The fundamental role of chromatin loop extrusion in physiological V(D)J recombination

Yu Zhang1,2,3,4,13,14, Xuefei Zhang1,2,3,4, Zhaoqing Ba1,2,3, Zhuoyi Liang1,2,3, Edward W. Dring1,2,3, Hongli Hu1,2,3, Jiangman Lou1,2,3, Nia Kyritsis1,2,3, Jeffrey Zurita1,2,3, Muhammad S. Shamim5,6,7,8, Aviva Presser Aiden5,6,9, Erez Lieberman Aiden5,7,10,11,12 & Frederick W. Alt1,2,3

The RAG endonuclease initiates IgH V(D)J assembly in B cell progenitors by joining D segments to JH segments, before joining upstream VH segments to DJH intermediates1. In mouse progenitor B cells, the CTCF-binding element (CBE)-anchored chromatin loop domain2 at the 3’ end of IgH contains an internal subdomain that spans the 5’ CBE anchor (IGCR1)3, the DJ segments, and a RAG-bound recombination centre (RC)4. The RC comprises the JH1-proximal D segment (DSQ52), four JH segments, and the intronic enhancer (iEH)5. Robust RAG-mediated cleavage is restricted to paired V(D)J segments flanked by complementary recombination signal sequences (12RSS and 23RSS)6. D segments are flanked downstream and upstream by 12RSSs that mediate deletional joining with convergently oriented 12RSSs and 23RSSs, respectively6. Despite 12/23 compatibility, inversionsal D-to-JH joining via upstream D-12RSS is rare5,6. Plasmid-based assays have attributed the lack of inversionsal D-to-JH joining to sequence-based preference for downstream D-12RSSs5, as opposed to putative linear scanning mechanisms8,11. As RAG linearly scans convergent CBE-anchored chromatin loops12,14,15, potentially formed by cohesin-mediated loop extrusion15,16, we revisited its scanning role. Here we show that the chromosomal orientation of JH1-23RSS programs RC-bound RAG to linearly scan upstream chromatin in the 3’ IgH subdomain for convergently oriented D-12RSSs and, thereby, to mediate deletional joining of all D segments except RC-based DSQ52, which joins by a diffusion-related mechanism. In a DSQ52-based RC, formed in the absence of JH1 segments, RAG bound by the downstream DSQ52-RSS scans the downstream constant region exon-containing 3’ IgH subdomain, in which scanning can be impeded by targeted binding of nuclease-dead Cas9, by transcription through repetitive IgH switch sequences, and by the 3’ IgH CBE-based loop anchor. Each scanning impediment focally increases RAG activity on potential substrate sequences within the impeded region. High-resolution mapping of chromatin interactions in the RC 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 proteins and two cofactor RAG2 proteins19,20 (Extended Data Fig. 1a). We tested the hypothesis that upon acquisition of a JH1-23RSS in one active site of RAG, the JH RC 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-JH recombination (Extended Data Fig. 1b–e, Supplementary Video). We tested this hypothesis by mutational analyses of the effect of D-RSS orientation on deletional versus inversionsal D-to-JH rearrangement in a physiological D-JH–RC-containing chromosomal domain (Fig. 1a–c). To facilitate analyses, we used Cas9-guide RNA (gRNA) targeting to delete the DJH–JH–RC domain on one allele of a v-Ab1 transformed, RAG2-deficient progenitor B cell (pro-B) line21, referred to as the DJH–JH+/− parent (Extended Data Fig. 2a–c). To activate V(D)J recombination, we introduced RAG2 into DJH–JH+/− cells or mutant derivatives of these cells, and treated them with a v-Ab1 kinase inhibitor to induce G1 arrest, stabilization of RAG2, and robust D-to-JH joining potential22. Each experiment was repeated at least three times and used multiple independent mutant DJH–JH+/− derivatives. We used high-throughput genome-wide translocation sequencing (HTGTS) V(D)J sequencing6 (V(D)J-seq) to analyse V(D)J junctions with a JH1 coding end (CE) primer, which revealed that junctions were overwhelmingly deletional DJH–JH+/− junctions (Extended Data Fig. 2d–g). Similar to primary pro-B cells23, JH-distal DJFL16.1 had the highest rearrangement frequency (66%), with JH-proximal DSQ2 had the second highest frequency (27%), and the seven (‘intervening’) Ds between DF16.1 and DSQ2 had lower rearrangement frequencies (Extended Data Fig. 2e).

