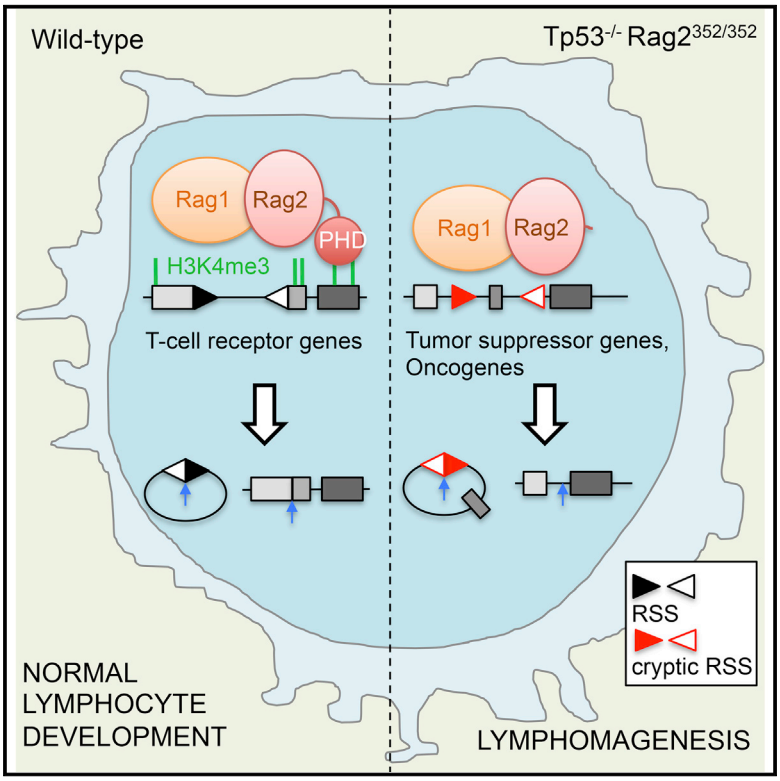
Article

Cell Reports

Off-Target V(D)J Recombination Drives Lymphomagenesis and Is Escalated by Loss of the Rag2 C Terminus

Graphical Abstract



Highlights

- V(D)J recombination errors are investigated in lymphomas from p53-deficient mice
- Numerous off-target rearrangements are found scattered across tumor genomes
- These rearrangements target known and suspected oncogenes and tumor suppressors
- Loss of Rag2 C terminus increases the frequency of off-target V(D)J recombination

Authors

Martina Mijušković, Yi-Fan Chou, Vered Gigi, ..., Olga Shestova, Susanna M. Lewis, David B. Roth

Correspondence

susle@mail.med.upenn.edu (S.M.L.), david.roth2@uphs.upenn.edu (D.B.R.)

In Brief

To investigate the role of V(D)J recombination errors in oncogenesis, Mijuskovic et al. analyzed whole genomes of lymphomas spontaneously arising in p53-deficient mice. The systematic analysis revealed numerous off-target, RAG-mediated rearrangements in known and suspected cancer genes, which were escalated by the loss of Rag2 C terminus.

Mijušković et al., 2015, Cell Reports *12*, 1842–1852
September 22, 2015
http://dx.doi.org/10.1016/j.celrep.2015.08.034

OPEN ACCESS
CellPress

Cell Reports
Article

Off-Target V(D)J Recombination Drives Lymphomagenesis and Is Escalated by Loss of the Rag2 C Terminus

Martina Mijušković,[1,2,3] Yi-Fan Chou,[1,2] Vered Gigi,[1,2,4] Cory R. Lindsay,[1,2,5] Olga Shestova,[1,2,6] Susanna M. Lewis,[1,2,7,*] and David B. Roth[1,2,7,*]

[1]Department of Pathology and Laboratory Medicine, Raymond and Ruth Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA
[2]Abramson Family Cancer Research Institute, Raymond and Ruth Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA
[3]Department of Genetics and Epidemiology, The Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK
[4]The Boston Consulting Group, 1735 Market Street, Philadelphia, PA 19103, USA
[5]Cardeza Foundation for Hematological Research, Thomas Jefferson University, 394 Jefferson Alumni Hall, 1020 Locust Street, Philadelphia, PA 19104, USA
[6]Translational Research Program, Abramson Family Research Cancer Institute, University of Pennsylvania, Philadelphia, PA 19104, USA
[7]Co-senior author
*Correspondence: susle@mail.med.upenn.edu (S.M.L.), david.roth2@uphs.upenn.edu (D.B.R.)
http://dx.doi.org/10.1016/j.celrep.2015.08.034

SUMMARY

Genome-wide analysis of thymic lymphomas from *Tp53*$^{-/-}$ mice with wild-type or C-terminally truncated Rag2 revealed numerous off-target, RAG-mediated DNA rearrangements. A significantly higher fraction of these errors mutated known and suspected oncogenes/tumor suppressor genes than did sporadic rearrangements ($p < 0.0001$). This tractable mouse model recapitulates recent findings in human pre-B ALL and allows comparison of wild-type and mutant RAG2. Recurrent, RAG-mediated deletions affected *Notch1*, *Pten*, *Ikzf1*, *Jak1*, *Phlda1*, *Trat1*, and *Agpat9*. Rag2 truncation substantially increased the frequency of off-target V(D)J recombination. The data suggest that interactions between Rag2 and a specific chromatin modification, H3K4me3, support V(D)J recombination fidelity. Oncogenic effects of off-target rearrangements created by this highly regulated recombinase may need to be considered in design of site-specific nucleases engineered for genome modification.

INTRODUCTION

Antigen receptor genes are assembled by controlled, large-scale deletions and inversions (Little et al., 2015). Errors in this process, known as V(D)J recombination, produce aberrant genomic rearrangements, which can fuel the development of T- and B-cell malignancies (Onozawa and Aplan, 2012). Normal V(D)J recombination entails double-strand DNA cleavage at pairs of gene segments by the heteromeric Rag1 and Rag2 endonuclease (hereinafter referred to as RAG). RAG recognizes conserved recombination signal sequences (RSSs) positioned adjacent to coding segments that are subsequently joined to form immunoglobulin (Ig) and T-cell receptor (TCR) variable region exons (Figure 1A). Enzymatic steps in joining are carried out by general DNA repair functions.

Mechanisms underlying aberrant V(D)J recombination comprise three general categories according to outcome and/or genetic dependencies. One is off-target cleavage (Figure 1B) at sequences fortuitously resembling an RSS (i.e., a cryptic RSS, or cRSS). A second type of mistake may occur at the point at which DNA ends are joined (Figure 1C). A third involves interchromosomal (*trans*) recombination (Figure 1D). Such errors, which can be compounded, leave distinctive traces in the structure of the junctions they produce in the genomes of lymphoid neoplasms (Marculescu et al., 2006; Onozawa and Aplan, 2012). A stunning example of a V(D)J recombination gone wrong was provided by recent genome-wide sequence analysis of pediatric acute lymphocytic leukemia (ALL), which revealed tumor genomes peppered with structural rearrangements caused by off-target V(D)J recombination (Papaemmanuil et al., 2014).

Studies in mice and in cultured cells indicate that multiple strategies direct RAG to appropriate loci in appropriate cells, but the degree to which these also serve to prevent aberrant genome-wide V(D)J recombination has not been determined. Possible failsafes include regulation of *Rag1* and *Rag2* expression, attenuation of RAG activity during S-phase, chromatin features that define possible recombination sites (transcription-associated accessibility, locus contraction, histone H3 methylation, etc.), a vigilant DNA damage response, and the RSS specificity of RAG endonuclease itself (Little et al., 2015). Two features in the Rag2 C-terminal domain have been explored at the molecular level. One is a specific phosphorylation site, targeted by cyclin A/CDK2, that confines RAG activity to the G0/G1 phase of the cell cycle. Another, also located in the C terminus, is a plant homeodomain-like domain (PHD) that specifically interacts with

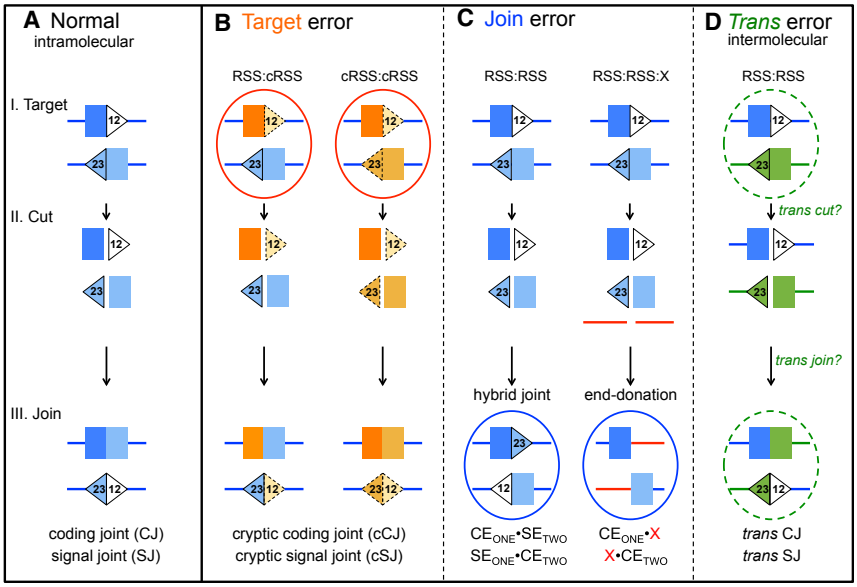


Figure 1. Normal and Aberrant V(D)J Recombination

(A) Normal V(D)J recombination. RAG binds recombination signal sequence (RSS) motifs, mediates synapsis, and cleaves the DNA target. Double-strand breaks occur 5′ of the RSS. Two coding ends (CE) lose and gain nucleotides in interim processing events and then join as a coding joint (CJ). Signal ends are ligated into a signal joint (SJ) with no or minimal interim modification. Products may contain junctional inserts (not diagrammed). GC-rich, "N regions," are nucleotide additions that may occur in either joint. Palindromic "P nucleotides" are found in CJs only.

