HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2020 March 11.

Published in final edited form as:

Nature. 2019 September ; 573(7775): 600–604. doi:10.1038/s41586-019-1547-y.

The Fundamental Role of Chromatin Loop Extrusion in Physiological V(D)J Recombination

Yu Zhang[1,2,*,10], **Xuefei Zhang**[1,10], **Zhaoqing Ba**[1], **Zhuoyi Liang**[1], **Eddie Dring**[1], **Hongli Hu**[1], **Jiangman Lou**[1], **Nia Kyritsis**[1], **Jeffrey Zurita**[1], **Muhammad S. Shamim**[3,4,5,6], **Aviva Presser Aiden**[3,4,7], **Erez Lieberman Aiden**[3,5,8,9], **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]Immunobiology Center, Department of Biomedical Sciences, Western Michigan University Homer Stryker M.D. School of Medicine, Kalamazoo, MI 49007, USA.

[3]The Center for Genome Architecture, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA.

[4]Department of Bioengineering, Rice University, Houston, TX 77030, USA.

[5]Center for Theoretical Biological Physics and Department of Computer Science, Rice University, Houston, TX 77030, USA

[6]Medical Scientist Training Program, Baylor College of Medicine, Houston, TX 77030, USA.

Correspondence and requests for materials should be addressed to F.W.A (alt@enders.tch.harvard.edu).

AUTHOR CONTRIBUTIONS

Y.Z., X.Z., and F.W.A. designed the study; Y.Z., X.Z., Z.L., H.H., J.L, and E.D performed experiments, except for Hi-C experiments which were performed by A.P-A and analyzed by A.P-A, M.S.S. and E.L-A. Z.B. provided critical reagents and advice on 3C-HTGTS. N.K. and J.Z. designed some bioinformatics pipelines; Y.Z., X.Z., and F.W.A analyzed and interpreted all data other than Hi-C data. Y.Z., X.Z., and F.W.A. designed figures and wrote the manuscript. Z.B, Z.L, H.H, J.L, A.P-A, and E.L-A helped polish the manuscript.

[10]These authors contributed equally: Yu Zhang, Xuefei Zhang.

[*]Present address: Immunobiology Center, Department of Biomedical Sciences, Western Michigan University Homer Stryker M.D. School of Medicine, Kalamazoo, MI 49007, USA.

The authors declare no competing financial interests.

Data availability

HTGTS V(D)J-seq, Hi-C, 3C-HTGTS, GRO-Seq and ChIP-Seq sequencing data reported in this study has been deposited in the GEO database under the accession number GSE130224. Specifically, HTGTS V(D)J-seq data is deposited in the GEO database under the accession number GSE130216 and is related to Fig. 1e-h; 2b, c; 3a-c, e; 4c; Extended Data Fig. 2c-e, g, h; 3e, f; 4a, c, d; 5a-c; 6c; 7d, e; 9b; and Supplementary Information Table 1&2. Hi-C data is deposited in the GEO database under accession number GSE134543 and is related to Extended Data Fig. 8a. 3C-HTGTS data is deposited in the GEO database under the accession number GSE130214 and is related to Fig. 3f; 4d; and Extended Data Fig. 8a; 9c; 10q, r. GRO-Seq data is deposited in the GEO database under the accession number GSE130215 and is related to Fig. 3d; 4b; and Extended Data Fig. 4e; 6d; 7f; 9b. ChIP-Seq data is deposited in the GEO database under the accession number GSE130213 and is related to Extended Data Fig. 8c, d; 9d.

Code availability.

HTGTS V(D)J-seq and 3C-HTGTS data was processed through published pipeline available at (http://robinmeyers.github.io/transloc_pipeline/). Code for Hi-C data process is available at (github.com/aidenlab). GRO-Seq and ChIP-Seq were aligned to mm9 genome with bowtie2 v2.2.8 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml), processed by samtools v1.8 (https://sourceforge.net/projects/samtools/files/samtools/1.8/) and generated graph files via RSeQC tool v2.6 (http://rseqc.sourceforge.net/#bam2wig-py).

[7]Department of Pediatrics, Texas Children's Hospital, Houston, TX 77030, USA.

[8]Broad Institute of MIT and Harvard, Cambridge, MA 02139, USA

[9]Shanghai Institute for Advanced Immunochemical Studies, ShanghaiTech University, Shanghai, China

Abstract

RAG endonuclease initiates IgH locus (*Igh*) V(D)J assembly in progenitor (pro)-B cells by joining Ds to J_Hs, before joining upstream V_Hs to DJ_H intermediates[1]. In mouse pro-B cells, the CTCF-binding element (CBE)-anchored chromatin loop domain[2] at the 3'end of *Igh* contains an internal sub-domain spanning the 5'CBE anchor (IGCR1)[3], the D_Hs, and a RAG-bound recombination center (RC)[4]. The RC comprises J_H-proximal D (DQ52), 4 J_Hs, and the intronic enhancer ("iEμ")[5]. Robust RAG cleavage is restricted to paired V(D)J segments flanked by complementary recombination signal sequences (12RSSs and 23RSSs)[6]. Ds are flanked downstream and upstream by 12RSSs that, respectively, mediate deletional joining with convergently-oriented J_H-23RSSs and V_H-23RSSs[6]. Despite 12/23 compatibility, inversional D to J_H joining via upstream D-12RSSs is rare[7,8]. Plasmid-based assays attributed lack of inversional D to J_H joining to sequence-based preference for downstream D-12RSSs[9], as opposed to putative linear scanning mechanisms[10,11]. Given recent findings that RAG linearly scans convergent CBE-anchored chromatin loops[4,12-14], potentially formed by cohesin-mediated loop extrusion[15-18], we revisited a scanning role. Here, we report that J_H-23RSS chromosomal orientation programs RC-bound RAG to linearly scan upstream chromatin in the 3'*Igh* sub-domain for convergently-oriented D-12RSSs and, thereby, to mediate deletional joining of all Ds, except RC-based DQ52 that joins by a diffusion-related mechanism. In a DQ52-based RC, formed in the absence of J_Hs, RAG bound by the downstream DQ52-RSS scans the downstream constant region exon-containing 3'*Igh* sub-domain in which scanning can be impeded by targeted nuclease-dead Cas9 (dCas9) binding, by transcription through repetitive *Igh* switch sequences, and by the 3'*Igh* CBE-based loop anchor. Notably, each scanning impediment focally increases RAG activity on potential substrate sequences within the impeded region. High resolution mapping of RC chromatin interactions reveals that such focal RAG targeting is associated with corresponding impediments to the loop extrusion process that drives chromatin past RC-bound RAG.

RAG comprises two catalytic RAG1 and two cofactor RAG2 proteins (Extended Data Fig. 1a)[19,20]. We tested the hypothesis that upon RAG acquisition of a J_H-23RSS in one active site, the J_HRC serves as a dynamic sub-loop anchor to promote loop extrusion-based presentation of predominately convergent D-12RSSs to the other active site, thereby, mediating deletional D to J_H recombination (Extended Data Fig. 1b-e; Supplementary Video). We first tested this hypothesis by mutational analyses of impact of D-RSS orientation on deletional versus inversional D to J_H rearrangement within a physiological D-J_H-RC-containing chromosomal domain (Fig. 1a-c). To facilitate analyses, we employed Cas9/gRNA targeting to delete the D_H-J_H-RC domain on one allele of a *v-Abl* transformed, RAG2-deficient pro-B line[21], referred to as "D_H-$J_H^{+/-}$" parental line (Extended Data Fig. 2a-c). To activate V(D)J recombination, we introduced RAG2 into D_H-$J_H^{+/-}$ cells or mutant derivatives and treated them with *v-Abl* kinase inhibitor to induce G1 arrest, RAG2-stabilization, and robust D to J_H joining potential[22]. Experiments had at least three repeats

and used multiple independent mutant D_H-$J_H^{+/-}$ derivatives. We employed HTGTS V(D)J-Seq[4] to analyze V(D)J junctions with a J_H1 coding end (CE) primer, which revealed that junctions were overwhelmingly deletional DJ_H junctions (Extended Data Fig. 2d-g). Similar to primary pro-B cells[23], J_H-distal DFL16.1 had highest rearrangement frequency (66%), J_H-proximal DQ52 had second highest frequency (27%), and the 7 ("intervening") Ds between DF16.1 and DQ52 had lower rearrangement frequency (Extended Data Fig. 2e).

To test impact of DFL16.1-12RSS orientation on D to J_H recombination, we separately inverted its downstream RSS ("RSS-DN") and upstream RSS ("RSS-UP") in the D_H-$J_H^{+/-}$ line (Fig. 1d-g). Inversion of DFL16.1-RSS-DN, placing it in the same orientation as DFL16.1-RSS-UP, made it essentially inert for J_H joining (Compare Fig. 1e and Fig. 1f). In contrast, inversion of DFL16.1-RSS-UP led to robust deletional J_H joining to the surrogate CE sequences (non-D upstream flanking sequences), with levels similar to those of deletional joins mediated by DFL16.1-RSS-DN (Fig. 1g). To rule out adjacent coding sequence impacts[24], we inverted DFL16.1, including both the D-RSS-UP and D-RSS-DN (Fig. 1h). DFL16.1-RSS-UP, in the downstream position convergent to the J_H-23 RSSs, mediated robust deletional joining; while the DFL16.1-RSS-DN in the upstream position in the same orientation as the J_H-23RSS had less than 2% of normal activity (Fig. 1h). The 8 unmodified Ds had little change in rearrangement patterns (Extended Data Fig. 2h). Therefore, recognition of the DFL16.1-RSS-DN, due to convergent orientation with the J_H-23RSS, prescribes deletional-orientation DFL16.1 to J_H joining and relative RSS-DN versus RSS-UP strength does not majorly impact this process (Extended Data Fig. 3a; Supplementary Discussion).

To further test impact of D-RSS orientation on D to J_H joining, we eliminated potential confounding effects of chromosomal or extra-chromosomal secondary joins (Extended Data Fig. 3b) by deleting the J_H2-4 sequence of the D_H-$J_H^{+/-}$ line to generate the "D_H-$J_H1^{+/-}$" line, which undergoes D to J_H1 recombination similarly to its D_H-$J_H^{+/-}$ parent (Extended Data Fig. 3c; Supplementary Information Table 1). We then inverted the region containing all 7 intervening Ds in the D_H-$J_H1^{+/-}$ line (Fig. 2a; Extended Data Fig. 3d). Analyses of J_H1 rearrangements in this line revealed greatly increased relative utilization of each D-RSS-UP in convergent orientation with the J_H-23RSS; and, correspondingly, decreased utilization of each D-RSS-DN when in same orientation as J_H1-23RSS (Fig. 2b, Extended Data Fig. 3e). As a control, predominant deletional DFL16.1 and DQ52 D-RSS-DN utilization was unchanged (Fig. 2b, Extended Data Fig. 3e). Utilization of most inverted intervening D-RSS-UPs was lower than that of D-RSS-DNs in normal position (Fig. 2b; Extended Data Fig. 3a, e; Supplementary Discussion). Regardless, both the D-RSS-DN and D-RSS-UP of these 7 Ds are far more highly utilized when in convergent orientation with J_H-23RSS, supporting a major role for RAG scanning in deletional joining (Extended Data Fig. 1c-e).

Cryptic RSSs within several kb of ectopic RCs are joined in either orientation by accessing RAG via diffusion[12]. Thus, given close proximity of DQ52 to J_Hs (Fig. 1a), both of its RSSs theoretically could similarly access RC-bound RAG (Extended Data Fig. 1f-h). To elucidate how overwhelmingly deletional DQ52 joining occurs, we inverted DQ52 and its RSSs in the D_H-$J_H1^{+/-}$ line (Fig. 2c). Strikingly, the vast majority of inverted DQ52 to J_H1 joins were mediated by the DQ52-RSS-DN and occurred by inversion (Fig. 2c; right panel). Thus, in

the RC location, DQ52-RSS-DN is much stronger than DQ52-RSS-UP for mediating D to J_H joining, allowing it to enforce deletional rearrangement by a sequence-based mechanism[9]. Studies of DFL16.1 in place of DQ52 further confirmed the need for an RSS-based mechanism to promote deletional joining of a RC-based D (Extended data Fig. 3f). We also replaced DFL16.1 with DQ52 in normal or inverted orientation (Fig. 2d). In this location, DQ52 in normal orientation was utilized similarly to endogenous DFL16.1, with joins overwhelmingly deletional (Fig. 2d; left panel). When inverted in DFL16.1 location, DQ52 joining was reduced; but, remarkably, the "weak" DQ52-RSS-UP predominated over the inverted "strong" DQ52-RSS-DN to generate deletional joins (Fig. 2d; right panel), confirming the major role for RAG scanning, versus RSS sequence, in enforcing deletional joining of Ds distal to the RC (See Extended Data Fig. 3a). Finally, low level joining of the inverted DQ52-RSS-DN in the DFL16.1 position may reflect accessing RC-bound RAG when brought into diffusion distance via loop extrusion (Fig. 2d; Extended Data Fig. 3g; Supplementary Discussion).

A nascent *Igh* RC forms in active chromatin over DQ52, J_Hs, and iEμ (Fig. 1a). RAG recruitment poises the RC for D to J_H joining[4]. To characterize RC function in RAG scanning, we deleted all J_Hs from the D_H-$J_H^{+/-}$ line to generate the "$J_H\Delta$" line, which lacks any *bona fide* 23RSSs within the 3'*Igh* domain for pairing/joining with D-12RSSs and forms a new DQ52-based RC from which upstream and downstream RAG scanning is readily detectable (Extended Data Fig. 4a-d; Supplementary Discussion). HTGTS V(D)J-Seq on the $J_H\Delta$ line revealed that DQ52-RSS-UP initiates RAG scanning to convergent cryptic RSSs within the RC-upstream D sub-domain, with robust "RAG cryptic scanning activity" at the transcribed heptamer of the non-12/23 compatible DFL16.1-RSS-DN and convergent CAC within D_H3-2-RSS-UP (Fig. 3a, c; Extended Data Fig. 4c, 5a; Supplementary Information Table 2). DQ52-RSS-DN initiated RAG cryptic scanning activity across the downstream constant region exon (C_H)-containing sub-domain, with robust activity at cryptic heptamers within the repetitive Sγ2b switch (S) region upstream of Cγ2b and in the 3'*Igh* CBEs anchor[25] (Fig. 3b, c; Extended Data Fig. 4d, 5b; Supplementary Information Table 2). While RAG scanning activity in Sγ2b coincided with robust transcription from the immediately upstream Iγ2b promoter[26], 3'CBEs RAG targets were only weakly transcribed (Fig. 3d, Extended Data Fig. 4e). Very low-level RAG cryptic activity occurred in the RC-upstream domain with the DQ52-RSS-DN bait (Fig. 3b; Extended Data Fig. 4d, 5c; Supplementary Discussion). We confirmed RSS orientation-mediated directional scanning in independent *v-Abl* pro-B lines with normal or inverted DFL16.1-J_H4 joins; found the latter lines lack Sγ2b transcription and corresponding Sγ2b RAG-scanning activity and that deletion of 3'CBEs in them relocated RAG scanning activity to downstream regions (Extended Data Fig. 6a-d). Overall, we conclude that chromosomal orientation of an RSS captured by RC-bound RAG determines upstream versus downstream scanning (Extended Data Fig. 4f-m).

Focal RAG downstream scanning activity from DQ52-RSS-DN in the $J_H\Delta$ line provided a system to further characterize mechanism. We asked whether introducing sequential sites of dCas9[27] generates a non-CBE-based scanning impediment. We targeted dCas9 to the repetitive Sγ1 that lies on the scanning path between the RC and the Sγ2b and 3'CBEs targets via an Sγ1-sgRNA that binds 16 sites within a 4kb portion of Sγ1 on the intact $J_H\Delta$ allele (Extended Data Fig. 7a). We derived multiple independent clones with stable dCas9

expression ("$J_H\Delta$-dCas9" lines) or with both dCas9 and Sγ1-sgRNA expression ("$J_H\Delta$-dCas9-Sγ1-sgRNA" lines; Extended Data Fig. 7b, c). HTGTS V(D)J-Seq with a DQ52-RSS-DN-primer confirmed RAG downstream scanning in multiple $J_H\Delta$-dCas9 lines with junction profiles similar to those of the $J_H\Delta$ line, including accumulation at Sγ2b and 3'CBEs (Fig. 3e; Extended Data Fig. 7d). Strikingly, $J_H\Delta$-dCas9-Sγ1-sgRNA lines had highly diminished RAG scanning downstream of the dCas9-targeted Sγ1, along with modestly decreased Sγ2b transcription (Fig. 3e; Extended Data Fig. 7d, f; Supplementary Discussion). In $J_H\Delta$-dCas9-Sγ1-sgRNA lines, we also observed substantially increased RAG scanning activity at cryptic targets in the dCas9-binding portion of Sγ1 and a modest increase at Sμ (Fig. 3e; Extended Data Fig. 7d, e). These findings indicate that dCas9 binding impedes RAG downstream scanning.

