Cell

Article

Convergent Transcription at Intragenic Super-Enhancers Targets AID-Initiated Genomic Instability

Fei-Long Meng,[1,7] Zhou Du,[1,2,7] Alexander Federation,[3,7] Jiazhi Hu,[1] Qiao Wang,[4] Kyong-Rim Kieffer-Kwon,[5] Robin M. Meyers,[1] Corina Amor,[1] Caitlyn R. Wasserman,[1] Donna Neuberg,[6] Rafael Casellas,[5] Michel C. Nussenzweig,[4] James E. Bradner,[3,8,*] X. Shirley Liu,[6,8] and Frederick W. Alt[1,8,*]

[1]Howard Hughes Medical Institute, Program in Cellular and Molecular Medicine, Boston Children's Hospital, and Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

[2]Department of Bioinformatics, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China

[3]Department of Medical Oncology, Dana-Farber Cancer Institute, and Department of Medicine, Harvard Medical School, Boston, MA 02115, USA

[4]Howard Hughes Medical Institute, Laboratory of Molecular Immunology, The Rockefeller University, New York, NY 10065, USA

[5]Genomics and Immunity, NIAMS, and Center of Cancer Research, NCI, National Institutes of Health, Bethesda, MD 20892, USA

[6]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, MA 02115, USA

[7]Co-first author

[8]Co-senior author

*Correspondence: james_bradner@dfci.harvard.edu (J.E.B.), alt@enders.tch.harvard.edu (F.W.A.)

http://dx.doi.org/10.1016/j.cell.2014.11.014

SUMMARY

Activation-induced cytidine deaminase (AID) initiates both somatic hypermutation (SHM) for antibody affinity maturation and DNA breakage for antibody class switch recombination (CSR) via transcription-dependent cytidine deamination of single-stranded DNA targets. Though largely specific for immunoglobulin genes, AID also acts on a limited set of off-targets, generating oncogenic translocations and mutations that contribute to B cell lymphoma. How AID is recruited to off-targets has been a long-standing mystery. Based on deep GRO-seq studies of mouse and human B lineage cells activated for CSR or SHM, we report that most robust AID off-target translocations occur within highly focal regions of target genes in which sense and antisense transcription converge. Moreover, we found that such AID-targeting "convergent" transcription arises from antisense transcription that emanates from super-enhancers within sense transcribed gene bodies. Our findings provide an explanation for AID off-targeting to a small subset of mostly lineage-specific genes in activated B cells.

INTRODUCTION

The B cell antigen receptor (BCR) is comprised of immunoglobulin (Ig) heavy (IgH) and light (IgL) chains. In response to antigen activation, B lymphocytes in peripheral lymphoid organs undergo somatic hypermutation (SHM) and IgH class switch recombination (CSR) and ultimately secrete their BCR as an antibody. SHM diversifies antibody repertoires by introducing high-frequency mutations into IgH and IgL variable region exons (Di Noia and Neuberger, 2007). SHM occurs in germinal centers (GCs) of peripheral lymphoid tissues, where B cells are selected for mutations that generate BCRs with increased antigen affinity (Victora and Nussenzweig, 2012). IgH CSR involves generation and joining of *IgH* locus DSBs in switch (S) regions that precede various sets of *IgH* C_H exons (C_Hs) to replace the initially expressed C_H with a downstream C_H, thereby producing antibodies with different effector functions (Matthews et al., 2014). Both SHM and CSR are initiated by activation-induced cytidine deaminase (AID) (Muramatsu et al., 2000), which deaminates cytosine to uridine on single-stranded DNA (ssDNA) (Di Noia and Neuberger, 2007). Mismatches created by these deaminated cytidines are processed into mutations or DSBs during SHM and CSR, respectively, through a process that employs activities of normal base excision or mismatch repair pathways (Di Noia and Neuberger, 2007).

Within target sequences, AID cytidine deamination focuses on 3–4 bp "SHM" motifs that are greatly enriched in S regions and in portions of variable region exons that encode antigen-binding sites (Di Noia and Neuberger, 2007). Transcription is required for AID targeting during SHM and CSR (Alt et al., 2013; Storb, 2014). In this regard, SHM of V(D)J exons in GC B cells begins ~150 bp downstream of the transcription start site (TSS) and tapers off 1–2 kb downstream (Liu and Schatz, 2009). Likewise, each C_H has a promoter upstream of the S region that upon induction by external signals generates transcription through the S region and, thereby, targets AID (Matthews et al., 2014). Mouse and human S regions also have a highly G-rich nontemplate strand that upon transcription forms stable R-loops that provide ssDNA to augment AID targeting (Matthews et al., 2014; Alt et al., 2013). RNA polymerase II (Pol II) has been implicated in directing AID to Ig gene SHM and CSR targets through a transcription

coupled mechanism (Storb, 2014) that involves AID association with the Spt5 transcription cofactor in the context of Pol II stalling (Pavri et al., 2010). R loops or other aspects of repetitive S region structure may augment AID access by promoting Pol II stalling (Rajagopal et al., 2009; Wang et al., 2009). Once AID is recruited to Ig targets, replication protein A (RPA) and the RNA exosome RNA degradation complex contribute to generating requisite ssDNA substrates (Basu et al., 2011; Matthews et al., 2014; Pefanis et al., 2014).

Beyond Ig gene targets, AID initiates recurrent mutations or DSBs in a small subset of non-Ig genes collectively termed AID "off-target" genes (Pasqualucci et al., 2001; Chiarle et al., 2011; Klein et al., 2011; Liu et al., 2008). Off-target AID activity promotes translocations between Ig loci and cellular oncogenes, as well as SHMs of oncogenes associated with B cell lymphomas (Alt et al., 2013; Küppers and Dalla-Favera, 2001). Identification of AID off-targets has been facilitated by genome-wide translocation cloning methods (Chiarle et al., 2011; Klein et al., 2011) and other large-scale approaches (Liu et al., 2008; Yamane et al., 2011). In general, AID activity occurs at much lower levels on off-targets than on Ig genes (Liu and Schatz, 2009; Yamane et al., 2011; Chiarle et al., 2011; Klein et al., 2011), likely due to specialized AID-targeting features of the latter. AID off-target sequences are not enriched in AID hotspot motifs relative to the genome in general (Duke et al., 2013). Consistent with a role for transcription, AID off-target activity is most abundant on transcribed genes downstream of their TSSs (Pasqualucci et al., 2001; Liu et al., 2008; Chiarle et al., 2011; Klein et al., 2011). However, transcription per se is not sufficient to target AID, as most transcribed genes are not AID off-targets (Alt et al., 2013; Liu and Schatz, 2009). Next-generation sequencing studies revealed unexpected transcriptional features, including divergent sense and antisense transcription at TSSs (Wu and Sharp, 2013; Adelman and Lis, 2012) and frequent promoter proximal Pol II pausing (Adelman and Lis, 2012). But, divergent transcription (DivT) from TSSs occurs in over half of all genes and generally does not map directly to sites of AID off-target activity (Chiarle et al., 2011; see below). Likewise, transcriptional pausing alone cannot explain AID off-targeting, because more than 30% of transcribed genes have paused Pol II (Adelman and Lis, 2012). Thus, mechanisms that lead to recurrent AID targeting may arise from previously unrecognized transcriptional or epigenetic determinants (Alt et al., 2013).

Global run-on sequencing (GRO-seq) detects nascent transcripts generated by transcriptionally engaged RNA polymerases (Core et al., 2008). GRO-seq revealed that a large fraction of intergenic regions are transcribed, with a subset emanating from transcriptional enhancers (Wang et al., 2011). Enhancers are sequence-defined, *cis*-regulatory elements that influence target gene expression irrespective of orientation (Levine et al., 2014). Both enhancers within genes (intragenic) and intergenic enhancers may regulate target promoters locally and over long distances (Levine et al., 2014). Active enhancer sequences are commonly transcribed by RNA Pol II generating so-called "enhancer RNAs" (eRNAs), and transcription arising from enhancers is often divergent, with both sense and antisense transcription emanating from enhancer elements (Natoli and Andrau, 2012; Wang et al., 2011). Various regulatory functions have been ascribed to eRNAs and other noncoding RNAs (Lam et al., 2014), however, much of noncoding RNA biology is not fully understood.