(B–D) V(D)J recombination errors. (B) A cRSS-like element (cRSS; orange symbols) may join to an RSS or to a second cRSS. Off-target cleavage gives rise to otherwise normal recombination products, yielding a cCJ and cSJ. (C) Alternatively, errors can occur at the joining stage. Coding and signal ends may be swapped in error to form "hybrid joints". Alternatively, RAG-cleaved ends can be erroneously connected to a third incidental break (end-donation; gapped red line). (D) Joining in *trans* (between two chromosomes) is a third distinguishable type of aberration and can occur along with recognition or joining errors. "*trans cut*?" and "*trans join*?" reflect that *trans* events could occur at either stage. A reciprocal "type I" translocation is diagrammed (reviewed in Marculescu et al., 2006).

H3K4me3, inducing a structural change in RAG that relieves a negative autoregulation of its activity (Lu et al., 2015).

Engineered mouse models bearing mutant *Rag2* alleles implicate the Rag2 C-terminal domain in curtailing the oncogenic potential of V(D)J recombination. A T490A mutation at the phosphorylation site, as well as either partial ($Rag2^{FS361}$) or complete ($Rag2^{352}$) deletion of the *Rag2* C terminus (to amino acids 361 and 352s respectively), contributes to lymphomagenesis when assayed in a p53-deficient background (Deriano et al., 2011; Gigi et al., 2014; Zhang et al., 2011). $Rag2^{T490A}$ $Rag2^{FS361}$ and $Rag2^{352}$, $Tp53^{-/-}$ lymphomas exhibit chromosomal translocations cytogenetically mapping near *Ig* and *TCR antigen receptor* loci (Deriano et al., 2011; Gigi et al., 2014; Zhang et al., 2011). Such translocations are rarely, if ever, seen in lymphomas from $RAG2^{wt/wt}$ $Tp53^{-/-}$ mice (Deriano et al., 2011; Zhang et al., 2011).

To investigate the role of Rag2's C terminus in suppressing oncogenic V(D)J recombination events, we analyzed structural genome variations in lymphomas from either $Rag2^{wt/wt}$, $Tp53^{-/-}$ or $Rag2^{352/352}$, $Tp53^{-/-}$ mice. The $Rag2^{352}$ mutation causes a significantly more rapid onset of lymphomagenesis than any of the other aforementioned *Rag2* mutations in the $Tp53^{-/-}$ background. This experimental model has several advantages over genome-wide characterizations of human tumors. It is possible to collect independent tumors from animals of the same genotype. Irrelevant "passenger" mutations are disfavored by the very short interval—within the first 14 weeks of life—in which tumors arise. Mouse studies avoid potentially confounding effects of therapy, which can be very difficult to control when examining human tumors. Additionally, there is a large body of genome-wide information on chromatin modification, Rag2 binding, and DNA accessibility as applied to purified subsets of primary developing T cells, a level of characterization not yet available for humans.

Analysis of nine thymic lymphoma genomes yielded 275 validated somatic structural rearrangements. Most chromosome translocations near *Ig* and *TCR* loci lacked evidence of V(D)J recombination error (Figure 1). Furthermore, the rearrangements showed no apparent connection to lymphomagenesis. In contrast, there were numerous deletions arising from off-target V(D)J recombination, many of which mutated genes with a known or strongly supported role in oncogenesis. It was evident that this type of aberrant V(D)J recombination is an inherent risk of the antigen receptor gene rearrangement process being observed in both $Rag2^{352}$ as well as wild-type Rag2-expressing mice. Results with the experimentally tractable model broadly recapitulate observations in human B-ALL. Furthermore, the distribution of off-target breakpoints relative to histone H3K4me3-differentiated chromatin suggests that a known Rag2-mediated negative autoregulation of RAG activity deters the endonuclease from generating oncogenic rearrangements.

RESULTS

Deletion Is the Predominant Structural Variant Class in p53-Deficient Lymphomas

Through genome-wide analyses, we compared the somatic structural variants (SVs) in T-cell lymphomas from $Tp53^{-/-}$ mice to those from $Rag2^{352/352}$, $Tp53^{-/-}$ double mutants (Experimental Procedures; tumor phenotypes are given in Figures S1A and S1B). Tumors were sequenced with an average coverage of 14×. 57 somatic SVs (range = 8–20) were identified in four $Tp53^{-/-}$ lymphomas, and 218 SVs (range = 23–68) were detected in five $Rag2^{352/352}$, $Tp53^{-/-}$ lymphomas (Figure 2). Breakpoints mapped to scattered positions throughout the genome (Figure 2). Over half of the junctions were deletions, the remaining being divided among inversions, translocations, and apparent duplications (Figure S1D).

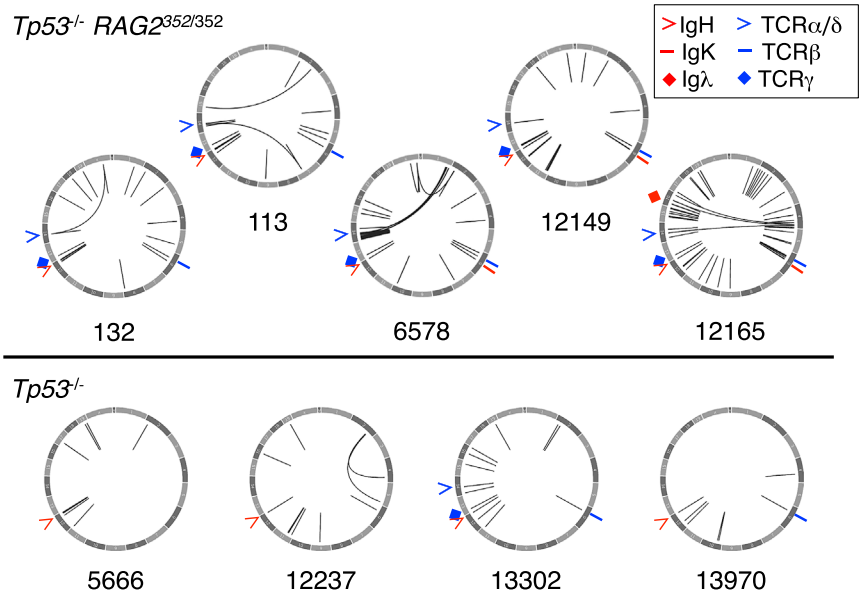


Figure 2. Structural Alterations in Thymic Lymphoma Genomes

Circos plot representation of all rearrangements (including normal *Ig/TCR* junctions) detected in the genome-wide analyses. *Ig* and *TCR* loci are marked by symbols. Translocations (seen as arcs) are more prevalent and show a tendency to localize in the region of *Ig/TCR* genes in $Rag2^{352/352}$, $Tp53^{-/-}$ tumors.

See also Figures S1 and S4.

Normal and Abnormal Rearrangements at *TCR* and *Ig* Loci

Among the 275 validated somatic rearrangements, 76 authentic V(D)J junctions were recovered (Figure S1; Table S1). Some junctions were supported by as few as two paired reads, confirming the high sensitivity of our analysis pipeline (Figure S1; Table S2). About 15% of the *Ig* or *TCR* locus recombinants were composed of abnormal J-to-J junctions; junctions between V gene segments and "locus-deleting elements"; incomplete (out-of-order) V-D recombination; directly joined Vδ-to-Jδ junctions (skipping D); and, in the case of the only detected signal joint, an atypical loss of sequence from the RSS of both signal ends (Table S1). Each type of abnormality is documented in wild-type, untransformed thymocytes. Rearrangements involving authentic V(D)J gene segments provided an internal control for various comparisons made in this study.