Hi-C analyses of $J_H\Delta$-dCas9 versus $J_H\Delta$-dCas9- Sγ1-sgRNA lines revealed that chromatin loops spanning the Sγ1 impediment in the latter were weakened, with new loops formed between the Sγ1 impediment and upstream RC or downstream 3'CBEs loop anchor (Extended Data Fig. 8a). Sensitive 3C-HTGTS on $J_H\Delta$-dCas9 lines revealed that iEμ robustly interacted with major RAG scanning targets including IGCR1, D_H3-2, Sγ2b, and 3'CBEs locales (Fig. 3f, upper panel; Extended Data Fig. 8b). In $J_H\Delta$-dCas9-Sγ1-sgRNA cells, iEμ gained robust interactions with dCas9-bound Sγ1 and had decreased interactions with downstream Sγ2b and 3'CBEs (Fig. 3f, lower panel, Extended Data Fig. 8b). Thus, the dCas9 impediment decreased RAG scanning activity at downstream regions in association with their decreased interaction with the RC. In $J_H\Delta$-dCas9- Sγ1-sgRNA lines, incomplete scanning inhibition downstream of Sγ1, along with broad RAG scanning activity and RC interactions across Sγ2b, indicates dynamic extrusion of Sγ2b across the RC that is impeded, but not abrogated, by Sγ1 dCas9 binding (Extended Data Fig. 10a-d, e-i; Supplementary Discussion). The greater effect of the Sγ1 dCas9 impediment on RAG scanning versus downstream interactions, with the latter done in RAG2-deficient cells, might reflect further inhibited extrusion of dCas9-bound Sγ1 chromatin past a RAG-bound RC (Extended Data Fig. 10j-l).

ChIP-Seq of $J_H\Delta$-dCas9-Sγ1 cells revealed strong binding of RAD21 cohesin subunit[18] at the IGCR1 and 3'*Igh* CBEs loop anchors and lower accumulation across transcribed iEμ/Sμ and Iγ2b/Sγ2b sequences (Extended Data Fig. 8c). In addition, a new RAD21 peak occurred at the impeded dCas9-bound Sγ1 in $J_H\Delta$-dCas9- Sγ1-sgRNA cells (Extended Data Fig. 8c). Furthermore, NIPBL, a cohesin-loading factor[18], accumulated across transcribed iEμ/Sμ and Iγ2b/Sγ2b sequences and downstream *Igh* regions including 3'CBEs. There was major additional accumulation of NIPBL at the non-transcribed Sγ1 in $J_H\Delta$-dCas9-Sγ1-sgRNA cells (Extended Data Fig. 8d), raising the possibility that dCas9 binding, beyond direct steric interference[27], may impede scanning-related extrusions via a mechanism involving increased cohesin loading at this ectopic site. These findings are consistent with a role for cohesin in loop-extrusion mediated RAG scanning.

We deleted the Iγ2b-promoter in $J_H\Delta$-dCas9 lines to test whether transcription targets RAG scanning activity at Sγ2b. This deletion abrogated constitutive Sγ2b transcription and, correspondingly RAG scanning activity, iEμ/RC interactions, and RAD21 accumulation associated with Sγ2b (Fig. 4a-d; Extended Data Fig. 9a-d). Moreover, inactivation of Sγ2b

transcription led to increased RAG activity at the downstream 3'CBEs, consistent with eliminating an upstream scanning impediment (Fig. 4c; Extended Data Fig. 9b). These findings indicate that transcription through Sγ2b impedes loop extrusion-mediated RAG scanning and that such impeded extrusion targets RAG activity to Sγ2b substrates by generating increased RC interactions. Again, Sγ2b transcription is not an absolute barrier, as RAG scanning activity at, and RC interaction with, 3'CBEs occurs in Sγ2b-transcribing cells (Fig. 4c, d; Extended Data Fig. 9b, c, Extended Data Fig. 10a-d, m-p). Elimination of transcription also might decrease RAG activity on RC-aligned targets by chromatin accessibility mechanisms[28].

We implicate a crucial role for loop extrusion-mediated RAG scanning in initiation of physiological D to J_H joining (Supplementary Video). During linear RAG scanning, downstream D-RSSs convergently-oriented with initiating RAG-bound J_H-RSSs are highly preferred for recruitment into the open RAG active site for deletional joining. Preferred use of convergent RSSs is an intrinsic property of linear RAG scanning, as it also is observed for convergent cryptic RSS utilization during RAG scanning from ectopic RCs in non-antigen receptor loop domains[12]. During scanning, loop extrusion impediments, including CBE anchors, transcription, and dCas9 binding focus RAG to targets within impeded regions. Robust DFL16.1 RSS-DN utilization correlates with location just downstream of IGCR1 CBE anchors, which impede extrusion-mediated RAG scanning, leading to strong interactions with the RC. Low, but significant, intervening D utilization may be promoted by location in an anti-sense-transcribed, repetitive region[29] that modestly impedes loop extrusion and increases accessibility to the RC (Extended Data Fig. 10q, r; Supplementary Discussion). Conversely, loop-extrusion also may frequently isolate DQ52 from the RC, preventing it from dominating overall D to J_H joining via diffusion. Finally, our dCas9 findings suggest that additional, yet to be defined, chromatin-based mechanisms may enhance synapsis of functional cis-elements via loop extrusion more generally.

METHODS

Experimental procedures.

No statistical methods were used to predetermine sample size. Experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Generation of mutant *v-Abl* pro-B cell lines.

CRISPR/Cas9 approach[30] was employed to generate the various mutant strains in this study. The D_H-$J_H^{+/-}$ line was derived from a previously described *Rag2*$^{-/-}$ *Eμ-Bcl2*-expressing *v-Abl* pro-B cell line[21] with a C57BL/6, 129/Sv mixed background. We deleted the entire D-J_H-RC region (from ~ 400 bp upstream of DFL16.1 to ~ 400 bp downstream of iEμ) on the 129/Sv allele, leaving the C57BL/6 allele intact and confirmed the deletion via Southern blotting. The D_H-$J_H^{+/-}$ line was served as a parental line to generate many mutant derivatives with at least two independent clones obtained for each. Specifically, CRISPR/Cas9 targeting was used to generate deletional and inversional mutations including J_H2-4 deletion, intervening D_Hs inversion, J_H1-4 deletion (all three mutations were confirmed via

Southern blotting) and Iγ2b deletion (confirmed via PCR genotyping followed by Sanger sequencing). CRISPR/Cas9 targeting combined with short single-stranded DNA oligonucleotide (ssODN) templates[31] was used to generate precise mutations including DFL16.1-RSS-UP inversion, DFL16.1-RSS-DN inversion, DFL16.1 inversion, DQ52 inversion, DQ52 or DQ52 inversion in place of DFL16.1, and DFL16.1 or DFL16.1 inversion in place of DQ52. These mutations were confirmed via PCR genotyping followed by Sanger sequencing.

The DFL16.1J_H4^{inv} lines were derived from the DFL16.1J_H4 line[4] via inversion of a ~ 1 kb region containing the DFL16.1J_H4 join (verified via Southern blotting). The DFL16.1J_H4^{inv} lines were used as parental lines to generate DFL16.1J_H4^{inv} 3'CBEs$^{-/-}$ lines by deleting the ~ 9 kb *Igh* 3'CBEs region containing all 10 CBEs (verified by Southern blotting) and DFL16.1J_H4^{inv} *Rag2*$^{-/-}$ lines by deleting the coding exon of RAG2 (verified by PCR genotyping and Sanger sequencing). At least two independent clones were obtained for each mutation for analysis.

The *v-Abl* pro-B cell lines were cultured in RPMI medium with 15% FBS (v/v). Cells were not tested for mycoplasma contamination. Information for sgRNAs and ssODN sequences are listed in Supplementary Information Table 3. Original gel scans for related Southern blotting confirmation in Extended Data Figures can be found in Supplementary Information Figure.

Generation of *v-Abl* lines with targeted dCas9-binding to Sγ1 region.

To introduce targeted dCas9 binding to Sγ1 region, we first generated the dCas9-expressing J_HΔ-dCas9 lines. We swapped the ORF of puromycin-resistant gene with that of neomycin-resistant gene on the retroviral pMSCV-dCas9 vector (Addgene, 44246) and transfected the modified vector into the J_HΔ line. Infected cells were selected with 1,600 ug/ml geneticin (Life technologies, 11811-031) 2 days post-infection at a concentration of 100 cell/well in 96 well plates. Selection was maintained for 8 to10 days until stable colonies appeared. Geneticin-resistant colonies were screened for dCas9 expression via Western blotting with Cas9 antibody (Diagenode, C15310258), using β-Actin antibody (Cell Signaling Technology, 3700S) as a loading control. Positive colonies were further sub-cloned and verified via Western blotting for Cas9 expression to generate the J_HΔ-dCas9 lines, which were then maintained in RPMI medium with 400 ug/ml geneticin. We then used the J_HΔ-dCas9 lines as parental lines to generate the J_HΔ-dCas9-Sγ1-sgRNA lines expressing both dCas9 and Sγ1-sgRNA. We swapped the ORF of puromycin-resistant gene with that of bleomycin-resistant gene on a lentiviral Sγ1-sgRNA expression vector (Addgene, 44248) and transfected the modified vector into the J_HΔ-dCas9 lines. Infected cells were selected with 800 ug/ml zeocin (ThermoFisher Scientific, R25001) 2 days post-infection at the concentration of 100 cell/well in 96 well plates. Selection was maintained for 8 to10 days until stable zeocin-resistant colonies appeared. Zeocin-resistant colonies were screened for Sγ1-sgRNA expression via RT-PCR. Positive colonies were further subcloned and verified via RT-PCR for Sγ1-sgRNA expression to obtain the J_HΔ-dCas9-Sγ1-sgRNA lines, which were then maintained in RPMI medium with 400 ug/ml geneticin and 400 ug/ml zeocin.

Original gel scans for Western blotting and RT-PCR confirmation in Extended Data Figures can be found in Supplementary Information Figure.

RAG complementation.

The RAG2 expressing vector pMSCV-FLAG-RAG2-GFP was generated by cutting out the FLAG-RAG2-GFP sequence from the shuttle vector pSP72-FLAG-RAG2^{WT}-GFP[21] via HpaI and XhoI digestion and cloning the sequence into the pMSCV-puro vector (Addgene, K1062-1) via the same restriction sites. RAG2 was reconstituted in RAG2-deficient *v-Abl* cells via retroviral infection of cells with the pMSCV-FLAG-RAG2-GFP vector followed by 3 days of puromycin selection to enrich for cells with virus integration.

HTGTS V(D)J-seq library preparation.

HTGTS V(D)J-seq libraries were prepared as described previously[32]. Genomic DNA was extracted from RAG2-complemented cells arrested in G1 for 4 days by treatment with 3 mM STI-571. Briefly, 10 ug DNA was fragmented via sonication on a Diagenode bioruptor and subjected to linear PCR amplification with a biotinylated primer. Single-stranded PCR products were purified via Dynabeads MyONE C1 streptavidin beads (Life Technologies, 65002) and ligated to bridge adaptors. Adaptor-ligated Products were amplified via nested PCR with indexed locus-specific primers and primer annealed to adaptor. The PCR products were further tagged with Illumina sequencing adaptor sequences, size-selected via gel extraction and loaded to Mi-Seq™ machine (Illumina) for paired-end 250 bp or 300 bp sequencing. Primer information can be found in the Supplementary Information Table 3.

HTGTS V(D)J-seq data processing.

HTGTS V(D)J-seq libraries were processed via a previously described pipeline[32]. Sequencing reads were aligned to either mm9 genome or modified mm9 genomes as indicated below. Duplicates were included for analysis as described previously[12]. In addition, since V(D)J junctions are normally processed through classical non-homologous end joining repair pathway without the involvement of long homology-mediated repair, junctions with long microhomology (> 5 bp) were excluded from analysis to avoid potential PCR artifacts.

HTGTS V(D)J-seq analysis of deletional and inversional D to J_H recombination.

A J_H1 CE primer was used as bait primer to detect D to J_H joining events. Libraries were size-normalized to total junctions of the smallest library within the set of libraries being compared. Utilization frequency of D-RSS-UPs and D-RSS-DNs was determined by counting number of junctions containing corresponding RSS-associated coding sequences within size-normalized libraries. As RSS-proximal coding nucleotides are frequently processed during CE joining and are often absent in the final junctions, we used the central 10 bp coding sequences of all D_Hs (16-23 bp) except DQ52 to represent the corresponding D_Hs for counting. DQ52 is a shorter D_H (11 bp) and thus we used the central 7 bp DQ52 coding sequence to represent DQ52 for counting. D_H2-5 and D_H2-6 share the same coding sequence and their utilization were counted together. In case of DFL16.1-RSS-UP and DFL16.1-RSS-DN inversions that use the non-D flanking sequences as the surrogate CEs for

the inverted D-RSSs, the 10 bp surrogate CE sequence lying 6 bp upstream or downstream of the predicted break sites were used to calculate the utilization of the corresponding inverted DFL16.1-RSSs. Considering resection and nucleotide addition at the break sites during V(D)J recombination, we included sequences within a 70 bp window across the predicted bait J_H1 break site to locate junctional D_H sequences for counting. Libraries were aligned to mm9 genome for D_H-$J_H^{+/-}$ and D_H-$J_H1^{+/-}$ lines. For strains harboring specific D_H mutations, libraries were aligned to modified mm9 genomes that replaced normal mm9 sequence with the modified D_H sequences. Specifically, for DFL16.1$^{RSS\text{-}UP\text{-}inv}$, chr12:114,720,404-114,720,436 in mm9 was replaced with the sequence "GCTTTTTGTGAAGGGATCTACTACTGTGgatc" ("mm9_DFL16.1-RSS-UP-inversion"); for DFL16.1$^{RSS\text{-}DN\text{-}inv}$, chr12:114,720,344-114,720,380 in mm9 was replaced with the sequence "cgcacaatgCACAGTGCTATATCCATCAGCAAAAACC" ("mm9_DFL16.1-RSS-DN-inversion"); for DFL16.1^{inv}, chr12:114,720,404-114,720,380 in mm9 was replaced with the sequence "cgcacaatgGCTTTTTGTGAAGGGATCTACTACTGTGTTTATTACTACGGTAGTAGCTACCACAGTGCTATATCCATCAGCAAAAACCggatc" ("mm9_DFL16.1-inversion"); for intervening D_H inversion, chr12:114,685,205-114,719,425 in mm9 was replaced with reverse complemented sequence of the same region ("mm9_DH-cluster-inversion"); for DQ52^{inv}, chr12:114,668,688-114,668,784 in mm9 was replaced with the sequence "gggctggagagctccaaacagaaGGTTTTGACTAAGCGGAGCACCACAGTGCTAACTGGGACCACGGTGACACGTGGCTCAACAAAAACCttgcagg" ("mm9_DQ52-inversion"); for DQ52Δ DFL16.1^{DQ52}, chr12:114,720,404-114,720,380 in mm9 was replaced with the sequence "cgcacaatgGGTTTTTGTTGAGCCACGTGTCACCGTGGTCCCAGTTAGCACTGTGGTGCTCCGCTTAGTCAAAACCggatc" ("mm9_DQ52-replace-DFL16.1"); for DQ52Δ DFL16.1$^{DQ52\text{-}inv}$, chr12:114,720,404-114,720,380 in mm9 was replaced with the sequence "cgcacaatgGGTTTTGACTAAGCGGAGCACCACAGTGCTAACTGGGACCACGGTGACACGTGGCTCAACAAAAACCggatc" ("mm9_DQ52inv-replace-DFL16.1"); for DFL16.1Δ DQ52$^{DFL16.1}$, chr12:114,668,688-114,668,784 in mm9 was replaced with the sequence "gggctggagagctccaaacagaaGGTTTTTGCTGATGGATATAGCACTGTGGTAGCTACTACCGTAGTAATAAACACAGTAGTAGATCCCTTCACAAAAAGCttgcagg" ("mm9_DFL16.1-replace-DQ52"); for DFL16.1Δ DQ52$^{DFL16.1\text{-}inv}$, chr12:114,668,688-114,668,784 in mm9 was replaced with the sequence "gggctggagagctccaaacagaaGCTTTTTGTGAAGGGATCTACTACTGTGTTTATTACTACGGTAGTAGCTACCACAGTGCTATATCCATCAGCAAAAACCttgcagg" ("mm9_DFL16.1inv-replace-DQ52").

