INTRODUCTION

Antigen receptor diversity enables lymphocytes to initiate effective immune responses against a virtually limitless array of pathogens. The diversity in the primary antigen receptor repertoire is achieved by V(D)J recombination, a site-specific reaction catalyzed by a heterotetrameric protein complex encoded by the recombination-activating genes RAG1 and RAG2 (Kim et al., 2015; Ru et al., 2015). The RAG recombinase (RAG1/2) joins randomly selected variable, diversity, and joining (V, D, and J) gene segments to assemble a V(D)J exon that encodes the variable region of antibodies and T cell receptors (Schatz and Ji, 2011; Schatz and Swanson, 2011). RAG1/2 does so in part by recognizing and cleaving conserved recombination signal sequences (RSSs) that flank each V, D, and J gene segment. RAG1 is the principal DNA binding and cleavage component of the recombinase. RAG2 is an essential cofactor and consists of a core portion (RAG2core) minimally required for its activity and a C-terminal region important for efficiency, fidelity, and ordering of V(D)J rearrangements (Sekiguchi et al., 2001; Liang et al., 2002; Akamatsu et al., 2003; Talukder et al., 2004; Curry and Schlissel, 2008).

RSSs are comprised of a conserved palindromic heptamer (consensus: 5′-CACAGTG-3′) that is required for DNA cleavage, a degenerate spacer of 12 or 23 bp, and a less-conserved A-rich nonamer (consensus: 5′-ACAAAA ACC-3′) that is important for RAG1/2 binding (Schatz and Ji, 2011; Schatz and Swanson, 2011). RSSs with 12- or 23-bp spacers are termed 12RSSs and 23RSSs, respectively. During V(D)J recombination, RAG1/2 first binds to a single 12- or 23RSS (signal complex) and then captures a “complementary” 23- or 12RSS (paired complex) according to the “12/23 rule.” Upon synopsis, the recombinase introduces DNA double-strand breaks between coding sequences and flanking RSSs by making a single-strand nick that is used to catalyze a transesterification that produces a hairpin-sealed coding end and a blunt-cut signal end. After cleavage, RAG1/2 remains associated with paired coding and signal ends in a post-cleavage complex, thereby scaffolding their repair by nonhomologous end joining (NHEJ). Coding ends are fused to produce V(D)J-coding exons, and ligation of signal ends generates noncoding signal joints. Depending on the orientation of paired RSSs, RAG1/2 catalyzes either inversionsal (head-to-tail RSSs) or deletional (convergent RSSs) recombination. During inversionsal recombination, signal joints remain in the genome, whereas they are excised as episomal signal joints during deletional recombination (Helmink and Sleckman, 2012).

In addition to its essential role in adaptive immunity, RAG1/2 has been implicated in the genesis of chromosome translocations and deletions associated with lymphoid malignancy (Roth, 2003; Lieber, 2016). Mice deficient for ataxia-telangiectasia mutated kinase (ATM) or both the tumor suppressor protein p53 and components of the NHEJ machinery develop RAG1/2-dependent chromosome translocations associated with pro-B cell lymphomas
(Nussenzweig and Nussenzweig, 2010; Alt et al., 2013). In humans, RAG1/2 is implicated in the genesis of follicular lymphoma (FL), mantle cell lymphoma, and acute lymphoblastic leukemia (ALL), all of which carry genome aberrations in the proximity of RSSs in antigen receptor genes or nonphysiological cryptic RSSs (cRSSs) with conserved heptamer motifs (Küppers and Dalla-Favera, 2001; Nussenzweig and Nussenzweig, 2010; Alt et al., 2013). Predicted cRSSs are broadly distributed throughout the genome, and so are RAG1/2 binding sites, as assayed by chromatin immunoprecipitation (Lewis et al., 1997; Ji et al., 2010; Merelli et al., 2010; Teng et al., 2015). Consistent with the idea that RAG1/2 can induce DNA damage at cRSSs, it causes chromosomal deletions, and in the context of ATM deficiency also translocations, between engineered RSSs and genomic cRSSs in primary pro–B cells and pro–B cell lines (Hu et al., 2015). The reported off-target mechanism involves directional, linear tracking of RAG1/2 within chromosomal loop domains to locate RSS/cRSS pairs (Hu et al., 2015).

Biochemical experiments as well as episomal assays in cell lines and yeast suggest that RAG1/2 can mediate DNA transposition by inserting RSS-containing donor sequences into target DNA (Agrawal et al., 1998; Hiom et al., 1998; Lee et al., 2002; Neiditch et al., 2002; Clatworthy et al., 2003; Elkin et al., 2003; Tasi et al., 2003; Chatterji et al., 2006; Posey et al., 2006; Reddy et al., 2006). Experiments with reporter cell lines indicate that RAG1/2 can also catalyze trans–V(D)J recombination, during which episomal signal joints are reinserted at an endogenous RSS or cRSS (Reddy et al., 2006). Nevertheless, only a few RAG1/2-mediated genomic insertions have been documented in vivo (Messier et al., 2003; Curry et al., 2007; Vanura et al., 2007). Moreover, RAG1/2-mediated DNA insertions contributing to cancer display characteristics that are not compatible with either DNA transposition or trans–V(D)J recombination (Navarro et al., 2015). Hence, how RAG1/2 causes genomic DNA insertions is still largely unknown.

Here we use translocation capture sequencing (TC-Seq) and insertion capture sequencing (IC-Seq) to analyze chromosomal rearrangements in primary murine developing B cells. We identify aberrant RAG1/2-dependent DNA deletions at immunoglobulin (Ig) genes, whose products are reinserted at DNA breaks generated by the I-SceI endonuclease on a heterologous chromosome. The existence of similar insertions in human cancer indicates that RAG1/2 also mobilizes genomic DNA into independent physiological breaks. Thus, our findings reveal a novel pathway through which RAG1/2 causes DNA insertions independent of DNA transposition and trans–V(D)J recombination. Importantly, this pathway has the potential to destabilize the lymphocyte genome by causing aberrant signal–end, hybrid–end, and coding–end insertions at RAG1/2-independent DNA breaks and shares features with reported oncogenic DNA insertions.

RESULTS

Chromosomal rearrangements in pro–B cells
To examine RAG1/2-induced chromosomal rearrangements in pro–B cells, we adapted a previously described next-generation TC-Seq method (Klein et al., 2011; Oliveira et al., 2012). TC-Seq captures genome-wide chromosomal rearrangements to a unique DNA double-strand break created by the I-SceI endonuclease. DNA rearrangements between the I-SceI break and the genome are amplified by PCR, deep-sequenced, and analyzed computationally. We prepared TC-Seq libraries from cell cultures of primary murine pro–B cells deficient for RAG2 and harboring I-SceI sites at c-myc (RAG2<sup>−/−</sup> Myc<sup>+</sup>) that were infected with retroviruses expressing either I-SceI alone (RAG2<sup>−/−</sup> TC-Seq libraries) or I-SceI together with murine RAG2<sup>core</sup> (RAG2<sup>core</sup> TC-Seq libraries; Fig. 1 A and Fig. S1 A; also see Materials and methods). RAG2<sup>core</sup> was used because it promotes aberrant V(D)J recombination and causes genomic instability at T cell receptor (TCR) loci in thymocytes (Sekiguchi et al., 2001; Tálukder et al., 2004; Curry and Schlissel, 2008; Deriano et al., 2011). Moreover, mice expressing RAG2<sup>core</sup> and deficient for either p53 alone or in combination with XRCC4-like factor develop thymic or pro–B cell lymphomas, respectively, with translocations involving antigen receptor genes (Deriano et al., 2011; Mijušković et al., 2015; Lescale et al., 2016).

In agreement with previous TC-Seq studies in other cell types, chromosomal rearrangements in pro–B cells were especially abundant near the I-SceI cleavage site on chromosome 15 (Fig. 1, B and C; Klein et al., 2011; Wang et al., 2014; Robbiani et al., 2015). Moreover, rearrangements were enriched at genic regions, highly transcribed genes, and early replication fragile sites (ERFSs), which define regions particularly susceptible to DNA damage during early replication (Fig. 1, D–F; Barlow et al., 2013).

