HHS Public Access

Author manuscript

Cell. Author manuscript; available in PMC 2016 November 05.

Published in final edited form as:

Cell. 2015 November 5; 163(4): 947–959. doi:10.1016/j.cell.2015.10.016.

Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes

Jiazhi Hu[1,3], **Yu Zhang**[1,3], **Lijuan Zhao**[1], **Richard L. Frock**[1], **Zhou Du**[1], **Robin M. Meyers**[1], **Fei-long Meng**[1], **David G. Schatz**[2], and **Frederick W. Alt**[1,*]

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

[2]Howard Hughes Medical Institute; Department of Immunobiology, Yale University School of Medicine, 300 Cedar Street, Box 208011, New Haven, CT 06520-8011, USA

SUMMARY

RAG initiates antibody V(D)J recombination in developing lymphocytes by generating "on-target" DNA breaks at matched pairs of *bona fide* recombination signal sequences (RSSs). We employ bait RAG-generated breaks in endogenous or ectopically-inserted RSS pairs to identify huge numbers of RAG "off-target" breaks. Such breaks occur at the simple CAC motif that defines the RSS cleavage-site and are largely confined within convergent CTCF-binding element (CBE)-flanked loop domains containing bait RSS pairs. Marked orientation-dependence of RAG off-target activity within loops spanning up to 2 megabases implies involvement of linear tracking. In this regard, major RAG off-targets in chromosomal translocations occur as convergent RSS pairs at enhancers within a loop. Finally, deletion of a CBE-based *IgH* locus element disrupts V(D)J recombination domains and, correspondingly, alters RAG on- and off-target distributions within *IgH*. Our findings reveal how RAG activity is developmentally focused and implicate mechanisms by which chromatin domains harness biological processes within them.

INTRODUCTION

During B and T lymphocyte development, exons encoding antigen-binding immunoglobulin (Ig) or T cell receptor (TCR) variable regions are assembled from V, D, and J gene segments

*Correspondence: alt@enders.tch.harvard.edu.
[3]Co-first author

AUTHOR CONTRIBUTIONS

J.H., Y.Z., D.G.S., and F.W.A. designed the study. Y.Z. performed the *IgH* experiments and J.H. performed all the other experiments. L.Z. and R.L.F. helped set-up the RAG-bait HTGTS assay and analyze the *IgH* data. Z.D. helped with statistic analyses. R.M.M. wrote the HTGTS pipeline. F.L.M. helped analyze the correlation between junctions and enhancers. J.H., Y.Z., D.G.S., and F.W.A. interpreted the data, designed the figures, and wrote the manuscript.

ACCESSION NUMBER

The Gene Expression Omnibus (GEO) accession number for the data sets reported in this paper is GSE73007.

by V(D)J recombination (Alt et al., 2013). V(D)J recombination is initiated by RAG endonuclease, which introduces DNA double-stranded breaks (DSBs) between a pair of V, D, and J coding gene segments and their flanking recombination signal sequences (RSSs) (Schatz and Swanson, 2011). A *bona fide* RSS comprises a conserved palindromic heptamer represented by the canonical CACAGTG sequence, a degenerate spacer of 12 or 23 basepairs (bp), and a less-conserved A-rich nonamer (Figure 1A; Schatz and Swanson, 2011). RSSs with 12- or 23-bp spacers are termed 12RSSs and 23RSSs, respectively. Efficient RAG cleavage is restricted to a pair of participating coding segments flanked, respectively, by a 12RSS and a 23RSS, referred to here as paired *bona fide* RSSs. This 12/23 RSS restriction helps direct RAG to appropriate targets within antigen receptor loci (Alt et al., 2013).

RAG cleavage generates a pair of blunt broken RSS ends and a pair of hairpin-sealed coding ends (Figure 1B; Schatz and Swanson, 2011). Classical non-homologous end joining (C-NHEJ) fuses the two RSS ends precisely to form RSS joins and opens the two coding-end hairpins and joins them to form coding joins, which may be "processed" to lose or gain several nucleotides from each end (Figure 1B; Alt et al., 2013). While potential *bona fide* RSS-related sequences occur frequently across the genome, only a small number are documented RAG off-targets ("cryptic RSSs") (Merelli et al., 2010). Such RAG off-target activity contributes to oncogenic deletions or translocations in immature B and T cell cancers (Boehm et al., 1989; Larmonie et al., 2013; Onozawa and Aplan, 2012; Papaemmanuil et al., 2014). While RAG1 and RAG2 bind several thousand genomic sites which mostly correspond to active promoters and enhancers (Ji et al., 2010; Teng et al., 2015), lower densities of *cryptic* RSS heptamers near transcription start sites may help limit RAG off-target activity (Teng et al., 2015).

The mouse *IgH* locus spans 2.7 megabases (Mb) with V_Hs and their downstream 23RSSs embedded in a 2.4-Mb distal portion separated by a 100-kb inter-genic region from D_Hs flanked on both sides by 12RSSs and J_Hs flanked upstream by 23RSSs. Even though 12/23 restriction should allow V_Hs to join to un-rearranged D_Hs, *IgH* V(D)J recombination is "ordered" with D_H to J_H joining occurring in early progenitor (pro)-B cells followed by appendage of a V_H to a DJ_H complex (Alt et al., 2013). Ordered rearrangement and other levels of *IgH* V(D)J recombination regulation are mediated by modulating gene segment accessibility to RAG (Yancopoulos and Alt, 1986). In this regard, *IgH* contains a critical regulatory element termed Inter-Genic Control Region 1 (IGCR1) within the V_H to D_H interval (Guo et al., 2011). IGCR1 suppresses proximal V_H transcription and rearrangement at the D_H to J_H joining stage and, thereby, mediates broad levels of V(D)J recombination control, including diversification of antibody repertoires by indirectly promoting increased utilization of distal V_Hs. The most D-proximal V_H (V_H81x), while preferentially utilized in wild-type (WT) pro-B cells (Yancopoulos et al., 1984), is even more frequently utilized upon IGCR1 inactivation (Guo et al., 2011).

The CTCF factor binds directionally to an approximately 14-bp DNA target (Nakahashi et al., 2013), referred to as a CTCF-binding element (CBE) (Guo et al., 2011). CTCF is implicated in transcriptional insulation through ability to mediate chromatin loops (Ong and Corces, 2014). IGCR1 function relies on two divergently-oriented CBEs within it (Guo et

al., 2011). Besides IGCR1 CBEs, the 3′ *IgH* boundary harbors a CBE cluster (termed "3′ CBEs"); and single CBEs occur just downstream of proximal V_Hs and in intergenic regions between distal V_Hs (Degner et al., 2011). V_H CBEs are convergently-oriented with respect to the upstream IGCR1 CBE, and 3′CBEs are convergently-oriented with respect to the downstream IGCR1 CBE (Guo et al., 2011). Mutational studies of individual IGCR1 CBEs indicated that loop(s) mediated by the downstream CBE focus RAG activity in early pro-B cells within a domain containing the D_Hs and J_Hs; while a second domain mediated by the upstream CBE sequesters proximal V_Hs from RAG activity (Lin et al., 2015).

Eukaryotic genomes are organized into a hierarchy of architectures. Hi-C shows that chromatin is organized into topologically associated domains (TADs) that occur on Mb or sub-Mb scales and which have high frequency chromatin interactions within them (Dixon et al., 2012; Nora et al., 2012). Boundaries of such domains are often co-anchored by long-range interactions of sites bound by CTCF in association with cohesin (Phillips-Cremins et al., 2013; Zuin et al., 2014). Recent higher resolution *in situ* Hi-C further revealed mammalian genomes are divided into contact domains at an average scale of 185 kb (Rao et al., 2014). Contact domains with endpoints that generate a loop are termed loop domains (Rao et al., 2014). Loop domains genome-wide are commonly associated with pairs of convergent CBEs bound by CTCF and cohesin (Rao et al., 2014; Vietri Rudan et al., 2015). TADs have been implicated in replication timing (Pope et al., 2014), super-enhancer-driven transcription (Dowen et al., 2014), and DSB synapsis during antibody class switch recombination (CSR) (Dong et al., 2015; Zarrin et al., 2007), as well as in promoting normal limb development (Lupiáñez et al., 2015). Mechanistic aspects of how loop domains and TADs function are are not well understood.

Our recent studies suggested an unanticipated source of RAG off-target activity within long chromatin domains. To study oncogenic consequences, we generated mice with *Tcrβ* Dβ1 and Jβ1-1 segments inserted into intron one of the *c-Myc* oncogene ("*c-Myc*-DJβ cassette"). Despite frequent *c-Myc*-DJβ cassette recombination in developing lymphocytes, these mice do not develop lymphoma (Ranganath et al., 2008). However, ATM deficient, *c-Myc*-DJβ cassette mice develop B cell lymphomas with *c-Myc* translocations/amplifications that fuse RAG-generated *IgH* DSBs to sequences over a several hundred kb region downstream of *c-Myc* (Tepsuporn et al., 2014). These downstream translocations occur exclusively on the cassette allele; but do not involve the cassette, suggesting RAG activity at *bona fide* RSS pairs within *c-Myc* promotes cutting at linked downstream cryptic RSSs (Tepsuporn et al., 2014). On this basis, we identify an immense number of previously unsuspected RAG off-targets generated by a mechanism that has broader implications for gene regulation within loop domains.