HTGTS V(D)J-seq analysis of RAG cryptic scanning activity from DQ52-RSSs.

5'DQ52 (DQ52-RSS-UP) (137 bp upstream of DQ52-RSS-UP break site) and 3'DQ52 (DQ52-RSS-DN) bait primers (145 bp downstream of DQ52-RSS-DN break site) were used to analyze the DQ52-RSS-UP and DQ52-RSS-DN joining profiles. The 5'DQ52 primer can simultaneously detect DSB ends joining to DQ52-RSS-UP signal ends (SEs) and DQ52-RSS-DN coding ends (CEs). Similarly, the 3'DQ52 primer can simultaneously detect DSB ends joining to DQ52-RSS-DN SEs and DQ52-RSS-UP CEs. To compare RAG cryptic scanning profiles of DQ52-RSS-UP and DQ52-RSS-DN, we extracted DSB ends joining to

the same type of RAG break ends for the two RSSs from HTGTS V(D)J-seq libraries. As such, we plotted the DQ52-RSS-UP SE junctions extracted from the 5'DQ52 primer libraries and DQ52-RSS-DN SE junctions extracted from the 3'DQ52 primer libraries. Junctions were plotted via Prism. Junctions are denoted as in ''+'' orientation if prey sequence reads in centromere-to-telomere direction and in ''-'' orientation if prey sequence reads in telomere-to-centromere direction. For DQ52-RSS-UP SE joining profiles, "-" junctions are deletions and "+" junctions are inversions. For DQ52-RSS-DN SE joining profiles, "+" junctions are deletions and "-" junctions are inversions. Note that although not shown, CE joining profiles showed very similar patterns of RAG targeting as that of SE joining profiles of the same RSS. We used coordinates of bait length to extract SE versus CE junctions from a given primer with criteria similar to those described previously[13]. Thus, taking into account potential processing of break ends, junctions with bait length from several nucleotides beyond the predicted break sites and several nucleotides downstream of the break sites were included for analysis. As such, junctions with bait length 134-139 bp were used for analysis of DQ52-RSS-UP SE profiles from the 5'DQ52 primer libraries; junctions with bait length 142-147 bp were used for analysis of DQ52-RSS-DN SE profiles from the 3'DQ52 primer libraries; junctions with bait length of 140-150 bp were used for analysis of DQ52-RSS-UP CE profiles from 5'DQ52 primer libraries; junctions with bait length 148-158 bp were used for analysis of DQ52-RSS-DN CE profiles from 3'DQ52 primer libraries. We included a larger bait length range for CE versus SE analysis due to more extensive end processing during CE joining.

Normalization of HTGTS V(D)J-seq libraries for RAG cryptic scanning activity analysis.

For DQ52-RSS-UP scanning activity analysis, DQ52-RSS-UP SE junctions were isolated from 5'DQ52 primer libraries and each library was normalized to 2,400 isolated junctions (Fig. 3a; Extended Data Fig. 4c; 5a). For DQ52-RSS-DN scanning activity analyses, DQ52-RSS-DN SE junctions were isolated from the 3'DQ52 primer libraries and normalized to 2,400 DQ52-RSS-UP CE junctions isolated from the same 3'DQ52 primer libraries (Fig. 3b, e; 4c; Extended Data Fig. 4d; 5b, c; 7d, e; 9b).

Analysis of HTGTS V(D)J-seq libraries from DFL16.1J_H4 line and its derivatives.

A bait primer 102 bp upstream of the DJ_H-RSS break site was used to generate the libraries for the DFL16.1J_H4 line and its derivatives. Sequencing reads were aligned to the modified mm9 genome harboring the DFL16.1J_H4 join. Specifically, chr12:114,666,771-114,720,401 in mm9 genome was replaced with the sequence "CCCCT" ("mm9_DFL16.1JH4"). Libraries were normalized to total bait aligned reads as previously described[4,12].

GRO-Seq library preparation and analysis.

GRO-Seq libraries were prepared from cells arrested in G1 for 4 days by 3 uM STI-571 and were generated as previously described[33] with slight modifications. Briefly, 10 million cells were permeabilized with permeabilization buffer (10 mM Tris-HCl, pH 7.4, 300 mM sucrose, 10 mM KCl, 5 mM $MgCl_2$, 1 mM EGTA, 0.05% Tween-20, 0.1% NP40 substitute, 0.5 mM DTT, one tablet of protease inhibitors per 50 ml and 4 units Rnase inhibitor per ml) and resuspended in 100 ul storage buffer (10 mM Tris-HCl, pHHH 8.0, 25% (vol/vol) glycerol, 5 mM $MgCl_2$, 0.1 mM EDTA and 5 mM DTT)[34]. Permeabilized cells were

subjected to nuclear run-on at 37°C for 5 min to incorporate BrdU into nascent transcribed RNA, followed by Trizol-based RNA extraction. Extracted RNA was hydrolyzed with NaOH (Final 0.2 N) on ice for 12 min, quenched by ice-cold Tris-HCl, PH 6.8 (Final 0.55 M), followed by buffer exchange with Bio-Rad P30 columns. Run-on samples were enriched with BrdU antibody-conjugated beads (Santa Cruz biotechnology, sc-32323-ac), followed by RNA 5'cap repair with RppH (NEB, M0356S) and hydroxyl repair with T4 PNK (NEB, M0201S). Samples were then subjected sequentially to 5'RNA adaptor ligation followed by second enrichment with BrdU antibody-conjugated beads, and 3'RNA adaptor ligation followed by third enrichment with BrdU antibody-conjugated beads. Adaptor-ligated RNA were subjected to RT-PCR. RT-PCR products were amplified with indexed Illumina sequencing adaptors for 6 cycles and the 200-500 bp products were selected via gel extraction to minimize adaptor dimers. Full-scale amplification of purified products was then performed with the appropriate number of PCR cycles (determined by test PCR amplification) followed by PAGE-purification to generate the final libraries. GRO-Seq libraries were sequenced via paired end 75 bp sequencing on a Next-Seq™550 (Illumina) or paired end 100 bp sequencing on a Hi-Seq™2500 (Illumina). Data were aligned to mm9 genome (J_HΔ-dCas9 lines, J_HΔ-dCas9-sγ1-sgRNA lines and J_HΔ-dCas9- Iγ2b-del lines) or the mm9_DFL16.1JH4 genome (DFL16.1J_H4^{inv} *Rag2*$^{-/-}$ lines). Libraries were normalized to a coverage of 10 million 100nt reads for display. Relative transcriptional activity of specific regions was calculated as Reads Per Million Reads (RPM).

ChIP-Seq library preparation and analysis.

ChIP was done with G1-arrested cells and performed based on a prior protocol[35] with modifications. Briefly, 20 million cells were crosslinked in 37°C prewarmed culture medium with 1% formaldehyde for 10 min at RT. Cells were then treated with cell lysis buffer (5 mM PIPES pH 8, 85 mM KCl, 0.5% NP-40) for 10 min on ice, followed by treatment with nuclei lysis buffer (50 mM TrisCl pH 8.1, 10 mM EDTA, 1% SDS) for 10 min at RT. Chromatin was subjected to sonication with Diagenode Bioruptor at 4°C to achieve an average size of 200-300 bp (30 sec on, 30 sec off, 20 cycles with high energy input). Chromatin was then precleared with Dynabeads Protein A at 4°C for 2 hours. 1/30 lysates were kept as input and the rest were incubated with 5 ug RAD21 antibody (Abcam, ab992) or 5 ug NIPBL antibody (Bethyl Laboratories, A301-779A) overnight at 4°C. IP samples were then captured by Dynabeads Protein A at 4°C for at least 2 hours, followed by bead washing and elution. IP and Input DNA were de-crosslinked at 65°C overnight and purified via Qiagen PCR purification columns. Purified DNA was subjected to ChIP-Seq library preparation with Illumina Truseq ChIP Sample Preparation Kit (Illumina, IP-202-1012). ChIP-Seq libraries were sequenced via paired end 75 bp sequencing on a Next-Seq™550 or paired end 100 bp sequencing on a Hi-Seq™2500. Data were aligned to mm9 genome. Libraries were normalized to a coverage of 10 million 100nt reads.

3C-HTGTS library preparation and analysis.

3C-HTGTS libraries were done with G1 arrested cells and performed as previously described[4]. Briefly, 10 million cells were cross-linked with 2% formaldehyde at room temperature for 10 minutes, quenched by glycine (final 0.125M), and then lysed with cold cell lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1%

Triton X-100, one tablet of protease inhibitor in 50 ml) on ice for 10 min. Nuclei were subjected to 0.3% SDS treatment (37°C, 1 hour) and 1.8% Triton X-100 treatment (37 °C, 1 hour) successively, followed by overnight DpnII (700 units, NEB, R0543L) or NlaIII (700 units, NEB, R0125L) digestion at 37°C. DpnII or NlaIII was inactivated at 65°C for 20 min and samples were subjected to ligation under diluted condition with T4 DNA ligase (100 units, NEB, M0202L). Ligated chromatin was de-crosslinked with Proteinase K (56°C, O/N) and treated with RNaseA (37°C, 1 hour). DNA was purified via phenol/chloroform extraction and resuspended in 200 ul 1XT.E. buffer. DNA templates were then subjected to Illumina library preparation via HTGTS V(D)J-seq method as described above. 3C-HTGTS libraries were sequenced using paired end 150 bp sequencing on a Next-Seq™550 or paired end 300 bp sequencing on a Mi-Seq™ machine. Data were processed as previously described[4]. In addition, as 3C-HTGTS junctions were generated by ligation of the restriction digestion products of the 4 bp-cutter that do not involve homology-mediated repair, junctions with long microhomology (> 5 bp) were excluded from analysis to avoid potential PCR artifacts. The overall 3C-HTGTS library profiles before and after removing the junctions with > 5 bp microhomology are very similar. Libraries were size-normalized to total junctions of the smallest library in the set of libraries for comparison. For 3C-HTGTS bait interaction frequency analysis, we counted the number of junctions within the indicated bait-interacting locales for both control and experimental samples. For bar graph presentations in Figures 3f and 4d, the junction number recovered from control (e.g. $J_H\Delta$-dCas9 lines) samples was normalized to represent 100% and relative experimental values are listed as a percentage of the control values. For bar graphs in Extended Data Figure 10q, r, values of D_H region interaction are set as 100% and relative values of the D_H-flanking regions are listed as a percentage of the D_H interaction values. DpnII digestion was used to generate libraries in Fig. 3f, 4d and Extended Data Fig. 8b, 9c, 10q (top panels); NlaIII digestion was used to generate libraries in Extended Data Fig. 10q (bottom panels) and 10r. Note that the DpnII digestion profiles for $J_H\Delta$-dCas9 lines in Extended Data Fig. 10q were derived from the same libraries of $J_H\Delta$-dCas9 lines presented in Fig. 3f and Extended Data Fig. 8b. Primers used for 3C-HTGTS are listed in Supplementary Information Table 3.

Hi-C analysis.

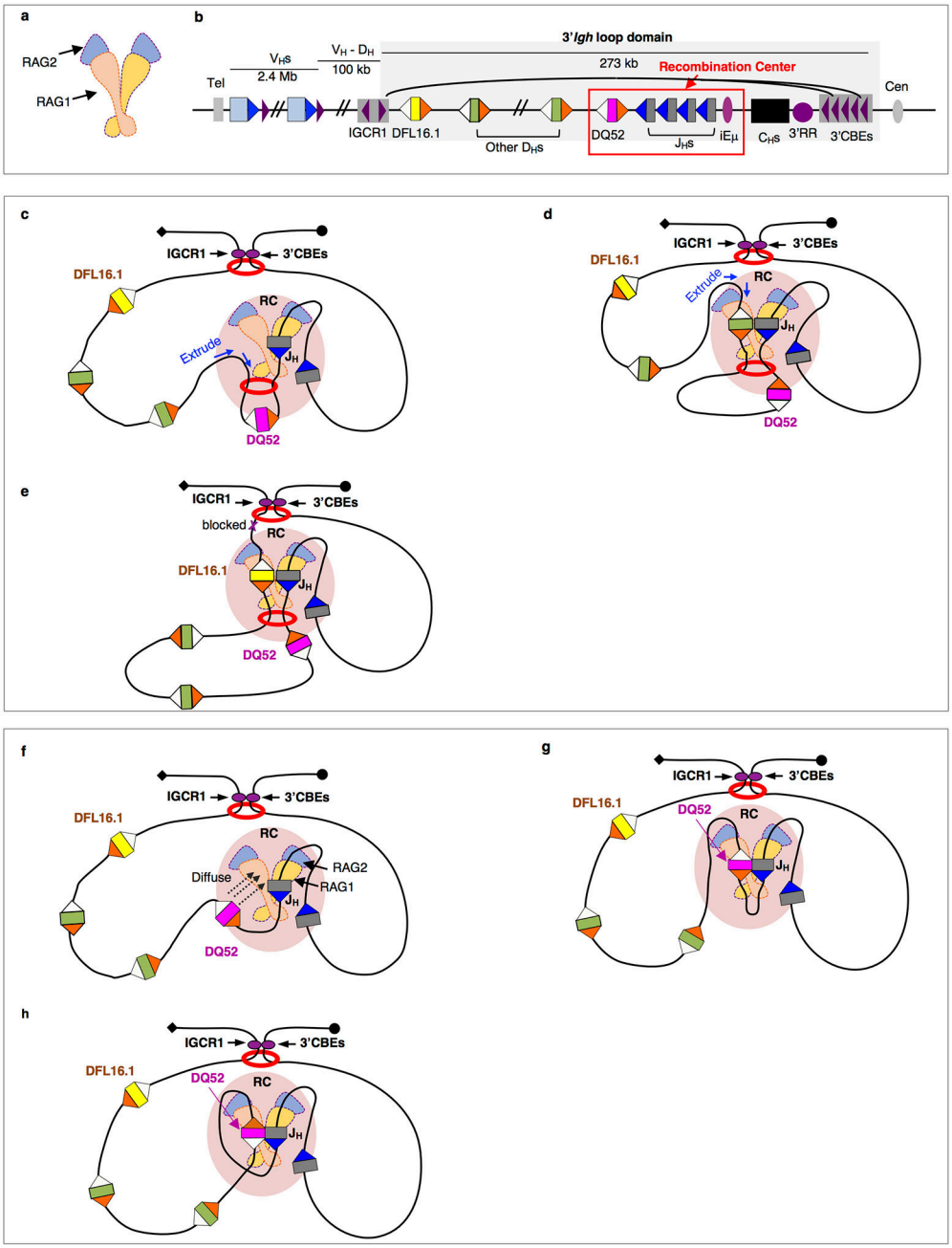
Hi-C libraries were generated using the in situ Hi-C protocol based on Rao and Huntley et al[2]. Approximately 1 million cells were crosslinked to create each library. Cells were then lysed with nuclei permeabilized while keeping them intact. DNA was then digested using the restriction enzyme MboI, and the overhangs filled in incorporating the biotinylated base bioU. Free ends were then ligated together in situ. Crosslinks were reversed, the DNA was sheared to 300-500 bp and then biotinylated ligation junctions were recovered with streptavidin beads. Small modifications in reagent volumes and incubation times were incorporated to optimize library quality for these cell types. The standard Illumina library construction protocol was utilized. Briefly, DNA was end-repaired using a combination of T4 DNA polymerase, E. coli DNA Pol I large fragment (Klenow polymerase) and T4 polynucleotide kinase. The blunt, phosphorylated ends were treated with Klenow fragment (32 to 52 exo minus) and dATP to yield a protruding 3- ‘A’ base for ligation of Illumina’s adapters which have a single ‘T’ base overhang at the 3’ end. After adapter ligation, DNA was PCR amplified with Illumina primers for 8-12 cycles and library fragments of 400-600

bp (insert plus adaptor and PCR primer sequences) were purified using SPRI beads. The purified DNA was captured on an Illumina flow cell for cluster generation. Libraries were sequenced on the Illumina sequencing platform following the manufacturer's protocols. We sequenced 2.3B Hi-C read pairs in the control $J_H\Delta$-dCas9 line, yielding 1.3B Hi-C contacts; we also sequenced 2.2B Hi-C read pairs in $J_H\Delta$-dCas9-Sγ1-sgRNA cells with the dCas9 impediment, yielding 1.2B Hi-C contacts. Hi-C libraries were analyzed using the Juicer pipeline[36], and visualized with Juicebox[37]. All the code used in the above steps is publicly available at (github.com/aidenlab). Note that while Hi-C analysis did not distinguish C57BL/6 and 129/Sv C_H alleles, it gave highly complementary results to the 3C-HTGTS with C57BL/6-specific iEμ bait with respect to interactions with or without the Sγ1 impediment.