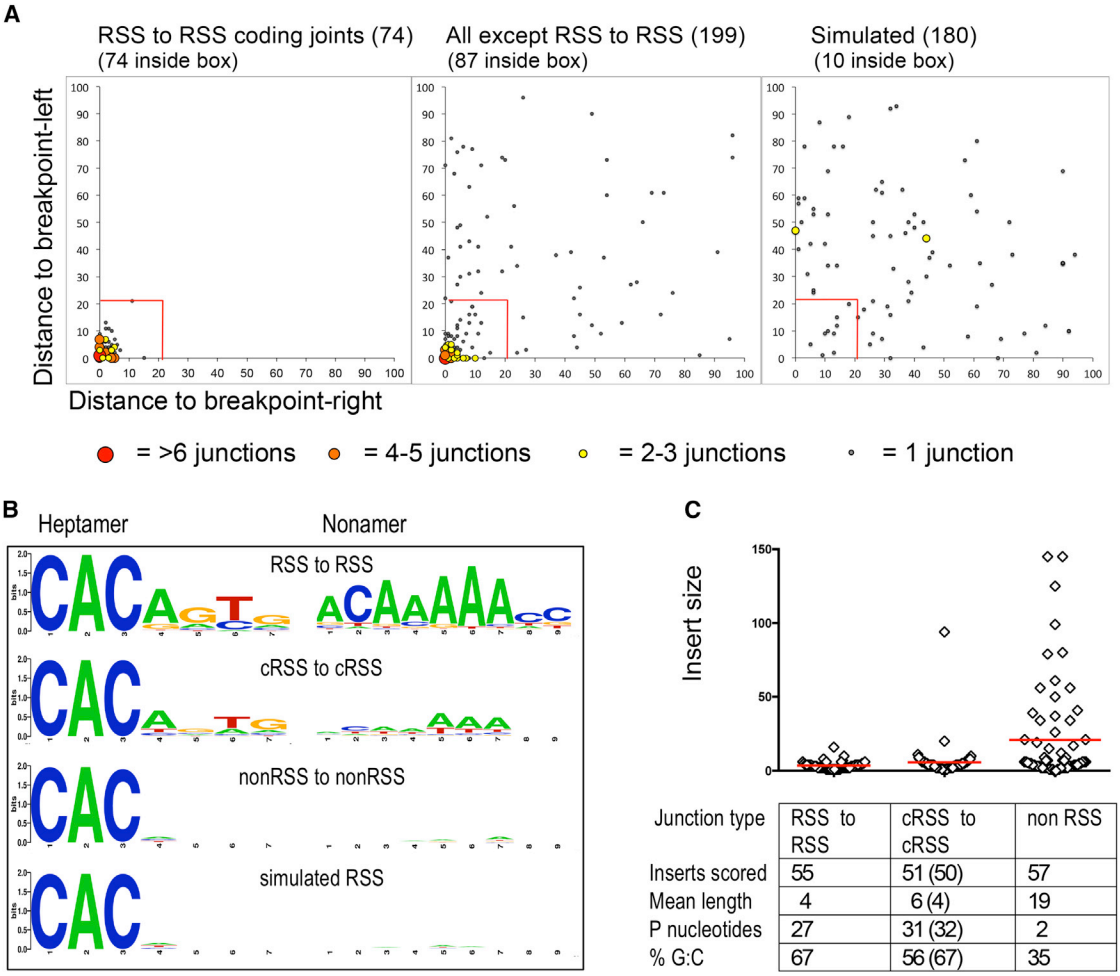
Off-Target V(D)J Recombination

To date, available systematic methods for identifying cRSS either are of unproved accuracy or use an algorithm of limited availability; nonetheless, an ability to identify (and quantify) off-target V(D)J recombination errors is a key necessity for the present study. We formulated and tested two simple criteria. First, a CAC must exist to the right (or GTG to the left) of both breakpoints (a point addressed more fully later), and, second, it must occur within a specified distance from the breakpoint. The conserved CAC of an RSS heptamer demarcates the position at which RAG makes its cut, and the distance between the CAC motif and a breakpoint reflects the limited number of base pairs that may be lost when an end is processed prior to ligation. For coding joints, the CAC distance-to-breakpoint value was set at 21 bp, the maximum observed for authentic RSS junctions in the data (Figure 3A). When each is plotted according to their two distance-to-breakpoint values, authentic V(D)J junctions fell into a tight cluster averaging 3 nt from either CAC motif. When the remaining 198 SVs from our dataset were similarly plotted, two distinct distributions emerged: junctions with CAC distance-to-breakpoint values like authentic recombination products and those with a wider, randomly distributed set of distances. The presence of two populations in our experimental data was confirmed by analyzing a set of 180 simulated junctions (see the Supplemental Experimental Procedures), which reconstructed only the broadly scattered distribution (Figure 3A).

Sequence logos (Crooks et al., 2004) were used to visually compare (1) the RSSs targeted in *Ig* and *TCR* junctions, (2) the cRSSs of cryptic junctions identified according to the CAC-to-breakpoint value, (3) the analogous "nonRSSs" of all junctions falling outside the 21-bp CAC-to-breakpoint range, and (4) the identically defined "simRSSs" of the simulated junction set (Figure 3B). Authentic targeted RSS elements gave the best match to the canonical RSS (CACAGTG [12/23 spacer] ACAAAAACC), most consistently at functionally important positions (Little et al., 2015). Similarly, the consensus sequence of the presumptive cRSSs reconstructed the canonical RSS except for the last two (CC) of the nonamer. There was a strong signal for the heptamer sequence beyond CAC and a signal at the functionally important nonamer positions 5, 6, and 7 (Little et al., 2015). The logos of the nonRSS and simRSS groups lacked these significant features.

A second test of the CAC distance-to-breakpoint method evaluates the cRSS junction group for two hallmark features of V(D)J recombination. These are "N regions," short G/C-rich insertions generated by terminal deoxynucleotidyl transferase and palindromic (P) nucleotides. These arise when hairpin-terminated RAG cleavage intermediates are nicked a few base pairs from the terminus and the opened ends happen to escape trimming before becoming joined. In both respects, cRSS junctions compared very well to the authentic V(D)J junctions (Figure 3C). In contrast, the insertions seen in nonRSS junctions were AT rich and more variable in size. P nucleotides were identified at the same frequency among cRSS as authentic V(D)J junctions, while almost no qualifying inserts were seen in the nonRSS group. These data indicate that the CAC distance-to-breakpoint criteria, indeed, identify off-target V(D)J recombination products.

The CAC distance-to-breakpoint method was largely in agreement with the computational approach (Papaemmanuil et al.,



Junction type	RSS to RSS	cRSS to cRSS	non RSS
Inserts scored	55	51 (50)	57
Mean length	4	6 (4)	19
P nucleotides	27	31 (32)	2
% G:C	67	56 (67)	35

Figure 3. Application of a "CAC-Distance-to-Breakpoint" Method to Identify Off-Target, RAG-Mediated "Coding Joint-like" Junctions

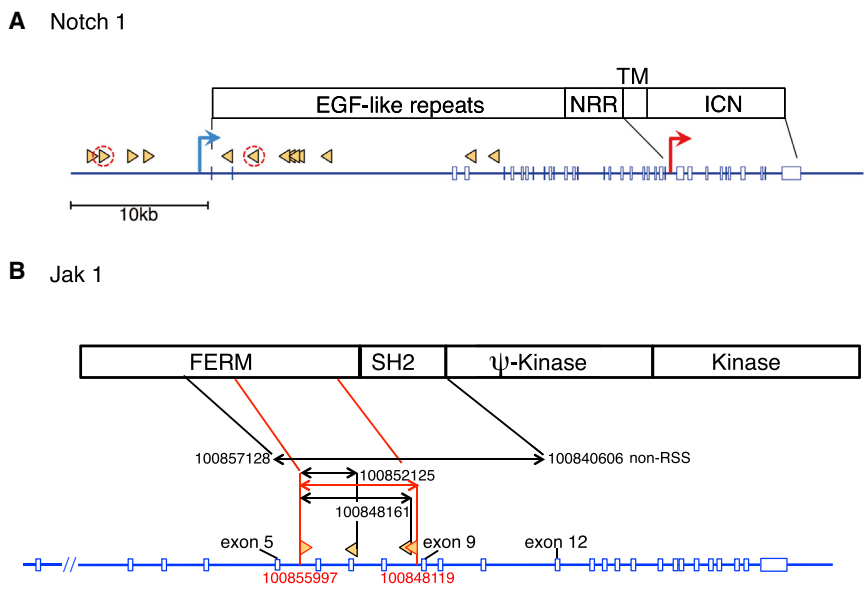
(A) CAC-distance-to-breakpoint values distinguish two populations of junctions in T-cell lymphomas. Left: tumor V(D)J coding joints are plotted according to CAC distance-to-breakpoint values. Middle: a plot of CAC distance-to-breakpoint values for all junctions other than authentic V(D)J coding joints shows a clustered and dispersed pattern. Right: a CAC-distance-to-breakpoint plot for a set of simulated junctions. The boxed area in each panel indicates the 21-bp cutoff used to discriminate between cRSS and nonRSS junctions as tested in (B) and (C). Numbers at the top refer to the number of junctions plotted within each set (junctions with distance to breakpoint values exceeding 100 bp on one or both sides are not shown). The 74 RSS-to-RSS junctions represent all the coding joints generated by V(D)J recombination, excluding1 hybrid joint and 1 signal joint. The "all except RSS-to-RSS junctions" set includes all other junctions (275 total junctions minus 74 coding joints, 1 hybrid joint, and 1 signal joint).