RESULTS

HTGTS Assay for RAG On-target and Off-target DSBs and Translocations

To test the hypothesis that the *c-Myc*-DJβ cassette promotes cutting at cryptic RSSs downstream of *c-Myc*, we generated a *v-Abl*-transformed pro-B cell line from mice homozygous for the *c-Myc*-DJβ cassette allele (referred to as "*c-Myc*-DJβ pro-B line"). In such lines, RAG expression can be induced in the context of G1 cell cycle arrest following

treatment with the *v-Abl* kinase inhibitor STI-571 (Bredemeyer et al., 2006). Due to propensity of cycling *v-Abl* transformants to form Dβ1 to Jβ1-1 cassette rearrangements at low level, we were able to isolate just one *v-Abl* pro-B clone with an un-rearranged cassette allele (Figure 1A). This clone had a second cassette allele in DJβ-rearranged configuration, which is inert for rearrangement (See below). Upon G1 arrest and RAG expression, the *c-Myc*-DJβ construct undergoes high frequency *bona fide* Dβ1 to Jβ1-1 rearrangements, which fuse the downstream coding end of Dβ1 (23RSS-associated) to the Jβ1-1 12RSS-associated coding end in the chromosome and, correspondingly fuses the Dβ1 23RSS to the Jβ1-1 12RSS within an excision circle (Figure 1A,B). To detect potential cryptic RSSs activated by the *c-Myc*-DJβ cassette in these *v-Abl* pro-B cells, we employed high-throughput genome-wide translocation sequencing (HTGTS). HTGTS is a highly sensitive DSB and translocation assay that identifies junctions between a broken end of a fixed "bait" DSB and ends of other prey DSBs genome-wide (Chiarle et al., 2011; Dong et al., 2015; Frock et al., 2015). For these analyses, we used an HTGTS bait primer termed "*c-Myc* E1" that anneals with sequences 213 bp upstream of the cassette Dβ1 23RSS. This primer detects Dβ1 downstream coding end joins to Jβ1-1 coding ends and also to other DSBs genome-wide (Figure 1B).

In the *c-Myc*-DJβ pro-B line, the vast majority of recovered HTGTS junctions represented expected *bona fide* cassette Dβ1 to Jβ1-1 coding joins. To enhance off-target detection, we experimentally suppressed recovery of *bona fide* cassette DJβ joins (Figure 1B; Supplemental Information). The vast majority of remaining Dβ1 downstream coding end junctions, representing 1–3% of total junctions, occurred to sequences up to 1.8 Mb downstream of *c-Myc*, with additional joins to sequences about 1 kb upstream. Notably, the junctions in this 1.8-Mb region abruptly ended in both directions (Figure 1C; see below). Indeed, the only other clear-cut hotspot region genome-wide occurred at about 0.02% of total junctions and involved low-level translocations to *Igκ* (Figure S1A), a major *bona fide* RAG target in *v-Abl* pro-B cells (Zhang et al., 2012). Approximately 20% of the apparent RAG off-target sites in the 1.8-Mb domain represented recurrent ("hotspot") junctions that, in some cases, were recovered dozens of times in independent libraries (Table S1). HTGTS analysis of bone marrow (BM) pro-B cells from *c-Myc*-DJβ mice gave similar results (Figure S1A–C).

We also isolated an ATM-deficient *v-Abl c-Myc*-DJβ pro-B line in which one allele had an inversion that joined the Dβ1 23RSS to a cryptic RSS (5′-CACAGTT) in the Jβ1-1 segment (Figure S1D). In this line, the second *c-Myc*$^{DJ\beta}$ allele was in the inert DJβ configuration. Following G1 arrest, HTGTS employing the *c-Myc* E1 primer revealed that the major "*bona fide*" V(D)J joining event in this line (> 97% of recovered junctions) was inversional joining of the Dβ1 12RSS (the upstream Dβ1 RSS) to the inverted Dβ1 23RSS 693bp downstream (Figure S1D). The vast majority of remaining joins (~3% of total junctions) fused the Dβ1 12RSS to other DSBs along the 1.8-MB cassette-containing domain with a distribution similar to that of Dβ1 downstream coding end joins in the ATM-proficient *c-Myc*-DJβ pro-B line and primary pro-B cells (Figure S1A–C; Table S1). Notably, there was increased, but still low, levels of translocations to *Igκ* (~0.2% of total junctions, Figure S1A) as compared to ATM-proficient proB cells. ATM-deficient BM *c-Myc*-DJβ pro-B cells also had similar

patterns of Dβ1 23RSS coding ends junctions to those of ATM-proficient pro-B lines, except that they had low-level translocations to *IgH* (~0.07%) and *TCRα/δ* (~0.05%) (Figure S1A–C). Finally, an ATM-deficient *v-Abl c-Myc*-DJβ pro-B line in which both cassette alleles were in the DJβ configuration generated few junctions, confirming that single 12RSS-containing alleles are inert (Figure S1E).

Abundant DSBs Across the 1.8-Mb *c-Myc*-DJβ Loop Domain

We investigated the orientation of the thousands of Dβ1 downstream coding end junctions within the 1.8-Mb *c-Myc* region in the *c-Myc*-DJβ *v-Abl* pro-B cell line. Junctions are denoted as in "+" orientation if prey sequence reads in a centromere to telomere direction and in "−" orientation if prey reads in the opposite direction (Chiarle et al., 2011). As the *c-Myc* E1 primer is centromeric to the bait Dβ1 downstream coding end, it captures junctions resulting in upstream excision circles and downstream deletions as + events and captures inversional junctions either upstream or downstream as − events (Figure 1D). Dβ1 downstream coding end junctions near (within 5kb) *c-Myc* occurred at similar frequency in + and − orientations (Figure S1F); strikingly, however about 95% of junctions to sequences further downstream of *c-Myc* occurred in deletional (+) orientation (Figure 1E). Similar results also were obtained with ATM-deficient *c-Myc*-DJβ pro-B cell lines (Figure S1C), even though their junctions involved Dβ1-12RSS ends (See Figure S1D legend).

To gain insight into the basis for the well-defined boundaries of the DSB hotspot region flanking the *c-Myc*-DJβ cassette, we analyzed existing ChIP-seq data from BM pro-B cells (Lin et al., 2012) and found a cluster of CTCF and cohesin subunit Rad21 binding sites on both boundaries of this 1.8-Mb domain (Figure 1F). Moreover, the two clusters of CBEs were in convergent orientation (Figure 1F). Indeed, recent high-resolution *in situ* Hi-C data performed in mouse CH12-LX B cell lines (Rao et al., 2014) confirmed that this 1.8-Mb region is a well-defined convergent CBE-based loop domain that contains within it a strong 840-kb sub-loop that also extends to a convergent CBE (Figure 1F, G). HTGTS junction density within the 1.8-Mb domain in both ATM-proficient (Figure 1E–G) and deficient *c-Myc*-DJβ pro-B cells (Figure S1C) correlated well with Hi-C interactions within the two loop domains.

RAG Generates DSBs in the 1.8-Mb *c-Myc*-DJβ Loop Domain

To test the relationship of frequent prey DSBs within the 1.8-Mb *c-Myc* loop domain to RAG-generated DSBs, we searched ATM-proficient and ATM-deficient *c-Myc*-DJβ junctions for sequence motifs in their vicinity. In this regard, the conserved 5′-CAC motif of the RSS heptamer is a position indicator for RAG cleavage, with cleavage invariably occurring 5′ to the CAC motif (Figure 1A). For convention, a CAC is considered in "forward" orientation if the presumed associated "coding" sequence is centromeric to the RSS and in "reverse" orientation if the presumed coding sequence is telomeric (Figure S2A). For widespread CACs, sequences in the coding position would not generally be gene segments; so we refer to them as surrogate coding ends. For analysis, we pooled and analyzed, respectively, all + junctions from the two *v-Abl* pro-B cell types and found that the majority occurred in putative surrogate coding sequences at or within 5 bp of a reverse CAC, with about 30% joined directly to the surrogate coding sequence immediately flanking

a CAC (Figure 2A,B and S2B, C). There was no significant correlation with forward CACs (Figure 2C and S2D). These results suggest that the frequent DSBs within the 1.8-MB *c-Myc* domain occur at "cryptic RSSs" represented predominantly by a conserved CAC. Moreover, surrogate coding ends fused to the bait ends were processed similarly to normal coding ends during V(D)J recombination. The most highly recurrent hotspot DSBs within the 1.8-Mb domain tended to involve CACs within more canonical heptamers (Figure S2E). Finally, remarkably similar results were obtained from ATM-proficient and ATM-deficient *c-Myc*-DJβ BM pro-B cells (data not shown).

To unequivocally test the role of RAG in generating DSBs in the 1.8-Mb *c-Myc* loop domains, we deleted *Rag2* in the ATM-deficient *c-Myc*-DJβ pro-B cell line (Figure S2F). For HTGTS bait, we employed a Cas9/gRNA to generate DSBs 519 bp downstream of the *c-Myc*-DJβ cassette and designed a primer that allowed 5′ broken ends of these DSBs to be used as bait ("5′Cas9 bait ends"; Figure 2D). We then performed HTGTS on RAG-sufficient and RAG2-deficient G1-arrested pro-B cells. Recovered 5′Cas9 HTGTS junctions from RAG-sufficient ATM-deficient *c-Myc*-DJβ *v-Abl* pro-B cells correlated with reverse CACs in the 1.8-Mb domain as expected; however, unlike RAG-generated bait broken ends, the Cas9/gRNA generated bait ends recovered junctions equally in + and − orientation (Figure 2D–F and S2G). Performing these assays in ATM-deficient *v-Abl* pro-B cells that either lacked the *c-Myc*-DJβ cassette or were RAG2-deficient generated only a very few junctions within the 1.8-Mb domain, and these had no correlation with CACs (Figure 2D–F and S2H). These findings confirm that RAG generates the off-target DSBs across the 1.8-Mb *c-Myc* domain in a *c-Myc*-DJβ cassette-dependent fashion and also demonstrate that the asymmetric prey joining preferences observed are specific to RAG-generated bait ends.