To test the effect of DJFL16.1-12RSS orientation on D-to-JH recombination, we separately inverted its downstream RSS (RSS-DN) and upstream RSS (RSS-UP) in the DJH–JH+/− line (Fig. 1d–g). Inversion of DJFL16.1-RSS-DN, placing it in the same orientation as DJFL16.1-RSS-UP, made it essentially inert for JH joining (Fig. 1e, f). By contrast, inversion of DJFL16.1-RSS-UP led to robust deletional JH joining to the surrogate CE sequences (non-D upstream flanking sequences), with levels similar to those of deletional joins mediated by DJFL16.1-RSS-DN (Fig. 1g). To rule out effects of adjacent coding sequences24, we inverted DJFL16.1, including both D-RSS-UP and D-RSS-DN (Fig. 1h). DJFL16.1-RSS-UP in the downstream position converged to the DJH–JH1-23RSSs mediated robust deletional joining: while DJFL16.1-RSS-DN in the upstream position in the same orientation as the DJH–JH1-23RSSs showed less than 2% of normal activity (Fig. 1h). The eight unmodified Ds showed little change in rearrangement patterns (Extended Data Fig. 2h).

Therefore, recognition of the DJFL16.1-RSS-DN, due to convergent orientation with the JH1-23RSS, preserves deletional-orientation DJFL16.1-to-JH joining and relative RSS-DN versus RSS-UP strength does not substantially affect this process (Extended Data Fig. 3a, Supplementary Discussion).

To further test the effect of D-RSS orientation on D-to-JH joining, we eliminated the potential confounding effects of chromosomal or extra-chromosomal secondary joins (Extended Data Fig. 3b) by deleting the JH1-2–4 sequence of the DJH–JH+/− line to generate the DJH–JH+/− line, which undergoes D-to-JH1 recombination similarly to its DJH–JH+/− parent (Extended Data Fig. 3c, Supplementary Table 1). We then inverted the region containing all seven intervening Ds in the DJH–JH+/− line (Fig. 2a, Extended Data Fig. 3d). Analyses of DJH+/−
rearrangements in this line revealed greatly increased relative utilization of D-RSS-DN versus D-RSS-UP (left) and D-RSS-DN (right) in the D_H–JH_1 line and mutant derivatives, showing relative utilization of D-RSS-DN and normal/inverted for D-RSS-DN. C. Relative utilization of DQ52-RSS-DN versus DQ52-RSS-UP for normal DQ52 (left) and DQ52 inversion (right) in the D_H–JH_1 line and its mutant derivative with inversion of intervening D_Hs (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-DN and as normal/inverted for D-RSS-UP.

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–JH_1 line with endogenous DQ52 deleted (n = 3 libraries for each genotype). Each library in b–d was normalized to 70,000 total junctions. Data are presented as mean ± s.d. from 3 biologically independent samples.
Fig. 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 IgHΔ line with DQ52-RSS-UP bait (red arrow). Plus and minus denote prey sequence read orientation relative to the centromere, which identifies deletional versus inversionsal joins (see Methods). Black dashed line indicates bait position. b, HTGTS V(D)J-seq of IgHΔ line with DQ52-RSS-UP bait (red arrow). c, 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, Global run-on sequencing (GRO-seq) of 3′ Igh, HTGTS V(D)J-seq of JH1 line. Bar graph compares HC interaction frequency with dCas9 only (green), dCas9 + Sy1-sgRNA (blue), and Δ-dCas9 versus JH1 line. Top, expanded view of I-1–Cγ2b region. Transparent blue box spanning e and f indicates location of dCas9 binding sites within C57BL/6 SYγ1. Transparent grey boxes indicate regions of evident RAG activity. Bar graph compares RAG junctions at indicated sites (n = 5 libraries for each genotype). e, 3C-HTGTS profiles showing RC interactions within 3′ Igh domain in IgHΔ–dCas9 versus IgHΔ–dCas9–Sy1-sgRNA line. Green asterisks indicate iE1-bait location. Bar graph compares RC interaction frequency with indicated regions for the two lines (n = 4 libraries for each genotype). Data represent mean ± s.d. from 3 (c), 5 (e) or 4 (f) biologically independent samples. P values were calculated via two-tailed paired t-test. NS, not significant, P ≥ 0.05. Repeat experiments for all panels are in Extended Data Fig. 4, 7, 8.

Fig. 4a–d, Supplementary Discussion). HTGTS V(D)J-seq on the IgHΔ line revealed that DQ52-RSS-UP initiates RAG scanning to convergent cryptic RSSs within the RC-upstream D subdomain, with robust RAG cryptic scanning activity at the transcribed heptamer of the non-12/23 compatible DFL16.1–RSS-DN and convergent CAC within DFL3-2-RSS-UP (Fig. 3a, c, Extended Data Fig. 4c, 5a, Supplementary Table 2). DQ52-RSS-DN initiated RAG scanning activity across the downstream constant region exon (CCH)–containing subdomain, with robust activity at cryptic heptamers within the repetitive Sγ2b switch region upstream of Cγ2b and in the 3′ Igh CBE anchor36 (Fig. 3b, c, Extended Data Fig. 4d, 5b, Supplementary Table 2). While RAG scanning activity in Sγ2b coincided with robust transcription from the immediately upstream I-2b promoter26, the 3′ CBE 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–Iγ14 joins, and found that the latter lines lacked Sγ2b transcription and corresponding Sγ2b RAG scanning activity and that deletion of 3′ CBEs in these lines 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).