DNA damage at physiological RSSs and cRSSs
To identify the DNA damage caused by RAG1/2<sup>core</sup>, we compared chromosomal rearrangements in RAG2<sup>core</sup> and RAG2<sup>−/−</sup> TC-Seq libraries. In brief, genomic hotspots of rearrangement were identified, and those unique to RAG1/2<sup>core</sup> were analyzed for the occurrence of breakpoint clusters (see Materials and methods). Overall, 33 RAG1/2<sup>core</sup>-dependent rearrangement breakpoint clusters were detected throughout the genome (Table S1).

In agreement with previous studies, we observed limited recombination of the I<sub>gh</sub> locus by RAG1/2<sup>core</sup> and consequently detected only few dispersed breakpoints at Vh, Dh, and Jh gene segments (Fig. S1 B and not depicted; Liang et al., 2002; Akamatsu et al., 2003). In contrast, 24 RAG1/2<sup>core</sup>-dependent breakpoint clusters were identified at I<sub>gh</sub> (Fig. 2 A and Table S1). Each functional J<sub>k</sub> segment (J<sub>k1</sub>, J<sub>k2</sub>, J<sub>k4</sub>, and J<sub>k5</sub>) had a single cluster at its 23RSS cleavage site (Fig. 2 B). Surprisingly, DNA at these clusters recombined with the I-SceI break in a biased manner. Although in principle both DNA ends of a RAG1/2<sup>core</sup>-induced break would have an
equal probability of joining to the cleaved I-SceI site, most rearrangements occurred with only one of the two ends for any RSS. For example, rearrangements between the I-SceI break and RAG1/2core breaks at Jκ1 exclusively involved the coding end (Fig. 2B, rearrangements in gray), whereas those at the neighboring Jκ2 predominantly (86%) contained the signal end (Fig. 2B, rearrangements in green). Moreover, rearrangements at Jκ1 did not extend beyond the 23RSS cleavage site of Jκ2, and vice versa. A similar phenomenon was observed for Jκ4/Jκ5 (Fig. 2B).
Figure 2. *Rearrangements between I-SceI and RAG1/2<sup>core</sup>-induced DNA breaks at Igκ.* (A) Overview of rearrangement breakpoints at the Igκ locus on chromosome 6. Histogram of the number of breakpoints in the presence or absence of RAG2<sup>core</sup> (red and black, respectively) in 10-kb intervals. RAG1/2<sup>core</sup>-dependent rearrangement breakpoint clusters are indicated by red triangles and labeled with the corresponding Jκ or Vκ gene segment. Asterisks mark breakpoint clusters with biased rearrangements (see B–D). Chromosome coordinates and scale bar are indicated on top. (B–D) Examples of RAG1/2<sup>core</sup>-dependent breakpoint clusters at Jκs and Vκs. On top is a diagram of the region, with gray boxes representing Ig segments, triangles indicating 12/23RSSs (green) or cRSSs (purple), and red bars indicating the breakpoint clusters. In the middle is a histogram showing the number and position of breakpoints (Bp, red). At the bottom, each horizontal line indicates a unique rearrangement (Rx), with its breakpoint represented by the vertical line and its sheared end (which determines the uniqueness of the event) shown by the box. Color coding indicates whether rearrangements contain RSSs/cRSSs (green/purple, signal ends) or not (gray, coding ends). Rearrangements in black are undefined. Chromosome coordinates and scale bar are indicated on top. Data analysis was performed with pooled RAG2<sup>core</sup> and RAG2<sup>−/−</sup> TC-Seq libraries (two independent experiments each). See also Fig. S2 (A–C) and Tables S1 and S2.
In addition to Jκs, breakpoint clusters were also found at 15 Vκ gene segments. Strikingly, although 10 of these had a single cluster at their physiological 12RSS cleavage sites, the other 5 (Vκ3-1, Vκ10-94, Vκ10-95, Vκ10-96, and Vκ1-110) revealed an additional cluster at a nearby cRSS (Fig. 2 A and Table S1). Overall, the heptamer sequences of these cRSSs were similar to the physiological consensus and to those identified in previous studies (Fig. S2 A; Hu et al., 2015). However, none of the cRSSs were detectable by computational tools because of their low RSS information content (RIC) scores (Table S2; Cowell et al., 2002; Merelli et al., 2010). Similar to the biased recombination pattern observed at Jκs, Vκ rearrangements at neighboring 12RSS/cRSS clusters were biased for coding or signal ends and limited in length by both cleavage sites (Fig. 2, C and D).

The remaining breakpoint clusters (nine) mapped to off-target regions outside of Ig loci (Table S1). Off-target sites (Fig. 2, C and D). The peculiar rearrangement pattern observed at Jκs, Vκ rearrangements at neighboring 12RSS/cRSS clusters were biased for coding or signal ends and limited in length by both cleavage sites (Fig. 2, C and D).

**Aberrant deletions at Igκ**

The peculiar rearrangement pattern observed at Jκ and some of the Vκ clusters suggested that RAG1/2core may mediate aberrant deletions by recombining neighboring RSSs and cRSSs at these sites. To examine this possibility, we searched for deletions by poison primer PCR (see Materials and methods; Edgley et al., 2002). Strikingly, aberrant deletions mediated by either RAG1/2core or endogenous wild-type RAG1/2 were readily detected at Jκs, where the RSSs at Jκ1 and Jκ4 were joined to the neighboring Jκ2 and Jκ5 exons, respectively (Fig. 3, A and B). The resulting deletion junctions (hybrid junctions) represent aberrant joining events because physiological recombination of head-to-tail RSSs induces inversions and only involves 12/23RSS pairs (Helmink and Sleckman, 2012). However, some of these joints could also originate from two sequential inversions involving nearby Vκs, similar to those observed between D segments at TCRδ in thymocytes and between engineered RSS/cRSS pairs in ATM-deficient pro-B cell lines (Hu et al., 2015; Zhao et al., 2016). In addition to those at Jκs, deletions mediated by either RAG1/2core or RAG1/2 wild type were also identified at Vκ3-1, where joining of the 12RSS to the nearby cRSS described above generated aberrant signal joints (Fig. 3 C). We conclude that both RAG1/2core and RAG1/2 wild type cause aberrant genomic deletions at Jκ and Vκ segments.

**Excised Igκ fragments insert into I-SceI breaks**

Based on the colocalization of biased rearrangements and aberrant deletions, we hypothesized that Jκ/Vκ fragments might be aberrantly excised by RAG1/2core and subsequently re-integrate at the I-SceI break (Fig. 4 A; see Discussion). To test this hypothesis, we searched TC-Seq libraries computationally for bona fide insertions, which would have been excluded from our initial bioinformatic analysis geared at identifying translocations. In brief, insertions at the I-SceI site are flanked by Myc⁺ sequence on both ends, whereas translocations contain Myc⁺ sequence only on one end (Fig. 4 A). Thus, all sequences with Myc⁺ on both ends were examined for intervening DNA originating from elsewhere in the genome (see Materials and methods).

We detected I-SceI insertions in both RAG2core and RAG2−/− TC-Seq libraries. Independent of RAG2core expression, inserted DNA fragments originated predominantly from a ±20-kb region around the I-SceI cleavage site on chromosome 15, similar to the chromosomal rearrangements described above (Fig. 1, B and C; and Fig. 4, B and C). Overall, inserted DNA fragments ranged from 36 to 354 bp in RAG2core and from 36 to 232 bp in RAG2−/− cells (36 bp being the minimum detection limit; see Materials and methods). Moreover, genic regions acted as preferred donors for insertions, particularly in RAG2core-expressing cells (Fig. 4 D). In contrast, insertions originating from highly transcribed regions and ERFSs were significantly enriched only in the absence of RAG2core, indicating that its expression alters the insertion landscape (Fig. 4, E and F). Thus, we found more insertions from chromosome 6 in RAG2core compared with RAG2−/− cells (140 vs. 8 events; Fig. 4 B), and with RAG2core nearly all of those (96%) originated from Igκ, whereas none derived from this locus in RAG2−/− cells (Fig. 5 A). Overall, Igκ insertions represented nearly half (43%) of all insertions in RAG2core cells and exclusively originated from regions flanked by RSSs and/or cRSSs (Fig. 5, B–D). Interestingly, donor regions included all of the Igκ gene segments displaying biased breakpoint clusters, suggesting that DNA insertions from these sites are responsible for the observed recombination pattern (see Figs. 2 and 5 and Discussion).

