Cell

Translocation-Capture Sequencing Reveals the Extent and Nature of Chromosomal Rearrangements in B Lymphocytes

Isaac A. Klein,[1] Wolfgang Resch,[5] Mila Jankovic,[1] Thiago Oliveira,[1,6] Arito Yamane,[3,5] Hirotaka Nakahashi,[3,5] Michela Di Virgilio,[1] Anne Bothmer,[1] Andre Nussenzweig,[4] Davide F. Robbiani,[1] Rafael Casellas,[3,5,7,*] and Michel C. Nussenzweig[1,2,7,*]

[1]Laboratory of Molecular Immunology
[2]Howard Hughes Medical Institute
The Rockefeller University, New York, NY 10065, USA
[3]Center for Cancer Research
[4]Experimental Immunology Branch
National Cancer Institute
[5]Genomics and Immunity, National Institute of Arthritis and Musculoskeletal and Skin, National Institutes of Health, Bethesda, MD 20892, USA
[6]Medical School of Ribeirao Preto/USP, Department of Genetics, 8 National Institute of Science and Technology for Stem Cells and Cell Therapy and Center for Cell-based Therapy, Ribeirao Preto, SP 14051-140, Brazil
[7]These authors contributed equally to this work
*Correspondence: casellar@mail.nih.gov (R.C.), nussen@mail.rockefeller.edu (M.C.N.)
DOI 10.1016/j.cell.2011.07.048

SUMMARY

Chromosomal rearrangements, including translocations, require formation and joining of DNA double strand breaks (DSBs). These events disrupt the integrity of the genome and are frequently involved in producing leukemias, lymphomas and sarcomas. Despite the importance of these events, current understanding of their genesis is limited. To examine the origins of chromosomal rearrangements we developed Translocation Capture Sequencing (TC-Seq), a method to document chromosomal rearrangements genome-wide, in primary cells. We examined over 180,000 rearrangements obtained from 400 million B lymphocytes, revealing that proximity between DSBs, transcriptional activity and chromosome territories are key determinants of genome rearrangement. Specifically, rearrangements tend to occur in *cis* and to transcribed genes. Finally, we find that activation-induced cytidine deaminase (AID) induces the rearrangement of many genes found as translocation partners in mature B cell lymphoma.

INTRODUCTION

Lymphomas, leukemias, and solid tumors frequently carry gross genomic rearrangements, including chromosomal translocations (Kuppers, 2005; Nussenzweig and Nussenzweig, 2010; Tsai and Lieber, 2010; Tsai et al., 2008; Zhang et al., 2010). Recurrent chromosomal translocations are key pathogenic events in hematopoietic tumors and sarcomas; they may juxtapose proto-oncogenes to constitutively active promoters, delete tumor suppressors, or produce chimeric oncogenes (Rabbitts, 2009). For example, the *c-myc/IgH* translocation, a hallmark of human Burkitt's lymphoma and mouse plasmacytomas, deregulates the expression of *c-myc* by bringing it under the control of Immunoglobulin (*Ig*) gene transcriptional regulatory elements (Gostissa et al., 2009; Kuppers, 2005; Potter, 2003). Alternatively, in chronic myeloid leukemia, the *Bcr/Abl* translocation fuses two disparate coding sequences to produce a novel, constitutively active tyrosine kinase (Goldman and Melo, 2003; Wong and Witte, 2004).

Chromosome translocation requires formation and joining of paired DNA double strand breaks (DSBs), a process that may be limited in part by the proximity of two breaks in the nucleus (Nussenzweig and Nussenzweig, 2010; Zhang et al., 2010). B lymphocytes are particularly prone to translocation-induced malignancy, and mature B cell lymphomas are the most common lymphoid cancer (Kuppers, 2005). This enhanced susceptibility appears to be the direct consequence of activation-induced cytidine deaminase (AID) expression in activated B cells (Nussenzweig and Nussenzweig, 2010). AID normally diversifies antibody genes by initiating *Ig* class switch recombination (CSR) and somatic hypermutation (SHM) (Muramatsu et al., 2000; Revy et al., 2000). It does so by deaminating cytosine residues in single-stranded DNA (ssDNA) exposed by stalled RNA polymerase II during transcription (Chaudhuri and Alt, 2004; Pavri et al., 2010; Storb et al., 2007). The resulting U:G mismatches are then processed by one of several repair pathways to yield mutations or DSBs, which are obligate intermediates in CSR, but may also serve as substrates for translocation (Di Noia and Neuberger, 2007; Honjo, 2002; Peled et al., 2008; Stavnezer et al., 2008). Although AID has a strong preference for targeting *Ig* genes, it also mutates a large number of non-*Ig* loci, including *Bcl6, Pax5, miR142*, *Pim1,* and *c-myc* (Gordon et al., 2003; Liu et al., 2008; Pasqualucci et al., 2001; Pavri et al., 2010; Robbiani

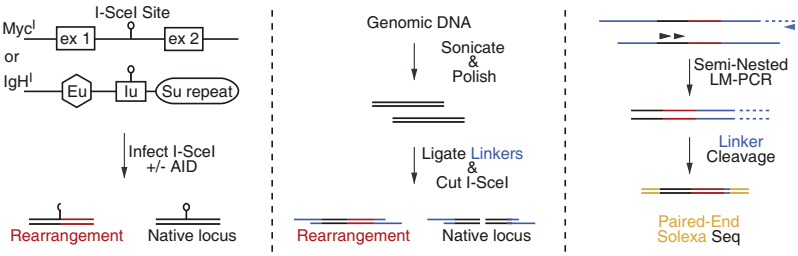

Figure 1. TC-Seq Schematic

IgH^{I} or Myc^{I} primary B cells are infected with retroviruses encoding I-SceI with or without AID. Genomic DNA is fragmented, blunted, A-tailed, ligated to T-tailed asymmetric linkers and native loci are eliminated by I-SceI digestion. Rearrangements are amplified by semi-nested ligation-mediated PCR followed by linker cleavage and paired-end deep sequencing.

et al., 2009; Shen et al., 1998; Yamane et al., 2011). While non-*Ig* gene mutation frequencies are low, it has been estimated that AID mutates as many as 25% of all genes expressed in germinal center B cells (Liu et al., 2008).

The full spectrum of potential AID targets was revealed by AID-chromatin immunoprecipitation studies, which showed AID occupancy at more than 5000 gene promoters bearing stalled RNA polymerase II (Yamane et al., 2011). AID is targeted to these genes through its interaction with Spt5, an RNA polymerase stalling factor (Pavri et al., 2010). Consistent with its genome-wide distribution, mice that overexpress AID exhibit chromosomal instability and develop translocation-associated lymphomas (Okazaki et al., 2003; Robbiani et al., 2009). Yet, *c-myc* is the only gene conclusively shown to translocate as a result of AID-induced DSBs (Ramiro et al., 2007; Robbiani et al., 2008). It has been estimated that up to 5% of activated primary B lymphocytes carry *IgH* fusions to unidentified partners which may or may not be selected during transformation (Franco et al., 2006; Jankovic et al., 2010; Ramiro et al., 2006; Robbiani et al., 2009; Wang et al., 2009; Yan et al., 2007). Additionally, recent deep-sequencing studies have revealed hundreds of genomic rearrangements within human cancers and documented their propensity to involve genes (Campbell et al., 2008; Pleasance et al., 2010a; Pleasance et al., 2010b; Stephens et al., 2009) However, the role of selection or other physiologic constraints in the genesis of these events is unclear because methods for mapping chromosomal translocations in primary cells do not yet exist.

Here, we describe a novel, genome-wide strategy to document primary chromosomal rearrangements. We provide insight into the effects of genomic position and transcription on the genesis of chromosomal rearrangements and DSB resolution. Our data also reveal the extent of recurrent AID-mediated translocations in activated B cells.

RESULTS

Translocation Capture Sequencing

To discover the extent and nature of chromosomal rearrangements in activated B lymphocytes we developed an assay to capture and sequence rearranged genomic DNA (TC-Seq). In this system, DSBs are induced at the *c-myc* (chromosome 15) or *IgH* (chromosome 12) loci, which were engineered to harbor the I-SceI meganuclease target sequence (Robbiani et al., 2008). c-myc$^{I\text{-}SceI/I\text{-}SceI}$ or IgH$^{I\text{-}SceI/I\text{-}SceI}$ (hereafter referred to as Myc^{I} and IgH^{I}) B cells were stimulated and infected with a retrovirus expressing I-SceI, in the presence or absence of AID. Rearrangements to I-SceI sites were recovered by semi-nested ligation-mediated PCR from genomic DNA that had been fragmented, A-tailed (to prevent intramolecular ligation) and ligated to asymmetric DNA linkers (Figure 1). Site-specific primers were placed at least 150 bp from the I-SceI site allowing for the capture of rearrangements involving moderate end-processing. PCR products were submitted for high-throughput paired-end sequencing and reads were aligned to the mouse genome. Identical reads were clustered as single events. Since sonication generates unique linker ligation points in each cell, this method allows for the study of independent events without sequencing through rearrangement breakpoints.