We used focal RAG downstream scanning activity from DQ52-RSS-DN in the IgHΔ line to further characterize the mechanism of RAG scanning. We investigated whether the introduction of sequential sites of nucleoside–dead Cas9 (dCas9)27 would generate 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 to 3′ CBE targets via an Sγ1-sgRNA that binds 16 sites within a 4-kb portion of Sγ1 on the intact IgHΔ allele (Extended Data Fig. 7a). We derived multiple independent clones with stable dCas9 expression (IgHΔ–dCas9 lines) or with both dCas9 and Sγ1-sgRNA expression (IgHΔ–dCas9–Sγ1-sgRNA lines; Extended Data Fig. 7b, c). HTGTS V(D)J-seq with a DQ52-RSS-DN primer confirmed downstream scanning by RAG in multiple IgHΔ–dCas9 lines with junction profiles similar to those of the IgHΔ line, including accumulation at Sγ2b and 3′ CBEs (Fig. 3e, Extended Data Fig. 7d). Strikingly, IgHΔ–dCas9–Sγ1-sgRNA lines had highly
diminished RAG scanning downstream of the dCas9-targeted Sγ1, along with modestly decreased Sγ2 transcription (Fig. 3e, Extended Data Fig. 7d, f, Supplementary Discussion). In J11Δ-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 Sγ2 (Fig. 3f, Extended Data Fig. 7d, e). These findings indicate that binding of dCas9 impedes downstream scanning by RAG.

High throughput chromosome conformation capture (Hi-C) analyses of J11Δ-dCas9 and J11Δ-dCas9-Sγ1-sgRNA lines revealed that transcription loops spanning the Sγ1 impediment in the latter were weakened, with new loops formed between the Sγ1 impediment and upstream RC or downstream 3′ CBE loop anchor (Extended Data Fig. 8a). Sensitive chromosome conformation capture (3C)-HTGTS5 on J11Δ-dCas9 lines revealed that iEμ interacted robustly with major RAG scanning target regions including IGCR1, D3j3–2, S2b, and 3′ CBEs (Fig. 3f, Extended Data Fig. 8b). In J11Δ-dCas9-Sγ1-sgRNA cells, iEμ gained robust interactions with dCas9-bound Sγ1 and had decreased interactions with downstream Sγ2b and 3′ CBEs (Fig. 3f, Extended Data Fig. 8b). Thus, the decreased RAG scanning activity at the dCas9-impediment-downstream regions correlates with their decreased interaction with the RC. In J11Δ-dCas9-Sγ1-sgRNA lines, incomplete scanning inhibition downstream of Sγ1, along with broad RAG scanning activity and RC interactions across Sγ2b, indicate that dynamic extrusion of Sγ2b across the RC is impeded, but not abrogated, by binding of Sγ1 to dCas9 (Extended Data Fig. 9a–i, Supplementary Discussion). The greater effect of the Sγ1 dCas9 impediment on RAG scanning when compared with 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. 9j–l).

Chromatin immunoprecipitation sequencing (ChIP-seq) of J11Δ-dCas9-Sγ1 cells revealed strong binding of the RAD21 cohesin subunit18 at the IGCR1 and 3′ Iγh CBE loop anchors, and lower accumulation across transcribed iEμ–Sγ1 and Iγ2b–Sγ2b sequences (Extended Data Fig. 8c). In addition, a new RAD21 peak occurred at the impeded dCas9-bound Sγ1 in J11Δ-dCas9-Sγ1-sgRNA cells (Extended Data Fig. 8c). Furthermore, NIPBL, a cohesin-loading factor19, accumulated across transcribed iEμ–Sγ1 and Iγ2b–Sγ2b sequences and downstream Iγh regions, including the 3′ CBEs. There was substantial additional accumulation of NIPBL at the non-transcribed Sγ1 in J11Δ-dCas9-Sγ1-sgRNA cells (Extended Data Fig. 8d), raising the possibility that dCas9 binding, beyond direct steric interference20, may impede scanning-related extrusions via a mechanism that involves 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 J11Δ-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. 10a–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. 10b). 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 occurred in Sγ2b-transcribing cells (Fig. 4c, d, Extended Data Figs. 9a–d, m–p, 10b, c). Elimination of transcription also might decrease RAG activity on RC-aligned targets by chromatin accessibility mechanisms21.