Enhancers are comprised of discrete or clustered transcription factor binding sequences. A common feature of active enhancers is chromatin that is characteristically modified by acetylation (e.g., histone 3 lysine 27; H3K27Ac) and methylation (e.g., histone 3 lysine 4 mono-methylation; H3K4me1) (Creyghton et al., 2010). An unexpected asymmetry in the regional allocation of enhancer factors and enrichment for enhancer marks within and unique to each mammalian cell type studied revealed a subset of so-called super-enhancers (SEs) that feature clusters of hyperacetylated and actively transcribed enhancers that, on average, are 10-fold longer than other "typical" enhancers (Whyte et al., 2013; Lovén et al., 2013). Like locus control regions, SEs regulate genes involved in specialized cellular function (Parker et al., 2013) and are found within or adjacent to lineage-specifying transcription factor genes (Whyte et al., 2013; Hnisz et al., 2013). In cancer, SEs frequently enforce oncogene expression (Lovén et al., 2013) and, thereby, contribute to tumor pathogenesis. For example, translocations that juxtapose *c-myc* to the IgH 3′ regulatory region, a known SE (Delmore et al., 2011; Chapuy et al., 2013), promote B cell lymphoma by activating *c-myc* over long distances (Gostissa et al., 2009). In this context, selectively blocking SE activity with bromodomain and extra-terminal domain (BET) inhibitors is a promising cancer therapeutic strategy (Delmore et al., 2011; Lovén et al., 2013; Chapuy et al., 2013).

Here, we report that the majority of detectable AID off-target activity in a variety of mouse and human lymphoid or nonlymphoid cell types occurs within focal regions of overlapping sense/antisense transcription within intragenic SEs.

RESULTS

Deep GRO-Seq Transcription Profiles of Naive, GC, and CSR-Activated B Cells

To elucidate transcriptional features that influence AID targeting genome-wide, we applied GRO-seq to splenic naive, GC, and CSR-activated B cells at much greater depth than done previously. Naive splenic B cells were purified (Figure S1A available online) and then cultured in the presence of αCD40 plus interleukin-4 (IL4) for 60 hr to stimulate AID induction and CSR to IgG1 and IgE (Figure S1A). Splenic GC B cells were purified from sheep red blood cell immunized mice (Figure S1A) and confirmed to be >90% pure (Figures S1B–S1D). Three independent GRO-seq biological replicates were performed for each cell type and gave highly reproducible results (Figure S1E). Transcription profiles of over 20,000 genes revealed distinct (but overlapping) gene expression patterns for each cell type that were further classified by gene ontology terms (Figure S1G; Table S1). As expected (Core et al., 2008; Chiarle et al., 2011), GRO-seq revealed divergent sense and antisense transcription at TSSs of over 50% of the genes in each of the three cell types (Figures 1 and S1F). In-depth examination of sense transcription profiles of several "signature" genes illustrates the specificity of purified cell populations. For example, *Aicda* sense transcription reflects AID protein expression in the three cell types, with high levels in GC B cells and activated B cells; but none detectable

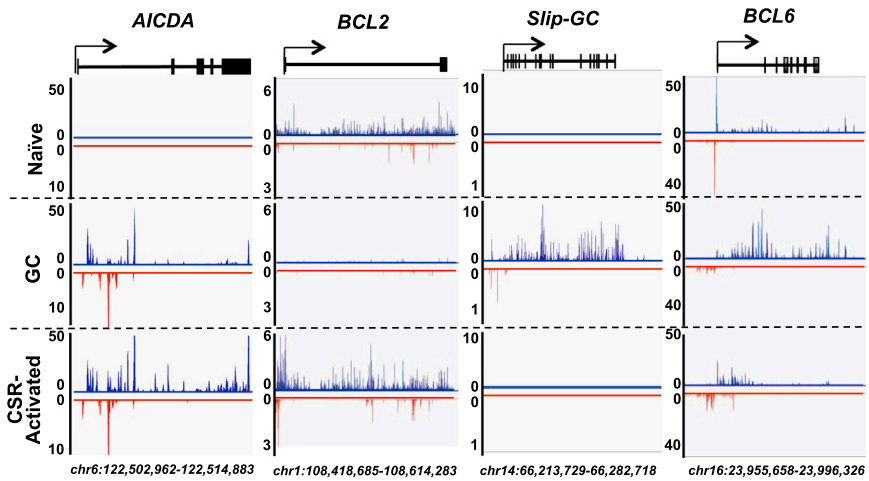

Figure 1. GRO-Seq Profiles of Naive, Germinal Center, and CSR-Activated B Cells

GRO-seq profiles of four representative genes are shown for different B cell types. The y axis indicates GRO-seq counts normalized to number of reads per million. Gene sense and antisense transcription are displayed in blue and red, respectively. Gene exons are illustrated by squares along gene bodies at the top of each panel. Arrows indicate TSSs and direction of sense transcription. Genome coordinates (mm9/NCBI37) are labeled at the bottom. All the profiles were generated from merged data of three independent experiments, which individually showed similar patterns.

See also Figure S1 and Table S1.

in naive B cells (Figure 1). In contrast, several GC B cell-specific genes, including *SLIP-GC* (Richter et al., 2009) and *Bcl6* (Basso and Dalla-Favera, 2010), had high sense transcription through their gene bodies in GC B cells, but not in naive or CSR-activated B cells (Figure 1). Finally, *Bcl2*, which is expressed in CSR-activated but not in GC B cells (Liu et al., 1991), showed corresponding sense transcription patterns (Figure 1).

While *IgH* C_H exons were appropriately transcribed in the three cell populations (Figure S1H), transcription within core S regions could not be mapped due to their abundant repetitive sequence (Pavri et al., 2010). All analyzed mice had a clonal knock-in V_H(D)J_H exon (V_HB1-8) (Sonoda et al., 1997), which showed active transcription at its upstream regions in all three cell types (Figure S1H). However, detailed analyses of transcription through the body of the V_HB1-8 allele was not possible (Figure S1H); because it uses a member of the V_HJ558 family, which contains many highly related, unexpressed upstream copies (Brodeur and Riblet, 1984).

Enhanced Identification of AID Off-Target Sites in αCD40 plus IL4-Stimulated B Cells

We developed high-throughput genome-wide translocation sequencing (HTGTS) to map, at the nucleotide level, translocation junctions between bait I-SceI nuclease generated DSBs in c-*myc* and other endogenous DSBs (Chiarle et al., 2011). Identification of DSB hotspots from a fixed chromosomal site is facilitated by ability of recurrent DSBs to dominate genome-wide translocation landscapes due to cellular heterogeneity in 3D genome organization (Zhang et al., 2012). Beyond expected Ig locus targets, our prior HTGTS studies revealed 15 non-Ig genes that are recurrent targets of AID-initiated DSBs and translocations (Chiarle et al., 2011) (Table S2). To increase the depth of HTGTS AID off-target data and allow better comparison with deeper GRO-seq transcription profiles, we further employed a modified, more sensitive HTGTS approach (Frock et al., 2015), coupled with ataxia telangiectasia mutated (ATM)-deficient CSR-activated B cells (Hu et al., 2014). This combined approach identified highly clustered AID-dependent off-target DSB sites within 36 additional genes (Figure S2A; Extended Experimental Procedures; Table S2). Overall, we now have identified 51 highly focal AID off-target DSB/translocation sites in αCD40 plus IL4-stimulated B cells (Table S2). Nearly 90% of the new off-target set was validated in WT B cells by HTGTS and/or by an independent method (Qian et al., 2014 in this issue of *Cell*) (Extended Experimental Procedures). As previously found for our more limited set of AID off-target sites (Chiarle et al., 2011), many of our new AID off-targets occurred within genes that have divergently transcribed TSSs; but the focal sites of HTGTS junctions within them were downstream of and distinct from divergently transcribed TSSs (Chiarle et al., 2011) (Figure 2). Thus, we were compelled to search for other factors that promote such focal AID off-targeting. As we found no enrichment for known AID targeting motifs in these regions (Extended Experimental Procedures), we focused our search on potentially novel transcriptional and/or epigenetic features and, as described below, identified both.