AID-Independent Translocations

In the absence of AID, DSBs arise as by-products of normal cellular metabolism including transcription and DNA replication (Branzei and Foiani, 2010). Consistent with a global distribution of DSBs, we mapped 28,548 unique rearrangements between the I-SceI site and every chromosome in Myc^{I}AID$^{-/-}$ B cells (100 million cells assayed, Figure 2A). To determine whether there is a genome-wide bias for rearrangement, these events were characterized based on location, transcription and histone modification of the locus.

We found a marked enrichment of intrachromosomal rearrangements on chromosome 15, with approximately 125 events per mappable megabase (11,066 rearrangements), or ~40% of all events (Figure 2B). Translocations between Myc^{I} and other chromosomes were evenly distributed throughout the genome (Figure 2B and Table S1). Notably, 86.7% (9591 of 11,066) of all intrachromosomal rearrangements were localized within a 350 kb domain surrounding the I-SceI site (from −50 kb to +300 kb; Figure 2C). This is consistent with the observation that 92% of intrachromosomal rearrangements in the breast cancer genome involve aberrant joining of DSBs within 2 Mb of each other (Stephens et al., 2009), and that 87% of RAG-mediated intrachromosomal rearrangements in Abl-transformed pre-B cells lie within 200 kb of a recombination substrate (Mahowald et al., 2009). The asymmetrical distribution of events in the direction of *c-myc* transcription and the adjacent *Pvt1* gene is also consistent with the idea that gene expression facilitates rearrangement (Thomas and Rothstein, 1989). I-SceI-proximal events may be the result of either resection and rejoining of I-SceI breaks, bona fide rearrangements between I-SceI and random DSBs, or a combination of DNA end resection and balanced translocations. Regardless of the precise molecular mechanism, the abundance of these events reveals a strong preference for DSBs to be resolved by ligation to a proximal sequence, a DNA repair strategy that may minimize gross genomic alterations.

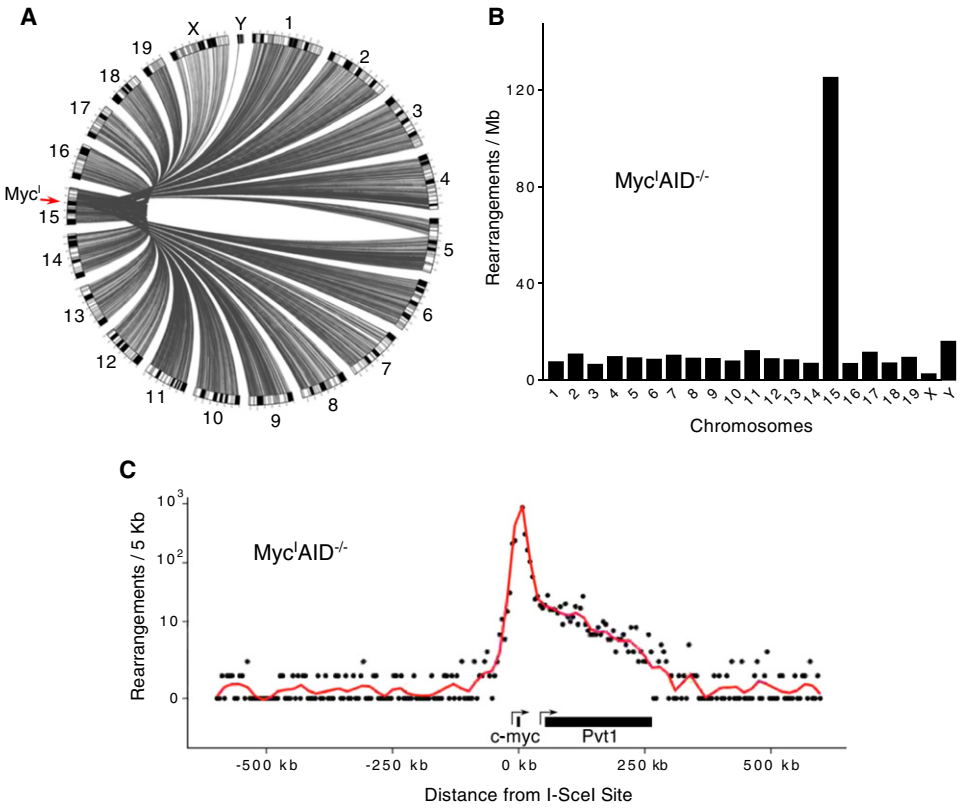

Figure 2. Rearrangements to a DSB Site Documented by TC-Seq

(A) Genome-wide view of rearrangements to Myc^{I} in AID$^{-/-}$ B cells.

(B) Rearrangements per mappable megabase to each chromosome in Myc^{I}AID$^{-/-}$ B cells.

(C) Profile of rearrangements around the I-SceI site in 5 kb intervals.

Recent cancer genome sequencing experiments uncovered a modest but highly significant preference for cancer-associated rearrangements to occur in genes, which compose only 41% of the human genome. For example, in 24 sequenced breast cancer genomes, 50% of all rearrangements involved genes (Stephens et al., 2009). Whether this bias resulted from selection or some inherent feature of DSB formation and repair specific to cancer cells could not be determined. To ascertain whether a similar bias is seen in primary cells in short term cultures, AID-independent rearrangements in Myc^{I}AID$^{-/-}$ B lymphocytes (excluding 1 Mb of DNA around the I-SceI site) were classified as genic or intergenic. Consistent with the human tumor studies, 51% (9677 of 19,246) of the events were associated with genes (Figure 3A). Because only 40% of the mouse genome is genic, this represents a small (1.25-fold) but significant difference (permutation test $p < 0.001$) relative to intergenic regions. Moreover, the genic rearrangements were particularly enriched at transcription start sites (Figure 3B).

Consistent with the preference for genic rearrangements, we also observed a bias to transcribed genes. Fewer rearrangements than expected occurred at silent (fe = 0.74, $p < 0.001$) and trace (fe = 0.95, $p < 0.001$) transcribed genes, while more than expected occurred at low (fe = 1.08 $p < 0.001$), medium (fe = 1.13, $p < 0.001$), and highly (fe = 1.14, $p < 0.001$) transcribed genes (Figure 3C and Figure S1). Additionally, rearrangements were enriched in genes bearing PolII and activating histone marks such as H3K4 trimethylation, H3 acetylation, and H3K36 trimethylation ($p < 0.001$, Figure 3D). Thus, there is a propensity for a DSB to recombine with gene rich regions of the genome and more specifically to transcription start sites of actively transcribed genes.

AID-Mediated Lesions Captured by TC-Seq

Processing of AID-induced U:G mismatches can result in DSBs in *Ig* and non-*Ig* genes such as *c-myc* (Robbiani et al., 2008). To determine whether AID-mediated DSBs can be captured by TC-Seq we examined the *IgH* and *c-myc* loci in B cells expressing retrovirally encoded AID (IgH^{I}AID^{RV} or Myc^{I}AID^{RV}). IgH^{I} B cells expressing both I-SceI and AID showed extensive AID-dependent rearrangement between the I-SceI site and downstream switch (S) regions (Figure 4A). The frequency of rearrangements resembled the pattern of AID-mediated CSR in LPS+IL-4 cultures (e.g., IgG1 ≫ IgG3>IgE), with 18,686 mapping to *Sγ1*, 3,192 to *Sγ3*, and 1433 to *Sε* (Table S2). Furthermore, translocations between *c-myc* and *IgH* were entirely dependent on AID (Figures 4B and 4C). In two biological replicate samples totaling 100 million B cells, we observed 45 translocations from IgH^{I} to *c-myc* (the I-SceI DSB was in *IgH*), and 5963 from Myc^{I}

A

		Total Captured Rearrangements		
		Genic	Intergenic	% genic
Myc^{I}	AID$^{-/-}$	9677	9569	50.3
	AID^{RV}	26221	25101	51.1
IgH^{I}	AID^{WT}	5651	5723	49.7
	AID^{RV}	9475	8026	53.9

P (permutation test) < 0.001 for all

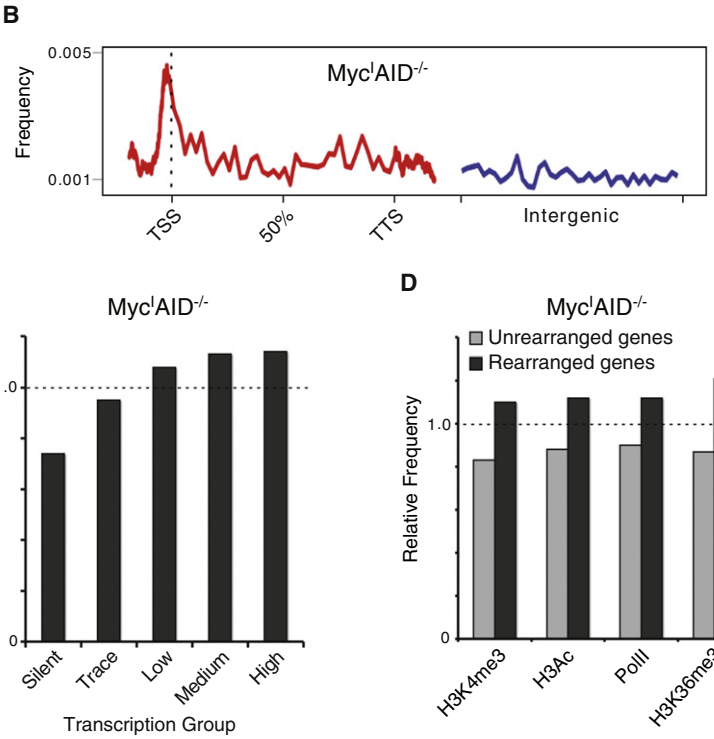

Figure 3. Rearrangements to Myc^{I} Occur Near the TSSs of Actively Transcribed Genes

(A) Rearrangements, excluding the 1 Mb around I-SceI, were categorized as genic or intergenic.