Statistical analysis

Statistical analyses for Fig. 3e, f, 4b-d, and Extended Data Fig. 7f, 9d, 10q were performed via two-tailed, paired t test. $P < 0.05$ is considered significant.

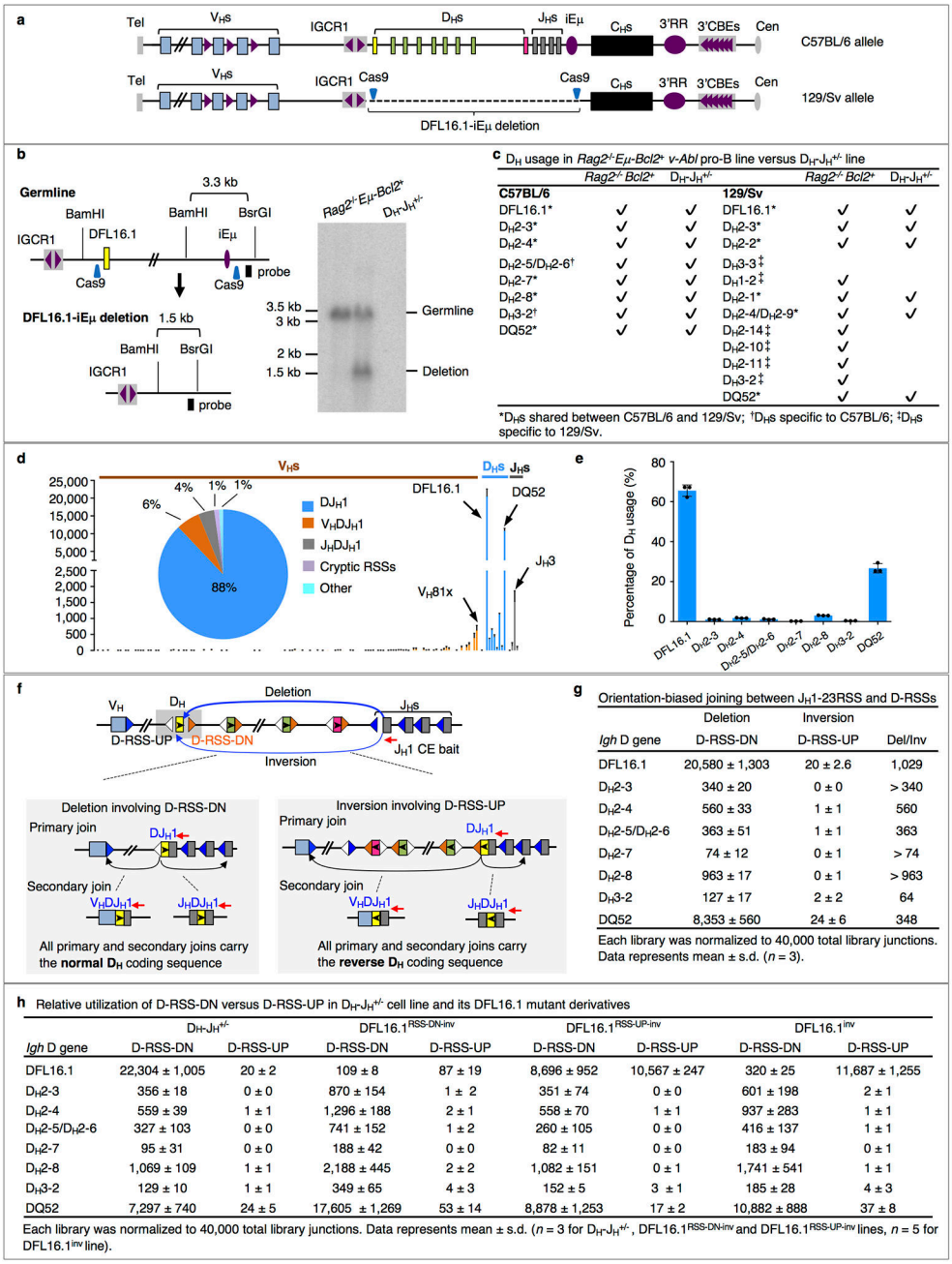
Extended Data



Extended Data Figure 1. Working model for role of loop-extrusion mediated RAG scanning in driving deletion-biased D to J_H recombination.

a, Illustration of the Y structured RAG heterodimer complex. **b,** Schematic of *Igh* highlighting the RC and 3'*Igh* loop domain bounded by IGCR1 and 3'CBEs. **c,** Working model for RAG scanning to Ds upstream of DQ52. Cohesin (red ring) initiates loop extrusion upon being loaded in the upstream portion of the RC within the IGCR1-iEμ/RC subdomain. Proximal downstream active RC chromatin impedes cohesin extrusion of downstream chromatin and, thereby, serves as a downstream sub-loop anchor allowing continued extrusion of upstream chromatin past RC-bound RAG. **d,** Continued upstream loop extrusion brings D_Hs upstream of RC-based DQ52 past the open RAG1 sub-unit active

site opposite the J_H-bound active site in the other RAG1subunit. This linear process aligns downstream D-12RSS with the RAG-bound J_H-23RSS for orientation-specific, deletional D to J_H recombination. **e,** Upstream Ds are frequently passed without being utilized and most loop extrusion-mediated RAG scanning continues until reaching the 5'CBE loop anchor (IGCR1) that strongly impedes (nearly blocks) loop extrusion and RAG scanning. The latter prolonged interaction may contribute to robust DFL16.1 utilization. **f-h,** Due to RC location, DQ52 can bind to the open RAG active site by diffusion[12] (f) which allows it to bind in both deletional (g) and inversional (h) configurations. In this case, deletion-biased usage of DQ52 is achieved through a much stronger RSS-DN that, in this location, dominates RAG binding/cleavage compared to its weaker RSS-UP. Other schematics in **b-h** are as described in Fig.1 legend.



Igh D gene	D_H-$J_H^{+/-}$ D-RSS-DN	D_H-$J_H^{+/-}$ D-RSS-UP	DFL16.1$^{RSS\text{-}DN\text{-}inv}$ D-RSS-DN	DFL16.1$^{RSS\text{-}DN\text{-}inv}$ D-RSS-UP	DFL16.1$^{RSS\text{-}UP\text{-}inv}$ D-RSS-DN	DFL16.1$^{RSS\text{-}UP\text{-}inv}$ D-RSS-UP	DFL16.1^{inv} D-RSS-DN	DFL16.1^{inv} D-RSS-UP
DFL16.1	22,304 ± 1,005	20 ± 2	109 ± 8	87 ± 19	8,696 ± 952	10,567 ± 247	320 ± 25	11,687 ± 1,255
D_H2-3	356 ± 18	0 ± 0	870 ± 154	1 ± 2	351 ± 74	0 ± 0	601 ± 198	2 ± 1
D_H2-4	559 ± 39	1 ± 1	1,296 ± 188	2 ± 1	558 ± 70	1 ± 1	937 ± 283	1 ± 1
D_H2-5/D_H2-6	327 ± 103	0 ± 0	741 ± 152	1 ± 2	260 ± 105	0 ± 0	416 ± 137	1 ± 1
D_H2-7	95 ± 31	0 ± 0	188 ± 42	0 ± 0	82 ± 11	0 ± 0	183 ± 94	0 ± 1
D_H2-8	1,069 ± 109	1 ± 1	2,188 ± 445	2 ± 2	1,082 ± 151	0 ± 1	1,741 ± 541	1 ± 1
D_H3-2	129 ± 10	1 ± 1	349 ± 65	4 ± 3	152 ± 5	3 ± 1	185 ± 28	4 ± 3
DQ52	7,297 ± 740	24 ± 5	17,605 ± 1,269	53 ± 14	8,878 ± 1,253	17 ± 2	10,882 ± 888	37 ± 8

Each library was normalized to 40,000 total library junctions. Data represents mean ± s.d. (n = 3 for D_H-$J_H^{+/-}$, DFL16.1$^{RSS\text{-}DN\text{-}inv}$ and DFL16.1$^{RSS\text{-}UP\text{-}inv}$ lines, n = 5 for DFL16.1^{inv} line).

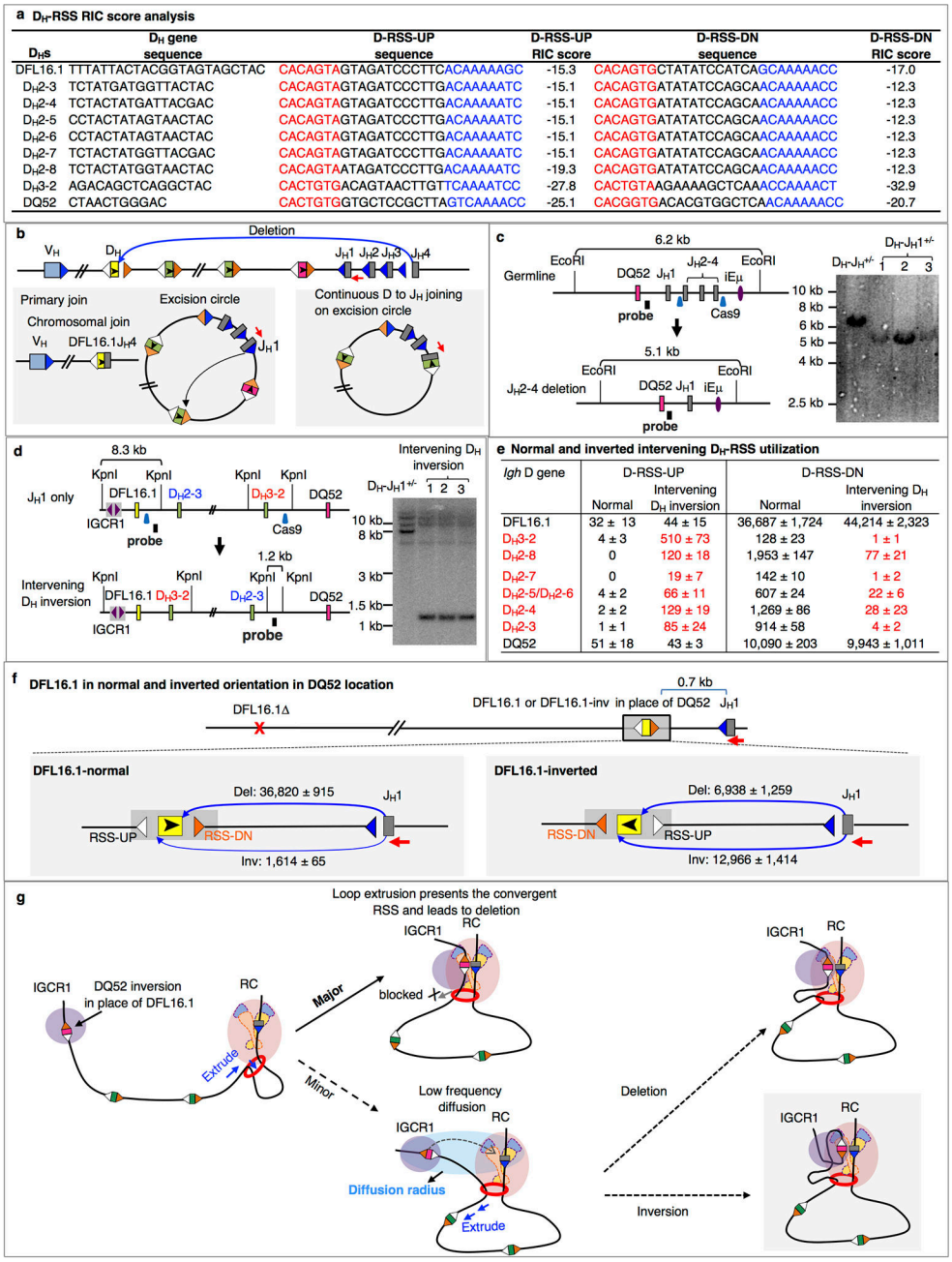
Igh D gene	Deletion D-RSS-DN	Inversion D-RSS-UP	Del/Inv
DFL16.1	20,580 ± 1,303	20 ± 2.6	1,029
D_H2-3	340 ± 20	0 ± 0	> 340
D_H2-4	560 ± 33	1 ± 1	560
D_H2-5/D_H2-6	363 ± 51	1 ± 1	363
D_H2-7	74 ± 12	0 ± 1	> 74
D_H2-8	963 ± 17	0 ± 1	> 963
D_H3-2	127 ± 17	2 ± 2	64
DQ52	8,353 ± 560	24 ± 6	348

Each library was normalized to 40,000 total library junctions. Data represents mean ± s.d. (n = 3).

Extended Data Figure 2. HTGTS V(D)J-seq analysis of V(D)J recombination outcomes in D_H-$J_H^{+/-}$ line and its mutant derivatives.

a, Schematic of the two *Igh* alleles of the D_H-$J_H^{+/-}$ *v-Abl* pro-B line. This C57BL/6, 129/Sv mixed background line was derived by deleting indicated region from the 129/Sv allele to inactivate it for V(D)J recombination. **b,** Southern blotting confirmation of deleted allele in D_H-$J_H^{+/-}$ line. Done twice with similar results. **c,** C57BL/6 versus 129/Sv D_H usage in parental versus D_H-$J_H^{+/-}$ line, as analyzed via HTGTS V(D)J-seq (J_H1 CE primer). Lack of 129/Sv-specific D_Hs in D_H-$J_H^{+/-}$ libraries confirmed retention of C57BL/6 and absence of 129/Sv allele in this line. **d,** Bar graph shows utilization frequency of each V_H, D_H and J_H

from J_H-distal to J_H-proximal locales (n = 3 independent libraries). Pie chart shows percentage of indicated V(D)J recombination products as fraction of total *Igh* junctions. Beyond predominant 'DJ_H1' junctions, both low 'V_HDJ_H1' joins [4,12] and inversional "$J_H(D)J_H1$" joins[38] were detected. Very low level J_H1 joins to 'cryptic RSSs', or a different J_H-RSS ("other") that likely occurs in extra-chromosomal excision circles[13] were also detected. **e,** Utilization of each D as percentage of total DJ_H1 joins (n = 3 independent libraries). **f,** Strategy for analysis of D-RSS-DN versus D-RSS-UP utilization. Orientation of D coding sequences relative to J_H1 CE primer is preserved in primary and secondary joins for both D-RSS-DN and D-RSS-UP, allowing calculation of relative utilization of D-RSS-DN versus D-RSS-UP. **g,** Utilization frequency of D-RSS-DN versus D-RSS-UP in D_H-J_H $^{+/-}$ line. **h,** Impact of DFL16.1-RSS mutations on utilization of D-RSS-DNs versus D-RSS-UPs. Libraries in **d, e, g, h** were normalized to 40,000 total junctions. Data represents mean ± s.d. Data for D_H-$J_H^{+/-}$ line in **d-g** and **h** are two sets of 3 repeats each, with the latter done along with DLF16.1 mutants.