AID Off-Targets Cluster at Sense/Antisense Transcription Sites Downstream of the TSS

With our present, substantially deeper, GRO-seq data sets, we further analyzed potential relationships between sense/antisense transcription and AID off-target sites in αCD40 plus IL4-activated B cells. Initially, we visually inspected three linked AID off-target sites, including sites in the previously characterized *IL4r* and *IL21r* genes (Chiarle et al., 2011) and a newly identified site in *Nsmce1*. In each of these linked genes, HTGTS translocation junctions were tightly clustered in a region downstream of the TSS (Figure 2A). Moreover, in each, translocation clusters fell within sites that exhibited enriched, overlapping sense and antisense transcription to which we heretofore apply the term "convergent transcription" (ConvT) (Figures 2A and 3A). We also found a robust AID off-target site within the AID gene (*Aicda*) itself (Figure 2B; Table S2). *Aicda* is associated with five enhancers that lie upstream, within, or downstream of the gene body (Kieffer-Kwon et al., 2013; Matthews et al., 2014) (Figure 2B). Four of these enhancers showed both sense and antisense transcription, likely at least in part in the context of generating eRNAs (Natoli and Andrau, 2012) (Figure 2B). Notably, the major focal cluster of AID off-target sites in and around *Aicda* fell within a ConvT region associated with enhancer 4 downstream of the TSS (Figure 2B).

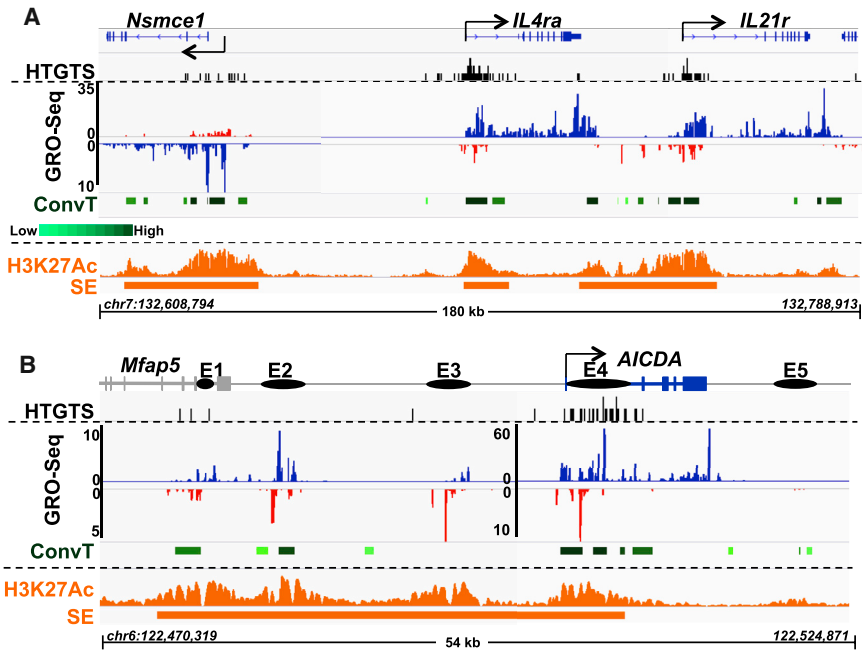

Figure 2. AID Off-Target Translocations Cluster within Regions of ConvT and SEs

(A) HTGTS, GRO-seq, ConvT, and H3K27Ac profiles in the vicinity of *Nsmce1*, *IL4ra*, and *IL21r* genes. Top: HTGTS junctions are indicated by black bars. Middle (GRO-seq): GRO-seq-determined sense and antisense transcription is displayed in blue and red, respectively. ConvT regions are shown as green bars at the bottom with the darkest shades corresponding to highest levels of ConvT as calculated by the geometric means of sense and antisense transcription reads (see Extended Experimental Procedures). A scale bar is shown below the ConvT label. Bottom (H3K27Ac and SE): the H3K27Ac ChIP-seq profile is shown in orange and identified SEs depicted below with orange bars. *Nsmce1* TSS is manually curated based on GRO-seq profile.

(B) Profile of *AICDA* gene. Known *AICDA* enhancers are represented as E1–E5 with solid circles. To represent lower level transcription of certain enhancers, a smaller scale is used for E1–E3. Genome coordinates (mm9/NCBI37) are at the bottom of each panel. Other details are the same as for (A).

See also Figure S2 and Table S2.

Genome-wide Association of ConvT and AID Off-Targets in CSR-Activated B Cells

Visual inspection of AID off-target sites in additional genes revealed similar coincidence of regions of robust sense/antisense (S/AS) ConvT downstream of the TSS with focal clusters of AID-dependent off-target translocations (see below, Figure S2), leading us to examine this potentially striking association genome-wide. While metagene profiles of GRO-seq data from αCD40 plus IL4-activated B cells confirmed expected DivT at many TSSs (Wu and Sharp, 2013), they did not reveal similarly abundant convergent transcription (Figure S1F). Thus, at least at robust levels, convergent transcription likely occurs in a much smaller fraction of genes (Figure S1F). For further analyses, we developed a computational pipeline to specifically identify S/AS ConvT regions genome-wide using deep GRO-seq data sets (Figure 3A; Extended Experimental Procedures). Strikingly, among the 51 AID off-target genes, 48 (94%) had their highly clustered AID off-target translocations within regions associated with S/AS convergent transcription (Figure 3B). We randomly sampled convergent transcription of regions, in the top three transcription-level deciles, that were similar in size to those of AID off-target regions and found a much lower association with convergent transcription than for AID off-target regions (Figure S3A). This finding shows that AID off-targets are highly enriched at ConvT sites. Finally, concurrency between S/AS convergent transcription and AID off-target translocations was much higher in αCD40 plus IL4-activated B cells (94%) than in naive (49%) or GC (63%) B cells, consistent the notion that not all AID off-targets would be shared among three cell types with overlapping, but clearly distinct, transcription profiles (Figures 3B and S3B; also see below).

To further examine the relationship between ConvT and AID targeting, we calculated the geometric mean of GRO-seq sense and antisense transcription reads in regions of interest to quantify degree of convergent transcription (Extended Experimental Procedures) and divided the values into deciles displayed by different shades of green bars below the GRO-seq profiles (Figures 2 and S2; dark green is highest and light green lowest levels). For most AID off-targets, HTGTS junctions clustered in regions with the most abundant ConvT (Figures 2 and S2). Furthermore, ConvT associated with AID off-targets was substantially greater than that at other genomic loci (Figure S3C). In addition, within AID off-target ConvT regions, the highest density of translocations occurred at sites with the most robust ConvT (Figure 3C). We further evaluated this relationship by determining how variations in sequencing depth influenced identification of ConvT. Even with our current very deep sequencing depth (>306 million mappable reads), we did not reach saturation of the total length of ConvT regions (Figure S3D), consistent with (at least low-level) pervasive transcription of the genome (Jacquier, 2009). In contrast, we reached saturation of the concurrency of AID off-targets with ConvT regions at ~40% of our current GRO-seq depth (120 million mappable reads; Figure S3D), confirming that most AID off-target DSB/translocation regions detectable by HTGTS in αCD40 plus IL4-stimulated B cells are associated with relatively strong convergent transcription (Figures 3C and S3D).