Paired *Bona Fide* RSSs Generate RAG Off-target Activity in Loop Domains Genome-wide

We next tested whether other loop domains genome-wide could similarly be targets for such widespread RAG-generated DSBs if they contain *bona fide* RSS pairs. To insert bait RSSs into multiple genomic sites, we infected ATM-proficient and deficient *v-Abl* pro-B lines with the pMX-DEL-SJ virus (referred to as "DEL-SJ"), which harbors a pair of divergent *bona fide* RSSs flanking an inverted GFP sequence (Figure 3A; Bredemeyer et al., 2006). V(D)J recombination between the divergent DEL-SJ RSSs fuses them in the chromosome and liberates the intervening GFP DNA within an excision circle generated via fusion of the surrogate coding ends (Figure 3A). We isolated 6 independent sub-clones from each genotype, each with a unique DEL-SJ-integration, treated them with STI-571, and generated HTGTS libraries with primers adjacent to either the construct 12RSS (12S primer) or 23RSS (23S primer) (Figure S3A). In all 12 DEL-SJ integration sites, the 12RSS and 23RSS junctions were confined within convergent CBE-based loop domains that ranged from 174 kb to 2.64 Mb in size (Table S2) and which often contained sub-domains flanked by convergent CBEs. For all integration sites, translocation junction density correlated well with interaction intensities revealed by Hi-C. Representative findings from chromosome X, 4, 12, and 19 integrations are shown (Figure 3B–K and S3B, C). Notably, junctions detected from either 12RSS- or 23RSS-specific primers mostly occurred in deletional orientation independent of the orientation in which the DEL-SJ was integrated relative to the

centromere (Figure 3B–K and S3B, C). As for the *c-Myc-DJβ* 1.8-Mb domain, hotspots also were apparent (Figure 3B–K and S3B, C).

We examined junction sequences within the two DEL-SJ loop domains on chromosome X and 4, respectively, for potential correlations with forward or reverse CACs. Deletional and excision circle junctions represented over 95% of events for any given integration site (Figure 3B–K). Strikingly, the vast majority of junctions were highly correlated with CACs; however, while bait RSSs joined to convergent upstream cryptic CACs, they joined to surrogate coding ends associated with downstream CACs in the same orientation to form apparent "hybrid" RSS to coding end joins (Figure 4A–C and S4A–C; but see below). Analysis of several other DEL-SJ-containing domains (on chromosomes 12 and 19) revealed precisely the same patterns, despite diverse locations and relative chromosomal orientations (data not shown). Notably, upstream CACs were generally joined precisely to bait RSSs, but downstream joins to surrogate coding ends was often imprecise with deletions of several nucleotides from the CAC border (Figure 4A–C and S4A–C). The latter result, together with junctional sequence analysis of bait RSS ends (Figure S4D), indicates that RSS ends from the DEL-SJ construct that join downstream behave like surrogate coding ends in a V(D)J recombination-type of joining reaction.

Normal DEL-SJ V(D)J recombination generates fused RSS pairs at high frequency (Figure 3A), which can be re-cleaved by RAG with one cleavage product then being treated as an RSS end and the other as a surrogate coding end (Figure 4D; Meier and Lewis, 1993). Thus, the apparent downstream "hybrid joins" observed with the bait 12RSS, consistent with their end-structure, could be generated from the fused intermediate. To test this possibility, we used as bait the 12RSSs of perfectly fused 12–23 joins of DEL-SJ within the chromosome X and 4 integrations, respectively. Indeed, this fused RSS pair faithfully recapitulated the joining patterns of the parental un-rearranged DEL-SJ construct in this location (Figure 4B,C, E and S4B, E), demonstrating that the joining orientation of the two fused RSSs determines whether one or the other acts as an RSS end or surrogate coding end in the off-target V(D)J recombination joining reaction. Finally, we also generated HTGTS libraries from the surrogate coding ends (GFP primer) associated with 12RSS of DEL-SJ integrated into chromosome X. Such surrogate coding end junctions would not be re-cleaved by RAG. Correspondingly, nearly 90% of the 12RSS-associated coding ends joined downstream of the GFP primer to surrogate coding ends adjacent to CACs (Figure 4F–H). Together our findings show that for *bona fide* RSS pairs within a loop domain, both the RSS and the associated coding sequence join to convergent cryptic RSSs (CACs) and associated surrogate coding ends within a loop domain via a V(D)J recombination-like reaction.

Robust Detection of RAG Off-targets Genome-wide Outside of Chromatin Domains

We further analyzed the 12RSS-associated coding end (GFP)-primed DEL-SJ HTGTS libraries from the X chromosome integration in ATM-deficient *v-Abl* lines and also additional libraries from an integration on chromosome 1 in a different ATM-deficient *v-Abl* line (Figure S5A). Beyond the expected joining patterns within the DEL-SJ-containing loop domains (Figure S5B–D), these libraries also revealed 107 translocation hotspots across the genome that all occurred at or near heptamers related to the canonical CACAGTG motif

(Figure 5A,B and S5E; Table S3). Notably, 60 of the 107 identified cryptic RSSs occurred in pairs in convergent orientation within less than 100 kb in the same domain (Figure 5A,C and S5E; Table S3). HTGTS employing a primer upstream of the cryptic RSS in one such pair on chromosome 1 (Figure 5C) revealed thousands of deletional junctions involving two cryptic RSSs (Figure S5F–H). We compared locations of these 107 cryptic RSSs with existing pro-B H3K4me3 ChIP-seq data, which marks promoters, or H3K27Ac data that marks promoters and enhancers (Lane et al., 2014; Whyte et al., 2013). Strikingly, 97 of the 107 RAG off-targets overlapped with H3K27Ac-marked regions, with 38 overlapping with super-enhancers and 59 with typical enhancers. Of these, 65 overlapped with regions marked by both H3K4me3 and H3K27Ac (Figure 5D,E). These remarkably high correlations demonstrate that accessibility, beyond RAG binding, also is important for efficient RAG cleavage at cryptic RSSs.

IgH Employs CBE-based Subdomains to Regulate RAG On- and Off-target Activity

We applied HTGTS to test whether RAG on- and off-target activity in *IgH* is confined within IGCR1 CBE-based domains (Figure 6A). We employed an ATM-deficient *v-Abl* pro-B cell line that harbors a DFL16.1-J_H3 rearrangement, providing a population of cells harboring a 5′D 12RSS expected to join to accessible upstream V_Hs 23RSSs (Alt et al., 2013). We used an HTGTS primer 82 bp upstream of the 5′ DFL16.1 12RSS to capture joins involving bait 5′DFL16.1-J_H3 RSS ends (Figure 6B). The majority of approximately 27,000 recovered *IgH* HTGTS junctions were on-targets at *IgH bona fide* RSSs (85%; Table S4), with most fusing the DFL16.1 5′RSS to a V_H 23RSS in physiologic (excision circle) orientation (Figure 6B,C). While such junctions involved multiple V_Hs across the 2.4-Mb V_H domain, they were biased towards proximal V_Hs, particularly V_H81x (38% of on-targets) (Figure 6C). We also observed substantial inversional (+) joining between the DFL16.1 5′RSS and J_H4 23RSS (20% of *IgH* on-targets) (Figure 6B,C). Strikingly, IGCR1 deletion dramatically increased the number of DFL16.1 5′RSS junctions recovered (28 fold) (Figure S6A, Table S4), largely from markedly increased utilization of proximal V_Hs (48 fold) and, in particular, V_H81x (92% of junctions) (Figure 6D and S6B,D). Correspondingly, there was an 18-fold decrease in distal/middle V_H utilization and a 20-fold decrease in J_H4 junctions (Figure 6D and S6C,F).

These *IgH* HTGTS studies also revealed low, but highly reproducible, off-target joining of DFL16.1 12RSS ends to DSBs within *IgH* that correlated with CACs (Figure S6G,H). Strikingly, about 95% of the off-target *IgH* junctions were within a tightly focused 12.3-kb region that contains the DFL16.1-J_H3 and is bounded upstream by IGCR1 and downstream by iEμ/Sμ (Figure 6E,F). We refer to this region as the iEμ/Sμ to IGCR1 “recombination domain”. Strikingly, deletion of IGCR1 from this ATM-deficient pro-B line dramatically changed the profile of off-target DSBs, permitting them to spread ~120 kb upstream into the proximal V_Hs, while decreasing the percentage of off-target junctions in the former iEμ/Sμ to IGCR1 domain to 13% (Figure 6G and S6I). Thus, IGCR1 deletion established a new iEμ/Sμ to proximal V_H recombination domain in which RAG activity on both cryptic RSSs and proximal V_Hs *bona fide* RSSs is re-focused. As in other domains, RAG off-target activity was highly dependent on convergent CAC orientation once several kb from the DFL16.1 5′RSS break-site (Figure S6H,J).