(B) Composite density profile of genomic rearrangements from Myc^{I} to genes (red line) and intergenic regions (blue line). TSS = transcription start site, TTS = transcription termination site.

(C) Relative frequency (fe) of rearrangements in genes that are either silent or display trace, low, medium or high levels of transcription in activated B cells as determined by RNA-Seq (Figure S1). Dashed line indicates the expected rearrangement frequency based on a random model. $p < 0.001$ for all (permutation test).

(D) Relative frequency of rearrangements in PolII- and activating histone mark-associated gene groups (Yamane et al., 2011). Dashed line indicates expected frequency based on a random model. $p < 0.001$ for all samples (permutation test).

Also see Figure S1.

to *IgH* (the I-SceI DSB was in *c-myc*) (Table S2). Additionally, TC-Seq tags mapping to *c-myc* from IgH^{I} correlate well with *c-myc*/*IgH* translocation breakpoints sequenced from primary B cells (Figure 4C) (Robbiani et al., 2008). This suggests that TC-Seq reads are an accurate proxy for breakpoints. Furthermore, the data corroborate previous findings showing that AID induced breaks at *c-myc* are rate limiting for *c-myc*/*IgH* translocations (Robbiani et al., 2008) and suggest that AID-dependent *IgH* breaks are two orders of magnitude more frequent than those at *c-myc*. We conclude that TC-Seq captures rearrangements and translocations between DSBs in IgH^{I} or Myc^{I} and known AID targets.

As was the case for AID deficient samples, Myc^{I}AID^{RV} and IgH^{I}AID^{RV} libraries were enriched in intrachromosomal rearrangements: 17% (10,633 of 63,772 total events) for Myc^{I} and 70% (36,019 of 51,312) for IgH^{I} (Table S1 and Figures S2A and S2B). The difference in enrichment between the two was mostly the result of AID activity on chromosome 12, which generated a large number of rearrangements to *IgH* variable and constant domains (Figure 4A). Expression of AID did not alter the distribution of events around Myc^{I}, with 72.5% (7707 of 10,633) mapping within −50 kb to 300 kb of the break (Figure S2C). A notable exception was an additional cluster of rearrangements associated with *Pvt1* exon 5 (Figure S2C). These events coincided precisely with documented chromosomal translocations isolated from AID sufficient mouse plasmacytomas (Cory et al., 1985; Huppi et al., 1990) and likely represent an AID hot spot.

In agreement with the Myc^{I}AID$^{-/-}$ samples, translocations between IgH^{I} or Myc^{I} and other chromosomes were evenly distributed throughout the genome, except for the Myc^{I} capture sample, which displayed a marked bias for chromosome 12 due to creation of DSBs at the *IgH* locus by AID (Figure S2A). Similar to Myc^{I}AID$^{-/-}$ samples, rearrangements in both cases were more likely to occur in regions that are genic, transcriptionally active, recruiting PolII, and associated with activating histone marks (Figure 3A and Figures 4D and 4E). Furthermore, intragenic rearrangements were enriched at transcription start sites of genes (Figure S2D). In contrast to recent studies that used Nbs1 as an indirect marker of AID mediated damage (Staszewski et al., 2011), we found little or no difference in rearrangements to genomic repeats in the presence of AID (Table S3). Thus, AID does not dramatically alter the general profile of rearrangements.

Next, we examined whether IgH^{I} and Myc^{I} capture DSB targets at similar rates. Indeed, in AID sufficient samples, total translocations from a given chromosome to IgH^{I} or Myc^{I} occurred at roughly similar frequencies (Figure 4F). This similarity could be explained by the close physical proximity of *IgH* and *c-myc*, as suggested by studies with EBV-transformed B lymphoblastoid cells (Roix et al., 2003). Alternatively, the correlation in transchromosomal joining might represent random ligation between I-SceI DSBs in *IgH* or *c-myc* and DSBs on other chromosomes. We conclude that extra-chromosomal DSBs ligate DSBs in IgH^{I} and Myc^{I} at similar rates.

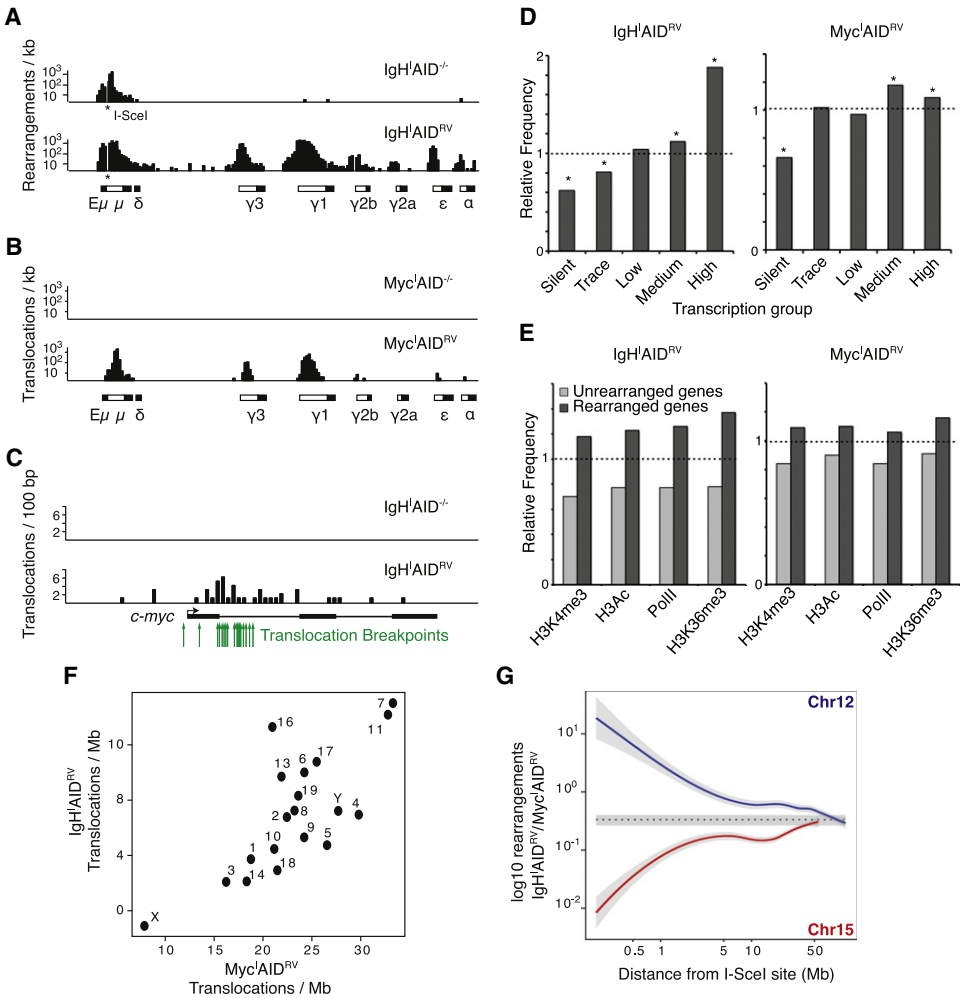

Figure 4. Rearrangements to IgH^I or Myc^I in Primary B Cells Expressing AID

(A and B) (A) Rearrangements per kb to I-SceI sites (indicated with an asterisk) in *IgH* or (B) *c-myc* in AID$^{-/-}$ (top panel) or AID^{RV} cells (bottom panel). White boxes in the schematics below each graph represent *Ig* switch domains while black boxes depict constant regions.

(C) Translocations per 100 bp from IgH^I to *c-myc* in AID$^{-/-}$ (top panel) or AID^{RV} cells (bottom panel). Green arrows indicate *c-myc*/*IgH* translocation breakpoints sequenced from primary B cells (Robbiani et al., 2008).

(D) Relative frequency of rearrangements in transcription-level gene groups (Figure S1), dashed line indicates expected frequency based on a random model. Asterisks highlight values with a $p < 0.001$ (permutation test).