Convergent Transcription at AID Off-Targets Arises from Intragenic SEs

ConvT of overlapping genes was first described in bacteriophage lambda (Ward and Murray, 1979) and has been associated with transcriptional gene silencing (Gullerova and Proudfoot, 2012) and RNA Pol II collision (Hobson et al., 2012). Considering that intragenic antisense transcription associated with AID-off target sequences may arise from enhancer elements, we explored whether intragenic SEs were enriched for

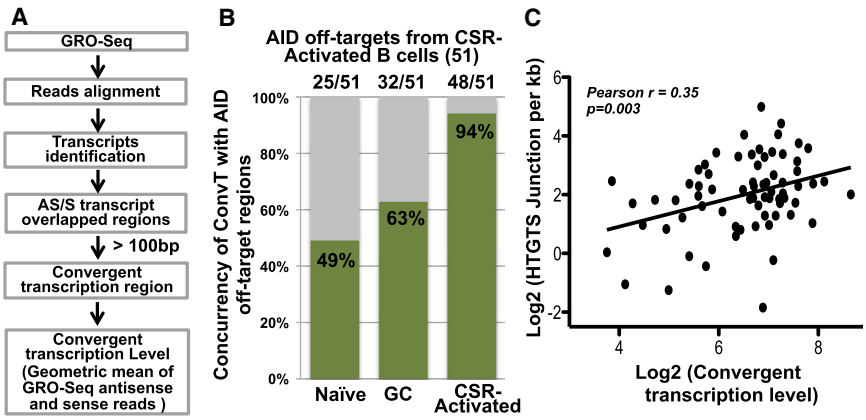

Figure 3. AID Off-Targets Correlate with ConvT in CSR-Activated B Cells

(A) Pipeline for identification of ConvT regions. Raw GRO-seq reads were aligned to the genome and transcripts were identified de novo. A "ConvT" region was defined as sense and antisense transcription overlaps that were longer than 100 bp. See Extended Experimental Procedures for details.

(B) The percentage of the 51 AID off-target regions identified in CSR-activated B cells that were associated with ConvT regions in the three listed cell populations is the indicated by the green bars.

(C) Numbers of translocation junctions per kb (y axis) plotted against ConvT levels (x axis) of all individual AID off-target regions except *Pvt1* (see Extended Experimental Procedures). Pearson's correlation coefficient and two-tailed p value are indicated.

See also Figure S3.

AID off-targets compared to typical enhancers. Enhancer regions were identified by triplicate chromatin immunoprecipitation sequencing (ChIP-seq) using an antibody to the active enhancer histone mark H3K27Ac in chromatin purified from αCD40 plus IL4-stimulated B cells (Figure S4A). SEs were called for regions of asymmetric, high enrichment for H3K27Ac, as previously described (Whyte et al., 2013). We found the *Aicda* locus to be largely encompassed within a SE in CSR-activated B cells with robust H3K27Ac signals over E1, E2, E3, and E4 (Figure 2B), the active enhancers in CSR-activated B cells (Kieffer-Kwon et al., 2013; Matthews et al., 2014). Notably, E4 also corresponds in position to a cluster of HTGTS junctions and robust ConvT (Figure 2B). Likewise, the *Nsmce1*, *IL4ra*, *Il21r*, and many other AID off-target genes were each associated with SEs and again the peak of HTGTS junctions and regions of robust ConvT occurred within regions of robust H3K27Ac SE signals (Figures 2A and S2).

We performed an unbiased association analysis between the 51 AID off-targets identified by HTGTS and the non-Ig 448 SEs that we identified in αCD40 plus IL4-activated B cells. These studies revealed that 50 of the 51 AID off-target genes in these cells are associated with SEs and that the discrete translocation clusters were within SEs (Figure 4A). Notably, the single AID off-target region not within a SE (under the current cutoff for SE identification; Extended Experimental Procedures) was in a typical enhancer (Table S2). In addition, 47 (92%) of the AID off-target translocation clusters were within regions of SEs that overlap with annotated gene bodies (Figure 4A). The other three HTGTS off-target translocation clusters occurred within transcribed regions of SEs that have not yet been assigned to a target gene (Table S3). As a comparison, random samplings of transcribed genomic regions corresponding in size to those of AID off-targets yielded at most three (6%) that overlapped with SEs. Independent analysis of the relationship between HTGTS hotspots and H3K27Ac ChIP-seq using an orthogonal computational method identified 41 AID off-targets within SE domains (Figures S4C and S4D), including additional off-targets that correlated with robust ConvT (Figure S2; Table S2; Extended Experimental Procedures). Finally, within a given AID off-target region, translocation junction frequency highly correlated with H3K27Ac abundance (Figure S4B). In this regard, SEs associated with AID off-target sequences were more enriched for H3K27Ac, compared to other SEs (Figure 4B). Thus, the relative activity of SEs, estimated by regional histone acetylation, correlates with the frequency of AID off-targets within them.

The majority (30 of 51) of the AID off-target genes had a SE that overlapped with the region just downstream of the TSS that was enriched in AID off-targets, as represented by the *CD83* gene (Figures 4C and S2). In addition, a number (12 of 51) of the AID targets were relatively small genes, such as *Pim1*, that were located within large SEs and, correspondingly, off-target translocations tended to span the gene body (Figures 4D and S2). Several AID off-target genes (3 of 51) were large genes, such as *Pvt1*, the well-known translocation target downstream of *c-myc*, in which translocations clustered within SEs that occurred inside the gene body (Figure 4E). Finally, the remainder (6 of 51) fell into a heterogeneous set in which AID off-target translocations clustered into convergently transcribed SE domains that, for various reasons were not yet assignable to a specific gene (e.g., *Gpr183*; Figure S2C).

Intragenic SEs with Robust ConvT Represent the Most Common AID Off-Targets

Nearly all AID off-target clusters identified by HTGTS in αCD40 plus IL4-activated B cells are associated with SEs; yet, only a subset of SEs are AID off-targets. Motivated by the putative contribution of S/AS eRNA transcription to translocation frequency, we compared regions of AID off-target genes where SEs overlap with the gene body (intragenic SEs) to regions where SEs lie outside the gene body (intergenic SEs) and to regions of gene bodies that do not overlap with SEs (nonoverlapping gene region), for translocation density (translocations per 1 kb; Figure 5A) and for ConvT levels (geometric means; Figure 5A). We observed that translocation junction density and ConvT levels in AID off-target regions are highly enriched among intragenic SEs compared to both intergenic SEs and nonoverlapping gene regions (Figure 5A; upper). Despite this enrichment, only ~10% of all intragenic SEs in the CSR-activated B cells are AID off-targets (Figure 4A; Table S3) and other SE-gene overlap regions exist that are not enriched in AID off-target activity

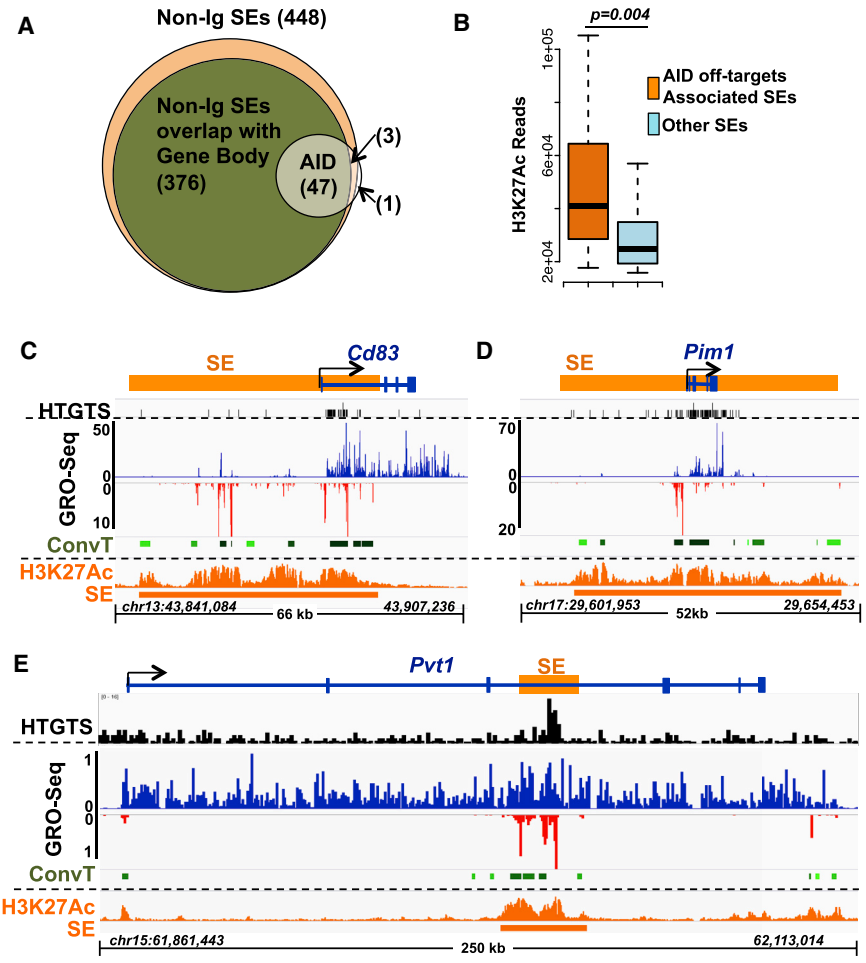

Figure 4. AID Off-Target ConvT Arising from Intragenic SEs

(A) Venn diagram showing the number of AID off-target regions that overlapped with total non-Ig SEs (448) and with non-Ig SEs overlapping with Gene Bodies (376).