DISCUSSION

Mechanism of RAG Off-target Activity

We report a major form of RAG off-target activity that eluded prior investigations. Remarkably, this activity is largely confined to loop domains containing paired *bona fide* RSSs, with cleavage requiring only recognition of a simple CAC motif. Also remarkable, this off-target RAG activity is directionally oriented such that RSS ends from paired *bona fide* RSSs join to convergent CAC-containing motifs; while coding ends from paired *bona fide* RSSs join to surrogate coding ends associated with a CAC. Thus, RSSs and corresponding coding ends join with the same patterns and in the same locations, consistent with V(D)J recombination (Figure 4). Such orientation dependence is most readily explained by a linear tracking mechanism (Yancopoulos et al., 1984; Figure 7). Based on RAG structural information (Kim et al., 2015), we propose a working model (Figure 7 and S7). This model assumes that formation and activation of the tetrameric RAG1/2 complex (Lapkouski et al., 2015) requires binding of paired *bona fide* RSSs (Figure 7B). We hypothesize that one or the other of these occasionally escapes the activated complex, allowing cryptic RSSs to replace them. The replacement process could involve diffusion of proximal cryptic RSSs or unidirectional tracking to more distal cryptic RSSs downstream. Once appropriately positioned in the activated complex, cryptic RSSs and surrogate coding ends could, likely at reduced frequency, be cleaved and joined to their remaining *bona fide* counterpart via a reaction that preserves most aspects of normal V(D)J recombination (Figure 7G). This general tracking model explains all aspects of our findings including tracking from the two *bona fide* RSSs of a pair in opposite directions around the loop.

Implications for Normal Loop Domain Functions

An obvious and important question arises as to why RAG activity is so highly restricted within loop domains containing the initiating paired *bona fide* RSSs. One contributing factor could be high interaction frequency of DNA in chromatin across these domains (Alt et al., 2013). In this regard, DSB ends find and join to ends of other DSBs within such domains at higher frequency than elsewhere in the genome (Alt et al., 2013; Zhang et al., 2012). During *IgH* CSR, this phenomenon promotes proper and frequent joining of AID-initiated DSBs (Dong et al., 2015; Zarrin et al., 2007). Such DSB interactions are evident in our current studies where Cas9/gRNA-generated DSBs frequently join to RAG off-target DSBs within the same loop. However, distinct from RAG-generated RSS or coding ends, a given Cas9/gRNA bait end joins to both cryptic RSS ends and surrogate coding ends of RAG off-target DSBs. Another apparent difference is that site-specific nuclease- or AID-generated DSBs appear to find off-targets in other regions across the genome much more readily than do RAG-generated DSBs, even in WT cells (Chiarle et al., 2011; Dong et al., 2015; Frock et al., 2015). The almost exclusive restriction of RAG off-targets to paired *bona fide* RSS-containing loops implies that an additional mechanism enforces such RAG activity.

The tracking mechanism can explain the additional restriction of RAG activity within a given loop (Figure 7). We do not know the mechanism that propels RAG tracking, although transcription and/or cohesin might be involved (e.g. Nichols and Corces, 2015). However, it is reasonable to assume that tracking is terminated when it encounters a block imposed by

the CTCF/cohesin-bound convergent CBE pair or similar loop-forming interactions (Figure 7F). Such blockage would terminate tracking in each direction from paired *bona fide* RSSs and limit off-target RAG activity to the loop. In support of this model, deletion of the CBE-based IGCR1 allows RAG off-target activity to extend from its initial highly restricted location in the D-J_H recombination domain to more than 100 kb upstream, where new boundaries may form via V_H CBEs and/or associated factors (Figure 6G). Beyond regulating V(D)J recombination, related loop domain functions might impact on other activities constrained within them, including replication (Pope et al., 2014) and promoter/enhancer interactions (Dowen et al., 2014).

IgH Locus Regulation

Regulated *IgH* V_H to DJ_H recombination depends on integrity of the two divergent CBEs within IGCR1, likely via formation of loop domains that focus RAG activity on D_Hs and J_Hs (Guo et al., 2011; Lin et al., 2015). Our HTGTS studies provide additional insights into *IgH* V(D)J recombination regulation (Figure 6). In a DJ_H rearranged pro-B cell line, on-target rearrangements of the 5′D RSS occurs to RSSs of V_Hs across the locus, but predominantly 3′ V_Hs (Figure 6C). As most RAG off-target activity is focused in a small 12.3-kb recombination domain from IGCR1 to the iEμ/Sμ boundary (Figure S6G), the recombination domain in these cells does not extend downstream to 3′CBEs as perhaps anticipated. This restriction could be due to IGCR1 CBE looping with non-CBE elements at iEμ/Sμ (Guo et al., 2011) and/or by tracking limitations imposed by a unidirectional mechanism. In the DJ_H rearranged cells, *bona fide* V(D)J recombination at upstream V_Hs in the absence of corresponding off-target activity, even in proximal portions of the locus, is consistent with V_Hs entering the recombination domain by a specialized mechanism operating subsequent to locus contraction (Bossen et al., 2012). Based on off-target activity as an assay, IGCR1 deletion extends the recombination domain linearly into proximal V_Hs, resulting in a huge overall V(D)J recombination increase, involving V_H81x and other proximal V_Hs (Figure 6D and S6D). This increase may be facilitated by increased interaction frequency gained by placing V_H 23RSSs in the same loop domain as the 5′ DFL16.1 12RSS and/or by a tracking contribution. Finally, a unidirectional RAG tracking mechanism also might explain why 3′D 12RSSs, but not 5′D 12RSSs are used developmentally in D to J_H rearrangements.

RAG Off-target Activity, Chromosomal Rearrangements, and Cancer

We prove our hypothesis that inserting paired *bona fid*e RSSs into *c-Myc* activates RAG-generated DSBs at cryptic RSSs over a long region downstream that, in the context of ATM-deficiency, promotes oncogenic translocations. These findings explain how paired *bona fide* RSSs within a *Tcra* excision circle fragment integrated into the *HPRT* locus in leukemia cells causes further genomic aberrations (Messier et al., 2006) and also support the hypothesis that translocations downstream of *c-Myc* in human B cell lymphomas involve cryptic RSSs (Kroenlein et al., 2011). Given that cryptic RSS targeting downstream of *c-Myc* occurs in both WT and ATM-deficient pro-B cells, one role of ATM in suppressing such translocations would be through stabilizing ends in RAG post-cleavage complexes to facilitate their joining via V(D)J recombination (Bredemeyer et al., 2006). Thus, ATM limits potential RAG-initiated translocations by promoting joining of RAG-initiated DSBs at RSSs and cryptic RSSs within a loop. Our findings also provide a mechanism for oncogenic

translocations to sequences far downstream of *c-Myc* in C-NHEJ/p53 double-deficient pro-B cells (Alt et al., 2013). In this regard, we find cryptic RSSs in the *c-Myc* 1.8-Mb domain that are closer to consensus (Merelli et al., 2010) and, therefore, may drive RAG-initiated DSBs at other cryptic RSSs in this domain that become liberated from post-cleavage complexes in the absence of C-NHEJ.

We also found 107 genome-wide cryptic RSSs, not related to antigen receptor loci or paired *bona fide* RSSs-containing domains, that were DSB and translocation targets in ATM-deficient *v-Abl* pro-B cells (Figure 5; Table S3). This set of cryptic RSSs tended to have heptamers even closer to consensus than recurrent hotspots within paired *bona fide* RSSs-containing loops (Figure 5B vs S2E). Many of these translocation target RSSs occurred in pairs separated by less than 100 kb (Table S3), with each member of the pair falling directly within enhancer and/or promoter regions (Figure 5C–E). Enhancer/promoter loops also might increase the frequency with which such paired cryptic RSSs are juxtaposed to form stable RAG synaptic complexes. Strikingly, all 30 pairs of these cryptic RSS translocation targets were in convergent orientation (Table S3), similar to most proximal paired *bona fide* RSSs within antigen receptor loci (Bossen et al., 2012) and the majority of cryptic RSSs captured by *bona fide* RSSs within loop domains. Thus, to serve as a strong genome-wide translocation target, cryptic RSS require a good heptamer, location in enhancers and/or promoters, and convergent pairing with another good cryptic RSS in the same loop. Finally, our findings provide a mechanistic basis for recurrent oncogenic chromosomal interstitial deletions in tumors arising from developing human lymphocytes (Larmonie et al., 2013; Mullighan et al., 2008; Papaemmanuil et al., 2014).

EXPERIMENTAL PROCEDURES

Cell Lines

BM pro-B cells were purified by αB220 selection from ATM-proficient and -deficient *c-Myc*-DJβ mice (Tepsuporn et al., 2014) and cultured in opti-MEM medium with 10% (v/v) FBS plus IL-7 (2ng/mL) and SCF (2ng/mL) for 4 days. The *v-Abl* pro-B cells were cultured in RPMI medium with 15% (v/v) FBS; cells were treated with STI-571 (3 μM) for 4 days to express RAG. WT and ATM-deficient *v-Abl* pro-B cell lines were described previously (Zha et al., 2010). ATM-proficient and deficient *c-Myc*-DJβ *v-Abl* pro-B cell lines were made specifically for this study from *Eμ-Bcl-2* transgenic mice of the corresponding genotypes. We included the *Eμ-Bcl-2* transgene in these cells to protect STI-571 treated (G1-arrested) *v-Abl* pro-B cells from apoptosis; prior work showed that Bcl-2 expression has no effect on V(D)J recombination (Zha et al., 2010).