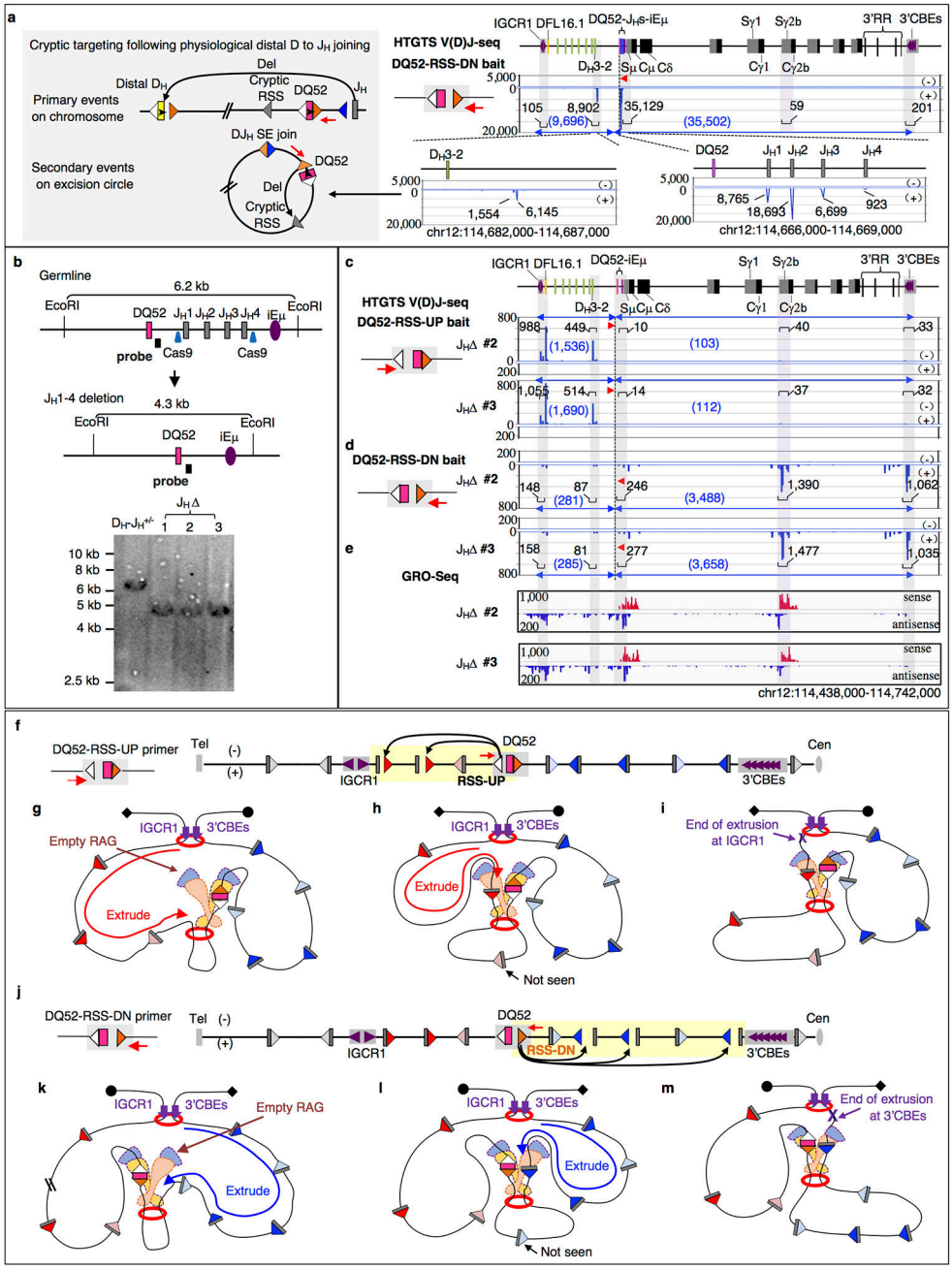
a D_H-RSS RIC score analysis

D_Hs	D_H gene sequence	D-RSS-UP sequence	D-RSS-UP RIC score	D-RSS-DN sequence	D-RSS-DN RIC score
DFL16.1	TTTATTACTACGGTAGTAGCTAC	CACAGTAGTAGATCCCTTCACAAAAAGC	-15.3	CACAGTGCTATATCCATCAGCAAAAACC	-17.0
D_H2-3	TCTATGATGGTTACTAC	CACAGTAGTAGATCCCTTGACAAAAATC	-15.1	CACAGTGATATATCCAGCAACAAAAACC	-12.3
D_H2-4	TCTACTATGATTACGAC	CACAGTAGTAGATCCCTTGACAAAAATC	-15.1	CACAGTGATATATCCAGCAACAAAAACC	-12.3
D_H2-5	CCTACTATAGTAACTAC	CACAGTAGTAGATCCCTTGACAAAAATC	-15.1	CACAGTGATATATCCAGCAACAAAAACC	-12.3
D_H2-6	CCTACTATAGTAACTAC	CACAGTAGTAGATCCCTTGACAAAAATC	-15.1	CACAGTGATATATCCAGCAACAAAAACC	-12.3
D_H2-7	TCTACTATGGTTACGAC	CACAGTAGTAGATCCCTTGACAAAAATC	-15.1	CACAGTGATATATCCAGCAACAAAAACC	-12.3
D_H2-8	TCTACTATGGTAACTAC	CACAGTAATAGATCCCTTGACAAAAATC	-19.3	CACAGTGATATATCCAGCAACAAAAACC	-12.3
D_H3-2	AGACAGCTCAGGCTAC	CACTGTGACAGTAACTTGTTCAAAATCC	-27.8	CACTGTAAGAAAAGCTCAAACCAAAACT	-32.9
DQ52	CTAACTGGGAC	CACTGTGGTGCTCCGCTTAGTCAAAACC	-25.1	CACGGTGACACGTGGCTCAACAAAAACC	-20.7

Extended Data Figure 3. Generation and analyses of D_H-$J_H1^{+/-}$ line and its mutant derivative lines.

a, Table shows coding and flanking D-RSS-UP and D-RSS-DN sequences and their RSS recombination information content (RIC) score [39, 40] generated from a publicly available program (http://www.itb.cnr.it/rss)[41]. Predicted "functional" 12RSSs have a RIC of at least −38.81, with increasing RIC scores proposed to reflect increasing RSS strength. **b,** Illustration of potential DJ_H1 recombination on excision circle. J_H1 joining to D_Hs downstream of DFL16.1 that occur on excision circles generated by primary joining between distal J_Hs (J_H2-J_H4) and distal D_Hs are not subject to the same mechanistic constraints as chromosomal D to J_H recombination[13]. To obviate such joins, we deleted J_H2-

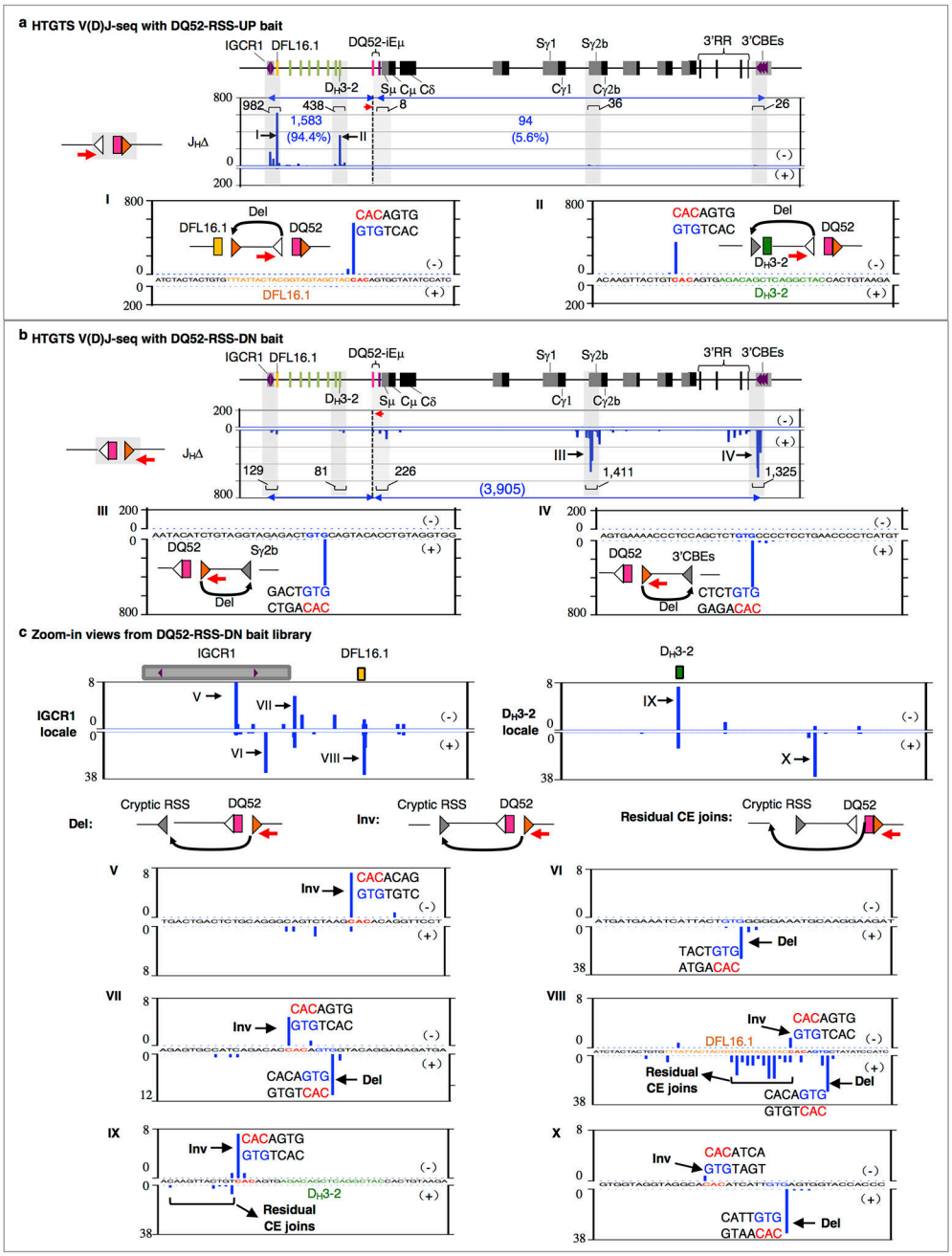
J_H4 in the D_H-$J_H^{+/-}$ line to generate the D_H-$J_H1^{+/-}$ line. **c, d,** Southern blotting confirmation of D_H-$J_H1^{+/-}$ (done once after PCR confirmation) (**c**) and intervening D_H inversion (done twice with similar results) (**d**) lines. **e,** Utilization of D-RSS-UP and D-RSS-DN in the D_H-$J_H1^{+/-}$ line and its mutant derivatives with intervening D_H inversion (n = 3 libraries for each genotype). **f,** Relative utilization of DFL16.1-RSS-DN versus DFL16.1-RSS-UP for normal DFL16.1 (left) or DFL16.1 inversion (right) located in place of DQ52 in D_H-$J_H1^{+/-}$ cells with endogenous DFL16.1 deleted ("DFL16.1Δ $DQ52^{DFL16.1}$" and "DFL16.1Δ $DQ52^{DFL16.1\text{-}inv}$") ($n$ = 3 libraries for each genotype). Data in panel **e** and **f** represents mean ± s.d from biologically independent samples and was normalized to 70,000 total junctions for each library. **g,** Model for low level inversional RC-distal D joining involving loop-extrusion based scanning, which could bring distal upstream D-RSSs into "diffusion radius" of the RC. See Supplementary Discussion for further discussion of findings and models in this figure.



Extended Data Figure 4. Directional RAG scanning from a DQ52-based RC within 3'*Igh* CBE-anchored loop domain.

a, HTGTS V(D)J-Seq analysis with DQ52-RSS-DN bait in D_H-$J_H^{+/-}$ line. Major junctional outcomes are deletional DQ52-RSS-DN to J_H joins (77%) and deletional DQ52-RSS-DN joins to cryptic RSSs near the immediately upstream D_H3-2 region (20%), with the latter likely resulting from secondary events on excision circles following primary J_H to distal D_H joins (illustrated on left panels; also, see below). **b,** Southern blot confirmation of $J_H\Delta$ lines (done once after PCR confirmation). **c,** Repeats of Fig. 3a HTGTS V(D)J-Seq experiments. Each library was normalized to the same number of DQ52-RSS-UP SE junctions. **d,** Repeats

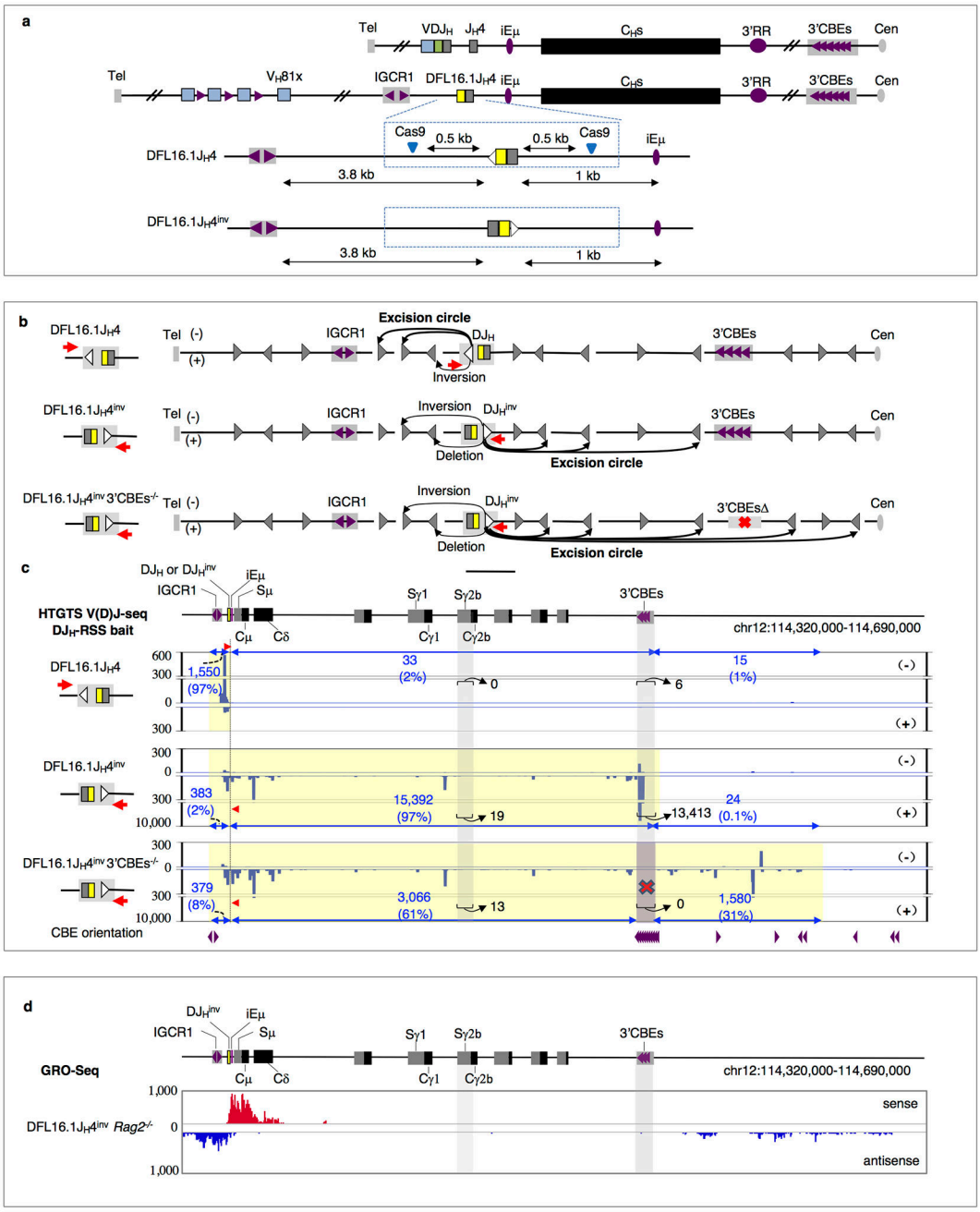
of Fig. 3b HTGTS V(D)J-Seq experiments. Each library was normalized to the same number of DQ52-RSS-UP CE junctions captured by the DQ52-RSS-DN bait (See methods). Note the near abrogation of cryptic deletional joins near D_H3-2 in $J_H\Delta$ lines, which is consistent with their excision circle origin. Also, unlike the D_H-$J_H^{+/-}$ line with germline J_Hs, robust RC downstream cryptic scanning activity is readily detected in the $J_H\Delta$ lines. **e,** Repeats of Fig. 3d GRO-Seq. Each library was normalized to a coverage of 10 million 100nt reads for display. **f-i,** Model for cohesin loop extrusion-meditated directional RAG scanning from RC DQ52-RSS-UP to upstream regions until reaching IGCR1 loop anchor. **j-m,** Model for extrusion-meditated directional RAG scanning from RC DQ52-RSS-DN to downstream regions until reaching 3'CBEs loop anchor. Transparent yellow rectangles in **f** and **j** indicate respectively the upstream and downstream RAG scanning regions with DQ52 upstream and downstream RSS joining to cryptic RSSs shown in schematic form. Other schematics are as described in Fig. 1 and Extended Data Fig. 1. The two models are supported by the directional RAG scanning activity in c, d and Fig. 3a, b.