(B) H3K27Ac signals of AID off-target-associated SEs (orange) and the other SEs (cyan) are plotted. AID off-target-associated SEs had a stronger H3K27Ac signal (Mann-Whitney U test, p value = 0.004).

For C, D, and E, representative AID off-targets are shown based on the SE location indicated in the diagram at the top of each panel.

(C) Many AID targets locate downstream of TSSs where SEs and genes overlap. *CD83* is shown as an example.

(D) For some relatively small genes located within a larger SE, nearly the whole gene body is an AID off-target, as shown for *Pim1*.

(E) SEs inside of very long genes, like *Pvt1* also provide focal AID off-targets. HTGTS, GRO-seq, and H3K27Ac/SE data is illustrated for each panel as described in Figure 2A. The relatively high HTGTS background in *Pvt1* results from long resections downstream of the HTGTS bait DSB in c-*myc* (Chiarle et al., 2011).

See also Figure S4 and Table S3.

(Figure 5B; upper). Comparison of ConvT levels in each of the three regions outlined above (Figures 5A and 5B, lower panels) revealed that intragenic SEs featuring high levels of ConvT were more frequently AID off-target regions than intragenic SEs lacking high-level S/AS transcription (Figures 5A and 5B, lower panels).

Finally, to further address why some SEs are AID targets and others are not, we grouped all intragenic SEs into deciles based on low to high convergent transcription (Figure 5C). We then calculated the percentage of the combined 228 unique AID off-targets revealed by HTGTS (this study) and by an independent RPA-ChIP study (Qian et al., 2014) in CSR-activated B cells in each decile. Strikingly, 60% of all SEs within the top two deciles (highest convergent transcription) were sites of clustered AID off-target DSBs and/or translocations. Comparative analysis of SEs in these top two deciles that were AID off-targets versus those that were not did not reveal any obvious sequence differences (e.g., GC content or WRCH and AGCT motifs density). However, ConvT regions associated with SEs in the top two deciles that were AID off-targets were significantly longer than those that were not (Figure 5D). These studies provide strong evidence that ConvT from intergeneic SEs generates a major class of focal AID off-target regions.

AID Off-Targets in GC B Cells Associate with Convergent Transcription

Prior studies of a selected set of AID off-targets divided them into three groups in GC B cells based on mutation frequency in Ung/Msh2 double-deficient B cells versus AID-deficient B cells, including 15 group A genes that had high levels of mutation, 21 group B genes that had substantially lower levels, and 47 group C genes that were infrequently mutated (Liu et al., 2008). Our GRO-seq analyses of GC B cells revealed that ~70% of the highly mutated group A gene off-target regions, including *Pim1*, *Ebf1*, *CD83*, and *Ocab*, overlapped with ConvT regions (Figures 6A and 6C) that were well above simulated background levels expected for the most highly transcribed genes (Figure S5A). In contrast, regions reported to have low level mutation frequency (groups B and C genes) showed low correlations with convergent transcription (33% and 32%, respectively; Figure 6A) that were not above simulated background concurrency (Figure S5A). Finally, of the five group A genes that did not associate directly with convergent transcription, SHMs in four occurred quite proximal to ConvT regions (Figure S5C). We identified SEs in GC B cells via H3K27Ac ChIP-seq analyses. We found that some SEs were shared between GC and CSR-activated B cells, while many others were found only in one or the other cell type (Figure S5B; Table S3), consistent with the overlapping but distinct GRO-seq profiles of these two cell types (Figure S3B; Table S1). Of the highly mutated group A gene regions, nearly half were associated with SEs (Figure 6B) and all were associated with H3K27Ac peaks (Figures 6C and S5C). For group B and C gene regions, concurrencies with SE were

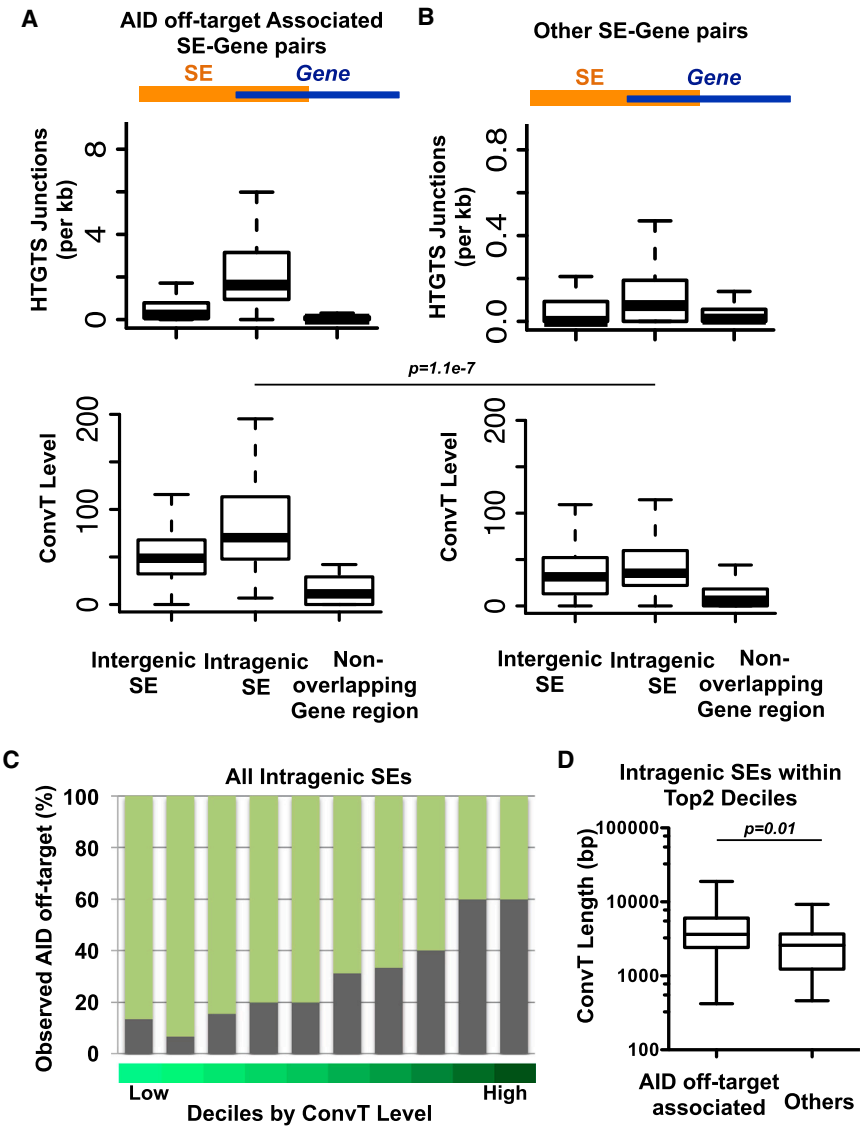

Figure 5. Convergently Transcribed Intragenic SEs Are Preferred AID Off-Targets

(A) Upper and lower: each SE associated with an AID off-target region and its overlapping gene body were divided into intergenic SEs, intragenic SEs, and nonoverlapping gene regions as described in the text and outlined at the top of the panels. For all AID off-targets, the number of translocation junctions per kb in each of the three regions (upper panel) and convergent transcription levels of each region (lower panel) are plotted.

(B) Upper and lower: each SE that was not associated with an AID off-target region and its overlapping gene body were divided into regions as described for (A) and translocation junction numbers per kb (upper panel) and convergent transcription levels (lower panel) plotted for each region. A Mann-Whitney U test was performed to compare two classifications of SEs for convergent transcription ratios within each of the three regions; the only significant difference found was that the AID-off-target intragenic SEs has a significantly higher convergent transcription ratio than non-AID off-target intragenic SEs (p value = 1.1×10^{-7}).