For 67% of Igκ insertions, we obtained sequence information on both junctions, providing insight into the original deletion events (Table S3). Overall, Igκ insertions originated from DNA excision between pairs of divergent, convergent, or head-to-tail RSSs, leading to insertions flanked by coding ends (coding-end insertions, 77), signal ends (signal-end insertions, 8), or both (hybrid-end insertions, 6), respectively (Fig. 5, B–D; and Table S3). Most deletions (87 out of 91) occurred between RSS/cRSS pairs, three resulted from excisions between two cRSSs, and one derived from a deletion between two 23RSSs. We conclude that RAG1/2core generates aberrant Igκ fragments that are mobile and can be reinserted into I-SceI breaks on a heterologous chromosome.
As demonstrated by our deletion PCR assays, RAG1/2 can produce aberrant Igκ deletions analogous to RAG1/2 core. Thus, mobilization and insertion of Igκ DNA could in principle also occur in wild-type B cells. To test this possibility, we developed a next-generation IC-Seq method that qualitatively documents chromosomal insertions at an I-SceI site under physiological conditions. We prepared IC-Seq libraries from primary bone marrow B cells expressing a tamoxifen-inducible I-SceI transgene and bearing I-SceI cleavage sites (ROSAerISCERIMycI/IIghI/IAID−/−; see Materials and methods; Robbiani et al., 2015) that were treated ex vivo with tamoxifen to induce I-SceI breaks in the presence of wild-type RAG1/2. DNA insertions at the I-SceI site in c-myc were amplified by PCR, deep-sequenced, and analyzed computationally (Fig. 6A; see Materials and methods). Overall, we detected I-SceI insertions from seven different Igκ gene segments (Jκ1, Jκ2, Jκ4, Jκ5, Vk1−110, Vk3−1, and Vk4−69), of which six were also involved in the aforementioned insertions mediated by RAG1/2core (Table S3). Moreover, similar to RAG1/2core, Igκ insertions in the presence of RAG1/2 originated exclusively from donor regions flanked by RSSs/and or cRSSs and were comprised of coding-, signal-, and hybrid-end insertions (Fig. 6, B–D; and Table S3). We conclude that DNA insertions from Igκ are not limited to RAG1/2core but also occur during physiological V(D)J recombination by wild-type RAG1/2.

Insertion of IG and TCR fragments at physiological DNA breaks

To determine whether RAG1/2 causes insertions at physiological DNA breaks in vivo, we searched published whole-genome sequences from ALL and FL patients for insertions deriving from IG and TCR loci (see Materials and methods). Overall, 5 out of 34 patients displayed genomic insertions of IG or TCR fragments at low frequency (Table S4). All insertions contained at least one RSS or cRSS motif and integrated near repetitive regions (Fig. 7A and Fig. S3). Interestingly, DNA flanking one of the inserts was inverted to form a putative cRSS/cRSS signal joint, and in another case, a TCR fragment inserted at a translocation junction (Fig. 7, A and B). We conclude that RAG1/2 has the potential to destabilize the lymphocyte genome by mobilizing DNA that then reinserts at RAG1/2-independent, physiological DNA breaks in vivo.
We used TC-Seq to examine chromosomal rearrangements in the pro–B cell genome and identified 33 RAG1/2core-dependent breakpoint clusters, of which 19 occurred at physiological RSS cleavage sites. Consistent with this finding, a previous study in ATM-deficient pro–B cell lines detected chromosomal rearrangements between I-SceI breaks at c-myc and RAG1/2-induced breaks at antigen receptor loci including Igκ (Zhang et al., 2012). Interestingly, off-target clusters at cRSSs were not detected in those experiments. In contrast, 14 of the 33 RAG1/2core-dependent breakpoint clusters identified herein were located near cRSS motifs at Vκs and off-target regions outside Ig loci. Off targets were not enriched in H3K4me3, an active chromatin mark that has been shown to colocalize with RAG1/2 binding and cleavage in developing B cells (Ji et al., 2010; Hu et al., 2015; Teng et al., 2015; unpublished data). Its absence might result from the fact that RAG2core lacks the C-terminal plant

**DISCUSSION**

**RAG1/2 damages the pro–B cell genome at physiological RSSs and cRSSs**

We used TC-Seq to examine chromosomal rearrangements in the pro–B cell genome and identified 33 RAG1/2core-dependent breakpoint clusters, of which 19 occurred at physiological RSS cleavage sites. Consistent with this finding, a previous study in ATM-deficient pro–B cell lines detected chromosomal rearrangements between I-SceI breaks at c-myc and RAG1/2-induced breaks at antigen receptor loci including Igκ (Zhang et al., 2012). Interestingly, off-target clusters at cRSSs were not detected in those experiments. In contrast, 14 of the 33 RAG1/2core-dependent breakpoint clusters identified herein were located near cRSS motifs at Vκs and off-target regions outside Ig loci. Off targets were not enriched in H3K4me3, an active chromatin mark that has been shown to colocalize with RAG1/2 binding and cleavage in developing B cells (Ji et al., 2010; Hu et al., 2015; Teng et al., 2015; unpublished data). Its absence might result from the fact that RAG2core lacks the C-terminal plant

![Figure 4.](image_url)
homeodomain, which normally mediates RAG1/2 binding to H3K4me3 (West et al., 2005; Liu et al., 2007; Matthews et al., 2007; Ramón-Maiques et al., 2007).  

Our results demonstrate that RAG1/2 core-mediated cleavage of cRSSs enables chromosomal rearrangements by producing cleaved ends that can recombine with RAG1/2-independent DNA breaks. Moreover, our data confirms that neighboring RSSs and cRSSs are substrates for aberrant genomic deletions, in agreement with previous studies using engineered RSSs (Mahowald et al., 2009; Hu et al., 2015). We speculate that the cRSSs at Vκs identified herein might also serve as beneficial substrates for secondary V-J rearrangements during V-gene replacement, similar to those described at Vhs (Rahman et al., 2006).

**Aberrantly excised Igκ DNA reinserts at I-SceI breaks**

We hypothesized that some of the observed rearrangements resembling translocations may actually represent insertions...
of deleted DNA into the I-SceI break. Because DNA is sonicated during preparation of TC-Seq libraries, a fraction of insertions would be randomly truncated and appear as translocations in the analysis (see comparison between translocation and insertion in Fig. 4 A). In agreement with this prediction, we identified bona fide insertions originating from all RAG1/2core breakpoint clusters with biased rearrangements. Furthermore, by using a novel next-generation sequencing method (IC-Seq), we confirmed Igκ insertions at I-SceI breaks in the presence of wild-type

Figure 6. Insertions at the I-SceI site induced by wild-type RAG1/2. (A) Detection of chromosomal insertions by IC-Seq. Primary ROSA<sup>α</sup>iSceI-Myc<sup>d</sup>Igh<sup>6</sup> (and ROSA<sup>α</sup>iSceI-Myc<sup>d</sup>Igh<sup>6</sup>AID<sup>−</sup>, see Materials and methods) bone marrow B cells were treated ex vivo with tamoxifen to induce I-SceI breaks at c-myc on chromosome 15 (black lightning). Mobilized DNA fragments, such as those excised by endogenous RAG1/2 from Igκ on chromosome 6 (blue lightning), reinsert into the cleaved I-SceI site and are subsequently amplified by PCR, deep-sequenced, and analyzed computationally. Two RAG1/2 wild-type IC-Seq libraries were independently prepared from tamoxifen-treated bone marrow B cells of 12 mice. (B–D) Qualitative comparison of insertions obtained by TC-Seq (RAG1/2<sup>core</sup>) and IC-Seq (RAG1/2 wild type). On top is a diagram of the region, with gray boxes representing Ig segments and triangles indicating 12/23RSSs (green) or cRSSs (purple). Below, insertions detected by TC-Seq (Ins, same as in Fig. 5 [B–D]) and IC-Seq (IC). Each horizontal line indicates a unique insertion, with its breakpoints represented by the vertical lines at the ends. Arrows represent insertions for which only one of the two breakpoints could be identified. Chromosome coordinates and scale bar are indicated on top. Data analysis was performed with pooled IC-Seq libraries (two independent experiments). See also Table S3.
RAG1/2. Igκ insertions mediated by RAG1/2core and RAG1/2 wild type were similar in that they both originated from donor regions with RSS/cRSSs and were comprised of all three insertion species (signal-end, coding-end, and hybrid-end).