RAG On- and Off-targets

HTGTS was performed and analyzed as previously described with modifications (Frock et al., 2015). Primers for HTGTS are listed in Table S5. Due to the very low junctional diversity of *bona fide* V(D)J recombination RSS joins and coding joins, we included duplicate junctions in our analyses of G1-arrested *v-Abl* cells to better reflect the actual frequencies of the various classes of *bona fide* and off-target junctions. Where approximate percentage and/or numbers of different classes of junctions are indicated (e.g. *c-Myc-DJβ* or

IgH), we controlled for reproducibility by performing at least three independent experiments (e.g. Table S4). RAG off-target hotspots were identified by MACS2 (Zhang et al., 2008) with extend size (extsize) at 20 bp and false discovery rate (FDR) cut-off at 10^{-9}. See Supplemental Information for more details.

ChIP-Seq and Hi-C Data

CTCF and Rad21 ChIP-seq data were extracted from Lin et al., 2012 (GSE40173); H3K4me3 and H3K27Ac ChIP-seq data were extracted from Lane et al., 2014 (GSE48555). These data are from BM pro-B cells. We re-analyzed ChIP-seq data with Chilin software (http://cistrome.org/chilin/) in the simple model against *mm9*. Enhancer annotation was either extracted directly from Whyte et al., 2013 (GSE44288) or identified by Homer software (Heinz et al., 2010) from re-analyzed H3K27Ac ChIP-seq data (Lane et al., 2014). In situ Hi-C data for CH12-LX B cells was extracted and displayed (KR normalization) by Juicebox software (Rao et al., 2014).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the Alt lab members and Richard Young (MIT) and Jay Bradner (DFCI) for stimulating discussions. This work was supported by NIH grant AI020047 and LLS SCOR 7009-12 to F.W.A, and NIH grant AI032524 to D.G.S. F.W.A. and D.G.S are Howard Hughes Medical Institute Investigators. J.H. is supported by a Robertson Foundation/Cancer Research Institute Irvington Fellowship. Y.Z. is supported by a career development fellowship from the Leukemia and Lymphoma Society. R.L.F. was supported by the NIH NRSA T32AI007512. F.L.M. is a Lymphoma Research Foundation postdoctoral fellow.

References

Alt FW, Zhang Y, Meng FL, Guo C, Schwer B. Mechanisms of programmed DNA lesions and genomic instability in the immune system. Cell. 2013; 152:417–429. [PubMed: 23374339]

Boehm T, Mengle-Gaw L, Kees UR, Spurr N, Lavenir I, Forster A, Rabbitts TH. Alternating purine-pyrimidine tracts may promote chromosomal translocations seen in a variety of human lymphoid tumours. EMBO J. 1989; 8:2621–2631. [PubMed: 2531086]

Bossen C, Mansson R, Murre C. Chromatin topology and the regulation of antigen receptor assembly. Annu Rev Immunol. 2012; 30:337–356. [PubMed: 22224771]

Bredemeyer AL, Sharma GG, Huang CY, Helmink BA, Walker LM, Khor KC, Nuskey B, Sullivan KE, Pandita TK, Bassing CH, et al. ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. Nature. 2006; 442:466–470. [PubMed: 16799570]

Chiarle R, Zhang Y, Frock RL, Lewis SM, Molinie B, Ho YJ, Myers DR, Choi VW, Compagno M, Malkin DJ, et al. Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. Cell. 2011; 147:107–119. [PubMed: 21962511]

Degner SC, Wong TP, Jankevicius G, Feeney AJ. Cutting edge: developmental stage-specific recruitment of cohesin to CTCF sites throughout immunoglobulin loci during B lymphocyte development. J Immunol. 2009; 182:44–48. [PubMed: 19109133]

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012; 485:376–380. [PubMed: 22495300]

Dong J, Panchakshari RA, Zhang T, Zhang Y, Hu J, Volpi SA, Meyers RM, Ho YJ, Alt FW. Orientation-specific joining of AID-initiated DNA breaks promotes antibody class switching. Nature. 2015; 525:134–139. [PubMed: 26308889]

Dowen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, Weintraub AS, Schuijers J, Lee TI, Zhao K, et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell. 2014; 159:374–387. [PubMed: 25303531]

Frock RL, Hu J, Meyers RM, Ho Y-J, Kii E, Alt FW. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nat Biotechnol. 2015; 33:179–186. [PubMed: 25503383]

Guo C, Yoon HS, Franklin A, Jain S, Ebert A, Cheng HL, Hansen E, Despo O, Bossen C, Vettermann C, et al. CTCF-binding elements mediate control of V(D)J recombination. Nature. 2011; 477:424–430. [PubMed: 21909113]

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010; 38:576–589. [PubMed: 20513432]

Ji Y, Resch W, Corbett E, Yamane A, Casellas R, Schatz DG. The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. Cell. 2010; 141:419–431. [PubMed: 20398922]

Kim MS, Lapkouski M, Yang W, Gellert M. Crystal structure of the V(D)J recombinase RAG1-RAG2. Nature. 2015; 518:507–511. [PubMed: 25707801]

Kroenlein H, Schwartz S, Reinhardt R, Rieder H, Molkentin M, Gökbuget N, Hoelzer D, Thiel E, Burmeister T. Molecular analysis of the t(2;8)/MYC-IGK translocation in high-grade lymphoma/leukemia by long-distance inverse PCR. Genes Chromosom. Cancer. 2011; 51:290–299.

Lane AA, Chapuy B, Lin CY, Tivey T, Li H, Townsend EC, van Bodegom D, Day TA, Wu S-C, Liu H, et al. Triplication of a 21q22 region contributes to B cell transformation through HMGN1 overexpression and loss of histone H3 Lys27 trimethylation. Nat Genet. 2014; 46:618–623. [PubMed: 24747640]

Lapkouski M, Chuenchor W, Kim M-S, Gellert M, Yang W. Assembly Pathway and Characterization of the RAG1/2-DNA Paired and Signal-end Complexes. J Biol Chem. 2015; 290:14618–14625. [PubMed: 25903130]

Larmonie NSD, Dik WA, Meijerink JPP, Homminga I, van Dongen JJM, Langerak AW. Breakpoint sites disclose the role of the V(D)J recombination machinery in the formation of T-cell receptor (TCR) and non-TCR associated aberrations in T-cell acute lymphoblastic leukemia. Haematologica. 2013; 98:1173–1184. [PubMed: 23904235]

Lin SG, Guo C, Su A, Zhang Y, Alt FW. CTCF-binding elements 1 and 2 in the Igh intergenic control region cooperatively regulate V(D)J recombination. Proc Natl Acad Sci USA. 2015; 112:1815–1820. [PubMed: 25624508]

Lin YC, Benner C, Mansson R, Heinz S, Miyazaki K, Miyazaki M, Chandra V, Bossen C, Glass CK, Murre C. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. Nat Immunol. 2012; 13:1196–1204. [PubMed: 23064439]

Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 2015; 161:1012–1025. [PubMed: 25959774]

Meier JT, Lewis SM. P nucleotides in V(D)J recombination: a fine-structure analysis. Mol Cell Biol. 1993; 13:1078–1092. [PubMed: 8380891]

Merelli I, Guffanti A, Fabbri M, Cocito A, Furia L, Grazini U, Bonnal RJ, Milanesi L, McBlane F. RSSsite: a reference database and prediction tool for the identification of cryptic Recombination Signal Sequences in human and murine genomes. Nucleic Acids Res. 2010; 38:W262–W267. [PubMed: 20478831]

Messier TL, O'Neill JP, Finette BA. V(D)J recombinase mediated inter-chromosomal HPRT alterations at cryptic recombination signal sequences in peripheral human T cells. Hum Mutat. 2006; 27:829–829. [PubMed: 16835902]

Mullighan CG, Miller CB, Radtke I, Phillips LA, Dalton J, Ma J, White D, Hughes TP, Le Beau MM, Pui CH, et al. BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. Nature. 2008; 453:110–114. [PubMed: 18408710]

Nakahashi H, Kwon K-RK, Resch W, Vian L, Dose M, Stavreva D, Hakim O, Pruett N, Nelson S, Yamane A, et al. A genome-wide map of CTCF multivalency redefines the CTCF code. Cell Rep. 2013; 3:1678–1689. [PubMed: 23707059]

Nichols MH, Corces VG. A CTCF Code for 3D Genome Architecture. Cell. 2015; 162:703–705. [PubMed: 26276625]

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012; 485:381–385. [PubMed: 22495304]

Ong C-T, Corces VG. CTCF: an architectural protein bridging genome topology and function. Nat Rev Genet. 2014; 15:234–246. [PubMed: 24614316]

Onozawa M, Aplan PD. Illegitimate V(D)J recombination involving nonantigen receptor loci in lymphoid malignancy. Genes Chromosom. Cancer. 2012; 51:525–535.