(C) All intragenic SEs were grouped into deciles based on the ConvT levels. The fraction of AID off-targets in each decile is indicated by gray bar.

(D) Intragenic SEs in the top two deciles are divided into those associated with AID off-targets (60%) and those that are not (40%). Length of ConvT regions was plotted and found to be significantly longer in the AID off-target-associated intragenic SEs (Mann-Whitney U test, p value = 0.01).

20% and 2%, respectively. Thus, under physiological conditions in the GC, AID often targets convergently transcribed intragenic SEs or, occasionally, typical enhancers.

Convergently Transcribed Intragenic SEs Target AID in Non-B Lymphoid and Human Cells

Ectopic manipulation of endogenous SEs and ConvT regions to assess affects on AID targeting would be problematic because these regions are the actual AID targets. As an alternative approach, we performed GRO-seq on mouse embryonic fibroblasts (MEFs) in which ectopic AID expression revealed a set of 29 AID off-target sequences, most of which were novel (Qian et al., 2014) (Table S4). Remarkably, we found that the great majority of these clustered MEF translocations occurred in ConvT regions (Figures 7A and S6A) that also were mostly also associated with SEs (Qian et al., 2014) (Table S4). We also tested the generality of our ConvT findings with respect to AID off-target events observed during SHM in the human Ramos Burkitt's lymphoma cell line. Strikingly, the majority of 54 AID off-targets identified in this line again were associated with SEs (Qian et al., 2014), and we found that most were clustered in regions of strong ConvT (Figures 7B and S6B; Table S4). As discussed below, we have also extended our findings to human B cell lymphoma translocations.

DISCUSSION

Off-Target AID Activity in Convergently Transcribed Intragenic SEs

We report that most AID off-target DSBs and translocations in CSR-activated B cells occur in and around ConvT regions within genes (Figure 3). Furthermore, most of these AID off-target sites in CSR-activated B cells occurred within portions of genes that overlapped with enhancers, the vast majority of which were SEs (Figure 4). Together, these findings implicate a role for SEs within genes in generating robust ConvT and, thereby, in creating susceptibility to AID off-target activity. Notably, we also found that the majority of the regions with highest levels of off-target AID activity in GC B cells or in human Ramos cells

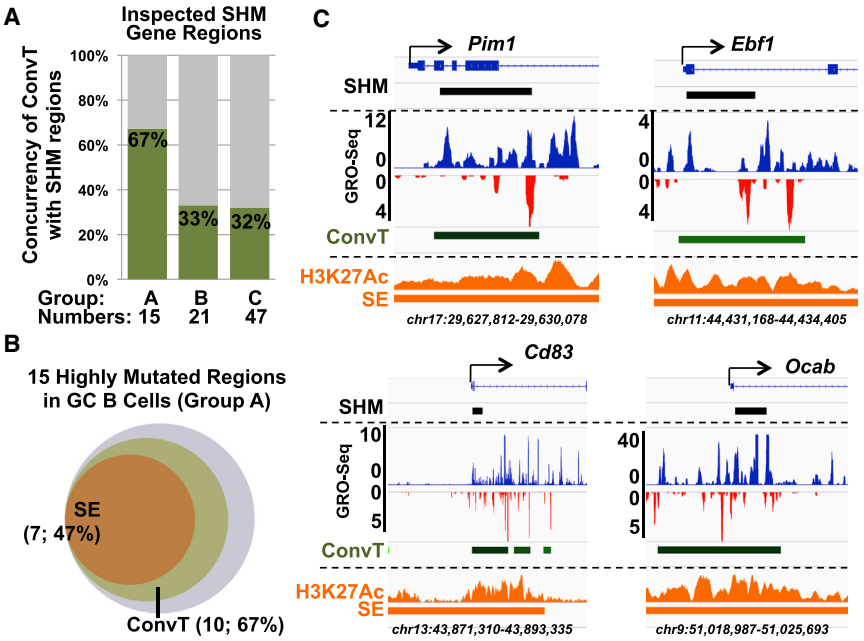

Figure 6. ConvT and SEs Correlate with AID Off-Targets in GC B Cells

Regions of genes containing SHMs in Ung/Msh2 double deficient GC B cells were analyzed for convergent transcription as determined by GRO-seq and outlined in Figure 3. GC AID off-target group A, B, and C genes include gene regions with high, intermediate, and low frequencies of AID-dependent mutations, respectively (Liu et al., 2008).

(A) Concurrency of group A, B, and C gene ConvT regions in GC B cells.

(B) Venn diagram showing the number of group A gene regions that overlapped with SEs and ConvT.

(C) Examples of group A gene regions are shown. Approximately 2–3 kb regions around the TSSs of the indicated genes are shown. The "SHM" diagram at the top of each subpanel indicates regions of these genes included in the prior SHM analyses (Liu et al., 2008) with a black bar. GRO-seq profile, ConvT, H3K27Ac ChIP-seq profile, and SEs are shown as in Figure 2A.

See also Figure S5 and Table S3.

undergoing SHM are in focal areas of target genes that contain SEs and undergo robust ConvT (Figures 6 and 7). Even in non-lymphoid cells (MEFs) in which AID was ectopically expressed, we found that the great majority of 29 AID-dependent translocation clusters occurred in regions that underwent robust ConvT (Figure 7), confirming our findings for a totally different set of genes in a different cell type. Together, these finding strongly support a mechanistic link between AID off-target sequences and S/AS convergent transcription. A role for SEs in AID off-targeting also has been revealed by a separate study (Qian et al., 2014).

Potential Mechanisms by which SEs and ConvT Contribute to AID Off-Target Activity

RNA polymerase II (Pol II) transcriptional pausing or stalling contributes to directing AID to Ig gene SHM and CSR targets via a process thought to involve AID association with the Spt5 transcription cofactor (Pavri et al., 2010; Storb, 2014). Ig gene V(D)J exons and S regions likely evolved specific features to promote AID targeting (Alt et al., 2013). As AID off-target genes lack consistent sequence features of Ig gene AID targets (Duke et al., 2013), the question of how they attract AID has been long-standing. Our current findings implicate a mechanism that answers this question for the majority of AID off-targets (Figure 7C). Thus, most robust AID off-target DSBs, SHMs, and translocations occur within intragenic SEs, where we find ConvT that includes sense gene transcription and antisense transcription emanating from the SEs. In such AID off-target regions, antisense eRNA transcription generally occurs at lower levels than sense transcription (Figures 2 and 4). Thus, most genic sense transcription likely proceeds unimpaired to generate full length mRNAs with only a small fraction encountering antisense transcription, consistent with ability of cells to generate products of these genes (Storb, 2014). Prior yeast studies showed that, within convergently transcribed sequences, Pol II elongation complexes proceeding in opposite directions cannot bypass each other, and consequential Pol II collisions lead to stalling or stopping (Hobson et al., 2012). We propose that such Pol II stalling due to convergent transcription leads to AID recruitment and further downstream events similar to those implicated in specialized Ig gene targets (Figure 7C) (Pavri et al., 2010; Basu et al., 2011). Beyond AID recruitment, convergent transcription could also generate ssDNA substrates for AID. Thus, following Pol II collisions, RNA exosome or other RNase activities could remove nascent transcripts (Basu et al., 2011; Pefanis et al., 2014; Andersson et al., 2014) to provide local ssDNA targets (Figure 7C).