(B) The CAC-distance-to-breakpoint method identifies junctions with and without extended matches to the canonical RSS sequence. Top: V(D)J junctions in tumors, as seen in (A), left panel, occur near RSS motifs that, as expected, exhibit canonical heptamer/nonamer sequences. The group of provisionally identified cRSS junctions, shown in the red boxed area in the middle panel of (A), reveals a consensus sequence that matches 11 of 13 canonical heptamer and nonamer positions. Putative cRSSs from nonRSS and simulated junctions (outside the red boxed area, including those not shown) do not resemble a canonical RSS. Some bias toward nonamer matches is introduced by the method of nonamer assignment (Supplemental Experimental Procedures) but would not artifactually enhance identities at functionally important positions. The CAC defining the heptamer start in all categories is shown for clarity, not for purposes of comparison.

(C) cRSS junctions as defined by the CAC-distance-to-breakpoint method have other, independent properties in common with V(D)J junctions. Junctional insertions are seen in RSS, cRSS, and nonRSS junctions. cRSS junction inserts have restricted length and high G:C composition, as well as P nucleotides, a unique feature of V(D)J recombination—similar to inserts observed in RSS junctions. Inserts in nonRSS junctions do not recapitulate these properties. Values in parentheses exclude an extreme outlier: length exceeding $Q3_{cRSS} + ([MAX_{RSS} - Q3_{RSS}] \setminus IQR_{RSS}) \cdot IQR_{cRSS}$. IQR, interquartile range.

See also Supplemental Experimental Procedures.

2014). In their study, 44 junctions were assigned with high confidence to off-target V(D)J recombination, 39 of which were also identified by our criteria (Table S3). For 214 junctions that were determined not to be cRSS related in the same study, there was agreement on 185—again, a good concordance (86%). The cryptic junctions detected by the distance-to-breakpoint method in the nonRSS group appear to be properly assigned because the sequence logo was well matched to the canonical nonamer sequence. The third cRSS discovery method, based on RIC (recombination information content) scores (Cowell



C cRSS signal joints

	location, mm9	nonamer	cryptic RSS end heptamer		(N)	cryptic RSS end heptamer		nonamer	location
Jak1	**chr4:100855997**	CTTTTTTGT	12	CACTGTG	TC	CACAGTG	23	CCTCTTCAA	**100848119**
		CTTTTTTGT	12	CACTGTG		CACAGTG	23	CCTCTTCAA	
		CTTTTTTGT	12	CACTGTG		CACAGTG	23	CCTCTTCAA	
Agpat9	**chr5:101283990**	GGTGTATGC	23	CTCTGTG	C	CACAGTA	12	CATATAACT	**101252196**
		GGTGTATGC	23	CTCTGTG	CCC	•••••••	23	CATATAACT	
Trat1	**chr16:48805735**	TGTTTCTGT	12	CACAGTG		CACAGAG	23	GCAAAATGT	**48770795**
		TGTTTCTGT	12	CACAGTG	CCACAA	CACAGAG	23	GCAAAATGT	
		TGTTTCTGT	12	CACAGTG		CACAGAG	23	GCAAAATGT	
		TGTTTCTGT	12	CACAG••		CACAGAG	23	GCAAAATGT	

Figure 4. Recurrent cRSS-Mediated Deletions

(A) Notch1 activation via cRSS-mediated recombination. Top: diagram of the Notch1 protein indicating relevant domains. Middle: representation of the *Notch1* locus with lines locating the corresponding translated regions in the protein diagram. Triangles indicate cRSSs identified in the present study as well as candidate cRSSs reported by Tsuji et al. (2004) (see Supplemental Experimental Procedures for inclusion details). The most commonly observed deletion removes the 5′ promoter (blue arrow), causing an internal promoter to be used (red arrow). The resulting protein lacks the extracellular epidermal growth factor 1 (EGF1)-like repeats, and the negative regulatory region (NRR) but retains most of the transmembrane (TM) domain. The truncated form is thought to insert into the membrane and become cleaved in a ligand-independent fashion, releasing constitutively activated, intracellular Notch1 (ICN). All other *Notch1* deletions likewise deleted the 5′ promoter. See Table S2 for junction sequences.

(B) Internal *Jak1* deletions suggest another instance of oncogene activation through cRSS-mediated recombination. A diagram of domains within the Jak1 protein is displayed above a representation of the locus. Individual cRSS-mediated deletions are shown as horizontal arrows. cRSS coordinates (corresponding to the heptamer border, mm9 genome assembly) are given for each deletion. All parts of the diagram in red indicate the most frequent cRSS deletion in *Jak1*. Clustering of the deletions suggests that deletions in this region may be activating (see Discussion; for sequences, see Figure S2).

(C) Predicted reciprocal cryptic signal joints can be detected in tumor DNA samples by PCR. Sequences of independently isolated cryptic signal joints from recurrent *Jak1*, *Agpat9*, and *Trat1* rearrangements. For evidence of ongoing recombination in thymoma, see Figure 3.

See also Supplemental Experimental Procedures, Figures S2 and S4, and Table S2.

et al., 2001), did not perform as robustly. Only 12 (31%) of the 39 concordant junctions—i.e., off-target V(D)J recombination products with strong support—bore a cRSS at both breakpoints according to RIC scores, and, individually, over half of the 39 cRSSs failed the RIC test (Table S3). Overall, our test and that of Papaemmanuil et al. (2014) appear equally discriminatory, given that the percent coincidence for both positive and negative assignments is about the same.

We note, however, that because the sensitivity of our method is based on the presence of a cRSS on both sides of the junction, it is not suited to the detection of "one-sided" events (which are predicted for end-donation errors; Figure 1B). Nevertheless, we found no evidence that end-donations could constitute more than a minor fraction of aberrant V(D)J recombination products. End-donations suffering extensive resection after cleavage would not have been detected by our analysis. Moreover, many one-sided junctions in the collection of Papaemmanuil et al. (2014) qualified as two-sided products according to the CAC distance-to-breakpoint test (Table S3).

Off-Target V(D)J Recombination Causes Recurrent Deletions in *Notch1* and *Jak1*

Mutations causing constitutive activation of the *Notch1* oncogene have been demonstrated in both human and mouse T-cell neoplasms. In the latter, recurrent cRSS-mediated deletions are found that produce a Notch1 protein activated via N-terminal truncation (Ashworth et al., 2010; Jeannet et al., 2010; Van Vlierberghe and Ferrando, 2012). The two most commonly observed *Notch1* deletions (Tsuji et al., 2004) were also seen in our dataset, along with additional examples (Figure 4A; Figure S2). Our results, along with data from Tsuji et al. (2004), identify 14 functional cRSSs in a 30-kb span of the *Notch1* locus (Figure 4A). This high density of functional cRSSs agrees with predictions based on artificial recombination substrates (Lewis et al., 1997).

PCR primers specific for each *Notch1* cRSS defined in the present study were used to assay for additional deletions that, due to coverage, may not have been detected in the genome-wide analysis. These were indeed seen, giving a total of eight

Tumor ID

Gene	Rag2^{352} 132	113	6578	12149	12165	Rag2-wt 5666	12237	13302	13970	
Notch1	O	✔	✔	O	✔✔✔	O✗		✔		Multiple Tumors
Jak1	O	✔	✔	✔✔	✔✔✔✔✗				O	
Pten	✔				✔		✔	✗✗✗		
Ikzf1				✔✔			✗✗			
Fhit			✗✗✔					✗		
Phlda1			OO		✔					
Nr3c1	✗				✔					
Agpat9	✔	O			O					
Trat1			OO		O		✔	O		
Akt1	✗✗✗									One Tumor Only
Gata3					✔					
Met							✔			
Pvt1					✔					
Runx1	✔									
Rarb			✗							
Tnc					✔					
Eya4					✔					
Kif1b					✔					
Lamc1					✔					
Siva1	✗									
Arhgap2	✗									
Cmah		✗								
Tnfaip8				✔						
Ano6			✗							
Rcor1				✔						
Cdh13	✔									
Epha1					✗					
Eya4					✔					
Kdm1a					✔					
Nell2					✔					
Pag1	✔									
Rock2					✔					
Tnc					✔					
Fmnl3					✔					
Angpt1					✔					

Key: ✔=cRSS ✗=nonRSS O=cRSS junction was detected by PCR only

Cosmic Cancer Gene | Likely Cancer gene | Unlikely Cancer Gene

Figure 5. Recurrently Disrupted Genes and Cancer-Associated Genes Are Found in Tumors of Both Genotypes

Genes mutated in tumors from different mice are listed above the thick line. The tumor harboring the detected mutations is indicated at the top of each column. Check marks and other symbols are defined in the key. There were no clustered, recurrent breakpoints shared between tumors occurring outside a gene. A comprehensive list of cancer gene mutations, whether multiply or singly observed in this study is given. Yellow boxes indicate genes that occur among the 572 currently listed in the COSMIC Cancer Gene Census (http://cancer.sanger.ac.uk/census). Green boxes indicate genes with a likely cancer association. wt, wild-type.

See also Table S4.

distinct *Notch1* deletions (Figure S2), distributed between all five *RAG2*$^{352/352}$, *Tp53*$^{-/-}$ tumors and two of the *Tp53*$^{-/-}$-only isolates (Figure 5). All deletions encompassed the 5′ region of the gene, deleting the 5′ promoter, as is the case for cRSS-mediated mutations previously demonstrated to result in constitutive Notch1 activation (Ashworth et al., 2010; Jeannet et al., 2010).