Overall, insertions detected by both TC-Seq and IC-Seq were short (354 bp or shorter). The absence of larger insertions is likely caused by technical limitations. During TC-Seq, which was originally designed to detect chromosomal translocations, the size of insertions is mainly limited by the sonication of genomic DNA (see Materials and methods). We therefore expect long insertions to be truncated and appear as “translocations” in the computational analysis. In this regard, some of the apparent translocations at VKs and JKs could in principle result from the insertion of large physiological excision fragments (tens to hundreds of kilobases). Moreover, because even short insertions can be truncated, it is possible that TC-Seq considerably underestimates the actual frequency of insertions. During IC-Seq, which omits DNA sonication, the major factor limiting the detection of large insertions is PCR amplification. DNA templates with large insertions are likely outcompeted by those with small or no insertions. Finally, both TC-Seq and IC-Seq use high-throughput sequencing (see Materials and methods), which is inefficient for DNA fragments >1.5 kb.

Igκ insertions at I-SceI breaks are not mediated by DNA transposition or trans-V(D)J recombination

We observed three distinct insertion species from Igκ: those flanked by RSS/cRSS pairs (signal-end insertions), those lacking RSSs altogether (coding-end insertions), and those bearing only one RSS or cRSS (hybrid-end insertions).

Signal-end insertions derive from DNA deletion between convergent RSSs, which are normally joined to form episomal signal joints. There is some in vivo evidence that RAG1/2 can induce genomic insertions by recleaving and subsequently reintegrating episomal signal joints through either trans-V(D)J recombination or DNA transposition (Messier et al., 2003; Curry et al., 2007; Vanura et al., 2007). However, the observed signal-end insertions are not compatible with these two pathways because they occur at RAG1/2-independent DNA breaks generated by I-SceI. In contrast, during trans-V(D)J recombination, it is RAG1/2 that cleaves the RSS/cRSS at the insertion site, and in DNA transposition, RAG1/2 is responsible for catalyzing the nucleophilic attack required for insertion. Thus, the RAG1/2-induced signal-end insertions observed in our study are mediated by a pathway distinct from these previously described mechanisms.

Coding-end insertions do not fit previously proposed RAG1/2 insertion mechanisms either because both trans-V(D)J recombination and DNA transposition require...
RSS-containing donor fragments (Agrawal et al., 1998; Hiom et al., 1998; Curry et al., 2007; Vanura et al., 2007). Coding-end insertions originate from DNA deletions between divergent RSSs, whose products are predicted to circularize into episomal coding joints. Because these cannot be recleaved by RAG1/2, coding-end insertions likely originate from noncircularized, linear deletion products.

Hybrid-end insertions derive from deletions between head-to-tail RSSs. In principle, such deletions produce episomal hybrid joints that contain a single RSS or cRSS. Although in vitro assays have shown that RAG1/2 can induce breaks at single RSSs, the extent to which this occurs in vivo is unclear (McBlane et al., 1995; Eastman and Schatz, 1997; Yu and Lieber, 2000; Rahman et al., 2006). Hence, similar to coding-end insertions, those with hybrid ends likely derive from linear deletion products.

Although Igκ insertions originate from distinct types of RAG1/2 deletions, we propose a model in which they all share a common intermediate: excised linear DNA fragments that escaped from the post-cleavage complex before end joining (Fig. 8). This model agrees with biochemical experiments and studies with reporter cell lines showing that cleaved ends can prematurely escape the post-cleavage complex upon destabilization by RAG2core, nonconsensus RSS heptamers, or absence of the DNA damage response kinase ATM (Bredemeyer et al., 2006; Arnal et al., 2010; Deriano et al., 2011; Coussens et al., 2013). Our data support this model in two ways. First, the occurrence of coding- and hybrid-end insertions speaks against DNA circularization and points to the existence of stable, linear DNA deletion products. Second, because DNA integration is independent of RAG1/2, neither donor fragments nor insertion sites require RSSs/cRSSs for the insertion process. Previous studies in reporter cell lines detected a few insertions of RSS-flanked donor substrates that were not mediated by DNA transposition or by trans-V(D)J recombination (Chatterji et al., 2006; Reddy et al., 2006). Similarly, a study in primary T cells reported a few cases in which the insertion of a specific RSS-flanked TCRβ fragment occurred independently of both pathways (Curry et al., 2007). We conclude that RAG1/2 likely mobilizes linear deletion products, which are stable and have the capacity to reinsert back into the genome at independently generated DNA breaks on heterologous chromosomes. Thus, our findings reveal a novel RAG1/2-mediated insertion pathway distinct from DNA transposition and trans-V(D)J recombination.
Insertions derived from non-Ig loci

Although RAG2<sup>core</sup> expression significantly alters the landscape of chromosomal insertions at I-SceI breaks, the majority of events originates from outside the Igκ locus in both RAG2<sup>core</sup> and RAG2<sup>±/−</sup> pro-B cells (57% and 100% of total, respectively). Those insertions possibly derive from regions prone to genomic instability caused by DNA transcription, replication, or other sources of DNA damage. Consistent with this possibility, chromosomal insertions in RAG2<sup>±/−</sup> cells preferentially originate from highly transcribed genes and ERFS. Alternatively or in addition, non-Igκ insertions may represent “templated-sequence insertions” that derive from reverse-transcribed RNA (Onozawa et al., 2014). Finally, we cannot exclude that some insertions originate from RAG1/2-mediated deletions at off-target sites. In this context, it is intriguing that insertions of non-Ig DNA into antibody receptor genes were recently shown (Navarro et al., 2015). In the reported T-ALL patient, a DNA fragment excised from the TCR locus and a nearby cRSS. These and other downstream re-combinations (e.g., deletions and translocations) might also represent “templated-sequence insertions” that derive from reverse-transcribed RNA (Onozawa et al., 2014).

RAG1/2 causes insertions at independent, physiological DNA breaks

As demonstrated by our computational analysis of human cancers, RAG1/2-induced DNA insertions are not limited to I-SceI breaks but also occur at physiological DNA breaks in vivo. The low number of IG/TCR insertions detected in our tumor analysis is likely caused by limitations of currently available datasets as well as general limitations of whole-genome sequencing techniques. Many of the publicly available tumor datasets either do not have a sufficient coverage or are not sequenced using long enough reads (e.g., 100 bp and longer) to allow for robust detection of insertion junctions. Moreover, the preparation of genomic libraries generally involves DNA fragmentation, which inevitably truncates existing insertions, thereby causing them to appear as translocations in the computational analysis.

Nevertheless, the detection of RAG1/2-induced insertions is particularly important because they pose a threat to genomic stability in at least two ways. First, they provide functional RSS and/or cRSS substrates for secondary re-arrangements. In fact, introducing a RSS outside of Igκ loci has been shown to cause aberrant RAG1/2-mediated deletions and inversions (Mahowald et al., 2009; Hu et al., 2015). Consistent with this, one of the tumor-associated insertions was accompanied by the formation of a putative cRSS/cRSS signal joint, which likely originated from a secondary RAG1/2-mediated DNA inversion between the cRSS in the insert and a nearby eRSS. These and other downstream re-combinations (e.g., deletions and translocations) might also render RAG1/2-induced insertions especially difficult to detect. Second, although none of the insertions in the patients we analyzed are cancer drivers, the oncogenic insertion of an excised TCR fragment was recently described (Navarro et al., 2015). In the reported T-ALL patient, a DNA fragment flanked by two RSSs was excised from the TCRβ locus and reinserted upstream of the TAL1 oncogene, causing its activation. Notably, the TCRβ fragment inserted at a RAG1/2-independent DNA break, analogous to the insertions detected in our study. Furthermore, the oncogenic insertion of an IGH fragment was described in a patient with diffuse large B cell lymphoma (Chaganti et al., 1998). In the reported patient, a rearranged DJ fragment inserted into a translocation junction involving the BCL6 oncogene led to the expression of an aberrant BCL6-IGH fusion transcript. Similarly, we detected an inserted TCR fragment at a translocation junction in our cancer analysis. Thus, RAG1/2 has the capacity to destabilize the lymphocyte genome by producing cancer-associated DNA insertions.