Papaemmanuil E, Rapado I, Li Y, Potter NE, Wedge DC, Tubio J, Alexandrov LB, Van Loo P, Cooke SL, Marshall J, et al. RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. Nat Genet. 2014; 46:116–125. [PubMed: 24413735]

Phillips-Cremins JE, Sauria MEG, Sanyal A, Gerasimova TI, Lajoie BR, Bell JSK, Ong CT, Hookway TA, Guo C, Sun Y, et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell. 2013; 153:1281–1295. [PubMed: 23706625]

Pope BD, Ryba T, Dileep V, Yue F, Wu W, Denas O, Vera DL, Wang Y, Hansen RS, Canfield TK, et al. Topologically associating domains are stable units of replication-timing regulation. Nature. 2014; 515:402–405. [PubMed: 25409831]

Ranganath S, Carpenter AC, Gleason M, Shaw AC, Bassing CH, Alt FW. Productive coupling of accessible Vbeta14 segments and DJbeta complexes determines the frequency of Vbeta14 rearrangement. J Immunol. 2008; 180:2339–2346. [PubMed: 18250443]

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014; 159:1665–1680. [PubMed: 25497547]

Schatz DG, Swanson PC. V(D)J recombination: mechanisms of initiation. Annu Rev Genet. 2011; 45:167–202. [PubMed: 21854230]

Teng G, Maman Y, Resch W, Kim M, Yamane A, Qian J, Kieffer-Kwon KR, Mandal M, Ji Y, Meffre E, et al. RAG represents a widespread threat to the lymphocyte genome. Cell. 2015; 162:751–765. [PubMed: 26234156]

Tepsuporn S, Hu J, Gostissa M, Alt FW. Mechanisms that can promote peripheral B-cell lymphoma in ATM-deficient mice. Cancer Immunol Res. 2014; 2:857–866. [PubMed: 24913718]

Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, Hadjur S. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep. 2015; 10:1297–1309. [PubMed: 25732821]

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013; 153:307–319. [PubMed: 23582322]

Yancopoulos GD, Desiderio SV, Paskind M, Kearney JF, Baltimore D, Alt FW. Preferential utilization of the most JH-proximal VH gene segments in pre-B-cell lines. Nature. 1984; 311:727–733. [PubMed: 6092962]

Yancopoulos GD, Alt FW. Regulation of the assembly and expression of variable-region genes. Annu Rev Immunol. 1986; 4(1):339–368. [PubMed: 3085692]

Zarrin AA, Del Vecchio C, Tseng E, Gleason M, Zarin P, Tian M, Alt FW. Antibody class switching mediated by yeast endonuclease-generated DNA breaks. Science. 2007; 315:377–381. [PubMed: 17170253]

Zha S, Guo C, Boboila C, Oksenych V, Cheng HL, Zhang Y, Wesemann DR, Yuen G, Patel H, Goff PH, et al. ATM damage response and XLF repair factor are functionally redundant in joining DNA breaks. Nature. 2010; 469:250–254. [PubMed: 21160472]

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008; 9:R137. [PubMed: 18798982]

Zhang Y, McCord RP, Ho YJ, Lajoie BR, Hildebrand DG, Simon AC, Becker MS, Alt FW, Dekker J. Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. Cell. 2012; 148:908–921. [PubMed: 22341456]

Zuin J, Dixon JR, van der Reijden MIJA, Ye Z, Kolovos P, Brouwer RWW, van de Corput MPC, van de Werken HJG, Knoch TA, van IJcken WFJ, et al. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc Natl Acad Sci USA. 2014; 111:996–1001. [PubMed: 24335803]

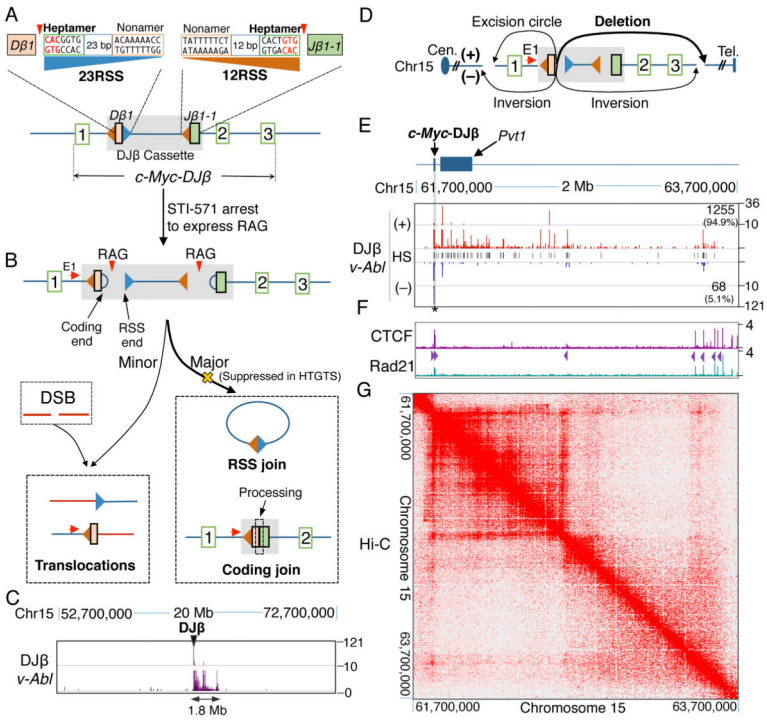

Figure 1. Abundant DSBs in the 1.8-Mb *c-Myc*-DJβ Loop Domain in *v-Abl* Pro-B Cells

(A) Diagram of *c-Myc*-DJβ and sequences of Dβ1 23RSS (blue) and Jβ1-1 12RSS (orange). Red triangles indicate RAG cleavage-sites. (B) RAG-initiated DSBs in *c-Myc*-DJβ cassette participate in cassette DJβ rearrangements but rarely to translocations involving DSBs on other chromosomes. Red arrows indicate HTGTS primer positions. (C) Linear plot with a broken *y* axis showing HTGTS junction profiles in indicated 20-Mb region containing *c-Myc*-DJβ. (D) Potential junctional outcomes between bait Dβ1 23RSS coding ends and other DSBs in *cis* include deletions, excision circles, and inversions. (E) HTGTS junction profiles in *v-Abl* cells within indicated 2-Mb region containing *c-Myc*-DJβ. For all panels, unmarked ticks represent 0. Black lines in the middle show hotspot (HS) positions (listed in Table S1). Junction numbers and percentages in + or − orientation downstream of *c-Myc*-DJβ are shown. Cassette location is shadowed in gray. Star indicates a good cryptic RSS. (F) ChIP-seq profiles of CTCF and Rad21 in the 2-Mb region defined in (E). CBE orientation is indicated by purple triangles. (G) Heat map showing the 1.8-Mb *c-Myc* loop domain defined by *in situ* Hi-C data in CH12-LX cell line. See also Figure S1 and Table S1.

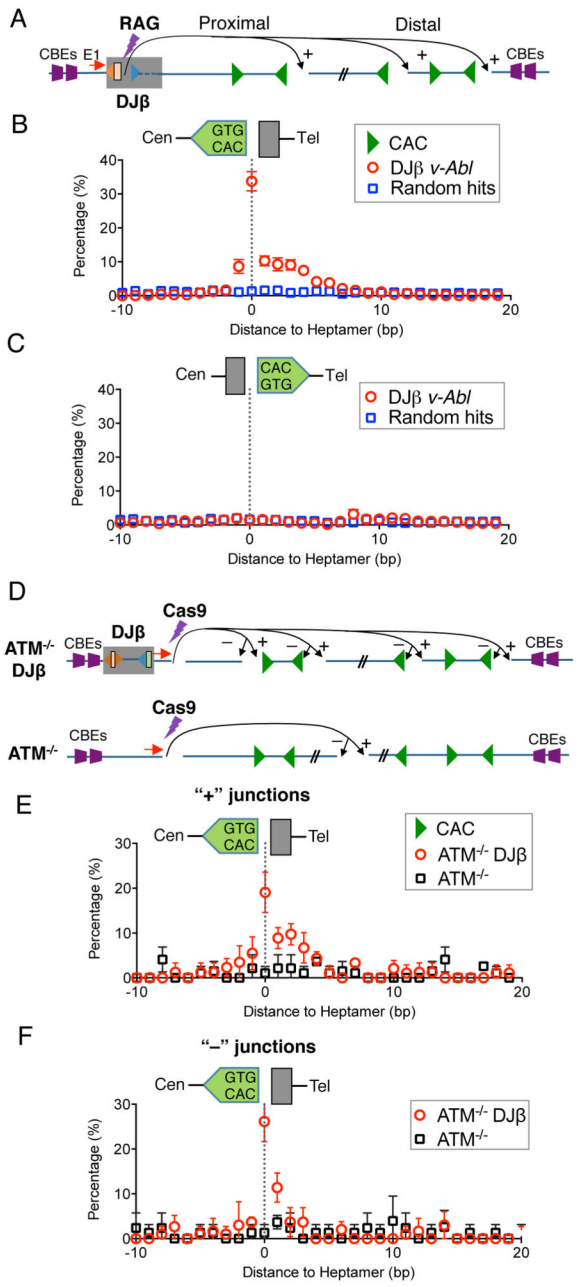

Figure 2. RAG Generates DSBs Across the 1.8-Mb *c-Myc*-DJβ Loop Domain

(A) Schematic of translocations between Dβ1 downstream coding ends and cryptic RSSs mostly represented by CAC motifs in the 1.8-Mb *c-Myc* domain in *c-Myc*-DJβ *vAbl* pro-B cells. (B, C) Distance of Dβ1 downstream coding end junctions to reverse CACs (B) or forward CACs (C) within the1.8-Mb domain in *c-Myc*-DJβ *vAbl* pro-B cells. Direct joining to the nucleotide immediately adjacent to CAC is defined as 0 in this and following panels. Mean ± SD, n=3. (D) Schematic of translocations between Cas9/gRNA-initiated bait DSBs and DSBs in the *c-Myc* domain in ATM-deficient pro-B cells with (upper) or without (lower) the *c-Myc*-DJβ cassette. (E, F) Distance of 5′Cas9 junctions in + (E) or − (F)

orientation to reverse CACs in the *c-Myc* domain in cells defined in (D). Means ± SD, n=3. See also Figure S2.