We also identified recurrent, cRSS-mediated rearrangements at a second known oncogene, *Jak1*. One specific recurrent deletion (8.2 kb) eliminated exons 6 through 8 (Figure 4B; Figure S2) and was found in every *Rag2*$^{352/352}$ thymic lymphoma, as well as in one *Rag2*$^{w/wt}$ tumor, by PCR. These *Jak1* deletions share several properties with the activating *Notch1* deletions. They recur in independent tumors, involve multiple sets of cRSSs, overlap one another over a specific area of the locus, and may provide for production of a modified (rather than absent) protein product. The similarities raise the possibility that, as for Notch1, these cRSS-mediated mutations might activate Jak1.

cRSS-Mediated Deletions and Candidate Driver Mutations

cRSS junctions affected genes with a potential oncogenic impact in that they govern cell proliferation, differentiation, or survival. Of these, 21 specific deletions were selected for PCR analysis in the hope of discovering additional recurrently mutated genes. We examined the nine original tumors and six additional *Rag2*$^{352/352}$ p53$^{-/-}$ tumors, revealing recurrent, cRSS-mediated deletions at three more loci: *Trat1*, *Phlda1*, and *Agpat9* (Figure S2). All recurrent mutations, except the *Phlda1* deletion, were verified in both *Tp53*$^{-/-}$, *RAG2*$^{352/352}$ and *Tp53*$^{-/-}$-only lymphomas (Figure 5). Recurring examples of the same cRSS junction in independent tumors suggest the possibility that, as in *Notch1*, these specific gene disruptions are driver mutations.

If off-target V(D)J recombination is a force in the oncogenic process, it should be possible to detect enrichment of cRSS junctions among SVs that interrupt genes of significance in cancer. "Cancer genes" were defined by a metric based on the co-occurrence of search terms in the Web of Science Core Collection (http://apps.webofknowledge.com/) (Experimental Procedures; Table S4). Almost 90% of cRSS junctions in the collection interrupted a gene, and well over half of these genes (65%) were associated with cancer (for additional details, see

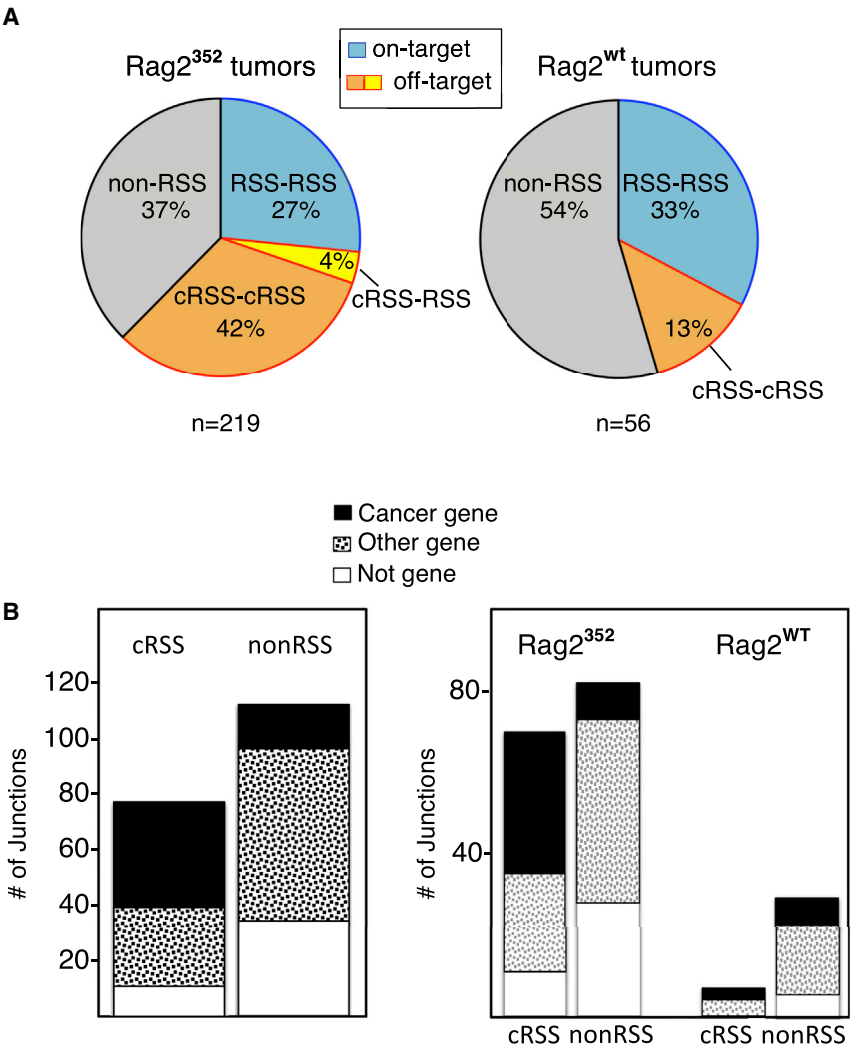


Figure 6. Overview and Role of Off-Target V(D)J Recombination

(A) SVs generated by either RSS, cRSS or nonRSS mechanisms in RAG2^{352} versus RAG2^{wt} tumors. Off-target V(D)J joints (cRSS-cRSS and cRSS-RSS), on-target RSS-RSS V(D)J, and non-RSS junctions are indicated by colors given in the key. cRSS-mediated events account for 46% of the structural genomic alterations in the *RAG2*$^{352/352}$, *Tp53*$^{-/-}$ tumors, but this is the case for only 15% of rearrangements in cRSS-mediated in thymic tumors.

(B) The role of cRSS recombination in tumor formation. The y axis in each plot represents junction numbers. Black, stippled, and white sections indicate junctions in cancer genes, non-cancer (Other) genes, and non-genes, respectively, for all junctions except those involving an RSS (RSS-RSS and RSS-cRSS). Assignment of cancer gene association is given in Table S4. There is a highly significant difference between cRSS versus nonRSS mechanisms with respect to observed mutations in cancer-associated genes ($p < 0.0001$). A further breakdown of the data in (B), right panels, represents the impact of the Rag2 C terminus RAG2 carboxyl-terminal domain with respect to three types of rearrangements. *RAG2*$^{352/352}$ tumors show a bias toward cRSS-mediated cancer gene mutations relative to non-RSS mechanisms that is absent in RAG2 wild-type (WT) tumors (compare the four black sections in each of the four bars, right panel; the p value for the difference is 0.0044).

See Tables S4 and S5.

Supplemental Experimental Procedures; Table S4). By comparison, 70% of nonRSS junctions had breakpoints located within a gene, and only 24% of these disrupted cancer genes, a highly significant difference ($p < 0.0001$; Table S5). These data, summarized in Figure 6, support the possibility that off-target V(D)J recombination contributes to oncogenesis.

To investigate whether the Rag2 C terminus reduces oncogenic mutation, the degree to which cRSS versus nonRSS junctions interrupt cancer genes was further compared between *Tp53*$^{-/-}$, *RAG2*$^{352/352}$ and *Tp53*$^{-/-}$-only tumors. The fraction of cancer genes that were "hit" by cRSS recombination with Rag2^{352} was significantly different than that observed with intact Rag2 ($p = 0.0044$; Figure 6B; Table S5C).

Excision Products Containing cRSS Signal Joints Indicate Recent Recombination Activity

V(D)J recombination co-generates reciprocal products, a coding joint, and a signal joint. When a coding joint is created by deletion, the signal joint will occur on an excised circular DNA molecule (Fujimoto and Yamagishi, 1987) (Figure 1A). The circular excision products, dubbed "Trecs" (*TCR* rearrangement excision circles) do not replicate and are progressively diluted by cell division. We sought evidence that cryptic Trecs (cTrecs) might be generated by off-target recombination. To maximize the possibility of detection, we focused the effort on recurrent cRSS rearrangements. Excised signal joints reciprocal to the *Jak1*, *Phlda*, and *Trat1* cryptic coding joints were recovered by PCR (Figure 4C).The demonstration of cTrecs corresponding to chromosomal cRSS rearrangement is a new observation. By analogy to Trecs, it appears that off-target V(D)J rearrangements occurred relatively recently in the history of the rapidly dividing tumor, indicating ongoing recombination activity. (Identification of cTrecs corresponding to potential oncogenic events does not, however, mean that they are related to founder mutations; rather, it reflects the evolving, multiclonal nature of the lymphomas, as noted in the Discussion.) Consistent with this possibility, RAG transcripts were detected in several tumors (Figure S3). Furthermore, a high level of V(D)J recombination was detected with a standard extrachromosomal recombination assay in a cell line established from a Rag2$^{352/352}$, Tp53$^{-/-}$ tumor (Figure S3). The presence of specific cTrecs could conceivably provide an additional parameter for phenotyping human tumors as well as be an indicator of recent mutation.