Extended Data Figure 5. RAG cryptic targeting activity from DQ52-RSS-UP and DQ52-RSS-DN in $J_HΔ$ lines.

a, HTGTS V(D)J-seq profile of upstream RAG cryptic scanning activity from DQ52-RSS-UP with indicated peak regions at IGCR1 and D_H3-2 locales (grey transparent bars). Upper panel: Junctions plotted at 100 bp bin size. Bottom panels: Examples of most robust peak near IGCR1 (I) and D_H3-2 (II) plotted at single bp resolution. Letters next to the peaks show DNA duplex sequences of the targeted cryptic heptamers. See text for more details. **b,** HTGTS V(D)J-seq of downstream RAG cryptic scanning activity from DQ52-RSS-DN with indicated peak regions in Sγ2b and 3'CBEs locales and lower frequency peaks in iEμ-Sμ,

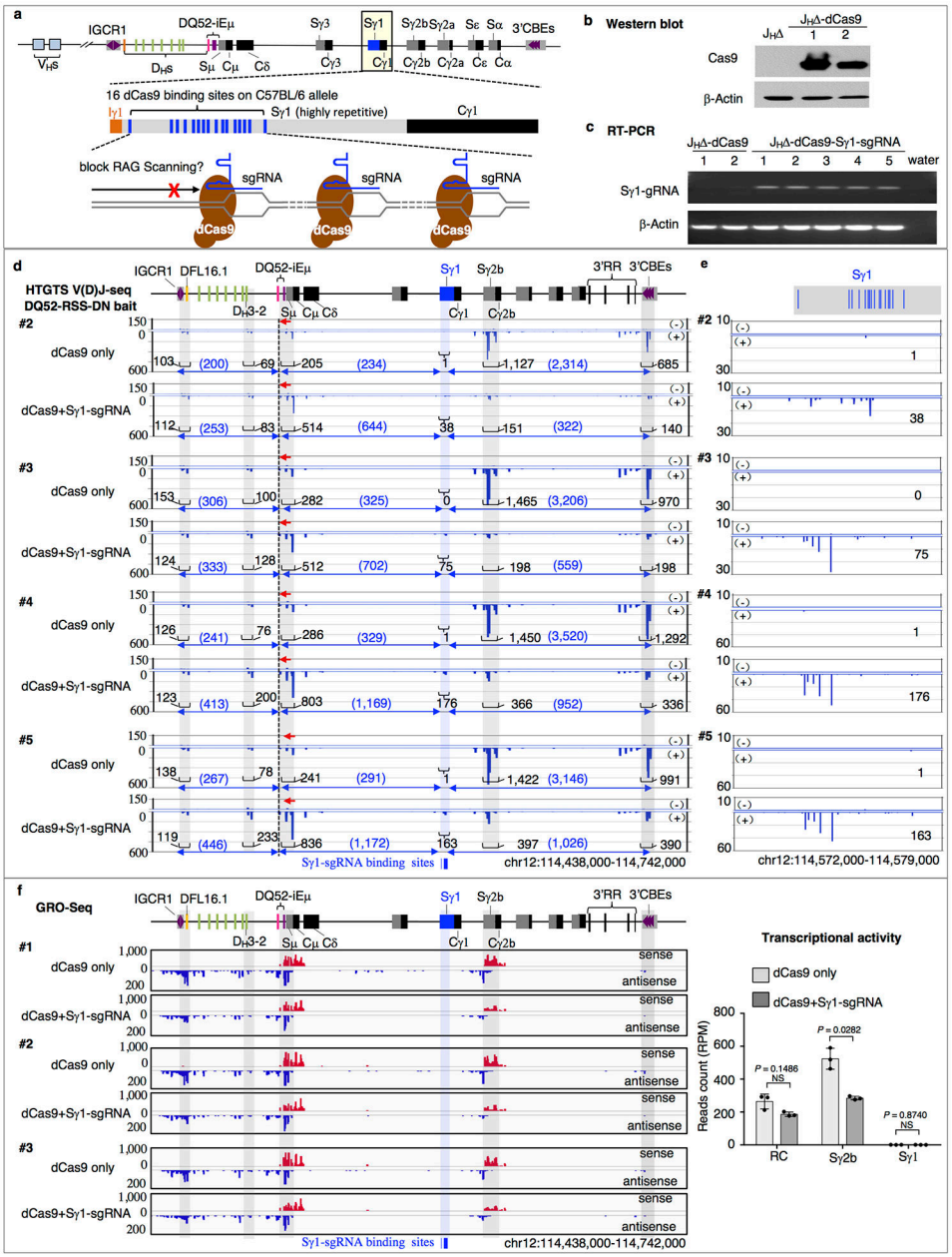
D_H3-2 and IGCR1(grey transparent bars). Upper panel: Junctions plotted at 100 bp bin size. Lower panels: Examples of most robust Sγ2b (III) and 3'CBEs (IV) locale peaks plotted at single bp resolution. **c,** Low frequency DQ52-RSS-DN junctions upstream of RC detected by DQ52-RSS-DN bait. Top panels: Zoom-in view of IGCR1 and D_H3-2 locales identified in panel b are plotted at 20 bp bin size with representative junctions labeled (V-X). Bottom panels: Single bp resolution of junctions for V-X. Deletions are mediated by cryptic RSSs in divergent orientation (forward "CAC") and inversions are mediated by cryptic RSSs in the same orientation (reverse "CAC") as DQ52-RSS-DN. Also illustrated are junctions resulting from joining DQ52 CEs to cryptic CEs[12], mediated by DQ52-RSS-UP and cryptic convergent RSSs. A likely explanation for these low level joins is that loop extrusion brings them into proximity with the RC where their location/transcription impedes extrusion, allowing them to access RC-bound RAG by local diffusion[12], analogous to diffusion-mediated DQ52 to J_H1 joining.



Extended Data Figure 6. RAG targeting and transcriptional activity analysis in the DFL16.1J_H4^{inv} lines.

a, Generation of the DFL16.1J_H4^{inv} line. Schematic shows two *Igh* alleles of DFL16.1J_H4 line and DFL16.1J_H4^{inv} line. In the DFL16.1J_H4 line, one *Igh* allele contains a nonproductive VDJ join involving V_H1-2P and J_H3, and the other allele harbors the DFL16.1J_H4 join. The DFL16.1J_H4^{inv} line was derived from DFL16.1J_H4 line by inverting a 1kb segment encompassing the DJ_H join via CRISPR/Cas9. **b,** Illustration of mechanism for RAG cryptic scanning activity from the RC DJ_H-RSS in DFL16.1J_H4 line (top), DFL16.1J_H4^{inv} line (middle) and DFL16.1J_H4^{inv} 3'CBEs$^{-/-}$ line (bottom). **c,** Representative

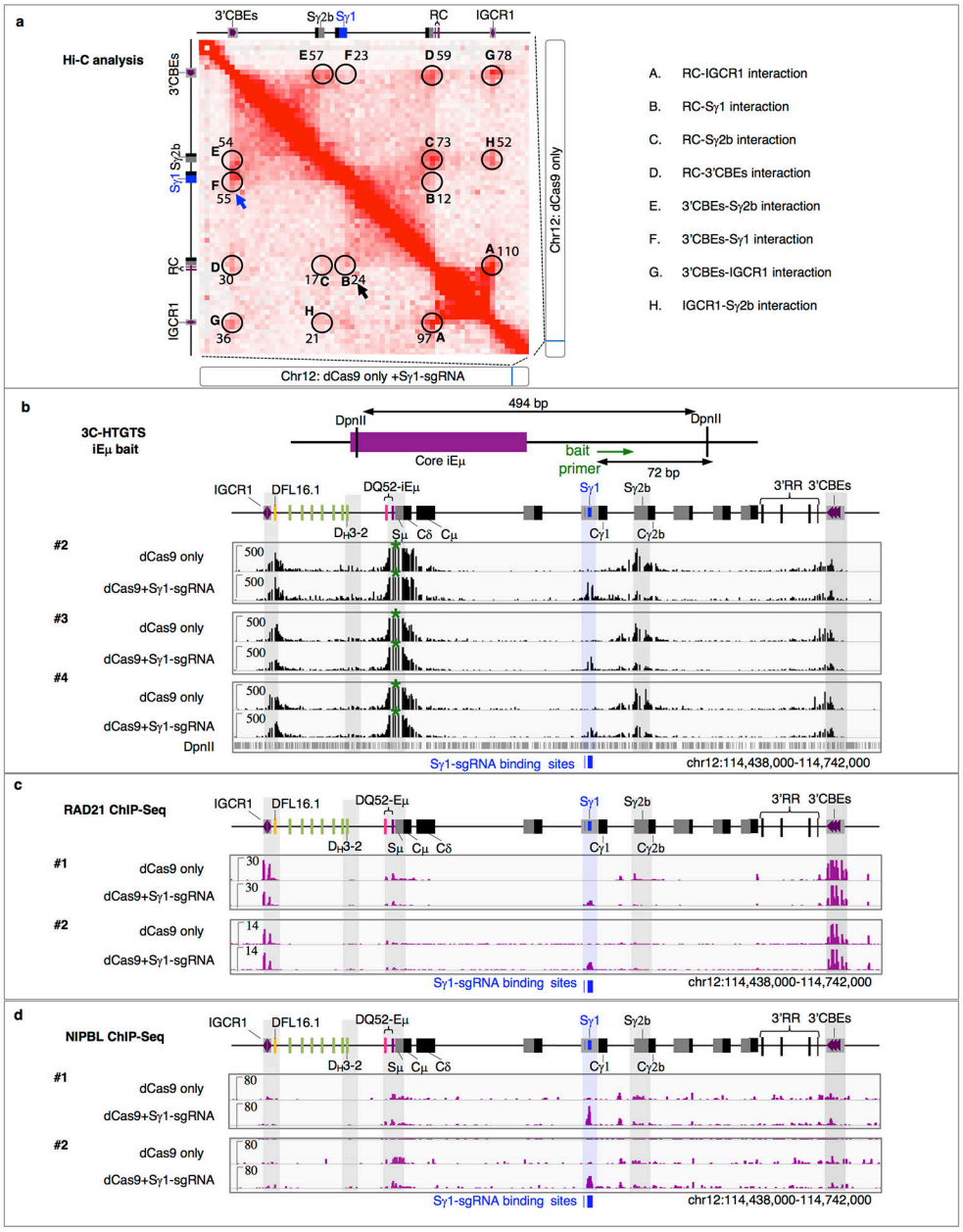
HTGTS V(D)J-seq profiles showing RAG cryptic scanning patterns of DFL16.1J_H4 line (top) (n = 3 technical repeats), DFL16.1J_H4^{inv} line (middle) (n = 3 biological replicates) and DFL16.1J_H4^{inv} 3'CBEs$^{-/-}$ line (bottom) (n = 3 biological replicates). Black line indicates bait primer position. Yellow shadows highlight RAG scanning regions. Purple arrows underneath the RAG targeting profiles indicate positions of forward and reverse CBEs. **d,** Representative GRO-Seq profile of 3 repeats of the DFL16.1J_H4^{inv} *Rag2*$^{-/-}$ line (n = 3 biological replicates).



Extended Data Figure 7. dCas9-binding impedes RAG scanning and corresponding loop formation.

a, Illustration of the dCas9-block system. An Sγ1-gRNA that has 16 binding sites (blue lines) within a 4kb highly repetitive Sγ1 region on the C57BL/6 allele was introduced into the $J_HΔ$-dCas9 line. **b,** Western blot confirmation of dCas9 expression in $J_HΔ$-dCas9 lines but not the parental $J_HΔ$ line (done twice with similar results). **c,** Reverse transcription PCR (RT-PCR) confirmation of Sγ1-gRNA expression in the $J_HΔ$-dCas9-Sγ1-sgRNA lines but not parental lines (done twice with similar results). **d,** Additional HTGTS V(D)J-seq repeats (DQ52-RSS-DN bait) for $J_HΔ$-dCas9 lines and $J_HΔ$-dCas9-Sγ1-sgRNA lines shown in Fig. 3e. Each library was normalized to the same number of DQ52-RSS-UP CE junctions captured by the DQ52-RSS-DN bait (See methods). **e,** Zoom-in view of Sγ1 region from

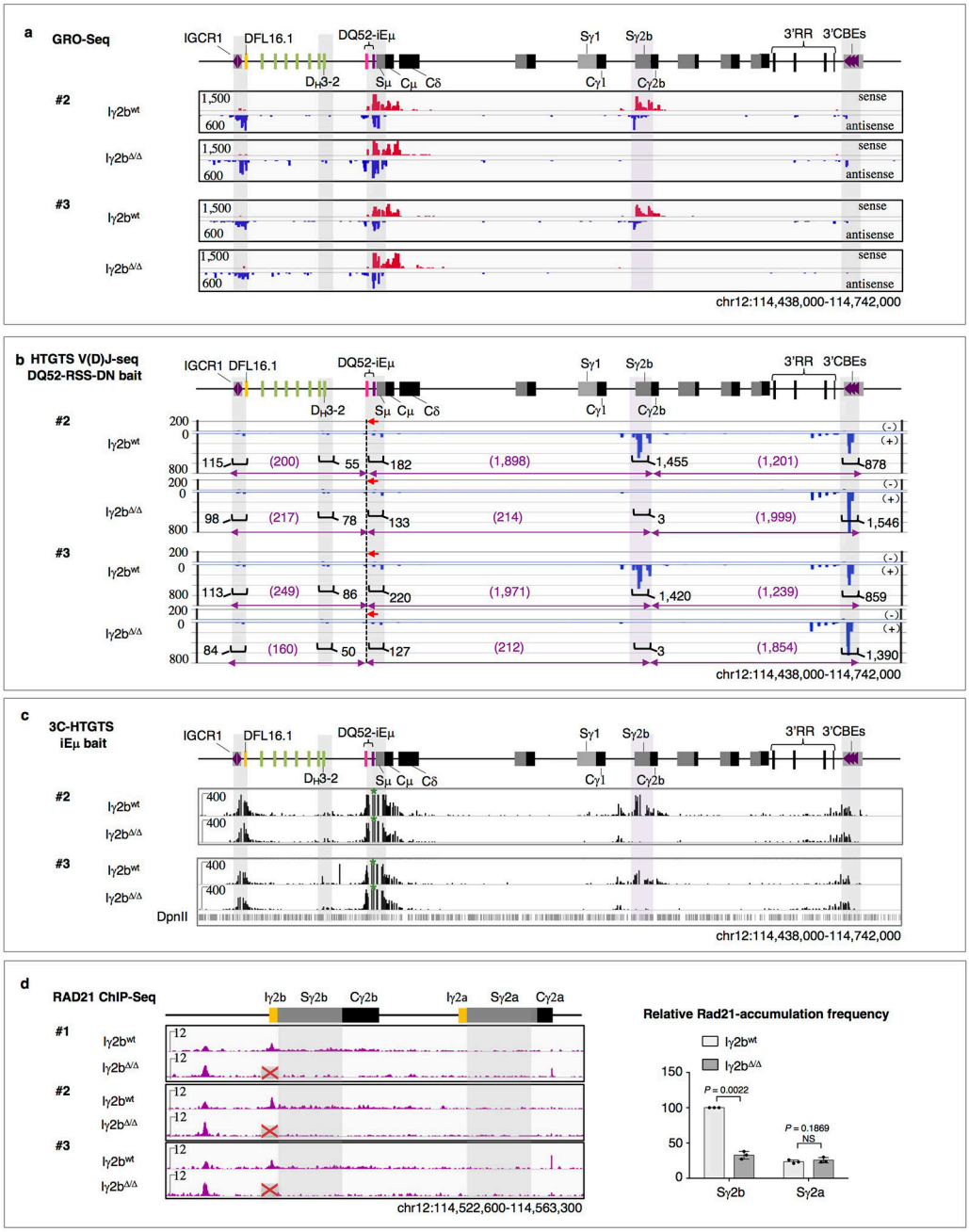
HTGTS V(D)J-seq profiles in **d**, showing accumulation of RAG activity at the dCas9-bound Sγ1 region. **f,** GRO-Seq analysis of J_HΔ-dCas9 and J_HΔ-dCas9-Sγ1-sgRNA lines. Each library was normalized to a coverage of 10 million100nt reads for display. Bar graph compares transcriptional activity of indicated regions (n = 3 libraries for each genotype). Data represents mean ± s.d from biologically independent samples. P values were calculated via two-tailed paired t-test. NS: $P \geq 0.05$. The modestly decreased Sγ2b transcription upon Sγ1 dCas9 binding is potentially due to impaired loop extrusion between Sγ2b and iEμ.