Our data suggest that loop-extrusion-mediated RAG scanning has a crucial role in the initiation of physiological D-to-Iγh joining (Supplementary Video). During linear RAG scanning, downstream D-RSSs that are convergently oriented with initiating RAG-bound J11-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 utilization of convergent cryptic RSSs during RAG scanning from ectopic RCs in non-antigen receptor loop domains12. During scanning, impediments to loop extrusion, including CBE anchors, transcription, and dCas9 binding, focus RAG to targets within impediment regions. Robust utilization of DFL16.1–RSS-DN correlates with its location just downstream of IGCR1 CBE anchors, which impede extrusion-mediated RAG scanning, leading to strong interactions with the RC. Low, but still relevant, utilization of intervening Ds may be promoted by their location in an antisense-transcribed, repetitive region12 that modestly impedes loop extrusion and increases accessibility to the RC (Extended Data Fig. 9q, r, Supplementary Discussion). Conversely, loop extrusion may also frequently isolate DQ52 from the RC, preventing it from dominating overall D-to-Iγh joining via diffusion. Finally, our findings with dCas9 suggest that additional, undefined, chromatin-based mechanisms may enhance the synopsis of functional cis-elements via loop extrusion more generally.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1547-y.

Received: 16 April 2019; Accepted: 12 August 2019; Published online 11 September 2019.

1. Alt, F.W. et al. Ordered rearrangement of immunoglobulin heavy chain variable region segments. EMBO J. 3, 1209–1219 (1984).
2. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680 (2014).

3. Guo, C. et al. CTCF-binding elements mediate control of V(D)J recombination. Nature 477, 424–430 (2011).

4. Jain, S., Ba, Z., Zhang, Y., Dai, H. Q. & Alt, F. W. CTCF-binding elements mediate accessibility of RAG substrates during chromatin scanning. Cell 174, 102–116 (2018).

5. Teng, G. & Schatz, D. G. Regulation and evolution of the RAG recombinase. Adv. Immunol. 128, 1–39 (2015).

6. Alt, F. W., 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).

7. Sollbach, A. E. & Wu, G. E. Inversions produced during V(D)J rearrangement at IgH, the immunoglobulin heavy-chain locus. Mol. Cell. Biol. 15, 671–681 (1995).

8. Bolland, D. J. et al. Two mutually exclusive local chromatin states drive efficient V(D)J recombination. Cell Rep. 15, 2475–2487 (2016).

9. Gauss, G. H. & Lieber, M. R. The basis for the mechanistic bias for deletional over inversional V(D)J recombination. Genes Dev. 6, 1553–1561 (1992).

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).

11. Yancopoulos, G. D. et al. Preferential utilization of the most JH-proximal VH gene segments in pre-B-cell lines. Nature 311, 727–733 (1984).

12. Hu, J. et al. Chromosomal loop domains direct the recombination of antigen receptor genes. Cell 163, 947–959 (2015).

13. Zhao, L. et al. Orientation-specific RAG activity in chromosomal loop domains contributes to Tcrd V(D)J recombination during T cell development. J. Exp. Med. 213, 1921–1936 (2016).

14. Lin, S. G., Ba, Z., Alt, F. W. & Zhang, Y. RAG chromatin scanning during V(D)J recombination and chromatin loop extrusion are related processes. Adv. Immunol. 139, 93–135 (2018).

15. Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl Acad. Sci. USA 112, E5456–E5465 (2015).

16. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. Cell Rep. 15, 2038–2049 (2016).

17. Vian, L. et al. The energetics and physiological impact of cohesin extrusion. Cell 173, 1165–1178 (2018).

18. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. Nat. Rev. Genet. 19, 789–800 (2018).

19. Kim, M. S., Lapkouski, M., Yang, W. & Gellert, M. Crystal structure of the V(D)J recombinase RAG1–RAG2. Nature 518, 507–511 (2015).

20. Gapud, E. J., Lee, B. S., Mahowald, G. K., Bassing, C. H. & Sleckman, B. P. 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).

21. Bredemeyer, A. L. et al. ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. Nature 442, 466–470 (2006).

22. Choi, N. M. et al. Deep sequencing of the murine IgH repertoire reveals complex regulation of nonrandom V gene rearrangement frequencies. J. Immunol. 191, 2393–2402 (2013).

23. Gerstein, R. M. & Lieber, M. R. CodiR end sequence can markedly affect the initiation of V(D)J recombination. Genes Dev. 7, 1459–1469 (1993).

24. Benner, C., Isoda, T. & Murre, C. New roles for DNA cytosine modification, eRNA, anchors, and superanchors in developing B cell progenitors. Proc. Natl Acad. Sci. USA 112, 12776–12781 (2015).

25. Lutzker, S., Rothman, P., Pollock, R., Coffman, R. & Alt, F. W. Mitogen- and IL-4-regulated expression of germ-line Igγ2b transcripts: evidence for directed heavy chain class switching. Cell 53, 177–184 (1988).

26. Qi, L. S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152, 1173–1183 (2013).

27. Bevington, S. & Boyes, J. Transcription-coupled eviction of histones H2A/H2B governs V(D)J recombination. EMBO J. 32, 1381–1392 (2013).