Implications of AID Off-Target Activity for AID On-Target Ig Gene Activity

AID activity generally occurs at much higher levels on specialized Ig gene targets than on off-targets (Liu and Schatz, 2009; Yamane et al., 2011; Chiarle et al., 2011; Klein et al., 2011). Whether or not the ConvT mechanism we propose for off-targets can be applied to on-targets remains to be determined. In CSR-activated B cells, we observed ConvT within the very 5′ Sμ region (Figure S1H). However, the transcription profile of core S regions cannot be obtained due to poor mappability of repetitive S regions (Pavri et al., 2010). Clearly, S regions evolved specialized structural features that facilitate AID recruitment and access to the ssDNA substrates (Alt et al., 2013). However, mechanisms by which AID specifically targets Ig variable region exons for SHM in GCs may be more relevant. In this regard, a long-standing paradox involves that fact that SHM of variable region exons occurs only in GC B cells and not in CSR-activated B cells, even though the variable region exons are transcribed in both (Liu and Schatz, 2009). Our preliminary analyses reveal potentially higher relative levels of

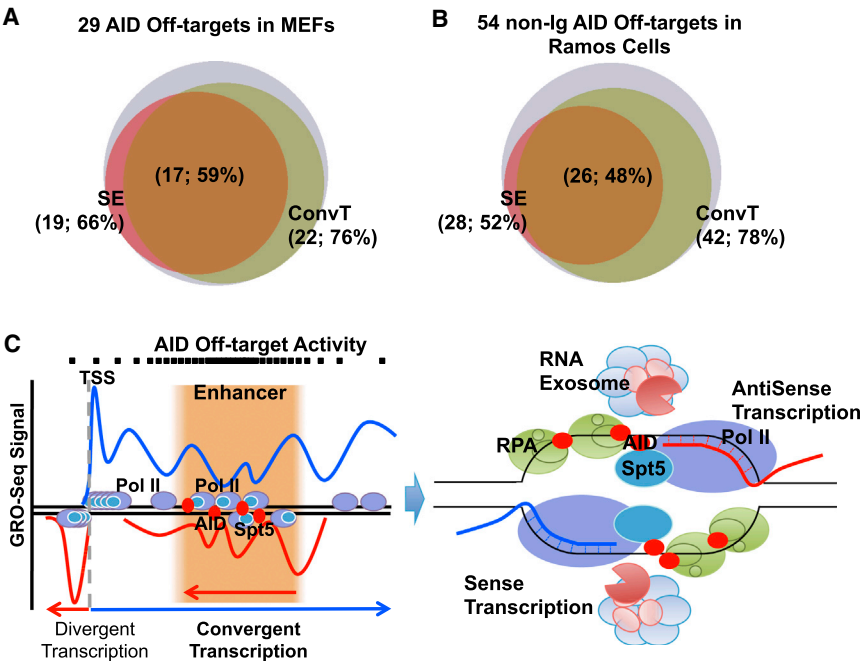

Figure 7. Model of AID Targeting at Off-Targets

(A) Venn diagram showing the number of AID off-target regions that overlapped with SEs and ConvT in MEFs with ectopic AID overexpression.

(B) Venn diagram showing number of AID off-target regions that overlapped with SEs and ConvT in Ramos Human Burkitt's lymphoma cell line.

(C) Model of AID "off-targeting." Left: at AID off-targets, SEs overlap with gene bodies and this combination generates regions of sense/antisense convergent transcription due to sense gene transcription encountering the enhancer antisense transcription. Right: stalled RNA polymerase with the help of Spt5 recruits AID and generates regions of ssDNA. RNA Exosome or other RNases degrade the aborted sense and antisense transcripts and works together with RPA to help AID access to the ssDNA substrates. Some aspects adapted from Basu et al. (2011). See Discussion for other details. See also Figure S6 and Table S4.

antisense to sense transcription on the downstream edge of the KI V(D)J (VB1-8) exon in GC versus naive or CSR-activated B cells (Figure S1H). However, as we cannot map transcription within the main body of the KI VB1-8 due to many highly related unexpressed, upstream V_HJ558 sequences, final testing of this potential mechanism for specific AID targeting of V(D)J exons will require additional mouse models that eliminate sequence redundancies.

Role of SE Transcription in Genome Instability and Cancer

SEs are important for establishment of cell lineage and expression of cell lineage-specific genes (Whyte et al., 2013; Hnisz et al., 2013). Correspondingly, SEs are associated frequently with genes highly expressed in activated B cells (Table S3). Many of the 51 genes that we have shown to have SEs that are AID off-targets are B cell-specific genes and a notably high proportion (25%) are known oncogenes (Figure S2B). In this regard, many human B cell lymphomas contain translocations or mutations of oncogenes that are initiated by off-target AID activity (Alt et al., 2013; Küppers and Dalla Favera, 2001. Reminiscent of the AID off-targeting pattern in mouse CSR-activated and GC B cells, human B cell oncogene translocation sites often occur downstream of TSSs (Migliazza et al., 1995; Pasqualucci et al., 2001; Shen et al., 1998). Indeed, we have analyzed SEs in human tonsil B cells (enriched in GC B cells) and found that many oncogene translocations in human B cell lymphoma, including those in *c-myc*, *Pax5*, *Bcl6*, *Bcl2*, *Pim1*, *Ocab*, *Lcp*, and *Bcl7a*, occur in regions downstream of TSSs where SEs overlap with gene bodies (Figure S6C). Thus, beyond contributing to deregulated oncogene expression (Chapuy et al., 2013), our findings suggest that SEs may target oncogenes for translocations in B cell lymphoma. Finally, AID also has been implicated in genomic instability and translocations in cells beyond those of the immune system (Lin et al., 2009; Marusawa et al., 2011). Our MEF studies suggest ConvT from SEs could play a role in such settings.

EXPERIMENTAL PROCEDURES

B Cell Purification

Splenic naive B cells were purified from V_HB1-8 heavy chain knock-in mice as described (Cato et al., 2011). Naive B cells were activated with αCD40 plus IL4 for 60 hr to generate CSR-activated B cells. V_HB1-8 knock-in mice were immunized with 5×10^8 sheep red blood cells (SRBCs) for 9 days. Splenic GC B cells were purified as described (Cato et al., 2011) (see Extended Experimental Procedures for details.) All animal experiments were performed under protocols approved by the Institutional Animal Care and Use Committee of Boston Children's Hospital.

GRO-Seq and ChIP-Seq

GRO-seq (Core et al., 2008) and H3K27Ac ChIP-seq (Chapuy et al., 2013) were performed as described. Three biological replicates of each mouse B cell type were performed. Two biological replicates of mouse MEF experiments and one biological replicate of Ramos experiments were performed.

AID Off-Targets

HTGTS was performed with αCD40 plus IL4 or RP105-activated ATM-deficient CSR-activated B cells as described (Hu et al., 2014) and also with a new HTGTS method (Frock et al., 2015). AID off-target coordinates were retrieved via a new HTGTS pipeline (Frock et al., 2015) (see Extended Experimental Procedures for details).

Data Analysis

GRO-seq and ChIP-seq data sets were aligned using Bowtie software (Langmead and Salzberg, 2012) to mouse genome build mm9/NCBI37 or human genome build hg19/NCBI37. Uniquely mapped, nonredundant sequence reads were retained. We used Homer software (Heinz et al., 2010) to de novo identify transcripts from both strands of the genome in the context of the GRO-seq data and considered broad sense/antisense overlap regions (>100 bp) as ConvT regions. We used the MACS1.4 software (Zhang et al., 2008) to identify regions of ChIP-seq enrichment over background with a p value threshold of 10^{-5}. We used ROSE software to

identify SEs (Whyte et al., 2013) (see Extended Experimental Procedures for details).

ACCESSION NUMBERS

The Gene Expression Omnibus databank accession number for all deep sequencing data reported in this paper is GSE62296.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2014.11.014.

AUTHOR CONTRIBUTIONS

F.L.M., Z.D., A.F., J.E.B., X.S.L., and F.W.A. designed the study. F.L.M, C.A., and C.R.W. purified B cells and performed GRO-seq. Z.D. analyzed GRO-seq data. F.L.M., A.F., and C.A performed H3K27Ac ChIP-seq. Z.D. and A.F. analyzed ChIP-seq data. J.H. performed HTGTS. J.H. and R.M. analyzed HTGTS data. Q.W. and M.C.N. prepared MEFs and supplied the AID off-target and SE list for MEFs. K.K. and R.C. supplied the AID off-target and SE list for Ramos cells and shared their AID off-target list for CSR-activated B cells. D.N. advised on statistical analysis and various aspects of data analysis. F.L.M., Z.D., A.F., and F.W.A. designed figures. F.W.A and F.L.M. drafted the manuscript. F.L.M., Z.D., A.F., J.E.B., X.S.L., and F.W.A. polished the manuscript.