Extended Data Figure 8. dCas9-binding impedes downstream loop formation in association with cohesin loading and accumulation at impediment locale.

a, Hi-C analysis of the 3'*Igh* domain interaction of J_HΔ-dCas9 line versus J_HΔ-dCas9-Sγ1-sgRNA line. We compared 1.3 billion (B) contacts in the control line versus 1.2B contacts in the dCas9-impediment line. Letters annotate the interactions between the two indicated loci, and the numbers next to the letters reflect relative interaction intensity. Black and blue arrows highlight Sγ1 interaction with RC (B) and 3'CBEs (F) locale, respectively, in the J_HΔ-dCas9-Sγ1-sgRNA line. **b,** 3C-HTGTS repeats with iEμ bait (green stars) for J_HΔ-dCas9 and J_HΔ-dCas9-Sγ1-sgRNA lines shown in Fig. 3f. The iEμ bait primer strategy is shown on the top. Each library was normalized to 192,000 total junctions for analysis. While these lines retain downstream C_H sequences on their 129/Sv allele (Extended Data Fig. 2b),

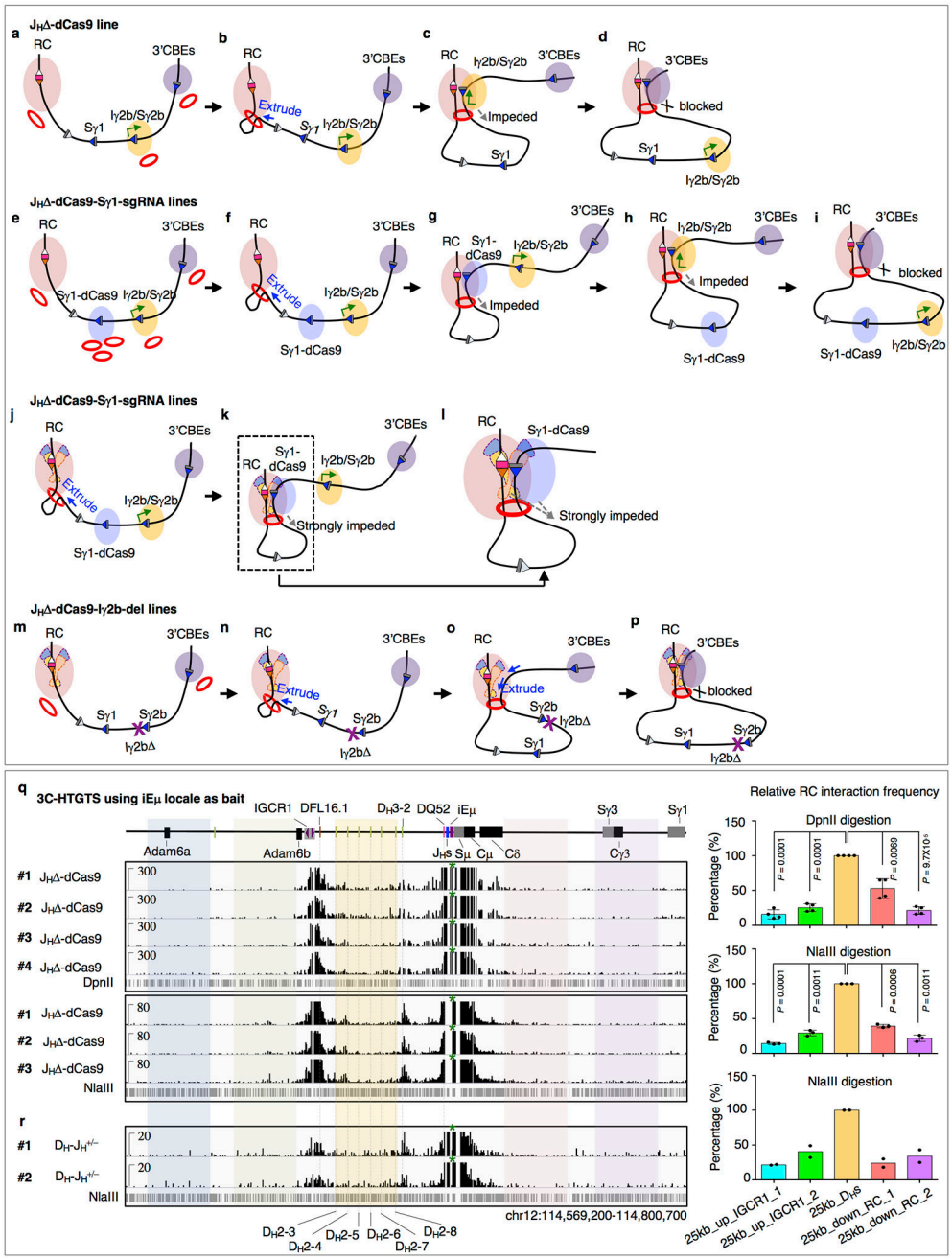
the iEμ bait should have very low interactions in trans[42]. Blue and grey transparent bars extending through all panels are as described in Fig. 3. In addition, an interaction between the RC and an Iγ2b upstream enhancer named hRE1, an enhancer of unknown activity[43, 44], was evident (see also Fig. 4) and was accompanied by RAD21 and NIPBL accumulation (see below) and low level of RAG scanning activity (Extended Data Fig. 7d). **c,** RAD21 ChIP-Seq profiles of J_HΔ-dCas9 lines versus J_HΔ-dCas9-Sγ1-sgRNA lines (*n* = 2 biological replicates). Each library was normalized to a coverage of 10 million 100ntreads. **d,** NIPBL ChIP-Seq profiles of J_HΔ-dCas9 lines versus J_HΔ-dCas9-Sγ1-sgRNA lines (*n* = 2 biological replicates). Each library was normalized to a coverage of 10 million 100nt reads.



Extended Data Figure 9. Ectopic transcription of Iγ2b-Sγ2b region impedes downstream loop formation and RAG scanning.

a, GRO-Seq repeats for $J_H\Delta$-dCas9 lines (Iγ2b^{wt}) and $J_H\Delta$-dCas9-Iγ2b-del lines (Iγ2b$^{\Delta/\Delta}$) shown in Fig. 4b. Each library was normalized to a coverage of 10 million 100nt reads. **b,** HTGTS V(D)J-seq repeats with DQ52-RSS-DN bait for Iγ2b^{wt} versus Iγ2b$^{\Delta/\Delta}$ lines shown in Figure 4c. Each library was normalized to the same number of DQ52-RSS-UP CE junctions captured by the DQ52-RSS-DN bait. **c,** 3C-HTGTS repeats from iEμ bait for Iγ2b^{wt} and Iγ2b$^{\Delta/\Delta}$ lines for data shown in Fig. 3d. Each library was normalized to 150,000 total junctions for analysis. **d,** RAD21 ChIP-Seq analysis for Iγ2b^{wt} and Iγ2b$^{\Delta/\Delta}$ lines. Each

library was normalized to a coverage of 10 million 100nt reads for display. Bar graph shows comparison of RAD21 accumulation at the Sγ2b region (Sγ2a region as control) in Iγ2b^{wt} lines versus Iγ2b$^{\Delta/\Delta}$ lines (n = 3 libraries for each genotype). Data represents mean ± s.d from biologically independent samples. For bar graph presentation, the junction number recovered from Sγ2b region of Iγ2b^{wt} control samples was normalized to represent 100%, relative values of Sγ2a region in the control and Sγ2b and Sγ2a regions in the Iγ2b$^{\Delta/\Delta}$ samples are listed as a percentage of the control Sγ2b values. P values were calculated by a two-tailed paired t-test. NS: $P \geq 0.05$.



Extended Data Figure 10. Working model for loop extrusion-mediated RAG downstream scanning.

a-i, Model for cohesin-mediated loop extrusion of chromatin past nascent *Igh* RC in $J_H\Delta$ *v-Abl* lines based on RAG2-deficient background analyses. For all examples, increased interactions of impediment sites with RC targets scanning activity in RAG-sufficient cells. **a**. Cohesin (red rings) are loaded at multiple sites in the RC-3'CBEs *Igh* sub-domain. Illustrations show cohesin loading at RC-downstream region. **b.** Cohesin-mediated extrusion promotes linear interaction of the nascent RC with downstream regions. **c.** Robust transcription (green arrow) across the Iγ2b/Sγ2b impedes loop extrusion. **d.** In a subset of

cells, loop extrusion proceeds past Iγ2b/Sγ2b impediment to 3'CBEs loop anchor. **e-i,** Loop extrusion in J_HΔ-dCas9-Sγ1-sgRNA lines is impeded, directly or indirectly, by the dCas9-bound Sγ1. As dCas9 impediment is not a complete block, loop extrusion in a subset of cells proceeds downstream, allowing dynamic sub-loop formation of RC with Iγ2b/Sγ2b or 3'CBEs. **j-l,** In RAG-sufficient cells, RC-bound RAG might enhance the dCas9-bound Sγ1 extrusion impediment. **m-p,** Elimination of Iγ2b-promoter-driven transcription permits unimpeded RAG-bound RC extrusion to 3'CBEs anchor, increasing RAG scanning activity there. **q-r,** 3C-HTGTS analysis of RC interactions with D_H and flanking regions in J_HΔ-dCas9 line (**q**) and D_H-$J_H^{+/-}$ line (**r**). DpnII (n = 4, biological replicates) and NlaIII (n = 3, biological replicates) digestions are shown for the J_HΔ-dCas9 line. NlaIII digestion more clearly reveals interaction peak near D_H3-2 due to paucity of DpnII sites in that region. NlaIII digestion of D_H-$J_H^{+/-}$ line shows a similar RC interaction pattern to that of J_HΔ-dCas9 line (**r,** n = 2, technical repeats). Bar graphs show relative RC interaction of the 25kb intervening D_H region (from D_H2-3 to D_H2-8) versus that of the same-size neighboring regions (n as indicated above). Data represents mean ± s.d (**q**) or mean (**r**). P values calculated via two-tailed paired t-test.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank Dr. Barry P. Sleckman for the *v-Abl* transformed, RAG2-deficient mouse pro-B cell line and RAG2 expression shuttle plasmid and Alt lab members for stimulating discussions. This work was supported by NIH R01 AI020047 (to F.W.A). F.W.A. is an investigator of the Howard Hughes Medical Institute. Y.Z. is a special fellow of the Leukemia and Lymphoma Society. Z.B. was a Cancer Research Institute Irvington fellow. E.L-A. was supported by an NSF Physics Frontiers Center Award (PHY1427654), the Welch Foundation (Q-1866), a USDA Agriculture and Food Research Initiative Grant (2017-05741), an NIH 4D Nucleome Grant (U01HL130010), and an NIH Encyclopedia of DNA Elements Mapping Center Award (UM1HG009375).

REFERENCES

1. Alt FW et al. Ordered rearrangement of immunoglobulin heavy chain variable region segments. EMBO J 3, 1209–1219 (1984). [PubMed: 6086308]
2. Rao SSP et al. A 3D map of the human genome at kilobase resolution reveals prin- ciples of chromatin looping. Cell 159, 1665–1680 (2014). [PubMed: 25497547]
3. Guo C et al. CTCF-binding elements mediate control of V(D)J recombination. Nature 477, 424–430 (2011). [PubMed: 21909113]
4. Jain S, Ba Z, Zhang Y, Dai H & Alt FW CTCF-Binding Elements Mediate Accessibility of RAG Substrates During Chromatin Scanning. Cell 174, 1–15 (2018). [PubMed: 29958101]
5. Teng G, & Schatz DG Regulation and evolution of the RAG recombinase. Adv. Immunol. 128, 1–39 (2015). [PubMed: 26477364]
6. Alt FW, Zhang Y, Meng F-L, Guo C & Schwer B Mechanisms of programmed DNA lesions and genomic instability in the immune system. Cell 152, 417–429 (2013). [PubMed: 23374339]
7. Sollbach AE & Wu GE Inversions produced during V(D)J rearrangement at IgH, the immunoglobulin heavy-chain locus. Mol Cell Biol 15, 671–681 (1995). [PubMed: 7823936]
8. Bolland DJ et al. Two Mutually Exclusive Local Chromatin States Drive Efficient V(D)J Recombination. Cell Rep 15, 2475–2487 (2016). [PubMed: 27264181]
9. Gauss GH & Lieber MR The basis for the mechanistic bias for deletional over inversional V(D)J recombination. Genes&Development 6,1553–1561 (1992). [PubMed: 1644296]

10. Wood C & Tonegawa S Diversity and joining segments of mouse immunoglobulin heavy chain genes are closely linked and in the same orientation: implications for the joining mechanism. Proc. Natl. Acad. Sci. USA 80, 3030–3034 (1983). [PubMed: 6407007]
11. Yancopoulos GD et al. Preferential utilization of the most J_H-proximal V_H gene segments in pre-B-cell lines. Nature 311, 727–733 (1984). [PubMed: 6092962]
12. Hu J et al. Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes. Cell 163, 947–959 (2015). [PubMed: 26593423]
13. Zhao L et al. Orientation-specific RAG activity in chromosomal loop domains contributes to Tcrd V(D)J recombination during T cell development. Journal of Experimental Medicine 213, 1921–1936 (2016). [PubMed: 27526713]
14. Lin SG, Ba Z, Alt FW & Zhang Y RAG Chromatin Scanning During V(D)J Recombination and Chromatin Loop Extrusion are Related Processes. Adv Immunol 139, 93–135 (2018). [PubMed: 30249335]
15. Sanborn AL et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. USA 112, E6456–E6465 (2015). [PubMed: 26499245]
16. Fudenberg G et al. Formation of Chromosomal Domains by Loop Extrusion. Cell Rep 15, 2038–2049 (2016). [PubMed: 27210764]
17. Via L, et al. The Energetics and Physiological Impact of Cohesin Extrusion. Cell 173, 1165–1178 (2018). [PubMed: 29706548]
18. Rowley MJ & Corces VG Organizational principles of 3D genome architecture. Nat Rev Genet 19, 789–800 (2018). [PubMed: 30367165]
19. Kim MS, Lapkouski M, Yang W & Gellert M Crystal structure of the V(D)J recombinase RAG1-RAG2. Nature 518, 507–511 (2015). [PubMed: 25707801]
20. Ru H et al. Molecular Mechanism of V(D)J Recombination from Synaptic RAG1-RAG2 Complex Structures. Cell 163, 1138–1152 (2015). [PubMed: 26548953]
21. Gapud EJ, Lee BS, Mahowald GK, Bassing CH & Sleckman BP Repair of chromosomal RAG-mediated DNA breaks by mutant RAG proteins lacking phosphatidylinositol 3-like kinase consensus phosphorylation sites. J Immunol 187, 1826–1834 (2011). [PubMed: 21742970]
22. Bredemeyer AL et al. ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. Nature 442, 466–470 (2006). [PubMed: 16799570]
23. Choi NM, et al. Deep sequencing of the murine IgH repertoire reveals complex regulation of nonrandom V gene re- arrangement frequencies. J. Immunol 191, 2393–2402 (2013). [PubMed: 23898036]
24. Gerstein RM & Lieber MR Coding end sequence can markedly affect the initiation of V(D)J recombination. Genes Dev 7, 1459–69 (1993). [PubMed: 8330743]
25. Benner C, Isoda T, & Murre C 2015. New roles for DNA cytosine modification, eRNA, anchors, and superanchors in developing B cell progenitors. Proc. Natl. Acad. Sci. USA 112, 12776–12781 (2015). [PubMed: 26417104]
26. Lutzker S, Rothman P, Pollock R, Coffman R & Alt FW Mitogen- and IL-4-regulated expression of germ-line Ig gamma 2b transcripts: evidence for directed heavy chain class switching. Cell 53, 177–84 (1988). [PubMed: 2834063]
27. Qi LS et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152, 1173–1183 (2013). [PubMed: 23452860]
28. Bevington S & Boyes J Transcription-coupled eviction of histones H2A/H2B governs V(D)J recombination. EMBO J 32, 1381–92 (2013). [PubMed: 23463099]
29. Bolland DJ, et al. Antisense intergenic transcription precedes Igh D-to-J recombination and is controlled by the intronic enhancer Emu. Mol Cell Biol 27, 5523–5533 (2007). [PubMed: 17526723]
30. Cong L, et al. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823 (2013). [PubMed: 23287718]
31. Yang L, et al. Optimization of scarless human stem cell genome editing. Nucleic Acids Res 41, 9049–9061 (2013). [PubMed: 23907390]