(E) Relative frequency of rearrangements in PolII-associated or activating histone mark-associated gene groups (Yamane et al., 2011). Dashed line indicates the expected frequency based on a random model. $p < 0.001$ for all samples (permutation test).

(F) Graph comparing the number of translocations per mappable megabase to each chromosome from IgH^I (*y* axis) or Myc^I (x axis).

(G) Ratio of IgH^I captured to Myc^I captured events in 500 kb bins moving away from the I-SceI capture site (both directions combined). Dotted line represents the average transchromosomal joining rate computed on all chromosomes other than 12 or 15. Gray areas show 2 standard deviations around the mean.

Also see Figure S2.

To examine the nature of the intrachromosomal rearrangement bias we calculated the ratio of IgH^I to Myc^I captured events for each 500 kb segment of the genome and compared the values for chromosome 12 and 15 to the transchromosomal average (Figure 4G and Figure S2E). This analysis revealed that DSBs are preferentially captured intrachromosomally and this effect diminishes at a rate inversely proportional to the distance from the I-SceI site ($d^{-1.29}$) (Figure S2F). This effect was most prominent locally but was evident at up to ~50 Mb away from the I-SceI break. We conclude that paired DSBs are preferentially joined intrachromosomally and that the magnitude of this effect decreases with increasing distance between the two lesions.

Translocation Hot Spots

To determine whether there are hot spots for rearrangement, we searched the B cell genome for local accumulations of reads in AID deficient and sufficient samples. TC-Seq hot spots were defined as a localized enrichment of rearrangements above what is expected from a uniform genomic distribution. We

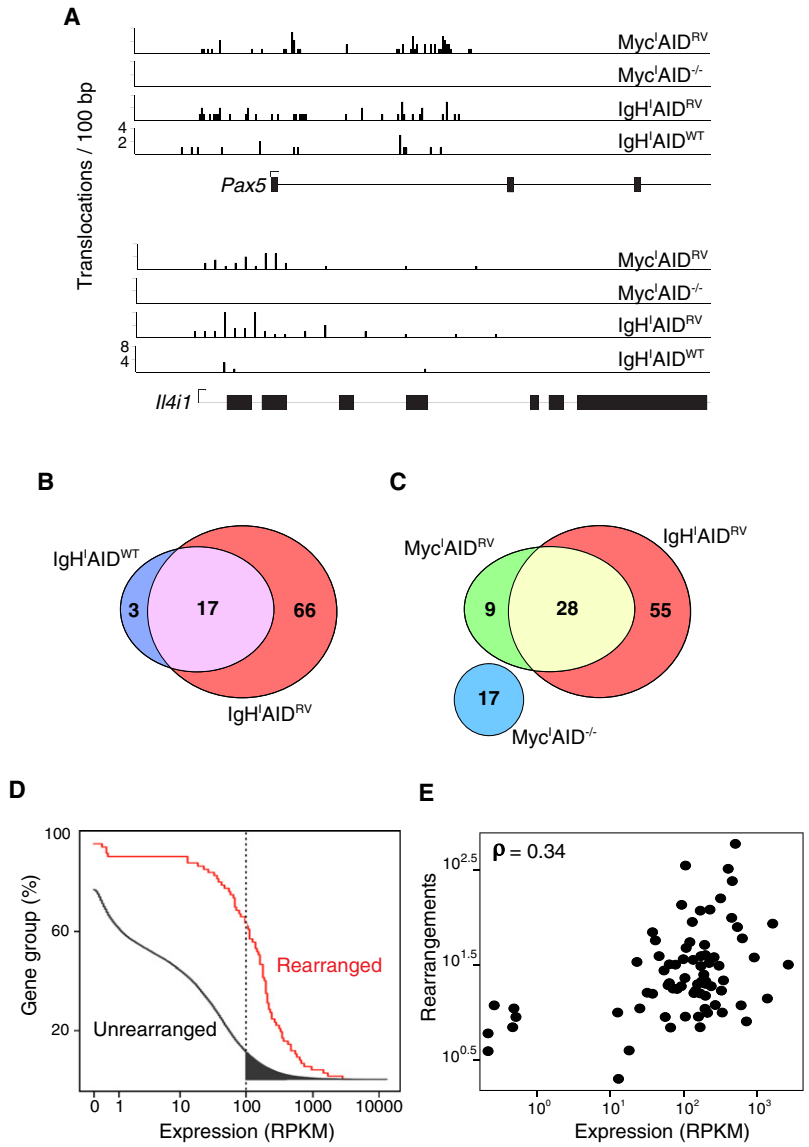

Figure 5. AID-Dependent Rearrangement Hot Spots

(A) Screenshots of translocations per 100 bp present at *Il4i1* and *Pax5* genes in all samples.

(B) Overlap of AID-dependent hot-spot-bearing genes in IgH^{I}AID^{RV} and IgH^{I}AID^{WT} experiments.

(C) Overlap of AID-dependent hot-spot-bearing genes in IgH^{I}AID^{RV}, Myc^{I}AID$^{-/-}$ and Myc^{I}AID^{RV} experiments.

(D) Empirical cumulative distribution showing transcript abundance in genes displaying (red) or lacking (black) rearrangement hot spots. Filled-in gray slice represents ~2000 highly transcribed unrearranged genes.

(E) Total rearrangements as a function of gene expression (RNA-Seq) (Yamane et al., 2011) in genes bearing rearrangement hot spots.

Also see Figure S3.

removed likely artifacts; namely hot spots containing >80% of reads within DNA repeats, and those with footprints of <100 nt (because translocations are amplified from randomly sonicated DNA (Figure 1), deep-sequence tags associated with bona fide rearrangements are unlikely to map within a small region).

We identified 34 hot spots captured by Myc^{I} in the absence of AID (Table S4). There were 31 hot spots in 17 genes and three in nongenic regions (Table S4). Seventeen of the hot spots were in *Pvt1*, within 500 kb of the I-SceI site. In addition, two hot spots occurred within 5 kb of cryptic I-SceI sites (each bearing one mismatch to the 18-base pair recognition sequence). For example, one such hot spot at chr15:16219195-16219312 containing a 1-off I-SceI recognition sequence bore 8 rearrangements (Figure S3). A genome-wide search for rearrangements within 5 kb of cryptic I-SceI sites (83 within the mouse genome with 1 or 2 mismatches to the canonical I-SceI recognition sequence) yielded 57 events, seven times more than expected in a random distribution model. When allowing up to 6 mismatches we find a total of five out of 34 AID-independent hot spots near putative cryptic I-SceI sites. Although I-SceI has been used to generate a unique DSB in gene targeting and DNA repair experiments, our data suggest that DNA recognition by I-SceI can be promiscuous in the mouse genome, as demonstrated for other yeast endonucleases (Argast et al., 1998).

In contrast to AID$^{-/-}$, we found 157 AID-dependent hot spots in 83 genes captured by IgH^{I}AID^{RV} and 60 hot spots in 37 genes by Myc^{I}AID^{RV} in 100 million B cells. (Table S4). 80% of the hot spots captured by *c-myc* and 90% of those captured by *IgH* were within genes. For example, we found robust AID-dependent hot spots on *Il4i1* and *Pax5* (a recurring *IgH* translocation partner in lymphoplasmacytoid lymphoma [Kuppers, 2005]) (Figure 5A and Table S4). AID-dependent hot spots were similar for IgH^{I} B cells expressing wild-type levels (WT) or retrovirally overexpressed (RV) AID, however the number of events was decreased in the former (Figures 5A and 5B and Table S4). Therefore, translocations to AID targets occur in cells expressing physiological levels of AID and hotspots are not dependent on AID overexpression. We conclude that AID produces substrates for translocations in a number of discreet sites throughout the genome, and these sites are mainly in genes.

Genes containing AID-dependent hotspots overlapped between IgH^{I}AID^{RV} and Myc^{I}AID^{RV} samples (Figure 5C). Consistent with the similar capture rates observed for transchromosomal targets (Figure 4F), we found that 28 of the frequently translocated targets were shared (Table S5). In contrast, we found a number of unique intrachromosomal AID-dependent hot spots. For example, rearrangements to *Inf2* on chromosome 12 (~850 kb from IgH^{I}) were only found by IgH^{I} capture while rearrangements near *Pvt1* on chromosome 15 (up to ~350 kb from Myc^{I}) were only found by Myc^{I} capture (Table S5). Thus, there is

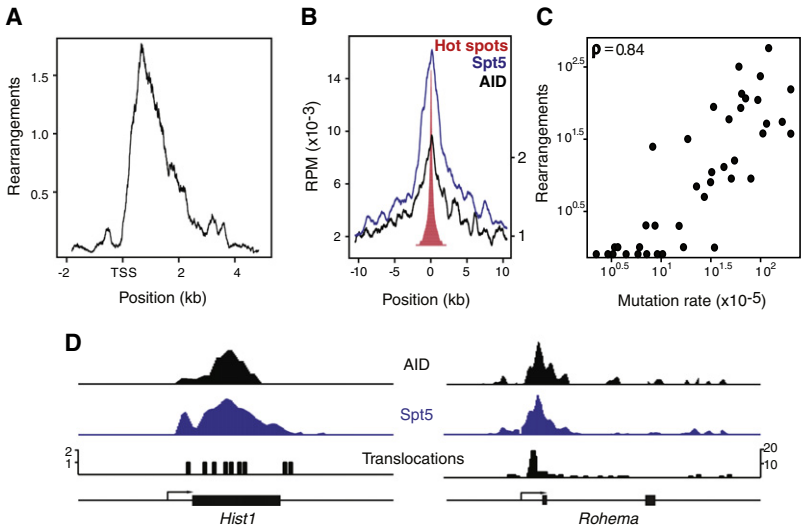

Figure 6. Characterization of AID-Dependent Hot Spots

(A) Composite density graph showing the distribution of rearrangements in genes associated with AID-dependent hotspots relative to the TSS.