28. Bolland, D. J. 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).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019
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-Ab1 pro-B cell lines. The CRISP-Cas9 approach was used to generate the various mutant strains in this study. The D13-JH1 line was derived from a previously described Rag2- Δβt-12C2-12E2-expressing mouse v-Ab1- pro-B cell line13 with pMSCV-RAG2-GFP vector transfected at 5 µg/ml geneticin. We deleted the entire D13-JH1-RC region (from −400 bp upstream of DFL1.6 to −100 bp downstream of Eµ1) on the 129/Sv allele, leaving the C57BL/6 allele intact, and confirmed the deletion via Southern blotting. The D13-JH1-Δ line 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 inversionsal mutations including JH1-Δ4 deletion, inversion of intervening Dµs, JH1-Δ4 deletion (all three mutations were confirmed via Southern blotting) and γ2b deletion (confirmed via PCR genotyping followed by Sanger sequencing). CRISPR-Cas9 targeting combined with short single-stranded DNA oligonucleotide (ssODN) templates was used to generate precise mutations including DFL1.6-RSS-UP inversion (DFL1.6-RSS-UP-inv), DFL1.6-RSS-DN inversion (DFL1.6-RSS-DN-inv), DFL1.6 inversion (DFL1.6-inv), DQ52 inversion (DQ52-inv), DQ52 or DQ52 inversion in place of DFL1.6 (DQ52-replace-DFL1.6 or DQ52-inv-replace-DFL1.6), and DFL1.6 or DFL1.6 inversion in place of DQ52 (DFL1.6-replace-DQ52 or DFL1.6-inv-replace-DQ52). These mutations were confirmed via PCR genotyping followed by Sanger sequencing.

The DFL1.6-JH1-Δ line was derived from the DFL1.6-JH1 line via inversion of a −1-kb fragment containing the DFL1.6-JH1-Δ region (verified via Southern blotting). The DFL1.6-JH1-Δ line served as parents to generate DFL1.6-JH1-Δ3 C BEs−/− lines by deleting in the −9-kb flanking region containing all 10 CBEs (verified by Southern blotting) and DFL1.6-JH1-Δ3 C BEs−/− lines by deleting the exon coding for Rag2 (verified by PCR genotyping and Sanger sequencing). At least two independent clones were obtained for each mutation for analysis. The v-Ab1 pro-B cell lines were cultured in RPMI medium with 15% FBS (v/v). Cells were not tested for mycoplasma contamination. Sequences of Cas9-gRNAs and ssODNs are listed in Supplementary Table 3. Original gel scans for related Southern blotting confirmation in Extended Data Figures can be found in Supplementary Figure 1.

Generation of v-Ab1 lines with targeted dsCas9-binding to Sγ1 region. To introduce targeted dsCas9 binding to the Sγ1 region, we first generated the dsCas9-expressing JH1-Δ-dsCas9 lines. We swapped the open reading frame (ORF) of the pyrcunin-resistance gene with that of the neomycin-resistance gene on the retroviral pdCas9-humanized vector (Addgene, 44246) and transduced the modified vector into the JH1 line. Infected cells were selected with 1,600 µg/ml genetin (Life Technologies, 11811-031) 2 days post-infection at a concentration of 100 cells/well in 96-well plates. Selection was maintained for 8–10 days until stable colonies appeared. Genetin-resistant colonies were screened for dsCas9 expression via western blotting with Cas9 antibody (Diagenode, C15310258), using β-actin antibody (Cell Signalling Technology, 3700S) as a loading control. Positive colonies were further subcloned and verified via western blotting for Cas9 expression to generate the JH1-Δ-dsCas9 lines, which were then maintained in RPMI medium with 400 µg/ml genetin. We then used the JH1-Δ-dsCas9 lines as parental lines to generate the JH1-Δ-dsCas9-Sγ1-sgRNA lines, which expressed both dsCas9 and Sγ1-sgRNA. We swapped the ORF of the pyrcunin-resistance gene with that of the bleomycin-resistance gene on a lentiviral mCherry-expressing, pgRNA-humanized vector (Addgene, 44248), cloned Sγ1-sgRNA-expressing oligos into the modified vector and transduced it into the JH1-Δ-dsCas9 lines. Infected cells were selected with 800 µg/ml zeocin (ThermoFisher Scientific, R25001) 2 days post-infection at the concentration of 100 cells/well in 96-well plates. Selection was maintained for 8 to 10 days until stable zeocin-resistant colonies appeared. Zeocin-resistant colonies with high mCherry expression were screened for Sγ1-sgRNA expression via reverse transcription PCR (RT–PCR). Positive colonies were further subcloned and verified via high mCherry expression and RT–PCR for Sγ1-sgRNA expression to obtain the JH1-Δ-dsCas9-Sγ1-sgRNA lines, which were then maintained in RPMI medium with 400 µg/ml genetin and 400 µg/ml zeocin. Original gel scans for western blotting and RT–PCR confirmation in Extended Data Figures can be found in Supplementary Figure 1.