32. Hu J, et al. Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. Nat Protoc 11, 853–871 (2016). [PubMed: 27031497]
33. Core LJ, Waterfall JJ & Lis JT Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322:1845–1848 (2008). [PubMed: 19056941]
34. Mahat DB, et al. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). Nat Protoc 11, 1455–76 (2016). [PubMed: 27442863]
35. Marinov GK ChIP-seq for the Identification of Functional Elements in the Human Genome. Methods Mol Biol 1543, 3–18 (2017). [PubMed: 28349419]
36. Durand NC et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst 3, 95–98 (2016). [PubMed: 27467249]
37. Durand NC et al. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst 3, 99–101 (2016). [PubMed: 27467250]
38. Alt FW & and Baltimore D Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions. Proc Natl Acad Sci U S A 79, 4118–22 (1982). [PubMed: 6287467]
39. Cowell LG, Davila M, Kepler TB, & Kelsoe G. Identification and utilization of arbitrary correlations in models of recombination signal sequences. Genome Biol. 3: research0072 (2002).
40. Cowell LG, Davila M, Yang K, Kepler TB, & Kelsoe G. Spective estimation of recombination signal efficiency and identification of functional cryptic signals in the genome by statistical modeling. J. Exp. Med. 197: 207–220 (2003). [PubMed: 12538660]
41. Merelli I et al. RSSsite: a reference database and prediction tool for the identification of cryptic Recombination Signal Sequences in human and murine genomes. Nucleic Acids Res 38, W262–W267 (2010). [PubMed: 20478831]
42. Lieberman-Aiden E et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289–93 (2009). [PubMed: 19815776]
43. Predeus AV et al. Targeted Chromatin Profiling Reveals Novel Enhancers in Ig H and Ig L Chain Loci. J Immunol 192, 1064–1070 (2014). [PubMed: 24353267]
44. Medvedovic J et al. Flexible long-range loops in the VH gene region of the Igh locus facilitate the generation of a diverse antibody repertoire. Immunity 39, 229–244 (2013). [PubMed: 23973221]

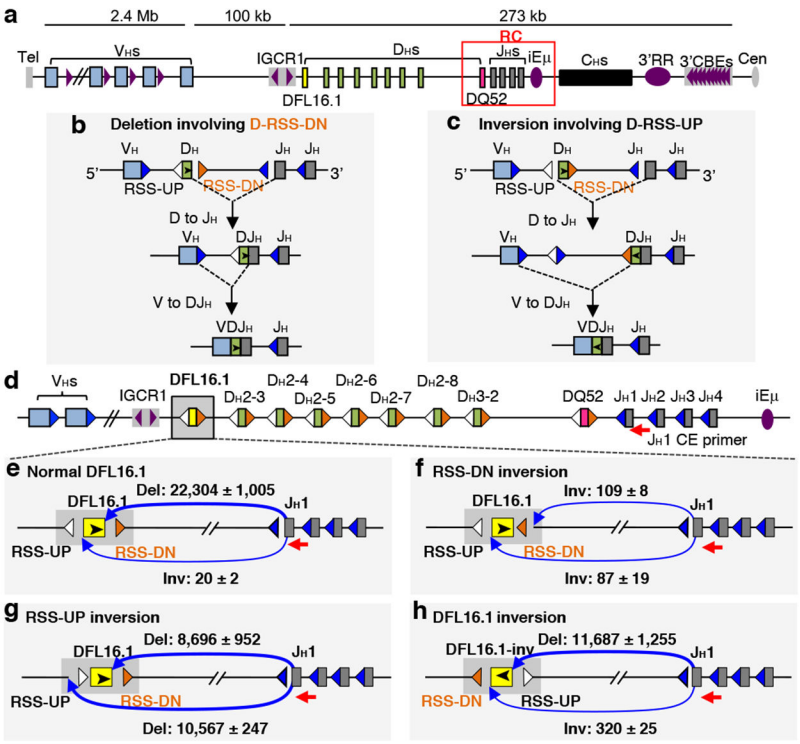


Figure 1. Role of RSS-based versus RAG scanning mechanisms in DFL16.1 deletional joining.

a, Schematic of the murine C57BL/6 *Igh* locus (not to scale) showing upstream V_Hs followed by the 273kb downstream 3'*Igh* domain anchored by IGCR1 and 3'CBEs. Approximate locations of Ds, RC, the C_H-containing region and 3'IgH regulatory region (3'RR) are indicated. Distal DFL16.1 and proximal DQ52 are indicated, respectively, by yellow and red boxes and 7 Ds in the 34kb region between them are indicated by green boxes. CBEs and their orientation are indicated by purple arrows. Tel: telomere. Cen: centromere. **b, c,** Illustration of deletional (b) and inversional (c) D to J_H V(D)J recombination mediated by joining, respectively, between D-RSS-DN (orange) or D-RSS-UP (white) and the J_H-RSSs (blue). Black arrow inside D_H coding segment (green box) denotes orientation of coding sequence. **d,** Schematic of RSSs flanking proximal V_Hs, 9 Ds, and 4 J_Hs. Names of the 7 Ds between DFL16.1 and DQ52 are indicated above them. Red arrow indicates J_H1 CE bait primer used in HTGTS V(D)J-seq; other symbols are as in panel **a**. **e-h**, HTGTS V(D)J-seq analysis of the D_H-$J_H^{+/-}$ line and mutant derivatives, showing relative utilization of D-RSS-DN versus D-RSS-UP for normal DFL16.1 joining to J_H1 (n = 3 libraries) (**e**), and effects of indicated DFL16.1 modifications including: **f,** DFL16.1-RSS-DN inversion ("DFL16.1$^{\text{RSS-DN-inv}}$") (n = 3 libraries); **g,** DFL16.1-RSS-UP inversion ("DFL16.1$^{\text{RSS-UP-inv}}$") (n = 3 libraries) and **h,** inversion of the entire DFL16.1("DFL16.1$^{\text{inv}}$") (n = 5 libraries). Del: deletional joins. Inv: inversional joins. Each library in panel **e-h** was normalized to 40,000 total library junctions. Data is presented as mean ± s.d from biologically independent samples.

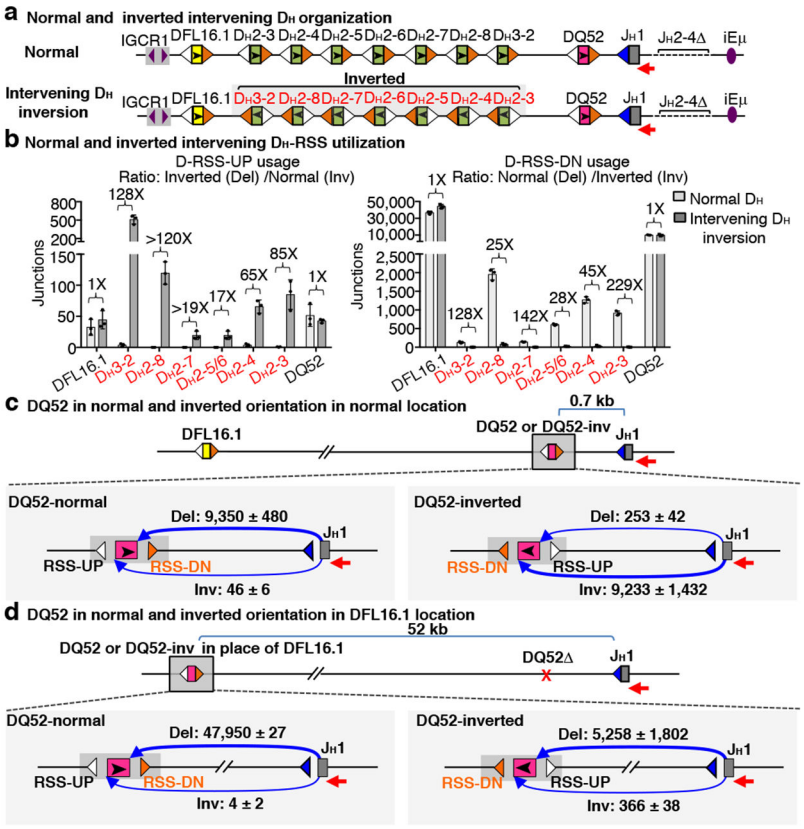


Figure 2. Mechanism of orientation-biased D to J_H joining of 7 Ds between DFL16.1 and DQ52.

a, Illustration of CRISPR/Cas9-mediated inversion of a 34kb D_H region between DFL16.1 and DQ52, which contains 7 functional Ds, in the D_H-$J_H1^{+/-}$ line. Other details are as in Fig. 1. **b,** HTGTS V(D)J-seq analysis (J_H1 CE primer) of utilization of D-RSS-UP (left panel) and D-RSS-DN (right panel) in the D_H-$J_H1^{+/-}$ line and its mutant derivative with the intervening D_H inversion ($n = 3$ libraries for each genotype). The fold change between mean usage level of each D_H in normal versus inverted locale indicated was calculated as inverted/normal for D-RSS-UP and calculated as normal/inverted for D-RSS-DN. **c,** Relative utilization of DQ52-RSS-DN versus DQ52-RSS-UP for normal DQ52 (left) and DQ52 inversion ("$DQ52^{inv}$", right) in D_H-$J_H1^{+/-}$ line ($n = 3$ libraries for each genotype). **d,** Relative utilization of DQ52-RSS-DN versus DQ52-RSS-UP for normal DQ52 (left) or DQ52 inversion (right) when located in place of DFL16.1 in D_H-$J_H1^{+/-}$ line with endogenous DQ52 deleted ("DQ52Δ $DFL16.1^{DQ52}$" and "DQ52Δ $DFL16.1^{DQ52\text{-}inv}$") ($n = 3$ libraries for each genotype). Each library in panel **b-d** was normalized to 70,000 total junctions. Data represents mean ± s.d from biologically independent samples.

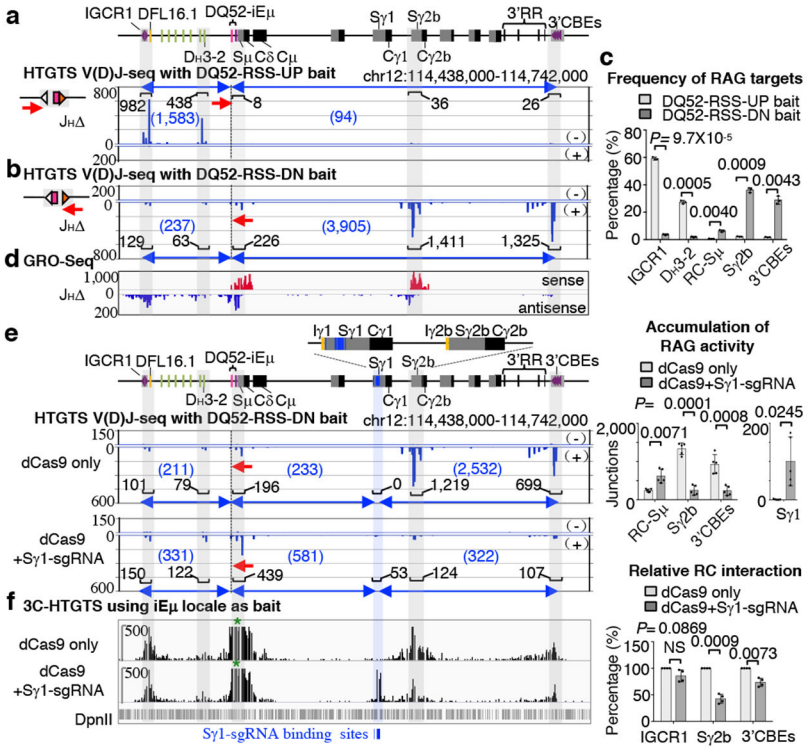


Figure 3. Binding of dCas9 impedes downstream RAG scanning and associated loop formation.

a-d, Characterization of upstream and downstream RAG scanning from DQ52-based RC. **a**, HTGTS V(D)J-seq profile of $J_HΔ$ line with DQ52-RSS-UP bait (red arrow). The ‘‘+’’ and “−” labels denote prey sequence read orientation relative to the centromere which identifies deletional versus inversional joins (see Methods). Black dashed line indicates bait position. **b.** HTGTS V(D)J-seq of $J_HΔ$ line with DQ52-RSS-DN bait (red arrow). **c**, Bar graph shows RAG scanning activity at indicated locales as percentage of total activity within 3’*Igh* domain (n = 3 libraries for both DQ52-RSS-UP bait and DQ52-RSS-DN bait). **d,** GRO-Seq of 3’*Igh* domain in $J_HΔ$ line. Transparent grey bars through a, b, and d panels indicate locations of the most robust RAG cryptic scanning activity. **e, f,** Characterization of dCas9 binding effects on downstream RAG scanning and chromatin looping. **e,** HTGTS V(D)J-seq of DQ52-RSS-DN joining in $J_HΔ$-dCas9 versus $J_HΔ$-dCas9-Sγ1-sgRNA line. Top: zoom-in of the Iγ1-Cγ2b region. Transparent blue and grey bars indicate, respectively, location of 16 dCas9 binding sites within C57BL/6 Sγ1 and regions of evident RAG activity. Bar graphs compare RAG junctions at indicated sites (n = 5 libraries for each genotype). **f,** 3C-HTGTS profiles showing RC interactions within 3’*Igh* domain in $J_HΔ$-dCas9 versus $J_HΔ$-dCas9-Sγ1-sgRNA line. Green star indicates iEμ bait location. Bar graphs compare RC interaction frequency with indicated regions for the two lines (n = 4 libraries for each genotype). Data represents mean ± s.d in panel **c, e, f** from biologically independent samples. P values were calculated via two-tailed paired t-test. NS: not significant, $P \geq 0.05$. Repeat experiments for all panels are in Extended Data Fig. 4, 7 and 8.

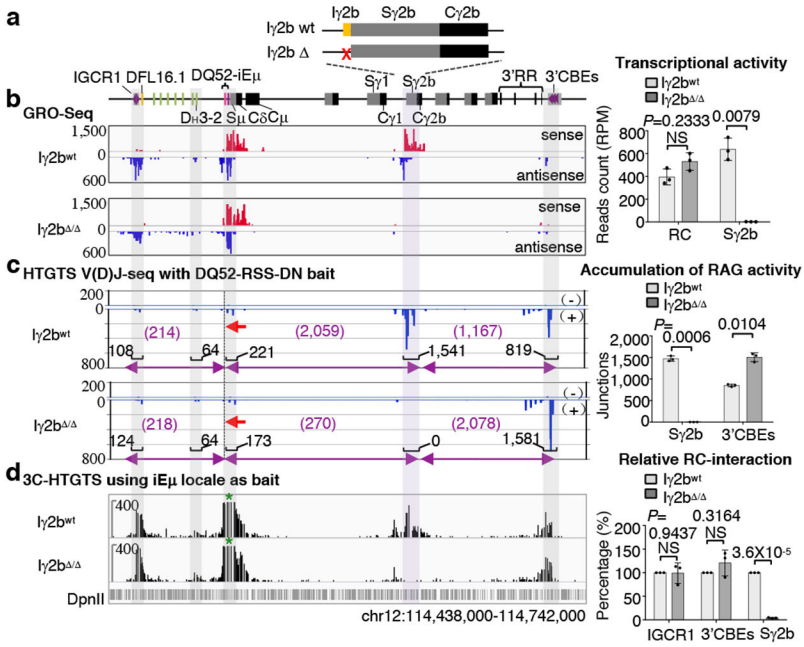


Figure 4. Active transcription across Sγ2b impedes loop extrusion-mediated RAG scanning.

a, Schematic of Iγ2b-Cγ2b region with normal or deleted Iγ2b. **b**, Representative GRO-Seq profiles of J_HΔ-dCas9 line ("Iγ2b^{wt}") and J_HΔ-dCas9-Iγ2b-del line ("Iγ2b$^{Δ/Δ}$"). Bar graph shows comparison of transcriptional activity of the indicated regions for Iγ2b^{wt} versus Iγ2b$^{Δ/Δ}$ (n = 3 libraries for each genotype). **c,** Representative HTGTS V(D)J-seq profiles showing breaks joining to DQ52-RSS-DN in the Iγ2b^{wt} line versus the Iγ2b$^{Δ/Δ}$ line. Bar graph shows comparison of RAG junctions at the indicated regions in Iγ2b^{wt} lines versus Iγ2b$^{Δ/Δ}$ lines (n = 3 libraries for each genotype). **d,** Representative 3C-HTGTS profiles showing RC interactions in Iγ2b^{wt} line versus Iγ2b$^{Δ/Δ}$ line. Bar graph shows comparison of RC interaction frequency with indicated regions in Iγ2b^{wt} lines versus Iγ2b$^{Δ/Δ}$ lines (n = 3 libraries for each genotype). Data represents mean ± s.d in panel **b-d** from biologically independent samples. P values were calculated via two-tailed paired t-test. NS: $P \geq 0.05$. Repeat experiments for all panels are shown in Extended Data Fig. 9.