(B) Spt5 and AID recruitment at genomic sites associated with translocation hotspots (Pavri et al., 2010; Yamane et al., 2011).

(C) Somatic hypermutation frequency versus number of rearrangements in genes bearing AID-dependent translocation hotspots.

(D) Distribution of AID, Spt5. and translocations in *Rohema* and *Hist1* genes.

Also see Figure S4.

a bias toward recombination between I-SceI breaks and AID hot spots within the same chromosome. Additionally, the finding that some hot spots are only captured in *cis* indicates that TC-Seq underestimates the number of AID mediated DSBs in the genome and suggests that we have not reached saturation.

Combined analysis of the IgH^{I} and Myc^{I} TC-Seq data sets shows that AID-dependent hot spots are primarily found in transcribed genes (Figure 5D). However, although nearly all of the translocated genes are actively transcribed, there is no clear correlation between transcript abundance and rearrangement frequency (Figure 5E). Furthermore, ~2000 highly transcribed genes are not rearranged (Figure 5D, shaded area). Therefore, transcription is necessary but not sufficient for AID targeting, and transcription levels alone cannot account for AID-dependent DSBs.

AID-dependent hot spots are biased to the region around the transcription start site (Figure 6A). This finding is consistent with the accumulation of AID and Spt5 around the promoters of stalled genes and the distribution of somatic hypermutation (Pavri et al., 2010; Yamane et al., 2011). Indeed, AID-dependent TC-Seq hot spots overlap with regions of AID (Figure S4A) and Spt5 accumulation (Figures 6B and 6D). This correlation prompted us to explore the relationship between AID activity and accumulation of chromosomal translocations by measuring somatic hypermutation at TC-Seq captured AID targets (Yamane et al., 2011 and Table S6). We found a positive correlation (Spearman coefficient = 0.84) between hypermutation and rearrangement frequency (Figure 6C). All genes analyzed with a mutation rate over $10x10^{-5}$ bear rearrangements, and all genes with AID-dependent TC-Seq hot spots show mutations (Figure 6C). Rearrangements were only seen rarely in genes with lower rates of mutation (Figure 6C). This suggests that the rate of hypermutation and the frequency of AID-induced DSBs are directly proportional. We conclude that AID-dependent TC-Seq hot spots occur on stalled genes that accumulate Spt5, AID, and high rates of hypermutation.

Among AID-dependent hot spot containing genes we find several that are translocated or deleted in mature B cell lymphoma. These include *Pax5/IgH*, *Pim1/Bcl6*, *Il21r/Bcl6*, *Gas5/Bcl6*, and *Ddx6/IgH* translocations and *Junb* and *Socs1* deletions in diffuse large B cell lymphoma, *Birc3/Malt1* translocation in MALT lymphoma, *Ccnd2/IgK* translocation and *Bcl2l11* deletion in mantle cell lymphoma, *Aff3/Bcl2* and *Grhpr/Bcl6* translocations in follicular lymphoma, *mir142/c-myc* translocation in B cell prolymphocytic leukemia as well as *c-myc/IgH* and *Pvt1/IgK* translocations in Burkitt's lymphoma (Table 1). Interestingly, we find that AID is capable of inducing DSBs in *Fli1* (Table S4), which is translocated to *EWS* in 90% of Ewing's sarcomas, a malignant tumor of uncertain origin (Riggi and Stamenkovic, 2007). We conclude that in addition to mutating many genes, AID also initiates DSBs in numerous non-*Ig* genes. These genes serve as substrates for translocations associated with mature B cell lymphoma, strongly implicating AID as a source of genomic instability in these cancers.

DISCUSSION

To date, the study of chromosomal aberrations has been primarily limited to events identified in tumors and tumor cell lines. Although we have learned a great deal about the importance of genomic rearrangements in cancer, it has not been possible to develop an understanding of the cellular and molecular requirements that govern their genesis. To examine genomic rearrangements in primary cells in short term cultures, we developed a technique to catalog these events by deep sequencing, TC-seq. Our results and analysis reveal the importance of transcription and physical proximity in recombinogenesis, and identifies hotspots for AID-mediated translocations in mature B cells.

Nuclear Proximity and Chromosomal Position

The existence of chromosome territories, regions in which individual chromosomes segregate, has been long proposed (Cremer and Cremer, 2001) and recently shown to be a key feature of genome organization (Lieberman-Aiden et al., 2009). Our analysis provides evidence that physical proximity and chromosome territories are partial determinants for joining of specific rearrangement partners. The effects of physical proximity are most evident in the 350 kb region around the DSB. In the absence of AID the plurality of rearrangements fall in this region.

Table 1. AID-Dependent Translocations in Human B Cell Lymphoma

TC-Seq Gene	Translocation in Mature BCL	Mature BCL Type	Reference
Birc3 (Api2)	t(11;18)(q21;q21)	MALT	(Rosebeck et al., 2011)
Il21r	t(3;16)(q27;p11)	DLBCL	(Ueda et al., 2002)
Pax5	t(9;14)(p13;q32)	DBLCL	(Iida et al., 1999)
Pim1	t(3;6)(q27;p21.2)	DLBCL	(Yoshida et al., 1999)
Aff3	t(2;18)(q11.2;q21)	FL	(Impera et al., 2008)
Gas5	t(1;3)(q25;q27)	DLBCL	(Nakamura et al., 2008)
Ccnd2	t(2;12)(p12;p13)	MCL	(Gesk et al., 2006)
c-myc	t(8;14)(q23;q32)	BL	(Kuppers, 2005)
Ddx6 (Rck)	t(11;14)(q23;q32)	DLBCL	(Lu and Yunis, 1992)
Grhpr	t(3;9)(q27;p11)	FL	(Akasaka et al., 2003)
Bcl2l11	Deleted	MCL	(Bea et al., 2009)
Socs1	Deleted	DLBCL	(Mottok et al., 2009)
Junb	Deleted	DLBCL	(Mao et al., 2002)
mir142	t(8;17)	B-PLL	(Gauwerky et al., 1989)
Pvt1	t(2;8)(p11.2;q24.1)	BL	(Einerson et al., 2006)
IgH	several	several	(Kuppers, 2005)
IgK	several	several	(Kuppers, 2005)
IgL	several	several	(Kuppers, 2005)

Genes bearing AID-dependent TC-Seq hotspots are shown with the associated translocation or deletion observed in human mature B cell lymphoma (BCL). BCL types are abbreviated as follows: MALT, mucosa-associated lymphoid tissue lymphoma; DLBCL, diffuse large B cell lymphoma; FL, follicular lymphoma; MCL, mantle cell lymphoma; BL, Burkitt's lymphoma; B-PLL, B cell prolymphocytic leukemia.

This observation is consistent with the analysis of rearrangements in the breast cancer genome and suggests that the abundance of these events is independent of cancer specific selection (Stephens et al., 2009). Additionally, a preference for DSB repair within 350 kb matches the range of gamma-H2AX spreading from a DSB (Bothmer et al., 2011). This is consistent with the idea that the DNA damage response facilitates proximal rearrangement, a phenomenon most prominent at the *IgH* locus during CSR.

The magnitude of the effect of chromosome territories on rearrangement is far less prominent than proximal joining, but is consistent with recent genome mapping data obtained by high-throughput chromosome conformation capture (Hi-C) (Lieberman-Aiden et al., 2009). Intrachromosomal joining bias is evident in the preferential joining of AID hotspots and nonhotspots on Chr12 and Chr15 with their respective I-SceI breaks. When compared to transchromosomal joining, the bias to intrachromosomal rearrangements is evident even when DSBs are separated by as much as 50 Mb. In mouse, the mean autosome size is ~130 Mb, so a 50 Mb preference for intrachromosomal joining on either side of a DSB will encompass nearly the entire average chromosome. We conclude that intrachromosomal joining is preferred to transchromosomal joining.

