci. USA *109*, 10972–10977.

Küppers, R. (2005). Mechanisms of B-cell lymphoma pathogenesis. Nat. Rev. Cancer *5*, 251–262.

Lada, A.G., Dhar, A., Boissy, R.J., Hirano, M., Rubel, A.A., Rogozin, I.B., and Pavlov, Y.I. (2012). AID/APOBEC cytosine deaminase induces genome-wide kataegis. Biol. Direct *7*, 47.

Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell *148*, 84–98.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

Lin, Y.C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., Chandra, V., Bossen, C., Glass, C.K., and Murre, C. (2012). Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. Nat. Immunol. *13*, 1196–1204.

Liu, M., Duke, J.L., Richter, D.J., Vinuesa, C.G., Goodnow, C.C., Kleinstein, S.H., and Schatz, D.G. (2008). Two levels of protection for the B cell genome during somatic hypermutation. Nature *451*, 841–845.

Lossos, I.S., Levy, R., and Alizadeh, A.A. (2004). AID is expressed in germinal center B-cell-like and activated B-cell-like diffuse large-cell lymphomas and is not correlated with intraclonal heterogeneity. Leukemia *18*, 1775–1779.

Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell *153*, 320–334.

Meng, F.-L., Du, Z., Federation, A., Hu, J., Wang, Q., Kieffer-Kwon, K.-R., Meyers, R.M., Amor, C., Wasserman, C.R., Neuberg, D., et al. (2014). Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. Cell *159*. Published online December 4, 2014. http://dx.doi.org/10.1016/j.cell.2014.11.014.

Müschen, M., Re, D., Jungnickel, B., Diehl, V., Rajewsky, K., and Küppers, R. (2000). Somatic mutation of the CD95 gene in human B cells as a side-effect of the germinal center reaction. J. Exp. Med. *192*, 1833–1840.

Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al.; Breast Cancer Working Group of the International Cancer Genome Consortium (2012). Mutational processes molding the genomes of 21 breast cancers. Cell *149*, 979–993.

Nussenzweig, A., and Nussenzweig, M.C. (2010). Origin of chromosomal translocations in lymphoid cancer. Cell *141*, 27–38.

Parker, S.C., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., Black, B.L., et al.; NISC Comparative Sequencing Program; National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program Authors; NISC Comparative Sequencing Program Authors (2013). Chromatin stretch

enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc. Natl. Acad. Sci. USA *110*, 17921–17926.

Pasqualucci, L., Migliazza, A., Fracchiolla, N., William, C., Neri, A., Baldini, L., Chaganti, R.S., Klein, U., Küppers, R., Rajewsky, K., and Dalla-Favera, R. (1998). 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.

Pasqualucci, L., Guglielmino, R., Houldsworth, J., Mohr, J., Aoufouchi, S., Polakiewicz, R., Chaganti, R.S., and Dalla-Favera, R. (2004). Expression of the AID protein in normal and neoplastic B cells. Blood *104*, 3318–3325.

Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina San-Martin, B., Barreto, V., et al. (2010). Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. Cell *143*, 122–133.

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Chao, J., Rabadan, R., Economides, A.N., and Basu, U. (2014). Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. Nature *514*, 389–393.

Petersen, S., Casellas, R., Reina-San-Martin, B., Chen, H.T., Difilippantonio, M.J., Wilson, P.C., Hanitsch, L., Celeste, A., Muramatsu, M., Pilch, D.R., et al. (2001). AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. Nature *414*, 660–665.

Rajagopal, D., Maul, R.W., Ghosh, A., Chakraborty, T., Khamlichi, A.A., Sen, R., and Gearhart, P.J. (2009). Immunoglobulin switch mu sequence causes RNA polymerase II accumulation and reduces dA hypermutation. J. Exp. Med. *206*, 1237–1244.