RAG complementation. The Rag2-expressing vector pMSCV-FLAG-RAG2-GFP was generated by cutting out the FLAG-RAG2-GFP sequence from the shuttle vector pSP27-FLAG-RAG2-KV7, GFP25 via Hpal and XhoI digestion and cloning the sequence into the pMSCV-puro vector (Addgene, K10602) via the same restriction sites. Rag2 was reconstituted in Rag2-deficient v-Ab1 cells via retroviral infection of cells with the pMSCV-FLAG-RAG2-GFP vector followed by 3 days of pyrurcin selection to enrich for cells with virus integration.

HTGTS (V(D)-seq) library preparation. HTGTS (V(D)-seq) libraries were prepared as described previously. Genomic DNA was extracted from Rag2-complemented cells arrested in G1 for 4 days with treatment of 3 mM STI-571. In brief, 10 µg DNA was fragmented via sonication on a Diagenode bio-bruptor and subjected to linear PCR amplification with a biotinylated primer. Single-stranded PCR products were purified using Dynabeads MyONE C1 streptavidin beads (Life Technologies, 65002) and ligated to bridge adapters. Adapter-ligated products were amplified by nested PCR with indexed locus-specific primers and primer annealed to adapter. The PCR products were further tagged with Illumina sequencing adapters and size-selected via gel extraction and loaded onto a MiSeq machine (Illumina) for paired-end 250-bp or 300-bp sequencing. Primer information can be found in Supplementary Table 3.

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

HTGTS (V(D)-seq) analysis of deletional and inversionsal D to JH recombination. A JH1 CE primer was used as a bait primer to detect D-to-JH joining events. Libraries were size-normalized to total junctions of the smallest library within the set of libraries being compared. The utilization frequency of D-RSS-UPs and D-RSS-DNs was determined by counting the number of junctions containing corresponding RSS-associated coding sequences within size-normalized libraries. As RSS-proximal coding nucleotides are frequently processed during CE formation, the utilization percentage was calculated by subtracting PCR artefacts.

HTGTS (V(D)-seq) analysis of JH1-JH4 deletion, inversion of intervening DHs, and JH inversion. HTGTS (V(D)-seq) analysis of size-selected genomic DNA extracted from DQ52-RSSs, 5′ DQ52-DSS-RSS-UP (137 bp upstream of DQ52-DSS-RSS break site) and 3′ DQ52 (DQ52-DSS-RSS) bait primers (145 bp downstream of DQ52-DSS-RSS break site) were used to analyse the DQ52-RSS-DQ52 and DQ52-DSS junctions.
profiles. The 5′ DQ52 primer can simultaneously detect DSβ ends joining to DQ52-SS-UP signal ends (SEs) and DQ52-SS-DN coding ends (CEs). Similarly, the 3′ DQ52 primer can simultaneously detect DSβ ends joining to DQ52-SS-SEs and DQ52-SS-CEs. To compare the RPG cystic scanning profiles of DQ52-SS-UP and DQ52-SS-DN, we extracted DSβ ends joining to the same type of RPG break ends for the two RSSs from HTGTS V(D)J-seq libraries. As such, we plotted the DQ52-SS-UP SE junctions extracted from the 5′ DQ52 primer libraries and DQ52-SS-SEs CE junctions extracted from the 3′ DQ52 primer libraries. Junctions were plotted via Prism. Junctions are denoted as in ‘plus’ orientation if the prey sequence reads in centromere-to-telomere direction and in ‘minus’ orientation if the prey sequence reads in telomere-to-centromere direction. For DQ52-SS-UP SE joining profiles, ‘minus’ junctions are deletions and ‘plus’ junctions are inversions. For DQ52-SS-DN SE joining profiles, ‘plus’ junctions are deletions and ‘minus’ junctions are inversions. Note that although not shown, CE joining profiles showed very similar patterns of RAG targeting to those of SE joining profiles of the same RSSs. We used coordinates of bait length to extract SE versus CE junctions from a given primer with criteria similar to those described previously. 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-SS-UP SE profiles from the 5′ DQ52 primer libraries; junctions with bait length 142–147 bp were used for analysis of DQ52-SS-UP SE profiles from the 3′ DQ52 primer libraries; junctions with bait length 148–158 bp were used for analysis of DQ52-SS-UP CE profiles from 5′ DQ52 primer libraries; and junctions with bait length 145–150 bp were used for analysis of DQ52-SS-UP CE profiles from 3′ DQ52 primer libraries. We isolated a larger bait length range for CE than for SE analysis because there was more extensive end processing during CE joining.

Normalization of HTGTS V(D)J-seq libraries for RAG cystic scanning activity analysis. For analysis of DQ52-SS-UP scanning activity, DQ52-SS-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-SS-DN scanning activity analyses, DQ52-SS-DN SE junctions were isolated from the 3′ DQ52 primer libraries and normalized to 2,400 DQ52-SS-UP CE junctions isolated from the same 5′ DQ52 primer libraries (Fig. 3b, e, 4c, Extended Data Fig. 4d, 5b, 7d, e, 10b).