Since this effect diminishes with distance, it is mediated by proximity, a likely consequence of local chromosome packing and nuclear chromosomal territories. A strong preference for proximal intrachromosomal rearrangement minimizes gross genomic alterations. We propose that this may be an important feature of DSB repair regulation that maintains genomic integrity.

Transcription

Transcription is associated with increased rates of DNA damage and genome instability; these effects are likely mediated by a number of different mechanisms (Gottipati and Helleday, 2009). Transcription may expose ssDNA, which is susceptible to chemical or oxidative damage (Aguilera, 2002). Additionally, head-on collision of the replication and transcription machinery has been implicated in fork stalling and genomic instability (Takeuchi et al., 2003). Consistent with these ideas, TC-Seq reveals that transcription facilitates DNA rearrangement. In the case of the *c-myc* locus, transcription increases the size of the local area around a DSB that is available for recombination from 50 kb to 300 kb. Moreover, I-SceI breaks rearrange predominantly to transcribed genes genome-wide and more specifically to the TSS. Thus, exposed ssDNA may serve as a primary source of genomic instability. AID expression further reinforces this phenomenon by creating U:G mismatches in ssDNA at sites of PolII stalling downstream of the TSS (Pavri et al., 2010).

A bias for rearrangement between genic regions was also reported in recent studies of the cancer genome, but the role of transcription, transformation or selection in these events could not be evaluated (Stephens et al., 2009). Our experiments demonstrate that transcribed genic regions are over-represented in chromosomal rearrangements in primary cells in short-term cultures. In addition to being more susceptible to damage, this effect may be due to the increased physical proximity of transcribed regions to each other in the nucleus (Lieberman-Aiden et al., 2009). We speculate that this phenomenon may have consequences for tumorigenesis. The rearrangement of proto-oncogenes to transcribed regions may lead to their deregulation or produce hybrid entities that alter cellular metabolism.

AID and Chromosome Translocation

AID initiates SHM, CSR, and chromosome translocation by deaminating cytosine residues in ssDNA exposed by transcription (Chaudhuri and Alt, 2004; Di Noia and Neuberger, 2007; Nussenzweig and Nussenzweig, 2010; Peled et al., 2008; Stavnezer et al., 2008). AID targets the *IgH* locus and the TSSs of stalled genes through direct interaction with Spt5, a PolII stalling factor (Pavri et al., 2010), resulting in widespread somatic mutations (Yamane et al., 2011). Additionally, AID has been shown to initiate DSBs in non-*Ig* targets such as *c-myc,* and generates diverse translocations and chromosome breaks (Robbiani et al., 2008; Robbiani et al., 2009). However the precise relationships between AID and Spt5 occupancy, mutation, and translocations have not previously been investigated.

By capturing and sequencing chromosomal rearrangements, a readout for aberrantly resolved DSBs, we have gained insight into the mechanisms by which AID targets DNA for chromosomal rearrangement. First, we show that AID targets discreet sites in the genome for DSB. These sites are predominantly genic and actively transcribed. A recent study using Nbs1-ChIP as a surrogate for DNA damage suggested that AID targets repeat rich sequences (Staszewski et al., 2011). In contrast, we find no

AID-dependent increase in rearrangements to repeats. Moreover, AID-dependent rearrangement hotspots predominantly occur in genes, not in or near repeat regions that are not transcribed. Hotspots that do fall in repeats (Figure S4B), are not AID-dependent and do not suffer somatic hypermutation (Table S6). While it is difficult to map short reads to repetitive sequences, these data suggest that rearrangements to repeats may be from AID-independent DSB.

While genes rearranged by AID are largely transcribed, expression and PolII accumulation do not correlate directly with rearrangement frequency suggesting that transcription is necessary but not rate-limiting for rearrangement. Reflecting the distribution of AID and its co-factor Spt5 in the genome (Pavri et al., 2010; Yamane et al., 2011), AID-dependent rearrangements occur mainly on transcription start sites of stalled genes that carry high levels of the PolII stalling factor Spt5. In addition, we find a strong and direct correlation between hypermutation and rearrangements, suggesting that genes susceptible to AID mediated recombinogenesis are a subset of the most highly mutated genes in the genome. Consistent with this notion, we show that *Pax5, Il21r, Gas5, Ddx6*, *Birc3*, *Ccnd2*, *Aff3, Grhpr, c-myc, Pvt1*, *Bcl2l11, Socs1, mir142, Junb*, and *Pim1*, which are translocated or deleted in mature B cell lymphomas (Table 1) are among the more highly mutated AID targets and bear AID-dependent translocation hotspots. Our experiments were performed on in vitro stimulated B cells. Germinal center B cells will have an alternate gene expression profile that might influence the number and position of AID target sites. We conclude that in addition to hypermutation, AID is also a source of genomic instability in mature B cell lymphomas.

Finally, we note that TC-seq can be adapted for use in other cell types to study translocation biology in any tissue.

EXPERIMENTAL PROCEDURES

B Cell Cultures, Infections, and Sorting

Resting B lymphocytes were isolated from mouse spleens by immunomagnetic depletion with anti-CD43 MicroBeads (Miltenyi Biotech) and cultured at 0.5 × 10^6 cells/ml in RPMI supplemented with L-glutamine, sodium pyruvate, antibiotic/antimycotic, HEPES, 50 μM 2-mercaptoethanol (all from GIBCO-BRL), and 10% fetal calf serum (Hyclone). B cells were stimulated in the presence of 500 ng/ml RP105 (BD PharMingen), 25 μg/ml lipopolysaccharide (LPS) (Sigma) and 5 ng/ml mouse recombinant IL-4 (Sigma). Retroviral supernatants were prepared by cotransfection of BOSC23 cells with pCL-Eco and pMX-IRES-GFP-derived plasmids encoding for I-SceI-mCherry or AID-GFP with Fugene 6, 72 hr before infection. At 20 and 44 hr of lymphocyte culture, retroviral supernatants were added, and B cells were spinoculated at 1150 g for 1.5 hr in the presence of 10 μg/ml polybrene. For dual infection, separately prepared retroviral supernatants were added simultaneously on both days. After 4 hr at 37°C, supernatants were replaced with LPS and IL-4 in supplemented RPMI. At 96 hr from the beginning of their culture, singly infected B cells were collected and frozen in 10 million cell pellets at −80C. Dually infected B cells were sorted for double positive cells with a FACSAria instrument (Becton Dickson) then frozen down.

Translocation Capture Sequencing

Genomic DNA Library Preparation

5 × 10 million B cell aliquots were lysed in Proteinase K buffer (100 mM Tris [pH 8], 0.2% SDS, 200 mM NaCl, 5 mM EDTA) and 50 μl of 20mg/ml Proteinase K. Genomic DNA was extracted by phenol chloroform precipitation and fragmented by sonication (Bioruptor - Diagenode) to yield a 500–1350 bp distribution of DNA fragments. DNA was divided into (5ug) aliquots in 1.5mL eppendorf tubes. Each experiment consisted of genomic DNA from 50 million B cells in 50 × 5 μg aliquots for a total of 250 μg of fragmented genomic DNA per experiment. Subsequent reactions were performed individually on 5 μg aliquots. DNA was blunted by End-It DNA Repair Kit (Epicenter), purified, then adenosine-tailed by Klenow fragment 3 → 5′ exo^- (NEB) and purified. Fragments were ligated to 200 pmol of annealed linkers (pLT + pLB) (Table S8A) and unrearranged loci were eliminated by I-SceI digestion. Reactions were purified and pooled.

Rearrangement Amplification

Pooled linker-ligated DNA was divided into two equal parts for semi-nested ligation-mediated PCR using either forward or reverse primers (to capture rearrangements to either side of the I-SceI break). All PCRs were performed using the Phusion Polymerase system (NEB). DNA was divided into 1ug aliquots and subjected to single-primer PCR with biotinylated pMycF1, pMycR1, pIghF1 or pIghR1 [1x(98C-1min) 12x(98C-15s, 65C-30s, 72C-45s) 1x(72C-1min)] (Table S8A). Each reaction was spiked with pLinker and subjected to additional cycles of PCR [1x(98C-1min) 35x(98C-15s, 65C-30s, 72C-45s) 1x(72C-5min)]. Forward and reverse PCR reactions were pooled separately. Higher molecular weight products were isolated by agarose gel electrophoresis and magnetic streptavidin bead purification. Semi-nested PCR was performed on the magnetic beads with pMycF2, pMycR2, pIghF2, or pIghR2 and pLinker (Table S8A) [1x(98C-1min) 35x(98C-10s, 65C-30s, 72C-40s) 1x(72C-5min)]. Higher molecular weight products were isolated by agarose gel electrophoresis.

Paired-End Library Preparation

Linkers were removed by AscI digestion. Fragments were blunted by End-It DNA Repair Kit (Epicenter), purified, adenosine-tailed and ligated to Illumina paired-end adapters. Higher molecular weight products were isolated by agarose gel electrophoresis and adaptor-ligated fragments were enriched by 25 cycles of PCR with Illumina primers PE1.0 and PE2.0. Forward and reverse libraries for the same sample were mixed in equimolar ratios and sequenced by 36 × 36 or 54 × 54 paired end deep sequencing on an Illumina GAII.