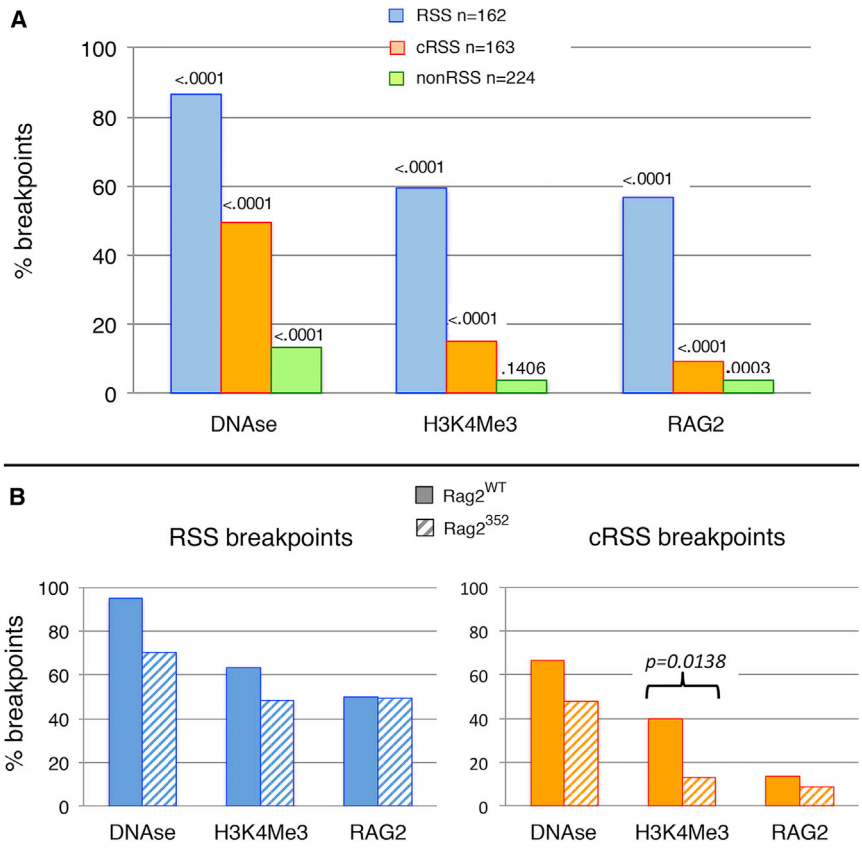


Figure 7. Truncation of the Rag2 C Terminus Disrupts a Correlation between cRSS Breakpoints and H3K4me3

(A) Distribution of breakpoints (blue, yellow, and green indicate RSS, cRSS, and nonRSS, respectively) relative to published genome-wide determinations of DNase I hotspots (left), H3K4me3 chromatin immunoprecipitation sequencing (ChIP-seq) analysis of a genetically isolated DP thymocyte population (middle), and RAG2 ChIP-seq analysis, same population (right). Bar heights represent the percentage of breakpoints in DNase I hotspots, tri-methylated H3K4 peaks, or RAG2 binding peaks. Correlations between breakpoint categories and chromatin features are significantly different from a random distribution as indicated for the evaluated RSS and cRSSs. The correlation between nonRSS breakpoints and DNase I hotspots reached significance, but nonRSS breakpoints were not markedly different from random with respect to H3K4me3 or RAG2 binding peaks. (B) Comparisons made between breakpoints in p53 null tumors with wild-type (WT; solid bars) versus C-terminally truncated RAG2. Left: blue bar graphs give the percentage of authentic RSS breakpoints co-localized with accessibility, H3K4me3 modification, and RAG binding. No statistically significant differences are seen between genotypes. Right: orange bars indicate the percentage of cRSS breakpoints co-localized with accessibility, H3K4me3 modification, and Rag2 binding. The cRSS breakpoints have a significantly reduced coincidence with H3K4me3 in $RAG2^{352}$ tumors when compared to the cRSS breakpoints in $RAG2^{wt}$ tumors (two-tailed Fisher's exact test). None of the other comparisons reach significance. See also Table S5.

Chromosomal cRSS Signal Joints Are Surprisingly Rare

If, as expected, cRSS targets are randomly oriented with respect to one another in the genome, recombination between any pair of cRSSs should be as likely to generate a signal joint as a coding joint. Rearrangement of a pair of cRSSs that happen to be oriented "heptamer to heptamer" produces a cryptic signal joint that remains in the chromosome. Surprisingly, chromosomal coding joints outnumbered chromosomal signal joints 6- to 7-fold in the tumor genomes (Tables S1 and S3). A marked deficit was also seen in data from human *ETV6-Runx1* ALL (Papaemmanuil et al., 2014): we found only one cryptic signal joint among 140 candidate junctions (Table S3). Indeed, very few signal joints representing off-target V(D)J recombination products have been reported thus far (e.g., Mendes et al., 2014). The rarity of cryptic signal joints in the tumor genomes is not due to a failure in their production, because as described earlier, cTrecs are found corresponding to *Trat1*, *Agpat9*, and *Jak1* rearrangements. It is not obvious what the chromosomal cSJ deficit means, but the consistency with which it is observed suggests that further investigation could be revealing.

Genomic Context of Functional cRSSs

Our data establish firmly that off-target V(D)J recombination occurs in tumors bearing wild-type RAG1 and RAG2, underscoring the importance of mechanisms that counteract recognition errors during V(D)J rearrangements. Such errors might be minimized by chromatin/RAG interactions. Three relevant chromatin properties have been measured at the whole-genome level in primary mouse cell populations corresponding to CD4+ CD8+ T-cell precursors or to unfractionated thymocytes (most of which are CD4+ CD8+ pre-T cells). The properties measure accessibility, as reflected by DNase I sensitivity (Sabo et al., 2006), methylation of lysine 4 in histone H3 (Ji et al., 2010; Vigano et al., 2014), and the location of chromatin-bound Rag2 (Ji et al., 2010).

The location of each 39-bp interval (a length that accommodates both 12- and 23-bp versions of a RAG target) corresponding to 162 RSSs, 163 cRSSs, and 224 nonRSSs (Table S6) in the thymic tumors was determined. The relationship of the target motif and the position of H3K4me3-modified chromatin is not known, so this choice minimized imprecision introduced by taking a larger interval to both sides of the breakpoint. The percentages of RSS targets overlapping DNase I-sensitivity hotspots, H3K4me3-modified chromatin, or Rag2-binding regions are shown in Figure 7B. Even in the case of authentic RSSs, slightly over half lay in chromatin bearing all three features, and 14% of RSS breakpoints occupied positions that lacked any of the properties. This is likely due to the fact that the distribution of marks is measured in aggregate and, thus, may not reflect minor subpopulations. Similarly, the degree to which these aggregate

measurements reflect the situation in a tumor cell at the time of cRSS rearrangement cannot be determined.

Despite these caveats, cRSS breakpoints exhibited a non-random distribution relative to DNase I hotspots, H3K4me3-modified chromatin, and RAG2 binding ($p < 0.0001$; Table S5E). Localization of cRSS breakpoints to the aforementioned markers may define the state of genomic regions susceptible to off-target V(D)J recombination. Alternatively, considering that all three properties, including Rag2 binding (Ji et al., 2010), are directly or indirectly associated with active gene expression, a more trivial possibility is that the distribution of cRSS targets nonspecifically reflects DNA target accessibility.

Because $Rag2^{352}$ lacks the PHD domain implicated in recognition of H3K4me3 marks, the location of breakpoints in relation to this chromatin mark was further examined. Comparing cRSS sites in wild-type Rag2, Tp53/ tumors with those in arising in $Rag2^{352/352}$, Tp53/ revealed no significant differences with respect to colocalization of authentic RSS targets and H3K4me3 (Figure 7). In contrast, the distribution of cRSS targets with respect to H3K4me3 was significantly different ($p = 0.0007$). Thus, C-terminally truncated RAG2 appears markedly less constrained by the presence of H3K4me3 than its wild-type counterpart.

Most Translocation Junctions Reveal No Apparent Link to V(D)J Recombination

In 22 translocation SVs, we detected 23 translocation junctions (one SV contained two junctions). Genome-wide SV analysis confirmed earlier findings that translocations with breakpoints near *TCR* loci were present in $Rag2^{352/352}$, $Tp53^{-/-}$ lymphomas and not the $Tp53^{-/-}$-only counterparts (Deriano et al., 2011) (Figure 2). Between the two BAC probes used in our previous study (Deriano et al., 2011) we identified a total of five junctions in two $RAG2^{352/352}$, $Tp53^{-/-}$ tumors (#132 and #12165). Of these, only one can be, without a doubt, classified as an off-target V(D)J rearrangement. This is the translocation between chr4 and chr14, creating a signal joint composed of the RSS of a *TCR* Vα gene segment (chr14) and a cRSS within a *Kdm1a* locus (chr4) (Figure S4). In the case of a simple *trans* joining error, a coding joint should be present on a reciprocal derivative chromosome, and this was, indeed, demonstrated by PCR of the tumor DNA (Figure S4). The 4:14 exchange illustrated that *trans* joining errors likely play a role in tumor formation, because there is abundant evidence that mutation of *Kdm1a*, a histone deacetylase gene, is significant in leukemia (Harris et al., 2012). This was in contrast to the other translocations, of which none interrupted a cancer-associated gene (Table S4). Furthermore, we found no evidence of RSS or cRSS in these junctions, suggesting that another mechanism is responsible. These results raise the question of whether apparent locus-specific structural rearrangements in other $Tp53^{-/-}$ mouse models can be attributed to V(D)J recombination, as few recombinant junctions have been analyzed at the DNA sequence level.