Analysis of HTGTS V(D)J-seq libraries from DFL16.1JH4 line and its derivat-ives. A bait primer 102 bp upstream of the DH1-SS break site was used to generate the libraries for the DFL16.1JH4 line and its derivatives. Sequencing reads were aligned to the modified mm9 genome harbouring the DFL16.1JH4 sequence. Specifically, chr12:114,666,771–114,720,401 in mm9 was replaced with the sequence CCCCT (mm9_DFL16.1JH4). Libraries were normalized to total bait-sequenced reads as previously described. Libraries were prepared from G1-arrested cells as previously described. In brief, 10 million cells were crosslinked with 2% formaldehyde at room temperature for 10 min, quenched with glycine (final 0.125 M), 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 inhibitors per 50 ml and 100 units RNase A per ml). DNA was sheared with a Covaris instrument and size-normalized to total bait-sequenced reads as previously described. In brief, 20 million DNA fragments were used to generate libraries in Figs. 3f, 4d and Extended Data Figs. 8b, 9q (top) and 10c; 3C-HTGTS libraries were sequenced using paired-end 150-bp sequencing on a NextSeq550 or paired-end 300-bp sequencing on an Mi-Seq machine. Data were processed as previously described. As such, 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 homomorphology (> 5 bp) were excluded from analysis to avoid potential PCR artefacts. The overall HTGTS V(D)J-seq profiles before and after removing junctions with > 5 bp microhomology were very similar. Libraries were size-normalized to total junctions of the smallest library in the set of libraries for comparison. In addition, HTGTS V(D)J-seq method as described above. HTGTS V(D)J-seq method as described above. HTGTS V(D)J-seq method as described above. HTGTS V(D)J-seq method as described above. HTGTS V(D)J-seq method as described above. HTGTS V(D)J-seq method as described above.

ChIP–seq library preparation and analysis. ChIP was done with G1-arrested cells as previously described. For analysis of DQ52-SS-UP scanning activity, DQ52-SS-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-SS-DN scanning activity analyses, DQ52-SS-DN SE junctions were isolated from the 3′ DQ52 primer libraries and normalized to 2,400 DQ52-SS-UP CE junctions isolated from the same 5′ DQ52 primer libraries (Fig. 3b, e, 4c, Extended Data Fig. 4d, 5b, 7d, e, 10b).
captured on an Illumina flow cell for cluster generation. Libraries were sequenced on the illumina sequencing platform according to the manufacturer’s protocols. We sequenced 2.3 billion Hi-C read pairs in the control JHΔ-dCas9 line, yielding 1.3 billion Hi-C contacts; we also sequenced 2.2 billion Hi-C read pairs in JHΔ-dCas9-S1-sgRNA cells with the dCas9 impediment, yielding 1.2 billion Hi-C contacts. Hi-C libraries were analysed using the Juicer pipeline36, and visualized with Juicebox37. All the code used in the above steps is publicly available at https://github.com/aidenlab. Note that while Hi-C analysis did not distinguish between C57BL/6 and 129/Sv C5H alleles, it gave highly complementary results to the 3C-HTGTS with C57BL/6-specific iE8 bait with respect to interactions with or without the S1 impediment.

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

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

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

Code availability
HTGTS V(D)J-seq and 3C-HTGTS data were processed through a published pipeline available at http://robinmeyers.github.io/transloc_pipeline/. Code for Hi-C data processing is available at https://github.com/aidenlab. GRO-seq and ChIP-seq data were aligned to the mm9 genome with Bowtie2 v.2.2.8 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml), processed using SAMtools v.1.8 (https://sourceforge.net/projects/samtools/files/samtools/1.8/) and graphs were generated using the RsEQC tool v.2.6 (http://rsseq.sourceforge.net/#bam2wig-py).

30. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823 (2013).
31. Yang, L. et al. Optimization of scarless human stem cell genome editing. Nucleic Acids Res. 41, 9049–9061 (2013).
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).
33. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322, 1845–1848 (2008).
34. Mahat, D. B. et al. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). Nat. Protoc. 11, 1455–1476 (2016).
35. Marinov, G. K. ChiP-seq for the identification of functional elements in the human genome. Methods Mol. Biol. 1543, 3–18 (2017).
36. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3, 95–98 (2016).
37. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 3, 99–101 (2016).
38. Alt, F. W. & Baltimore, D. Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions. Proc. Natl Acad. Sci. USA 79, 4118–4122 (1982).
39. Cowell, L. G., Davila, M., Kepler, T. B. & Kelsoe, G. Identification and utilization of arbitrary correlations in models of recombination signal sequences. Genome Biol. 3, research0072 (2002).
40. Cowell, L. G., Davila, M., Yang, K., Kepler, T. B. & Kelsoe, G. Prospective estimation of recombination signal efficiency and identification of functional cryptic signals in the genome by statistical modeling. J. Exp. Med. 197, 207–220 (2003).
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).
42. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 329, 289–293 (2009).
43. Predeus, A. V. et al. Targeted chromatin profiling reveals novel enhancers in Ig H and Ig L chain loci. J. Immunol. 192, 1064–1070 (2014).
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).