MATERIALS AND METHODS

Mice

Mutant mice used in this study include RAG2<sup>−/−</sup>Myc<sup>Δ/Δ</sup> (B6(Cg)-Rag2<sup>mu2.1Cpo</sup>/J (The Jackson Laboratory; Robbiani et al., 2008), ROSA<sup>ΔBCE/Δ</sup>Myc<sup>Δ/Δ</sup>Igh<sup>Δ/Δ</sup>, and ROSA<sup>ΔBCE/Δ</sup>Myc<sup>Δ</sup>Igh<sup>Δ/Δ</sup>AID<sup>Δ/Δ</sup> (Robbiani et al., 2015). All mice were in a C57BL/6 background or backcrossed to it for at least 10 generations. All experiments were performed in accordance with protocols approved by the Rockefeller University Institutional Animal Care and Use Committee.

Retroviruses

Murine RAG2 (RAG2<sup>full</sup>) and RAG2<sup>core</sup> sequences were amplified from mouse genomic DNA using primers p2/p6 and p3/p6, respectively (Table S5). I-SceI was amplified from pMX-I-SceI-EGFP using primers p4/p5 (Table S5; Robbiani et al., 2008). Overlap extension PCRs of the above products with primers p2/p4 and p3/p4 generated I-SceI-P2A-RAG2<sup>full</sup> and I-SceI-P2A-RAG2<sup>core</sup>, respectively (Table S5). Finally, both constructs were cloned into pMX-EGFP to generate pMX-I-SceI-P2A-RAG2<sup>full</sup>-EGFP and pMX-I-SceI-P2A-RAG2<sup>core</sup>-EGFP, respectively.

Cell culture and infection for TC-Seq

Pro-B cells were isolated from tibias, femurs, and humeri of RAG2<sup>−/−</sup>Myc<sup>Δ/Δ</sup> mice at 4–10 wk of age by immunomagnetic enrichment with anti-B220 MicroBeads (Miltenyi Biotec). Cells were cultured at 2.0 × 10<sup>6</sup> cells/ml in the presence of IL-7 (5 ng/ml; Sigma-Aldrich) in complete RPMI (RPMI-1640 supplemented with 1-glutamine [Gibco], sodium pyruvate [Gibco], antibiotic/antimycotic [Gibco], Hepes [Gibco], 55 μM β-mercaptoethanol [Gibco], and 10% fetal calf serum [HyClone]). IL-7 was replenished on day 2. On days 3 and 4, cell supernatants were replaced with retroviral supernatants resulting from cotransfection (Fugene-6; Roche) of BOSC23 cells with pCL-Eco and pMX-I-SceI-P2A-RAG2<sup>core</sup>-EGFP or pMX-I-SceI-EGFP plasmids 3 d before (Robbiani et al., 2008). Spinoculation was at 1,111 g for 1.5 h in the presence of 2.5 μg/ml polybrene, 5 ng/ml IL-7, and 20 mM Hepes. After 6–8 h at 37°C, on day 3, retroviral supernatants were replaced with original supernatants, whereas on day 4, cells were collected for IL-7 washout and replating in fresh com-
Cell culture for IC-Seq
Bone marrow B cells were isolated from tibias, femurs, and humeri of \( \text{ROSA}^{\text{SceI}} \)/\( \text{MyC}^{\text{J/J}} \)/\( \text{Igh}^{\text{J/J}} \)/\( \text{IgL}^{\text{J/J}} \)/\( \text{AI}^{\text{J/J}} \) mice at 6–8.5 mo of age by immunomagnetic enrichment with anti-B220 MicroBeads (Miltenyi Biotec). Cells were pooled and cultured at 2.0 \( \times \) 10^6 cells/ml in the presence of IL-7 (5 or 10 ng/ml; Sigma-Aldrich) and tamoxifen (1 µM; Sigma-Aldrich) in complete RPMI. On day 1, cells were collected for IL-7 washout and replated in fresh complete RPMI with 1 µM tamoxifen. On day 2, cultures were harvested, and cell pellets were snap-frozen on dry ice.

TC-Seq library preparation
TC-Seq libraries of \( \text{RAG2}^{\text{sce}} \) and \( \text{RAG2}^{\text{−/−}} \) pro-B cells were prepared in duplicates from each of 50 million sorted cells, as previously described (Klein et al., 2011; Robbiani et al., 2015) with the exception that sonication of genomic DNA was performed with Covaris S220 (power 105, duty factor 20%). Paired reads that both aligned to Myc at their 5' end were clipped from both ends of the reads, and the remaining sequences were mapped to the mouse genome (mm10) with SMALT (v0.7.6; parameters: -c 11 -x -O; Sanger Institute). Paired reads that both aligned to \( \text{MyC} \) at their 5' end were analyzed in “insertion mode” otherwise, they were processed in “rearrangement mode.”

In rearrangement mode (Figs. 1 and 2), bases aligning to \( \text{MyC} \) were clipped from either the beginning or the end of the reads, and the remaining sequences were mapped to the mouse genome (mm10) with SMALT (parameters: -O -r -1). Only alignments with at least 36 bp and a Phred score of 20 were accepted. Reads with the same sheared ends, which derive from sonication during library preparation, were merged into one event, and single reads were preserved. Rearrangements that did not yield breakpoints were discarded. Finally, reads that crossed the I-SceI site by >3 bp were excluded.

In insertion mode (Figs. 4 and 5), bases aligning to \( \text{MyC} \) were clipped from both ends of the reads, and the remaining sequences were mapped to the mouse genome (mm10) with SMALT (parameters: -O -r -1). Only alignments with at least 36 bp and a Phred score of 20 were accepted. Pairs with incorrect genomic orientation (+/+ and −/−) were excluded. The alignment of insertions yielded either both genomic breakpoints (double junctions) or only one (single junctions). Because of saturation at \( \text{MyC} \), events were merged if they possessed all of the following features: identical shears, genomic breakpoints within 5 bp, and same orientation. Events based on single reads were preserved. Finally, reads that crossed the I-SceI site by >3 bp were excluded.

IC-Seq analysis
Data from two independent libraries were pooled and analyzed similar to the “insertion mode” in TC-Seq, with minor
modifications (Fig. 6). Only genomic alignments with at least 25 bp and a Phred score of 20 were accepted. Insertions were merged if they possessed genomic breakpoints within 5 bp of each other and occurred in the same orientation. Finally, reads that crossed the I-SceI site by more than 3 bp were excluded.

**Analysis of rearrangements (TC-Seq) and insertions (IC-Seq)**

To characterize chromosomal rearrangements and insertions derived from distal regions (Figs. 1 D–F; and Fig. 4, D–F), the following portions of the genome were excluded: 50 or 20 kb surrounding the I-SceI site at Myc′ (rearrangements or insertions, respectively), 2 kb surrounding cryptic I-SceI sites (consensus [TCA][AT][GGGATA][AC][CAGG][GCT] [TC][ATCG][AG][TAC]), RAG2 (likely representing retroviral integrations), 3 Mb at each centromere, and chromosome M (mitochondrial DNA).