TC-Seq Computational Analysis

Read Alignment

Each end of the paired end sequences was matched against the relevant bait primer plus genomic sequences allowing up to two mismatches with bowtie (V 0.12.5; command line options: –v2). For read pairs longer than 2 × 36 nts, 10 nts were trimmed of the 3′ end of each read. Each read pair with a single match to one of the primers was then checked for a perfect match to the linker on the second arm. If the linker was present, this arm was designated a target arm, linker sequence was trimmed, and the remainder was aligned against the mouse genome (NCBI 37/mm9) with bowtie allowing up to 2 mismatches and requiring unique alignments in the best alignment stratum (command line options: -v2 –all –best –strata -m1:). Exactly identical alignments (same position, same strand) were combined into a single putative translocation event and events supported by a single alignment were not considered in any analyses. We also removed putative translocation events closer than 1 kb to their respective bait. For hotspot analyses the exclusion limit was increased to 50 kb. Translocation positions were given as the position of the 5′ end of the read in the alignment. Data from technical and biological repeats were pooled to increase saturation (Table S7).

Mapping of Translocation Hot Spots

A translocation hotspot was defined as a localized enrichment of translocation events above what is expected from the null hypothesis of uniform distribution of translocation events along the genome. To identify such hotspots, candidate regions were defined as locations containing consecutive translocations with distances shorter than expected from the mappable size of the mm9 genome assembly ($p < 0.01$ each as determined by a negative binomial test). For a candidate region to be called a hotspot it had to (1) have more than 3 translocations and (2) have at least one read from each of the two sides of the bait and (3) have at least 10% of the translocations come from each side of the bait and (4) have a combined *P* value less than 10^{-9} given the number of translocations and length of the region as determined by a negative binomial test. Hotspots with a large degree (>80%) of overlap with repeat regions, small

footprints (<100nt) or less than 10-fold enrichment over the AID$^{-/-}$ control were removed. Analyses of RNA-Seq, chromatin modifications, AID-, PolII-, and Spt5-ChIP as well as the identification of cryptic I-SceI sites, TSSs, genic and intergenic domains were carried out in R (http://www.R-project.org).

Hypermutation Analysis

CD43^{-} splenocytes from *Igk*AID-*Ung*$^{-/-}$ or *Aicda*$^{-/-}$ mice were cultured at 0.1 × 10^{6} cells/ml with LPS+IL-4, and 0.5 mg/ml of aCD180 (RP105) antibody (RP/14, BD PharMingen). At 72 hr cells were diluted 1:4 and cultured for another 48 hr. 50 ng of genomic DNA was amplified for 30 cycles with Phusion DNA polymerase (New England Biolabs) and specific primers (Table S8B). For nested PCR, two-20 cycle amplifications were performed with DMSO. The amplicon was cloned using PCR Zero blunt (Invitrogen) and sequenced.

ACCESSION NUMBERS

The TC-Seq datasets are deposited in SRA (http://www.ncbi.nlm.nih.gov/sra) under accession number SRA039959.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and eight tables and can be found with this article online at doi:10.1016/j.cell.2011.07.048.

ACKNOWLEDGMENTS

I.A.K. designed and performed experiments and analysis and wrote the manuscript. W.R. designed and performed data analysis. M.J. performed TC-Seq experiments. T.O. designed and performed data analysis. A.Y. and H.N. performed hypermutation sequencing and analysis. M.D.V. and A.B. assisted with TC-Seq experiments. D.F.R. assisted with TC-Seq experiments and contributed mice. A.N. made suggestions on the manuscript. R.C. and M.C.N. designed experiments and analysis and wrote the manuscript.

We thank all the members of the Nussenzweig and Casellas labs for valuable input and advice; Klara Velinzon and Svetlana Mazel for FACSorting; and David Bosque and Thomas Eisenreich for animal management. We also thank Scott Dewell of the Rockefeller Genomics Resource Center and Gustavo Gutierrez of the NIAMS genome facility for high-throughput sequencing and guidance; as well as Christopher Mason of the Weill Cornell Medical College for assistance with data analysis. I.A.K. was supported by NIH MSTP grant GM07739, and is a Cancer Research Institute Predoctoral Fellow and a William Randolph Hearst Foundation Fellow. A.B. is a Cancer Research Institute Predoctoral Fellow. This work was supported by NIH grant #AI037526 to M.C.N., NYSTEM #C023046, The Starr Cancer Consortium and the Intramural Research Program of the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health. M.C.N. is an HHMI investigator.

Received: May 24, 2011
Revised: July 14, 2011
Accepted: July 27, 2011
Published: September 29, 2011

REFERENCES

Aguilera, A. (2002). The connection between transcription and genomic instability. EMBO J. *21*, 195–201.

Akasaka, T., Lossos, I.S., and Levy, R. (2003). BCL6 gene translocation in follicular lymphoma: a harbinger of eventual transformation to diffuse aggressive lymphoma. Blood *102*, 1443–1448.

Argast, G.M., Stephens, K.M., Emond, M.J., and Monnat, R.J., Jr. (1998). I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential in vitro enrichment. J. Mol. Biol. *280*, 345–353.

Bea, S., Salaverria, I., Armengol, L., Pinyol, M., Fernandez, V., Hartmann, E.M., Jares, P., Amador, V., Hernandez, L., Navarro, A., et al. (2009). Uniparental disomies, homozygous deletions, amplifications, and target genes in mantle cell lymphoma revealed by integrative high-resolution whole-genome profiling. Blood *113*, 3059–3069.

Bothmer, A., Robbiani, D.F., Di Virgilio, M., Bunting, S.F., Klein, I.A., Feldhahn, N., Barlow, J., Chen, H.T., Bosque, D., Callen, E., et al. (2011). Regulation of DNA End Joining, Resection, and Immunoglobulin Class Switch Recombination by 53BP1. Mol. Cell *42*, 319–329.

Branzei, D., and Foiani, M. (2010). Maintaining genome stability at the replication fork. Nat. Rev. Mol. Cell Biol. *11*, 208–219.

Campbell, P.J., Stephens, P.J., Pleasance, E.D., O'Meara, S., Li, H., Santarius, T., Stebbings, L.A., Leroy, C., Edkins, S., Hardy, C., et al. (2008). Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat. Genet. *40*, 722–729.

Chaudhuri, J., and Alt, F.W. (2004). Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. Nat. Rev. Immunol. *4*, 541–552.

Cory, S., Graham, M., Webb, E., Corcoran, L., and Adams, J.M. (1985). Variant (6;15) translocations in murine plasmacytomas involve a chromosome 15 locus at least 72 kb from the c-myc oncogene. EMBO J. *4*, 675–681.

Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat. Rev. Genet. *2*, 292–301.

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

Einerson, R.R., Law, M.E., Blair, H.E., Kurtin, P.J., McClure, R.F., Ketterling, R.P., Flynn, H.C., Dogan, A., and Remstein, E.D. (2006). Novel FISH probes designed to detect IGK-MYC and IGL-MYC rearrangements in B-cell lineage malignancy identify a new breakpoint cluster region designated BVR2. Leukemia *20*, 1790–1799.

Franco, S., Gostissa, M., Zha, S., Lombard, D.B., Murphy, M.M., Zarrin, A.A., Yan, C., Tepsuporn, S., Morales, J.C., Adams, M.M., et al. (2006). H2AX prevents DNA breaks from progressing to chromosome breaks and translocations. Mol. Cell *21*, 201–214.

Gauwerky, C.E., Huebner, K., Isobe, M., Nowell, P.C., and Croce, C.M. (1989). Activation of MYC in a masked t(8;17) translocation results in an aggressive B-cell leukemia. Proc. Natl. Acad. Sci. USA *86*, 8867–8871.

Gesk, S., Klapper, W., Martin-Subero, J.I., Nagel, I., Harder, L., Fu, K., Bernd, H.W., Weisenburger, D.D., Parwaresch, R., and Siebert, R. (2006). A chromosomal translocation in cyclin D1-negative/cyclin D2-positive mantle cell lymphoma fuses the CCND2 gene to the IGK locus. Blood *108*, 1109–1110.

Goldman, J.M., and Melo, J.V. (2003). Chronic myeloid leukemia–advances in biology and new approaches to treatment. N. Engl. J. Med. *349*, 1451–1464.

Gordon, M.S., Kanegai, C.M., Doerr, J.R., and Wall, R. (2003). Somatic hypermutation of the B cell receptor genes B29 (Igbeta, CD79b) and mb1 (Igalpha, CD79a). Proc. Natl. Acad. Sci. USA *100*, 4126–4131.

Gostissa, M., Yan, C.T., Bianco, J.M., Cogne, 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.

Gottipati, P., and Helleday, T. (2009). Transcription-associated recombination in eukaryotes: link between transcription, replication and recombination. Mutagenesis *24*, 203–210.