ACKNOWLEDGMENTS

F.W.A was supported by NIH grants R01AI077595 and P01CA109901, and a Leukemia and Lymphoma Society (LLS) SCOR grant, and he is an investigator of the Howard Hughes Medical Institute. X.S.L was supported by NIH grant 1R01GM099409. A.F. and J.E.B. were supported by a Leukemia and Lymphoma Society SCOR, the National Science Foundation, and NIH grants 1R01 CA176745-01 and P01 CA109901. M.C.N. was supported by NIH grants AI072529 and AI037526 and is an investigator of the Howard Hughes Medical Institute. K.K. and R.C. were supported by the intramural research program of NIAMS, NIH. F.L.M. is a Lymphoma Research Foundation postdoctoral fellow and was a Cancer Research Institute postdoctoral fellow. J.Z. is supported by a Robertson Foundation/Cancer Research Institute Irvington Fellowship. Z.D. was supported by the National Science Foundation of China grant NSFC 31329003. The authors thank Drs. Yi Zhang and Li Shen for assistance with DNA sequencing and Dr. David Schatz (Yale University) for providing primers for GC SHM targets. Dr. Bradner is a scientific founder of Syros Pharmaceuticals, which is developing drugs targeting super-enhancers.

Received: August 1, 2014
Revised: October 1, 2014
Accepted: October 27, 2014
Published: December 4, 2014

REFERENCES

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. *13*, 720–731.

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.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al.; FANTOM Consortium (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461.

Basso, K., and Dalla-Favera, R. (2010). BCL6: master regulator of the germinal center reaction and key oncogene in B cell lymphomagenesis. Adv. Immunol. *105*, 193–210.

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.

Brodeur, P.H., and Riblet, R. (1984). The immunoglobulin heavy chain variable region (Igh-V) locus in the mouse. I. One hundred Igh-V genes comprise seven families of homologous genes. Eur. J. Immunol. *14*, 922–930.

Cato, M.H., Yau, I.W., and Rickert, R.C. (2011). Magnetic-based purification of untouched mouse germinal center B cells for ex vivo manipulation and biochemical analysis. Nat. Protoc. *6*, 953–960.

Chapuy, B., McKeown, M.R., Lin, C.Y., Monti, S., Roemer, M.G., Qi, J., Rahl, P.B., Sun, H.H., Yeda, K.T., Doench, J.G., et al. (2013). Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. Cancer Cell *24*, 777–790.

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.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845–1848.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936.

Delmore, J.E., Issa, G.C., Lemieux, M.E., Rahl, P.B., Shi, J., Jacobs, H.M., Kastritis, E., Gilpatrick, T., Paranal, R.M., Qi, J., et al. (2011). BET bromodomain inhibition as a therapeutic strategy to target c-Myc. Cell *146*, 904–917.

Di Noia, J.M., and Neuberger, M.S. (2007). Molecular mechanisms of antibody somatic hypermutation. Annu. Rev. Biochem. *76*, 1–22.

Duke, J.L., Liu, M., Yaari, G., Khalil, A.M., Tomayko, M.M., Shlomchik, M.J., Schatz, D.G., and Kleinstein, S.H. (2013). Multiple transcription factor binding sites predict AID targeting in non-Ig genes. J. Immunol. *190*, 3878–3888.

Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.-J., Kii, E., and Alt, F.W. (2014). Genome-wide detection of specific and non-specific DNA DSB activities of custom nucleases in human cells. Nat. Biotechnol. Published online January 4, 2015. http://dx.doi.org/10.1038/nbt.3101.

Gostissa, M., Yan, C.T., Bianco, J.M., Cogné, M., Pinaud, E., and Alt, F.W. (2009). Long-range oncogenic activation of Igh-c-myc translocations by the Igh 3′ regulatory region. Nature *462*, 803–807.

Gullerova, M., and Proudfoot, N.J. (2012). Convergent transcription induces transcriptional gene silencing in fission yeast and mammalian cells. Nat. Struct. Mol. Biol. *19*, 1193–1201.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

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.

Hobson, D.J., Wei, W., Steinmetz, L.M., and Svejstrup, J.Q. (2012). RNA polymerase II collision interrupts convergent transcription. Mol. Cell *48*, 365–374.

Hu, J., Tepsuporn, S., Meyers, R.M., Gostissa, M., and Alt, F.W. (2014). Developmental propagation of V(D)J recombination-associated DNA breaks and translocations in mature B cells via dicentric chromosomes. Proc. Natl. Acad. Sci. USA *111*, 10269–10274.

Jacquier, A. (2009). The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. Nat. Rev. Genet. *10*, 833–844.

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, 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.

Küppers, R., and Dalla-Favera, R. (2001). Mechanisms of chromosomal translocations in B cell lymphomas. Oncogene *20*, 5580–5594.

Lam, M.T., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. Trends Biochem. Sci. *39*, 170–182.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. Cell *157*, 13–25.

Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B.G., Ohgi, K., Zhang, J., Rose, D.W., Fu, X.D., Glass, C.K., and Rosenfeld, M.G. (2009). Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell *139*, 1069–1083.

Liu, M., and Schatz, D.G. (2009). Balancing AID and DNA repair during somatic hypermutation. Trends Immunol. *30*, 173–181.

Liu, Y.J., Mason, D.Y., Johnson, G.D., Abbot, S., Gregory, C.D., Hardie, D.L., Gordon, J., and MacLennan, I.C. (1991). Germinal center cells express bcl-2 protein after activation by signals which prevent their entry into apoptosis. Eur. J. Immunol. *21*, 1905–1910.

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.

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.

Marusawa, H., Takai, A., and Chiba, T. (2011). Role of activation-induced cytidine deaminase in inflammation-associated cancer development. Adv. Immunol. *111*, 109–141.

Matthews, A.J., Zheng, S., DiMenna, L.J., and Chaudhuri, J. (2014). Regulation of immunoglobulin class-switch recombination: choreography of noncoding transcription, targeted DNA deamination, and long-range DNA repair. Adv. Immunol. *122*, 1–57.

Migliazza, A., Martinotti, S., Chen, W., Fusco, C., Ye, B.H., Knowles, D.M., Offit, K., Chaganti, R.S., and Dalla-Favera, R. (1995). Frequent somatic hypermutation of the 5′ noncoding region of the BCL6 gene in B-cell lymphoma. Proc. Natl. Acad. Sci. USA *92*, 12520–12524.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell *102*, 553–563.

Natoli, G., and Andrau, J.C. (2012). Noncoding transcription at enhancers: general principles and functional models. Annu. Rev. Genet. *46*, 1–19.

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. (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., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R.S., Küppers, R., and Dalla-Favera, R. (2001). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. Nature *412*, 341–346.

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.

Qian, J., Wang, Q., Dose, M., Pruett, N., Kieffer-Kwon, K., Resch, W., Liang, G., Tang, Z., Mathé, E., Benner, C., et al. (2014). B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. Cell *159*. Published online December 4, 2014. http://dx.doi.org/10.1016/j.cell.2014.11.013.

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.

Richter, K., Brar, S., Ray, M., Pisitkun, P., Bolland, S., Verkoczy, L., and Diaz, M. (2009). Speckled-like pattern in the germinal center (SLIP-GC), a nuclear GTPase expressed in activation-induced deaminase-expressing lymphomas and germinal center B cells. J. Biol. Chem. *284*, 30652–30661.

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.

Sonoda, E., Pewzner-Jung, Y., Schwers, S., Taki, S., Jung, S., Eilat, D., and Rajewsky, K. (1997). B cell development under the condition of allelic inclusion. Immunity *6*, 225–233.

Storb, U. (2014). Why does somatic hypermutation by AID require transcription of its target genes? Adv. Immunol. *122*, 253–277.

Victora, G.D., and Nussenzweig, M.C. (2012). Germinal centers. Annu. Rev. Immunol. *30*, 429–457.

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.

Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K.A., et al. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature *474*, 390–394.

Ward, D.F., and Murray, N.E. (1979). Convergent transcription in bacteriophage lambda: interference with gene expression. J. Mol. Biol. *133*, 249–266.

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.

Wu, X., and Sharp, P.A. (2013). Divergent transcription: a driving force for new gene origination? Cell *155*, 990–996.

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.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

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.