To determine the enrichment at genic regions (Figs. 1 D and 4 D), the portion of DNA from −2 kb of the most upstream transcription start site to the end of the last exon was considered as genic. For transcription analysis (Figs. 1 E and 4 E), RNA-seq data (Revilla-i-Domingo et al., 2012) were mapped with STAR aligner (v2.4.2a; default parameters; Dobin et al., 2013) using the mouse genome (mm10) and removing multiple alignments. Transcripts were quantified and annotated using cufflinks (v2.2.1); cuffdiff parameters: –upper-quartile-norm –dispersion-method per-condition; Trapnell et al., 2013) and Ensembl annotation (release 80). Annotation of V(D)J segments was based on Ensembl (release 84). D segments were excluded, and repeat regions were masked. Second, whole-genome sequences from published human cancer datasets (Table S4; Wang et al., 2011; Holmfeldt et al., 2013; Okosun et al., 2014) were mapped with IG/TCR baits as references. Third, paired reads aligning to the baits were mapped against the human genome (hg38) using bwa mem (v0.7.12-r1039; default parameters). Only alignments with a Phred score of at least 20 were accepted. Finally, reads containing junctions (chimeric alignments) were filtered to yield insertions that were then manually verified using Geneious (Kearse et al., 2012). The analysis of publicly available human cancer datasets was classified as exempt activity by the Rockefeller University Institutional Review Board.

**Deletion PCR assays**

Genomic DNAs of TC-Seq (RAG2\textsuperscript{core} and RAG2\textsuperscript{−/−}) and IC-Seq (RAG1/2 wild type) cultures were used for deletion PCR assays. Duplicates for RAG2\textsuperscript{core} and RAG2\textsuperscript{−/−} originated from cell cultures with modified conditions: control was infected with pMX-EGFP; cells were transferred onto irradiated S17 stroma cells after IL-7 was reagent on day 4 and depleted for 1.5 d. To detect small and rare deletion events, nested PCRs with a “poison” primer were performed (Edgley et al., 2002). For PCR, 100 ng (Jκ1/2, Jκ4/5) or 200 ng (Vκ3–1) genomic DNA was amplified in 20-μl reactions with HotStarTaq polymerase (QIAGEN). For PCRII, 1 μl of PCR I was used as template. For deletions at Jκ1/2, primers p195/p256/p258 (PCR I) and p196/p257 (PCR II) were used with the following conditions: PCR I, 95°C for 15 min; 30× (95°C for 45 s, 63°C for 45 s, and 72°C for 50 s); and 72°C for 5 min; PCR II, 95°C for 15 min; 30× (95°C for 45 s, 63°C for 45 s, and 72°C for 10 s); and 72°C for 5 min (Table S5). For deletions at Jκ4/5, primers p199/p205/p255 (PCR I) and p200/p206 (PCR II) were used with the same cycling conditions as for Jκ1/2 (Table S5). For deletions at Vκ3–1, primers p243/p244/p245 (PCR I) and p207/p210 (PCR II) were used with the following conditions: PCR I, 95°C for 15 min; 30× (95°C for 45 s, 63°C for 45 s, and 72°C for 5 s); and 72°C for 5 min; PCR II, 95°C for 15 min; 30× (95°C for 45 s, 63°C for 45 s, and 72°C for 20 s); and 72°C for 5 min (Table S5). PCR II products were separated
on 2% agarose gels stained with ethidium bromide. Fragments shorter than the expected size from the germline locus (Jk1/2: <592 bp, Jk4/5: <575 bp, and Vk3-1: <635 bp) were extracted (MACHEREY-NAGEL) and sequenced (Geneviz). Deletion products were confirmed using Geneious (Kearse et al., 2012).

V(D)J PCR assays
Genomic DNAs of TC-Seq (RAG2core and RAG2full-P2A-RAG2full-EGFP) cultures were used for V(D)J PCR assays. For RAG2full, cells were cultured as for TC-Seq but infected with pMX-I-SceI-P2A-RAG2full-EGFP. Duplicates originated from cell cultures with modified conditions: control was infected with pMX-EGFP; all cells were transferred onto irradiated S17 stroma cells after IL-7 washout on day 4 and depleted for 1.5 d. Semiquantitative V(D)J PCRs were performed as previously described (Schlissel et al., 1991; Dudley et al., 2003) with modifications: 100, 50, or 25 ng of template DNA was amplified in 20-μl reactions with HotStarTaq polymerase (QIAGEN). For V(D)J PCRs, primers p58/p96 (Dh-Jh PCR), p96/p98 (VhQS52-Djh PCR), and p305/p306 (Vκ-Jk PCR) were used with the following conditions: 95°C for 15 min; 32× (95°C for 45 s, 62°C for 45 s, and 72°C for 2 min); and 72°C for 5 min (Table S5). For control PCRs (MycI) primers p113/p114 were used with the following conditions: 95°C for 15 min; 32× (95°C for 45 s, 62°C for 45 s, and 72°C for 2 min); and 72°C for 5 min (Table S5). PCR products were separated on 1.5% agarose gels stained with ethidium bromide.

Accession numbers
The TC-Seq and IC-Seq sequencing data generated in this study can be accessed from the SRA database (SRP077983).

Online supplemental material
Fig. S1 describes the RAG2-expressing retroviruses used in this study. Fig. S2 characterizes RAG1/2core-dependent rearrangement breakpoint clusters at cRSSs. Fig. S3 shows the sequences of inserted IG/TCR fragments identified in human cancers. Table S1 lists features of RAG1/2core-dependent breakpoint clusters detected by TC-Seq. Table S2 summarizes the RIC score analysis of identified cRSSs. Table S3 displays sequences of inserted IG/TCR fragments identified in human cancers. Table S4 summarizes the insertion analysis of human cancers. Table S5 lists primers used in this study. Tables S1–S5 are Accession numbers

REFERENCES
Agrawal, A., Q.M. Eastman, and D.G. Schatz. 1998. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature. 394:744–751. http://dx.doi.org/10.1038/29457
Akamatsu, Y., R. Monroe, D.D. Dudley, S.K. Elkin, F. Gärtner, S.R. Talukder, Y. Takahama, F.W. Alt, C.H. Basing, and M.A. Oettinger. 2003. Deletion of the RAG2 C terminus leads to impaired lymphoid development in mice. Proc. Natl. Acad. Sci. USA. 100:1209–1214. http://dx.doi.org/10.1073/pnas.0237043100
Alt, F.W., Y. Zhang, F.-L. Meng, C. Guo, and B. Schwer. 2013. Mechanisms of programmed DNA lesions and genomic instability in the immune system. Cell. 152:417–429. http://dx.doi.org/10.1016/j.cell.2013.01.007
Arnal, S.M., A.J. Holub, S.S. Salus, and D.B. Roth. 2010. Non-consensus heptamer sequences destabilize the RAG post-cleavage complex, making ends available to alternative DNA repair pathways. Nucleic Acids Res. 38:2944–2954. http://dx.doi.org/10.1093/nar/gkp2152
Barlow, J.H., R.B. Faryahi, E. Callén, N. Wong, A. Malhowski, H.T. Chen, G. Gutierrez-Cruz, H.-W. Sun, P. McKinnon, G. Wright, et al. 2013. Identification of early replicating fragile sites that contribute to genome instability. Cell. 152:620–632. http://dx.doi.org/10.1016/j.cell.2013.01.006
Bredemeyer, A.L., G.G. Sharma, C.-Y. Huang, B.A. Helmink, L.M. Walker, K.C. Khor, B. Nuskey, K.E. Sullivan, T.K. Pandita, C.H. Basing, and B.P. Sleckman. 2006. ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. Nature. 442:466–470. http://dx.doi.org/10.1038/nature04866
Chaganti, S.R., P.H. Rao, W. Chen, V. Dyomin, S.C. Jhanwar, N.Z. Parsa, R. Dalla-Favera, and R.S.K. Chaganti. 1998. Deregulation of BCL6 in non-Hodgkin lymphoma by insertion of IGH sequences in complex translocations involving band 3q27. Genes Chromosom. Cancer. 23:328–336. http://dx.doi.org/10.1002/(SICI)1099-2264(199812)23:4<328::AID-GCC8>3.0.CO;2-M
Chatterji, M., C.L. Tsai, and D.G. Schatz. 2006. Mobilization of RAG-generated signal ends by transposition and insertion in vivo. Mol. Cell. Biol. 26:1558–1568. http://dx.doi.org/10.1128/MCB.26.4.1558-1568.2006
Clatworthy, A.E., M.A. Valencia, J.E. Haber, and M.A. Oettinger. 2003. V(D)J recombination and RAG-mediated transposition in yeast. Mol. Cell. 12:489–499. http://dx.doi.org/10.1016/S1097-2765(03)00305-8
Cousens, M.A., R.L. Wendland, L. Deriano, C.R. Lindsay, S.M. Arnal, and D.B. Roth. 2013. RAG2’s acidic hinge restricts repair-pathway choice to M.C. Nussenzweig. P.C. Rommel was supported by a fellowship of the German Academic Exchange Service (DAAD), and M.C. Nussenzweig is a Howard Hughes Medical Institute Investigator. The cancer datasets used in this study (dbGaP: phs000341.v2.p1 and phs000340.v3.p1; EBI: EGAS000001000399) were generated with the financial support of the National Cancer Institute, the St. Baldrick’s Foundation, Partners for Cures, the American Lebanese Syrian Associated Charities of St. Jude Children’s Research Hospital as part of the St. Jude/Washington University Pediatric Cancer Genome Project, Cancer Research UK, Bloodwise (now Leukemia and Lymphoma Research), and the Hungarian Scientific Research Fund (OTKA).