Honjo, T. (2002). Does AID need another aid? Nat. Immunol. *3*, 800–801.

Huppi, K., Siwarski, D., Skurla, R., Klinman, D., and Mushinski, J.F. (1990). Pvt-1 transcripts are found in normal tissues and are altered by reciprocal(6;15) translocations in mouse plasmacytomas. Proc. Natl. Acad. Sci. USA *87*, 6964–6968.

Iida, S., Rao, P.H., Ueda, R., Chaganti, R.S., and Dalla-Favera, R. (1999). Chromosomal rearrangement of the PAX-5 locus in lymphoplasmacytic lymphoma with t(9;14)(p13;q32). Leuk. Lymphoma *34*, 25–33.

Impera, L., Albano, F., Lo Cunsolo, C., Funes, S., Iuzzolino, P., Laveder, F., Panagopoulos, I., Rocchi, M., and Storlazzi, C.T. (2008). A novel fusion 5′AFF3/3′BCL2 originated from a t(2;18)(q11.2;q21.33) translocation in follicular lymphoma. Oncogene *27*, 6187–6190.

Jankovic, M., Robbiani, D.F., Dorsett, Y., Eisenreich, T., Xu, Y., Tarakhovsky, A., Nussenzweig, A., and Nussenzweig, M.C. (2010). Role of the translocation partner in protection against AID-dependent chromosomal translocations. Proc. Natl. Acad. Sci. USA *107*, 187–192.

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

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.

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.

Lu, D., and Yunis, J.J. (1992). Cloning, expression and localization of an RNA helicase gene from a human lymphoid cell line with chromosomal breakpoint 11q23.3. Nucleic Acids Res. *20*, 1967–1972.

Mahowald, G.K., Baron, J.M., Mahowald, M.A., Kulkarni, S., Bredemeyer, A.L., Bassing, C.H., and Sleckman, B.P. (2009). Aberrantly resolved RAG-mediated DNA breaks in Atm-deficient lymphocytes target chromosomal breakpoints in cis. Proc. Natl. Acad. Sci. USA *106*, 18339–18344.

Mao, X., Lillington, D., Child, F., Russell-Jones, R., Young, B., and Whittaker, S. (2002). Comparative genomic hybridization analysis of primary cutaneous B-cell lymphomas: identification of common genomic alterations in disease pathogenesis. Genes Chromosomes Cancer *35*, 144–155.

Mottok, A., Renne, C., Seifert, M., Oppermann, E., Bechstein, W., Hansmann, M.L., Kuppers, R., and Brauninger, A. (2009). Inactivating SOCS1 mutations are caused by aberrant somatic hypermutation and restricted to a subset of B-cell lymphoma entities. Blood *114*, 4503–4506.

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.

Nakamura, Y., Takahashi, N., Kakegawa, E., Yoshida, K., Ito, Y., Kayano, H., Niitsu, N., Jinnai, I., and Bessho, M. (2008). The GAS5 (growth arrest-specific transcript 5) gene fuses to BCL6 as a result of t(1;3)(q25;q27) in a patient with B-cell lymphoma. Cancer Genet. Cytogenet. *182*, 144–149.

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

Okazaki, I.M., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K., and Honjo, T. (2003). Constitutive expression of AID leads to tumorigenesis. J. Exp. Med. *197*, 1173–1181.

Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R.S., Kuppers, 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.

Peled, J.U., Kuang, F.L., Iglesias-Ussel, M.D., Roa, S., Kalis, S.L., Goodman, M.F., and Scharff, M.D. (2008). The biochemistry of somatic hypermutation. Annu. Rev. Immunol. *26*, 481–511.

Pleasance, E.D., Cheetham, R.K., Stephens, P.J., McBride, D.J., Humphray, S.J., Greenman, C.D., Varela, I., Lin, M.L., Ordonez, G.R., Bignell, G.R., et al. (2010a). A comprehensive catalogue of somatic mutations from a human cancer genome. Nature *463*, 191–196.

Pleasance, E.D., Stephens, P.J., O'Meara, S., McBride, D.J., Meynert, A., Jones, D., Lin, M.L., Beare, D., Lau, K.W., Greenman, C., et al. (2010b). A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature *463*, 184–190.

Potter, M. (2003). Neoplastic development in plasma cells. Immunol. Rev. *194*, 177–195.

Rabbitts, T.H. (2009). Commonality but diversity in cancer gene fusions. Cell *137*, 391–395.

Ramiro, A.R., Jankovic, M., Callen, E., Difilippantonio, S., Chen, H.T., McBride, K.M., Eisenreich, T.R., Chen, J., Dickins, R.A., Lowe, S.W., et al. (2006). Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. Nature *440*, 105–109.

Ramiro, A., Reina San-Martin, B., McBride, K., Jankovic, M., Barreto, V., Nussenzweig, A., and Nussenzweig, M.C. (2007). The role of activation-induced deaminase in antibody diversification and chromosome translocations. Adv. Immunol. *94*, 75–107.

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A., et al. (2000). Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). Cell *102*, 565–575.

Riggi, N., and Stamenkovic, I. (2007). The Biology of Ewing sarcoma. Cancer Lett. *254*, 1–10.

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 breaks 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.

Roix, J.J., McQueen, P.G., Munson, P.J., Parada, L.A., and Misteli, T. (2003). Spatial proximity of translocation-prone gene loci in human lymphomas. Nat. Genet. *34*, 287–291.

Rosebeck, S., Madden, L., Jin, X., Gu, S., Apel, I.J., Appert, A., Hamoudi, R.A., Noels, H., Sagaert, X., Van Loo, P., et al. (2011). Cleavage of NIK by the API2-MALT1 fusion oncoprotein leads to noncanonical NF-kappaB activation. Science *331*, 468–472.

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.

Staszewski, O., Baker, R.E., Ucher, A.J., Martier, R., Stavnezer, J., and Guikema, J.E. (2011). Activation-induced cytidine deaminase induces reproducible DNA breaks at many non-Ig Loci in activated B cells. Mol. Cell *41*, 232–242.

Stavnezer, J., Guikema, J.E., and Schrader, C.E. (2008). Mechanism and regulation of class switch recombination. Annu. Rev. Immunol. *26*, 261–292.

Stephens, P.J., McBride, D.J., Lin, M.L., Varela, I., Pleasance, E.D., Simpson, J.T., Stebbings, L.A., Leroy, C., Edkins, S., Mudie, L.J., et al. (2009). Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature *462*, 1005–1010.

Storb, U., Shen, H.M., Longerich, S., Ratnam, S., Tanaka, A., Bozek, G., and Pylawka, S. (2007). Targeting of AID to immunoglobulin genes. Adv. Exp. Med. Biol. *596*, 83–91.

Takeuchi, Y., Horiuchi, T., and Kobayashi, T. (2003). Transcription-dependent recombination and the role of fork collision in yeast rDNA. Genes Dev. *17*, 1497–1506.

Thomas, B.J., and Rothstein, R. (1989). Elevated recombination rates in transcriptionally active DNA. Cell *56*, 619–630.

Tsai, A.G., and Lieber, M.R. (2010). Mechanisms of chromosomal rearrangement in the human genome. BMC Genomics *11* (*Suppl 1*), S1.

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.

Ueda, C., Akasaka, T., Kurata, M., Maesako, Y., Nishikori, M., Ichinohasama, R., Imada, K., Uchiyama, T., and Ohno, H. (2002). The gene for interleukin-21 receptor is the partner of BCL6 in t(3;16)(q27;p11), which is recurrently observed in diffuse large B-cell lymphoma. Oncogene *21*, 368–376.

Wang, J.H., Gostissa, M., Yan, C.T., Goff, P., Hickernell, T., Hansen, E., Difilippantonio, S., Wesemann, D.R., Zarrin, A.A., Rajewsky, K., et al. (2009).

Mechanisms promoting translocations in editing and switching peripheral B cells. Nature *460*, 231–236.

Wong, S., and Witte, O.N. (2004). The BCR-ABL story: bench to bedside and back. Annu. Rev. Immunol. *22*, 247–306.

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.

Yan, C.T., Boboila, C., Souza, E.K., Franco, S., Hickernell, T.R., Murphy, M., Gumaste, S., Geyer, M., Zarrin, A.A., Manis, J.P., et al. (2007). IgH class switching and translocations use a robust non-classical end-joining pathway. Nature *449*, 478–482.

Yoshida, S., Kaneita, Y., Aoki, Y., Seto, M., Mori, S., and Moriyama, M. (1999). Identification of heterologous translocation partner genes fused to the BCL6 gene in diffuse large B-cell lymphomas: 5′-RACE and LA - PCR analyses of biopsy samples. Oncogene *18*, 7994–7999.

Zhang, Y., Gostissa, M., Hildebrand, D.G., Becker, M.S., Boboila, C., Chiarle, R., Lewis, S., and Alt, F.W. (2010). The role of mechanistic factors in promoting chromosomal translocations found in lymphoid and other cancers. Adv. Immunol. *106*, 93–133.