Ramiro, A.R., Jankovic, M., Eisenreich, T., Difilippantonio, S., Chen-Kiang, S., Muramatsu, M., Honjo, T., Nussenzweig, A., and Nussenzweig, M.C. (2004). AID is required for c-myc/IgH chromosome translocations in vivo. Cell *118*, 431–438.

Robbiani, D.F., Bothmer, A., Callen, E., Reina San-Martin, B., Dorsett, Y., Difilippantonio, S., Bolland, D.J., Chen, H.T., Corcoran, A.E., Nussenzweig, A., et al. (2008). AID is required for the chromosomal translocations in c-myc that lead to c-myc/IgH translocations. Cell *135*, 1028–1038.

Robbiani, D.F., Bunting, S., Feldhahn, N., Bothmer, A., Camps, J., Deroubaix, S., McBride, K.M., Klein, I.A., Stone, G., Eisenreich, T.R., et al. (2009). AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. Mol. Cell *36*, 631–641.

Rogozin, I.B., and Kolchanov, N.A. (1992). Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. Biochim. Biophys. Acta *1171*, 11–18.

Sakofsky, C.J., Roberts, S.A., Malc, E., Mieczkowski, P.A., Resnick, M.A., Gordenin, D.A., and Malkova, A. (2014). Break-induced replication is a source of mutation clusters underlying kataegis. Cell Reports *7*, 1640–1648.

Sale, J.E., and Neuberger, M.S. (1998). TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line. Immunity *9*, 859–869.

Seifert, M., Scholtysik, R., and Küppers, R. (2013). Origin and pathogenesis of B cell lymphomas. Methods Mol. Biol. *971*, 1–25.

Shen, H.M., Peters, A., Baron, B., Zhu, X., and Storb, U. (1998). Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. Science *280*, 1750–1752.

Takizawa, M., Tolarová, H., Li, Z., Dubois, W., Lim, S., Callen, E., Franco, S., Mosaico, M., Feigenbaum, L., Alt, F.W., et al. (2008). AID expression levels determine the extent of cMyc oncogenic translocations and the incidence of B cell tumor development. J. Exp. Med. *205*, 1949–1957.

Taylor, B.J., Nik-Zainal, S., Wu, Y.L., Stebbings, L.A., Raine, K., Campbell, P.J., Rada, C., Stratton, M.R., and Neuberger, M.S. (2013). DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. eLife *2*, e00534.

Tsai, A.G., Lu, H., Raghavan, S.C., Muschen, M., Hsieh, C.L., and Lieber, M.R. (2008). Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. Cell *135*, 1130–1142.

Tumas-Brundage, K., and Manser, T. (1997). The transcriptional promoter regulates hypermutation of the antibody heavy chain locus. J. Exp. Med. *185*, 239–250.

van de Werken, H.J., Landan, G., Holwerda, S.J., Hoichman, M., Klous, P., Chachik, R., Splinter, E., Valdes-Quezada, C., Oz, Y., Bouwman, B.A., et al. (2012). Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nat. Methods *9*, 969–972.

Wang, L., Wuerffel, R., Feldman, S., Khamlichi, A.A., and Kenter, A.L. (2009). S region sequence, RNA polymerase II, and histone modifications create chromatin accessibility during class switch recombination. J. Exp. Med. *206*, 1817–1830.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H.W., Robbiani, D.F., McBride, K., Nussenzweig, M.C., and Casellas, R. (2011). Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. Nat. Immunol. *12*, 62–69.

Yamane, A., Robbiani, D.F., Resch, W., Bothmer, A., Nakahashi, H., Oliveira, T., Rommel, P.C., Brown, E.J., Nussenzweig, A., Nussenzweig, M.C., and Casellas, R. (2013). RPA accumulation during class switch recombination represents 5′-3′ DNA-end resection during the S-G2/M phase of the cell cycle. Cell Reports *3*, 138–147.

Zhang, Y., McCord, R.P., Ho, Y.J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., Alt, F.W., and Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. Cell *148*, 908–921.