DISCUSSION

The genomic lesions described here bear striking similarities to those recently reported in pediatric *ETV6-Runx1* B-ALL (Papaemmanuil et al., 2014). Common to both is a preponderance of deletions, many resulting from off-target V(D)J recombination. In both cases, RAG-mediated deletions between pairs of cRSSs create known and suspected driver mutations. *trans* events involving V(D)J recombination are much less frequent, and in both studies, misrecognition of RSSs is seen to occur in tumors with wild-type RAG. In human *ETV6-Runx1* B-ALL, initiating events are thought to occur prenatally, but tumors were not available for investigation until they reached, on average, 57 months of age (Papaemmanuil et al., 2014). By comparison, $Rag2^{352/352}$, *Tp53* null mice die of thymic lymphomas at a median age of 13 weeks. The status of the genome in the murine system provides as close a reflection of the "smoking guns" in the oncogenic process as is currently possible.

The paucity of chromosome translocations involving antigen receptor loci in our dataset is striking, given our previous identification of such translocations in thymic lymphomas from $Rag2^{352/352}$, *Tp53* null mice (Deriano et al., 2011). This discrepancy likely arises because the translocations detected by cytogenetic techniques in our earlier work, while recurrent, were not uniformly present in all tumors and, in many cases, were present in minor subpopulations (Deriano et al., 2011; see Table S1). Thus, those translocations—which, based on their frequency, are unlikely to represent driver mutations—might not have been present in the tumor samples analyzed here or were at a low enough abundance to have escaped detection by whole-genome sequencing.

Rag2 C Terminus-H3K4me3 Interactions May Enhance the Fidelity of V(D)J Recombination

Comparison of genomic lesions in lymphomas from $Rag2^{wt/wt}$, $Tp53^{-/-}$ and $Rag2^{352/352}$, $Tp53^{-/-}$ mice revealed that, although off-target V(D)J recombination events are found in both genotypes, they are much more prevalent when Rag2 is C-terminally truncated. cRSSs identified in $Rag2^{352/352}$, $Tp53^{-/-}$ tumors occur significantly more frequently outside H3K4me3-modified regions of the genome (as measured in normal pre-T cells) (Ji et al., 2010; Vigano et al., 2014) than those in $Tp53^{-/-}$ mice. The significance of this bias is underscored by the fact that comparisons of a less specific "accessibility" (DNase I sensitivity and Rag2 binding) show inconsistent trends between genotypes and fail to reach statistical significance. These data establish a mechanistic link between the PHD domain located in RAG2's C terminus, H3K4me3-modified chromatin, and suppression of off-target V(D)J recombination. The possibility of such a link has been suggested most recently in a biochemical study showing that RAG autoinhibition was relieved when the PHD finger of the Rag2 C terminus was allowed to interact with H3K4me3 peptides (Lu et al., 2015). The observations reported here support a role for H3K4me3-mediated reversal of autoinhibition in minimizing the oncogenic impact of faulty RSS recognition.

Numerous Rearrangements Lacking Identifiable Features of V(D)J Recombination Junctions in $Rag2^{352/352}$ Rumors

Curiously, all classes of SV, whether arising from authentic V(D)J recombination, cRSS rearrangement, or apparently nonRSS-mediated accidents (sporadic deletions, inversions, and duplications included) were quantitatively increased in $Rag2^{352/352}$,

Tp53$^{-/-}$ tumors. We considered several possible explanations for this observation. An overall increase in SVs might be caused by the loss of the T490 phosphorylation site. Rag-mediated breaks arising inappropriately during S phase could induce DNA repair functions that are error prone when handling other types of damage. Also, if coding ends bearing unresolved hairpin termini are abnormally available during S phase, they may be replicated, resulting in the generation of DNA palindromes. Such sequences are very labile and are known to provoke large-scale rearrangements in mice, humans, and yeast (Coté and Lewis, 2008). Another possible explanation for an increase in apparently non-specific sporadic junctions is that these may actually arise from some yet-unrecognized form of aberrant V(D)J recombination. For example, if uncontrolled, extensive resection of cleaved ends is possible (on a kilobase-to-megabase scale), this would obliterate any signature of off-target RAG cleavage in any resulting junction.

Off-Target V(D)J Recombination Generates Known and Candidate Driver Mutations

Where mutation is repeatedly discovered in the same gene in independent tumors, it is usually taken as evidence of biological selection in the pathological process. If, however, V(D)J recombination is the mutation mechanism, a relationship is less certain. The repeated observation of a cryptic junction could be due to properties that fortuitously favor recombination rather than biological selection. Whereas there is little question that *Notch1*, *Jak1*, *Ikzf1*, *Fhit*, and *Pten*—all included in the short, expertly curated list of 572 genes in the COSMIC Cancer Gene Census (http://cancer.sanger.ac.uk/census/)—are drivers in the oncogenic process, the four remaining recurrently mutated genes (*Nr3c1*, *Trat1*, *Phlda1*, and *Agpat9*) occupy a gray zone.

To address this, we developed a system for stratifying all the genes mutated in the present study according to cancer relevance (Experimental Procedures; Table S4). By this approach, *Nr3c1* and *Phlda1* belong to the same group as the aforementioned genes and can also be considered cancer associated. *Agpat9* (Table S4B) and *Trat1*, however, have modest or no support for a cancer role in the literature (Table S4C). Although *Agpat9* and *Trat1* are low-rated cRSS rearrangements, it may be significant that both operate in pathways that are connected to oncogenesis. Trat1 (TCR-associated protein 1) is upstream in the generation of PIP3, which has regulatory interactions with both Akt1 and Pten. The Pten pathway is evidently an Achilles heel, given that *Pten* itself is recurrently mutated by cRSS mechanisms here (Figure 1) and in human T-ALL (Mendes et al., 2014). Agpat9 is a 1-acylglycerol-3-phosphate o-acyltransferase involved in the production of phosphatidic acid, a cofactor in the AKT1/Tor pathway (Agarwal, 2012). It remains to be seen whether cRSS-mediated deletions pinpoint new cancer genes, but given the potential applications, this question bears further investigation.

Potential Mechanism for Jak1 Activation

Analyzing off-target mutations may enhance understanding of cancer development mechanisms. It has been demonstrated that cRSS rearrangements at *Notch1* cause the production of a truncated, activating form of the protein (Ashworth et al., 2010; Jeannet et al., 2010). It would appear, given the number of similarities between the two, that the *Jak1* mutations seen here, like the deletions at *Notch1*, may be activating. Activated Jak1, caused by point mutations or other small changes that increase protein stability, is seen in both human and murine lymphoid tumors (Chen et al., 2012; Bellanger et al., 2014; Flex et al., 2008). The commonly observed 8.2-kb cRSS-mediated deletion here is compatible with the possibility of an in-frame splice between exons 5 and 9 in the mRNA. This would delete a large part of the FERM domain (Figure 4), eliminating functions thought, at present, to be a prerequisite for mediating the oncogenic effects of Jak1 activation.

Are RAG-Mediated Off-Target Events Relevant to Nuclease-Mediated Genome Editing?

Suppressing off-target cleavage is an exceptionally challenging problem in the V(D)J recombination system and, as such, may inform strategies for increasing the accuracy of genome editing in clinical applications. During V(D)J recombination, several layers of regulation converge to limit RAG action at inappropriate target sites. Such measures include the tempo with which T-cell intermediates traverse the vulnerable developmental stage (Haines et al., 2006), the destruction of Rag2 upon entry into S phase of the cell cycle, and the physical accessibility of the underlying cRSS sequences upon transcription, histone modification, and chromatin looping (Little et al., 2015). Additionally, highlighted here is the possible importance of chromatin-modulated Rag2 auto-inhibition in the suppression of off-target activity. Finally, it goes without saying that a highly important requirement for fidelity is an intact p53 pathway. These safety nets, the product of millions of years of evolution, counteract the potential for disaster inherent in a process that occurs in cells with high proliferative ability throughout life. In the short term, it may not be a simple task to computationally predict the unintended targets of manmade genome editing nucleases or to reduce their numbers to a safe level by engineering ever more stringent enzymes. The strategies used in the V(D)J recombination system may inform these efforts.

EXPERIMENTAL PROCEDURES

Mouse husbandry and experimentation were reviewed and approved by the University of Pennsylvania Institutional Animal Care and Uses Committee. Whole-genome tumor and normal paired-end libraries were sequenced and analyzed with a custom analysis pipeline (Mijušković et al., 2012). Candidate SVs that passed all filters were tested for tumor specificity by PCR (Table S2). The DNA sequence of all 275 confirmed junctions was then determined by Sanger sequencing. Detailed methods, including downstream analysis, identification of cRSS junctions, creation of arbitrary junctions, and stratification of cancer genes, are presented in the Supplemental Experimental Procedures.