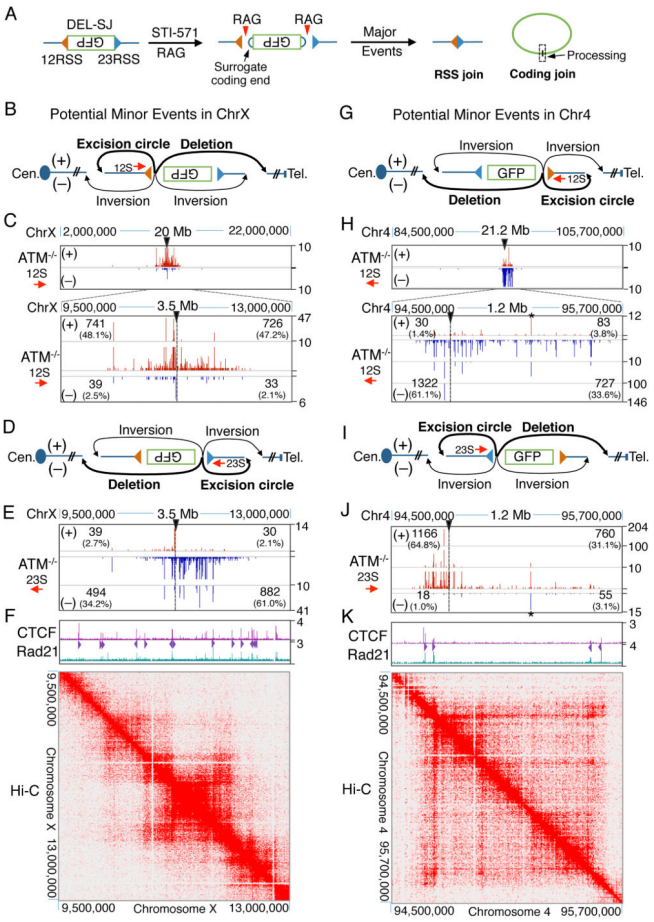

Figure 3. DSBs Restricted in Genome-wide DEL-SJ-containing Loop Domains

(A) Diagram showing major RAG-initiated joins in DEL-SJ-containing *v-Abl* pro-B cells. (B) Potential junctional outcomes between bait 12RSS and DSBs in *cis* include deletions, excision circles, and inversions. (C) Profiles of 12RSS junctions within chromosome X in ATM-deficient *vAbl* pro-B cells. Black triangles indicate insertion site of DEL-SJ. Junction numbers and percentages in + or − orientation upstream or downstream of bait 12RSS are shown separately. (D, E) Profiles of 23RSS junctions in the indicated 3.5-Mb region containing DEL-SJ on chromosome X. (F) ChIP-seq profiles of CTCF/Rad21 (upper panel) and heat map of *in situ* Hi-C (lower panel) in this 3.5-Mb region. (G–K) 12RSS and 23RSS junctions across the 1-Mb DEL-SJ-containing loop domain on chromosome 4. Stars indicate stronger cryptic RSSs. See also Figure S3 and Table S2.

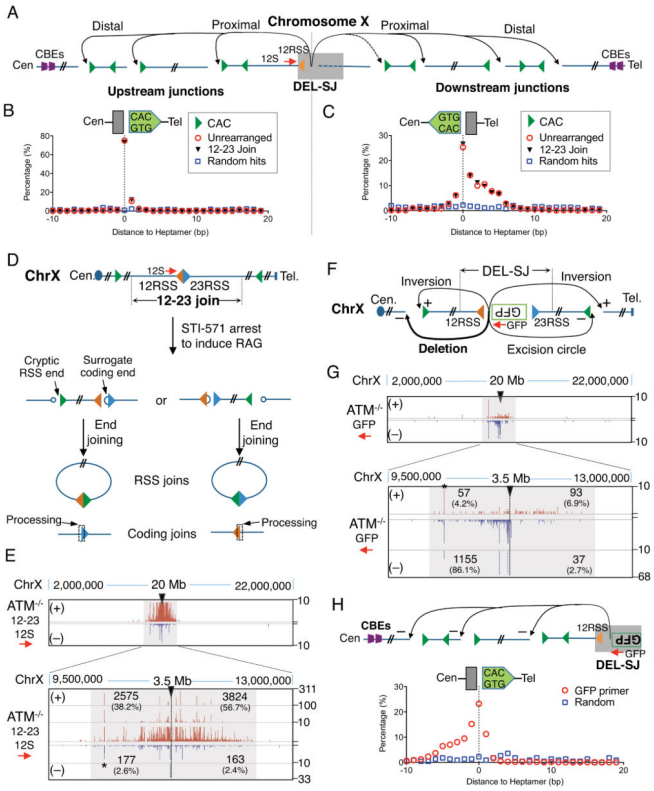

Figure 4. Orientation-biased Joining of RAG-initiated DSBs in Loop Domains

(A) Schematic of translocation between bait 12RSS and CACs within the DEL-SJ-containing loop domain on chromosome X in ATM-deficient *v-Abl* pro-B cells. (B) Distance of 12RSS junctions upstream of bait 12RSS to forward CACs; no correlation was found with reverse CACs. (C) Distance of 12RSS junctions downstream of bait 12RSS to reverse CACs; no correlation was found with forward CACs. (D) Diagram of recombination output generated by RAG re-cleavage at perfect 12–23RSS joins. (E) Profiles of 12RSS junctions of 12–23RSS join within chromosome X in ATM-deficient *vAbl* pro-B cells. Star indicates a relatively stronger cryptic RSS and loop domain is shadowed in gray, also in (G). (F) Diagram of bait surrogate coding ends associated with DEL-SJ 12RSS and the potential outcomes. (G) Profiles of GFP primer junctions within chromosome X in ATM-deficient *vAbl* pro-B cells. (H) Distance of GFP primer junctions downstream of bait surrogate coding ends to forward CACs. See also Figure S4.

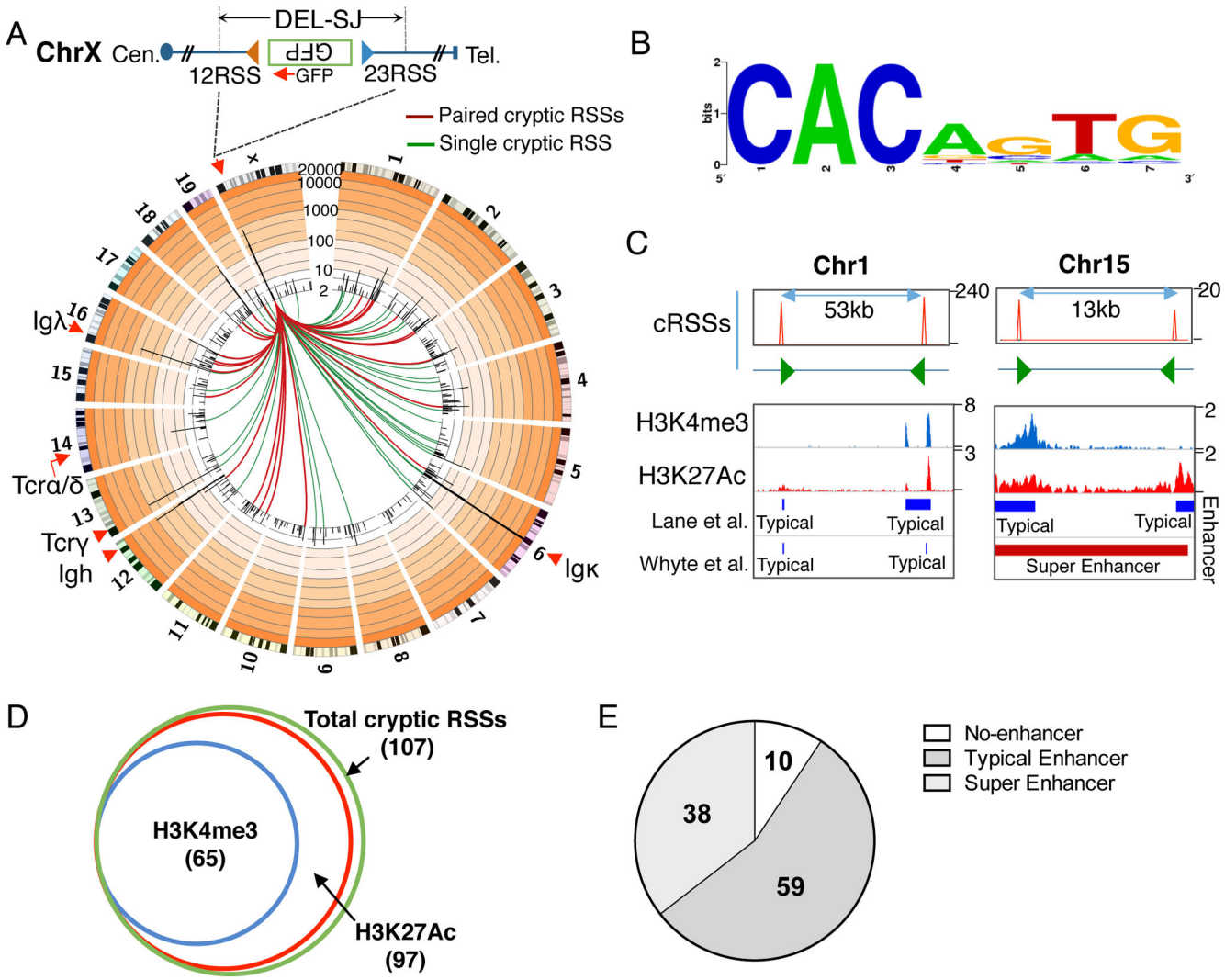

Figure 5. Genome-wide RAG Off-targets

(A) Circos plot on a custom log scale showing genome-wide translocation profile of GFP primer junctions in DEL-SJ-integrated ATM-deficient *v-Abl* pro-B cells. Bin size is 2 Mb and 50,000 HTGTS junctions are shown. Colored lines link break-site to identified cryptic RSSs. (B) Consensus sequence of cryptic RSS heptamer extrapolated from the 107 identified cryptic RSSs. (C) Paired cryptic RSSs on chromosome 1 (left panel) and 15 (right panel). Green arrows indicate the position and orientation of cryptic RSSs in these four translocation hotspots. (D) Venn diagram showing number of identified cryptic RSSs that overlap with H3K27Ac and H3K4me3. (E) Pie chart showing number of identified cryptic RSSs that overlap with typical enhancers and super-enhancers. See also Figure S5 and Table S3.