The authors declare no competing financial interests.

Author contributions: P.C. Rommel designed and performed experiments as well as data analysis, wrote the manuscript, and prepared the figures. T.Y. Oliveira designed and performed data analysis. M.C. Nussenzweig and D.F. Robbiani designed experiments and wrote the manuscript.

Submitted: 28 September 2016
Revised: 9 November 2016
Accepted: 12 December 2016
Curry, J.D., and M.S. Schlisel. 2008. The response to and repair of RAG-1-mediated DNA damage. *Nat. Rev. Mol. Cell Biol.* 9:1439–1450. http://dx.doi.org/10.1038/nrm2451

Rummel, M.R. 2016. Mechanisms of human lymphoid chromosomal translocations. *Nat. Rev. Cancer.* 16:387–398. http://dx.doi.org/10.1038/nrc.2016.40

Liu, Y., R. Subrahmanyan, T. Chakraborty, R. Sen, and S. Desiderio. 2007. A plant homeodomain in Rag-2 that binds hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. *Immunity.* 27:561–571. http://dx.doi.org/10.1016/j.immuni.2007.09.005

Mahowald, G.K., J.M. Barna, M.A. Mahowald, S. Kulkarni, A.L. Bredemeyer, C.H. Basing, and B.P. Sleckman. 2009. Aberrantly resolved RAG-mediated DNA breaks in Atm-deficient lymphocytes target chromosomal breakpoints in cis. *Proc. Natl. Acad. Sci. USA.* 106:18339–18344. http://dx.doi.org/10.1073/pnas.0902545106

Matthews, A.G.W., A.J. Kuo, S. Ramón-Maíques, S. Han, K.S. Champagne, D. Ivanov, M. Gallardo, D. Carney, P. Cheung, D.N. Ciccone, et al. 2007. RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature.* 450:1106–1110. http://dx.doi.org/10.1038/nature06431

McBlane, J.F., D.C. van Gent, D.A. Ramsden, C. Romeo, C.A. Cuomo, M. Gellert, and M.A. Oettinger. 1995. Cleaveage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell.* 83:387–395. http://dx.doi.org/10.1016/0092-8674(95)81011-6

Merelli, I., A. Guiffanti, M. Fabbrì, A. Cocito, L. Furia, U. Grazini, R.J. Bonnal, L. Milanesi, and F. McBlane. 2010. RSSSite: A reference database and prediction tool for the identification of cryptic Recombination Signal Sequences in human and murine genomes. *Nucleic Acids Res.* 38:W262–W267. http://dx.doi.org/10.1093/nar/gkq791

Messier, T.L., J.P. O'Neill, S.-M. Hou, J.A. Nicklas, and B.A. Finette. 2003. In vivo transposition mediated by V(D)J recombinase in human T lymphocytes. *EMBO J.* 22:1381–1388. http://dx.doi.org/10.1038/sj/emboj.cdlg137

Mikušković, M., Y.-F. Chou, V. Gigi, C.R. Lindsay, O. Shestova, S.M. Lewis, and D.B. Roth. 2015. Off-target V(D)J recombination drives lymphomagenesis and is escalated by loss of the Rag2 C terminus. *Cell Reports.* 12:1842–1852. http://dx.doi.org/10.1016/j.celrep.2015.08.034

Navarro, J.-M., A. Touzart, L.C. Pradel, M. Loosveld, M. Koubi, R. Fenouil, S. Le Noir, M.A. Majoobi, E. Morgado, C. Gregoire, et al. 2015. Site- and allele-specific polycomb dysregulation in T-cell leukemias. *Nat. Commun.* 6:6094. http://dx.doi.org/10.1038/ncomms7094

Neiditch, M.B., G.S. Lee, L.E. Huy, V.L. Brandt, and D.B. Roth. 2002. The V(D)J recombination efficiently cleaves and transposes signal joints. *Mol. Cell.* 9:871–878. http://dx.doi.org/10.1016/S1097-2765(02)00494-X

and promotes genomic stability. *Cell Reports.* 4:870–878. http://dx.doi.org/10.1016/j.celrep.2013.07.041

Cowell, L.G., M. Davia, T.B. Kepler, and G. Kelsoe. 2002. Identification and utilization of arbitrary correlations in models of recombination signal sequences. *Genome Biol.* 3:research0072.1. http://dx.doi.org/10.1186/gb-2002-3-12-research0072

Curry, J.D., and M.S. Schlisel. 2008. RAG2's non-core domain contributes to the ordered regulation of V(D)J recombination. *Nucleic Acids Res.* 36:5750–5762. http://dx.doi.org/10.1093/nar/gkn553

Curry, J.D., D. Schulz, C.J. Gudis, J.S. Danska, L. Nutter, A. Nussenzweig, and M.S. Schlisel. 2007. Chromosomal reinsertion of broken RSS ends during T cell development. *J. Exp. Med.* 204:2293–2303. http://dx.doi.org/10.1084/jem.20070583

Deriano, L., J.J. Chaumel, M. Coussens, A. Multani, Y. Chou, A.V. Alekseyenko, S. Chang, J.A. Skok, and D.B. Roth. 2011. The RAG2 C terminus suppresses genomic instability and lymphomagenesis. *Nature.* 471:119–123. http://dx.doi.org/10.1038/nature09755

Dobin, A., C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Iha, P. Batut, M. Chaisson, and T.J. Gingeras. 2013. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics.* 29:15–21. http://dx.doi.org/10.1093/bioinformatics/bts365

Dudley, D.D., J. Sekiguchi, C. Zhu, M.J. Sadofsky, S. Whitlow, J. DeVido, R.J. Dudley, D.D., J. Sekiguchi, C. Zhu, M.J. Sadofsky, S. Whitlow, J. DeVido, R.J.
Ru, H., M.G. Chambers, T.-M. Fu, A.B. Tong, M. Liao, and H. Wu. 2014. Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. *Nat. Genet.* 46:176–181. http://dx.doi.org/10.1038/ng.2856

Oliveira, T.Y., W. Resch, M. Jankovic, R. Casellas, M.C. Nussenzweig, and Okosun, J., C. Bödör, J. Wang, S. Araf, C.-Y. Yang, C. Pan, S. Boller, D. Cittaro, Nussenzweig, A., and M.C. Nussenzweig. 2010. Origin of chromosomal recombination events in early and late B lymphopoiesis. *EMBO J.* 29:176–181. http://dx.doi.org/10.1038/emboj.2012.155

Robbiani, D.F., A. Bothmer, E. Callen, B. Reina-San-Martin, Y. Dorsett, S. Difilippantonio, D.J. Bolland, H.T. Chen, A.E. Corcoran, A. Nussenzweig, and M.C. Nussenzweig. 2008. AID is required for the chromosomal breaks in c-myc that lead to c-myc/H3 histone methylated at both lysine-4 and arginine-2. *Proc. Natl. Acad. Sci. USA.* 105:1575–1582. http://dx.doi.org/10.1073/pnas.0709170104