Analysis of the genomic context of breakpoints in Figure 7 was based on publicly accessible genome-wide DNase I hypersensitivity data (wgEncodeUwDgfThymusC57bl6MAdult8wksHotspotsRep1), H3K4me3 peaks (GSM530317), and RAG2 binding peaks (GSM530316) (Ji et al., 2010; Sabo et al., 2006).

To visually confirm overlaps relative to H3K4me3 peaks (Ji et al., 2010), the track was uploaded to the University of California, Santa Cruz (UCSC) browser along with WIG (wiggle track format) files of genome-wide data from early T-cell precursors (Vigano et al., 2014) (GSM1164638, GSM1164639, GSM1164643, and GSM1164644). There was close agreement at the monitored breakpoints, despite the use of different cell purification methods and

different sources for H3K4me3 antibodies. The p values were calculated using a two-tailed Fisher's exact test.

ACCESSION NUMBERS

The accession number for the data reported in this paper is SRA:SRP043567.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2015.08.034.

AUTHOR CONTRIBUTIONS

M.M. and D.B.R. conceived the study. M.M. designed the experiments. Y.-F.C. isolated tumor cells and performed FACS analysis. M.M. and Y.-F.C. prepared libraries, conducted the bioinformatics and cRSS analysis, validated candidate junctions, and performed qPCR. V.G. assisted with the *Notch1* analyses and FACS profiling. C.R.L. prepared libraries. S.M.L. guided and performed cRSS analysis. O.S. isolated tumor cells and assisted with validation. M.M., S.M.L., and D.B.R. wrote the paper.

ACKNOWLEDGMENTS

Funding: NIH grants CA-104588 and PN1EY018244 to D.B.R. and 1R21AI097825-01 to S.M.L. We thank Ellen Rothenberg (Caltech), and David Schatz (Yale University) for making alternatively formatted data available for analysis. SRA: SRP043567.

Received: May 26, 2015
Revised: July 7, 2015
Accepted: August 8, 2015
Published: September 10, 2015

REFERENCES

Agarwal, A.K. (2012). Lysophospholipid acyltransferases: 1-acylglycerol-3-phosphate O-acyltransferases. From discovery to disease. Curr. Opin. Lipidol. *23*, 290–302.

Ashworth, T.D., Pear, W.S., Chiang, M.Y., Blacklow, S.C., Mastio, J., Xu, L., Kelliher, M., Kastner, P., Chan, S., and Aster, J.C. (2010). Deletion-based mechanisms of Notch1 activation in T-ALL: key roles for RAG recombinase and a conserved internal translational start site in Notch1. Blood *116*, 5455–5464.

Bellanger, D., Jacquemin, V., Chopin, M., Pierron, G., Bernard, O.A., Ghysdael, J., and Stern, M.H. (2014). Recurrent JAK1 and JAK3 somatic mutations in T-cell prolymphocytic leukemia. Leukemia *28*, 417–419.

Chen, E., Staudt, L.M., and Green, A.R. (2012). Janus kinase deregulation in leukemia and lymphoma. Immunity *36*, 529–541.

Coté, A.G., and Lewis, S.M. (2008). Mus81-dependent double-strand DNA breaks at in vivo-generated cruciform structures in S. cerevisiae. Mol. Cell *31*, 800–812.

Cowell, L.G., Davila, M., Kepler, T.B., and Kelsoe, G. (2001). Identification and utilization of arbitrary correlations in models of recombination signal sequences. Genome Biol. *3*, RESEARCH0072.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res. *14*, 1188–1190.

Deriano, L., Chaumeil, J., Coussens, M., Multani, A., Chou, Y., Alekseyenko, A.V., Chang, S., Skok, J.A., and Roth, D.B. (2011). The RAG2 C terminus suppresses genomic instability and lymphomagenesis. Nature *471*, 119–123.

Flex, E., Petrangeli, V., Stella, L., Chiaretti, S., Hornakova, T., Knoops, L., Ariola, C., Fodale, V., Clappier, E., Paoloni, F., et al. (2008). Somatically acquired JAK1 mutations in adult acute lymphoblastic leukemia. J. Exp. Med. *205*, 751–758.

Fujimoto, S., and Yamagishi, H. (1987). Isolation of an excision product of T-cell receptor alpha-chain gene rearrangements. Nature *327*, 242–243.

Gigi, V., Lewis, S., Shestova, O., Mijušković, M., Deriano, L., Meng, W., Luning Prak, E.T., and Roth, D.B. (2014). RAG2 mutants alter DSB repair pathway choice in vivo and illuminate the nature of 'alternative NHEJ'. Nucleic Acids Res. *42*, 6352–6364.

Haines, B.B., Ryu, C.J., and Chen, J. (2006). Recombination activating genes (RAG) in lymphoma development. Cell Cycle *5*, 913–916.

Harris, W.J., Huang, X., Lynch, J.T., Spencer, G.J., Hitchin, J.R., Li, Y., Ciceri, F., Blaser, J.G., Greystoke, B.F., Jordan, A.M., et al. (2012). The histone demethylase KDM1A sustains the oncogenic potential of MLL-AF9 leukemia stem cells. Cancer Cell *21*, 473–487.

Jeannet, R., Mastio, J., Macias-Garcia, A., Oravecz, A., Ashworth, T., Geimer Le Lay, A.S., Jost, B., Le Gras, S., Ghysdael, J., Gridley, T., et al. (2010). Oncogenic activation of the Notch1 gene by deletion of its promoter in Ikaros-deficient T-ALL. Blood *116*, 5443–5454.

Ji, Y., Resch, W., Corbett, E., Yamane, A., Casellas, R., and Schatz, D.G. (2010). The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. Cell *141*, 419–431.

Lewis, S.M., Agard, E., Suh, S., and Czyzyk, L. (1997). Cryptic signals and the fidelity of V(D)J joining. Mol. Cell. Biol. *17*, 3125–3136.

Little, A.J., Matthews, A., Oettinger, M., Roth, D.B., and Schatz, D.G. (2015). The mechanism of V(D)J recombination. In Molecular Biology of B Cells, F.W. Alt, T. Honjo, A. Radbruch, and M. Reth, eds. (Academic Press), pp. 13–34.

Lu, C., Ward, A., Bettridge, J., Liu, Y., and Desiderio, S. (2015). An autoregulatory mechanism imposes allosteric control on the V(D)J recombinase by histone H3 methylation. Cell Rep. *10*, 29–38.

Marculescu, R., Vanura, K., Montpellier, B., Roulland, S., Le, T., Navarro, J.M., Jäger, U., McBlane, F., and Nadel, B. (2006). Recombinase, chromosomal translocations and lymphoid neoplasia: targeting mistakes and repair failures. DNA Repair (Amst.) *5*, 1246–1258.

Mendes, R.D., Sarmento, L.M., Canté-Barrett, K., Zuurbier, L., Buijs-Gladdines, J.G., Póvoa, V., Smits, W.K., Abecasis, M., Yunes, J.A., Sonneveld, E., et al. (2014). PTEN microdeletions in T-cell acute lymphoblastic leukemia are caused by illegitimate RAG-mediated recombination events. Blood *124*, 567–578.

Mijušković, M., Brown, S.M., Tang, Z., Lindsay, C.R., Efstathiadis, E., Deriano, L., and Roth, D.B. (2012). A streamlined method for detecting structural variants in cancer genomes by short read paired-end sequencing. PLoS ONE *7*, e48314.

Onozawa, M., and Aplan, P.D. (2012). Illegitimate V(D)J recombination involving nonantigen receptor loci in lymphoid malignancy. Genes Chromosomes Cancer *51*, 525–535.

Papaemmanuil, E., Rapado, I., Li, Y., Potter, N.E., Wedge, D.C., Tubio, J., Alexandrov, L.B., Van Loo, P., Cooke, S.L., Marshall, J., et al. (2014). RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. Nat. Genet. *46*, 116–125.

Sabo, P.J., Kuehn, M.S., Thurman, R., Johnson, B.E., Johnson, E.M., Cao, H., Yu, M., Rosenzweig, E., Goldy, J., Haydock, A., et al. (2006). Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. Nat. Methods *3*, 511–518.

Tsuji, H., Ishii-Ohba, H., Katsube, T., Ukai, H., Aizawa, S., Doi, M., Hioki, K., and Ogiu, T. (2004). Involvement of illegitimate V(D)J recombination or microhomology-mediated nonhomologous end-joining in the formation of intragenic deletions of the Notch1 gene in mouse thymic lymphomas. Cancer Res. *64*, 8882–8890.

Van Vlierberghe, P., and Ferrando, A. (2012). The molecular basis of T cell acute lymphoblastic leukemia. J. Clin. Invest. *122*, 3398–3406.

Vigano, M.A., Ivanek, R., Balwierz, P., Berninger, P., van Nimwegen, E., Karjalainen, K., and Rolink, A. (2014). An epigenetic profile of early T-cell development from multipotent progenitors to committed T-cell descendants. Eur. J. Immunol. *44*, 1181–1193.

Zhang, L., Reynolds, T.L., Shan, X., and Desiderio, S. (2011). Coupling of V(D)J recombination to the cell cycle suppresses genomic instability and lymphoid tumorigenesis. Immunity *34*, 163–174.