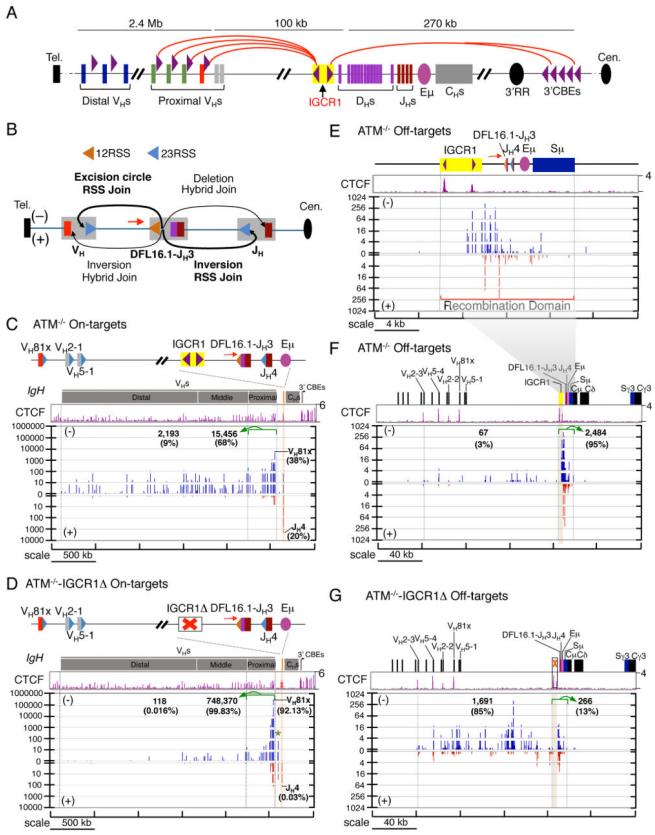

Figure 6. Effect of IGCR1 Deletion on RAG Targeting in the *IgH* Locus

(A) Schematic of murine *IgH* locus. Purple arrowheads indicate position and orientation of CBEs. Red arches show proposed looping between convergent *IgH* CBEs (Lin et al., 2015). (B) Illustration of possible joining outcomes between *bona fide* RSSs from DFL16.1 5′ RSS bait end to upstream V_H and downstream J_H broken ends. Red arrows in all panels indicate the position and orientation of HTGTS primer. (C,D) Upper panels are *IgH* annotation track. Middle panels show CTCF ChIP-seq profiles. Lower panels are pooled HTGTS junction profiles for *IgH bona fide* RSSs (n=3). Junctions are displayed as stacked tracks (log scale between tracks, linear scale within each track). Junction numbers and percentages as of total *IgH* on-targets in indicated regions are shown. We note that beyond junctions described in the text we also detected junctions between DFL16.1 5′RSS and pseudo D_H RSSs in the V_H to D_H intervening region in ATM$^{-/-}$-IGCR1Δ cells (Green star; see also Figure S6E). We also found a very small number of hybrid joins to V_H coding ends (+ joins) or J_H4 coding end (− joins). (E) Pooled RAG off-target junction profile of a 24-kb region including the recombination domain for ATM-deficient cells (n=3). (F, G) show pooled RAG off-target junction profiles of a 240-kb region including the recombination domain for ATM-deficient cells with (F) or without (G) IGCR1, respectively (n=3). Brown shadowed region marks the location of IGCR1. Junction numbers and percentages as of total *IgH* off-targets in indicated regions are shown. See also Figure S6 and Table S4.

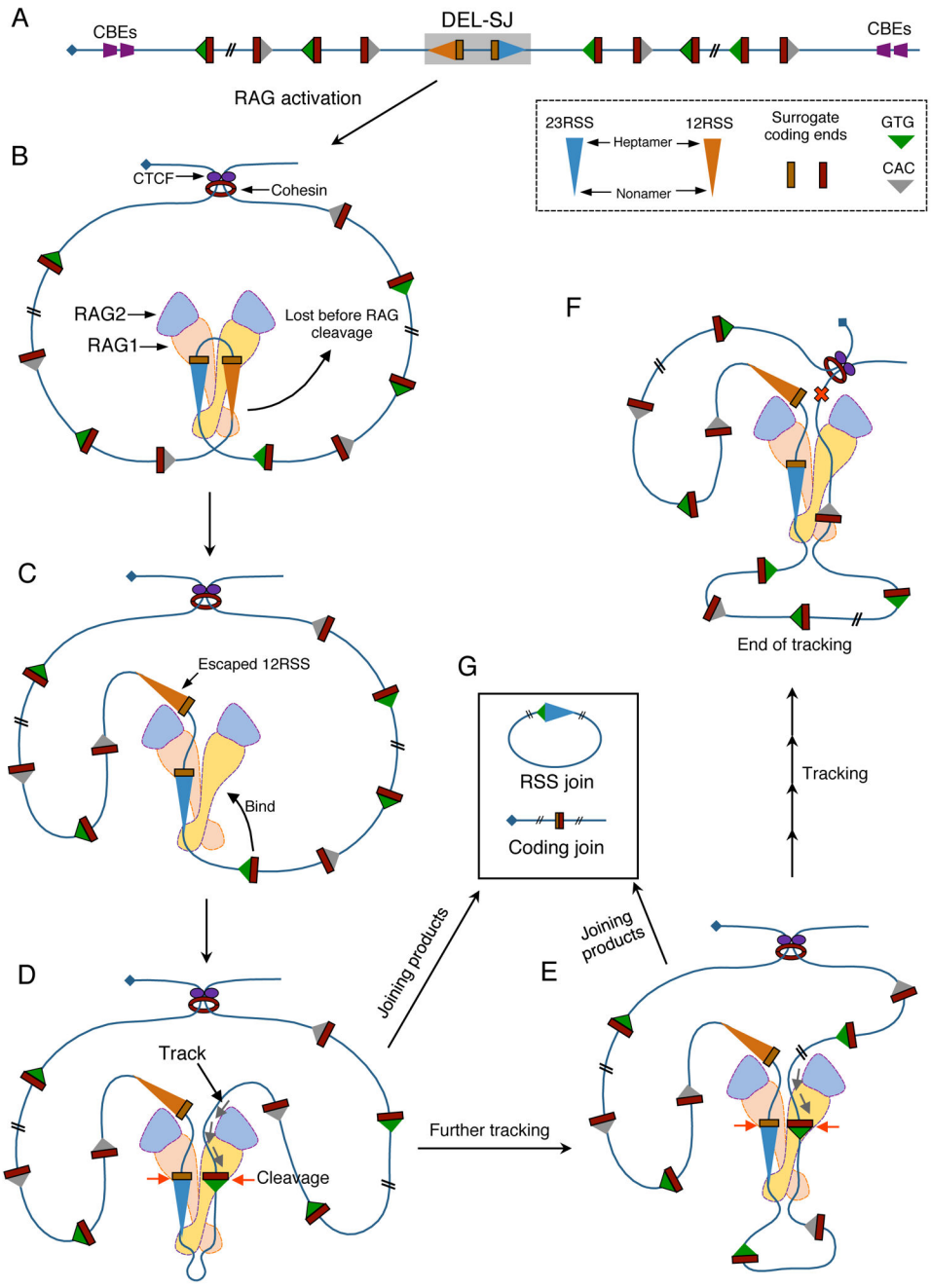

Figure 7. Linear Tracking Model for Orientation-biased Usage of CACs in the Paired *bona fide* RSSs-containing Loop Domains

(A) Linear map of a DEL-SJ-containing loop domain. (B) Activation of RAG by paired *bona fide* RSSs in the CTCF/cohesin-anchored loop domain. (C) One *bona fide* RSS may escape at a very low frequency leaving the activated complex. (D) A CAC in convergent orientation of the remaining *bona fide* RSS binds to RAG to facilitate cleavage or initiate unidirectional tracking. (E) Further tracking resulting in usage of another convergent CAC. (F) RAG tracking in the loop in either direction is stopped by the boundary formed by the CTCF/cohesin complex. (G) Joining products from pairing and cleavage between remaining

bona fide RSS and convergent CACs. More details are in the text. Various modifications of this basic tracking model can be conceived to explain joining patterns of proximal V, D, and J segments in endogenous antigen receptor loci (not shown). See also Figure S7.