Reddy, Y.V.R., E.J. Perkins, and D.A. Ramsden. 2006. Genomic instability due to V(D)J recombination-associated transposition. *Genes Dev.* 20:1575–1582. http://dx.doi.org/10.1101/gad.1432706

Revilla-i-Domingo, R., I. Bilic, B. Vilagos, H. Tagoh, A. Ebert, I.M. Tamin, L. Smeeck, J. Trupke, A. Sommer, M. Jaritz, and M. Buslinger. 2012. The B-cell identity factor Pax5 regulates distinct transcriptional programmes in early and late B lymphopoiesis. *EMBO J.* 31:3130–3146. http://dx.doi.org/10.1038/emboj.2012.155

Schlissel, M.S., L.M. Corcoran, and D. Baltimore. 1991. Virus-transformed pre-B cells show ordered activation but not inactivation of immunoglobulin gene rearrangement and transcription. *J. Exp. Med.* 173:711–720. http://dx.doi.org/10.1084/jem.173.3.711

Seikjusci, J.A., S. Whitlow, and F.W. Alt. 2001. Increased accumulation of hybrid V(D)J joins in cells expressing truncated versus full-length RAGs. *Mol. Cell.* 8:1383–1390. http://dx.doi.org/10.1016/S1097-2765(01)00423-3

Talukder, S.R., D.D. Dudley, F.W. Alt, Y. Takahama, and Y. Akamatsu. 2004. Increased frequency of aberrant V(D)J recombination products in core RAG-expressing mice. *Nucleic Acids Res.* 32:4539–4549. http://dx.doi.org/10.1093/nar/gkh778

Tan, J., K. Peiper, L. Piccoli, A. Abdii, M. Fogliarini, R. Geiger, C.M. Tully, D. Jarrossay, F.M. Ndungu, J. Wambua, et al. 2016. A LAIR1 insertion generates broadly reactive antibodies against malaria variant antigens. *Nature.* 529:105–109. http://dx.doi.org/10.1038/nature16450

Teng, G., Y. Maman, W. Resch, M. Kim, A. Yamane, J. Qian, K.-R. Kieffer-Kwon, M. Mandal, Y. J, E. Meffre, et al. 2015. RAG represents a widespread threat to the lymphocyte genome. *Cell.* 162:751–765. http://dx.doi.org/10.1016/j.cell.2015.07.009

Trapnell, C., D.G. Hendrickson, M. Sauvageau, L. Goff, J.L. Rinn, and L. Pachter. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* 31:46–53. http://dx.doi.org/10.1038/nbt.2450

Tsai, C.-L., M. Chatterji, and D.G. Schatz. 2003. DNA mismatches and GC-rich motifs target transposition by the RAG1/RAG2 transposase. *Nucleic Acids Res.* 31:6180–6190. http://dx.doi.org/10.1093/nar/gkg819

Vanura, K., B. Montpellier, T. Le, S. Spicuglia, J.-M. Navarro, O. Cabaud, S. Rouillard, E. Vachez, I. Prinz, P. Ferrier, et al. 2007. In vivo reinsertion of excised episomes by the V(D)J recombinase: A potential threat to genome stability. *PLoS Biol.* 5:e43. http://dx.doi.org/10.1371/journal.pbio.0050043

Wang, J., C.G. Mullighan, J. Easton, S. Roberts, J.A. Hettley, J. Ma, M.C. Rusch, K. Chen, C.C. Harris, L. Ding, et al. 2011. CREST maps somatic structural variation in cancer genomes with base-pair resolution. Nat. Methods. 8:652–654. http://dx.doi.org/10.1038/nmeth.1628

Wang, Q., T. Oliveira, M. Jankovic, I.T. Silva, O. Hakim, K. Yao, A. Gazuuny, C.T. Mayer, R. Pavli, R. Casellas, et al. 2014. Epigenetic targeting of activation-induced cytidine deaminase. *Proc. Natl. Acad. Sci. USA.* 111:18667–18672. http://dx.doi.org/10.1073/pnas.1405751111

West, K.L., N.C. Singhla, P. De Ioannes, L. Lacomis, H. Erdjument-Bromage, P. Tempst, and P. Cortes. 2005. A direct interaction between the RAG2 C terminus and the core histones is required for efficient V(D)J recombination. *Mol. Cell.* 19:7914–7921. http://dx.doi.org/10.1016/j.mcb.2005.10.055

Schatz, D.G., and Y. Ji. 2011. Recombination centres and the orchestration of V(D)J recombination. *Nat. Rev. Immunol.* 11:251–263. http://dx.doi.org/10.1038/nri2941

Schatz, D.G., and P.C. Swanson. 2011. V(D)J recombination: Mechanisms of initiation. *Annu. Rev. Genet.* 45:167–202. http://dx.doi.org/10.1146/annurev-genet-110410-132552

Yu, K., and M.R. Lieber. 2000. The nicking step in V(D)J recombination is independent of synopsis: Implications for the immune repertoire. *Mol. Cell.* 6:7914–7921. http://dx.doi.org/10.1128/MCB.20.21.7914-7921.2000

Zhao, L., R.L. Frock, Z. Du, J. Hu, L. Chen, M.S. Krangel, and F.W. Alt. 2016. Orientation-specific RAG activity in chromosomal loop domains contributes to T cell V(D)J recombination during T cell development. *J. Exp. Med.* 213:1921–1936. http://dx.doi.org/10.1084/jem.20160670
Correction: RAG1/2 induces genomic insertions by mobilizing DNA into RAG1/2-independent breaks
Philipp C. Rommel, Thiago Y. Oliveira, Michel C. Nussenzweig, and Davide F. Robbiani
Vol. 214, No. 3, March, 2017. https://doi.org/10.1084/jem.20161638

The authors regret that in the original version of their paper, the following funding information was not included. This text has been added to the second paragraph of their Acknowledgments section:

The cancer datasets used in this study (dbGaP: phs000341.v2.p1 and phs000340.v3.p1; EBI: EGAS00001000399) were generated with the financial support of the National Cancer Institute, the St. Baldrick’s Foundation, Partners for Cures, the American Lebanese Syrian Associated Charities of St. Jude Children’s Research Hospital as part of the St. Jude/Washington University Pediatric Cancer Genome Project, Cancer Research UK, Bloodwise (now Leukemia and Lymphoma Research), and the Hungarian Scientific Research Fund (OTKA).

In addition, a reference was not included in the reference list. This correction affects Materials and methods section Analysis of insertions in human tumors and Table S4. The corrected section and full reference appear below. Table S4, an online Excel file, has been replaced.

Analysis of insertions in human tumors
We designed a novel pipeline to search whole-genome sequences for insertions derived from IG/TCR loci. First, IG/TCR baits were generated that correspond to regions spanning 150 bp upstream and downstream from each physiological RSS cleavage site of human V and J segments (Ensembl, release 84). D segments were excluded, and repeat regions were masked. Second, whole-genome sequences from published human cancer datasets (Table S4; Wang et al., 2011; Holmfeldt et al., 2013; Okosun et al., 2014) were mapped with bwa mem (v0.7.12-r1039; default parameters) using the IG/TCR baits as references. Third, paired reads aligning to the baits were mapped against the human genome (hg38) using bwa mem (v0.7.12-r1039; default parameters). Only alignments with a Phred score of at least 20 were accepted. Finally, reads containing junctions (chimeric alignments) were filtered to yield insertions that were then manually verified using Geneious (Kearse et al., 2012). The analysis of publicly available human cancer datasets was classified as exempt activity by the Rockefeller University Institutional Review Board.

Wang, J., C.G. Mullighan, J. Easton, S. Roberts, S.L. Heatley, J. Ma, M.C. Rusch, K. Chen, C.C. Harris, L. Ding, et al. 2011. CREST maps somatic structural variation in cancer genomes with base-pair resolution. Nat. Methods. 8:652–654. http://dx.doi.org/10.1038/nmeth.1628

These errors have been corrected in the online HTML and PDF versions. The errors remain only in the print version.