Acknowledgements
We thank B. P. Sleckman for the v-Abl transforming, RAG2-deficient mouse pro-B cell line and Rag2-expressing shuttle plasmid, and Alt laboratory 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. A.P.A. was supported by a Thrasher Research Fund Early Career Award; 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).

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 analysed 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. analysed and interpreted all data other than Hi-C data. Y.Z., X.Z. and F.W.A. designed figures and Supplementary Video and wrote the manuscript. Z.B., Z.L., H.H., J.L., A.P.A. and E.L.A. helped to polish the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1547-y.
Correspondence and requests for materials should be addressed to F.W.A.
Peer review information
Nature thanks Andre Nussenzweig and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints.
Extended Data Fig. 1 | See next page for caption.
Extended Data Fig. 1 | Working model for role of loop-extrusion-mediated RAG scanning in driving deletion-biased D-to-J<sub>H</sub> 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 into the upstream portion of the RC within the IGCR1–iEμ/RC subdomain. Proximal downstream active RC chromatin impedes cohesion-mediated 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<sub>1H</sub> upstream of RC-based DQ52 past the open RAG1 subunit active site opposite the J<sub>H</sub>-bound active site in the other RAG1 subunit. This linear process aligns a downstream D-12RSS with the RAG-bound J<sub>H</sub>-23RSS for orientation-specific, deletional D-to-J<sub>H</sub> recombination. 

**e**, Upstream Ds are frequently passed without being used and most loop extrusion-mediated RAG scanning continues until reaching the 5′ CBE loop anchor (IGCR1), which strongly impedes (nearly blocks) loop extrusion and RAG scanning. The latter prolonged interaction may contribute to robust DFL16.1 utilization. 

**f–h**, Owing to the location of the RC, DQ52 can bind to the open RAG active site by diffusion (**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 and cleavage compared to its weaker RSS-UP. Other schematics in **b–h** are as described in Fig. 1 legend.
Extended Data Fig. 2 | See next page for caption.
Extended Data Fig. 2 | HTGTS V(D)J-seq analysis of V(D)J recombination outcomes in \( \text{DH}^{\text{+/-}} \) line and its mutant derivatives.  

**a**, Schematic of the two \( \text{Igh} \) alleles of the \( \text{DH}^{\text{+/-}} \) v-Abl pro-B line. This C57BL/6, 129/Sv mixed background line was derived by deleting the indicated region from the 129/Sv allele to inactivate it for V(D)J recombination.  

**b**, Southern blotting confirmation of allele deletion in the \( \text{DH}^{\text{+/-}} \) line. Done twice with similar results.  

**c**, C57BL/6 versus 129/Sv \( \text{D}_{\text{H}} \) usage in parental versus \( \text{DH}^{\text{+/-}} \) lines, as analysed via HTGTS V(D)J-seq (\( \text{J}_{\text{H1}} \) CE primer). The lack of 129/Sv-specific \( \text{D}_{\text{H}} \) in \( \text{DH}^{\text{+/-}} \) libraries confirmed the retention of C57BL/6 and absence of 129/Sv allele in this line.  

**d**, Bar chart shows utilization frequency of each \( \text{VH} \), \( \text{DH} \) and \( \text{JH} \) from \( \text{JH} \)-distal to \( \text{JH} \)-proximal locales (\( n = 3 \) independent libraries). Pie chart shows indicated V(D)J recombination products as percentage of total \( \text{Igh} \) junctions. Beyond predominant \( \text{DJ}_{\text{H1}} \) junctions, both low-level \( \text{V}_{\text{H}} \text{DJ}_{\text{H1}} \) joins\(^4,12 \) and inversional \( \text{J}_{\text{H1}} \text{DJ}_{\text{H1}} \) joins\(^38 \) were detected. There were very low levels of \( \text{J}_{\text{H1}} \) joins to ‘cryptic RSSs’, or to a different \( \text{J}_{\text{H1}} \)-RSS (other) that is likely to occur in extra-chromosomal excision circles\(^13 \).  

**e**, Utilization of each \( \text{D} \) as a percentage of total \( \text{DJ}_{\text{H1}} \) joins (\( n = 3 \) independent libraries).  

**f**, Strategy for analysis of D-RSS-DN versus D-RSS-UP utilization. The orientation of D coding sequences relative to the \( \text{J}_{\text{H1}} \) CE primer is preserved in primary and secondary joins for both D-RSS-DN and D-RSS-UP, allowing calculation of the relative utilization of D-RSS-DN versus D-RSS-UP.  

**g**, Utilization frequency of D-RSS-DN versus D-RSS-UP in the \( \text{DH}^{\text{+/-}} \) line.  

**h**, Effect 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 represent mean ± s.d. Data for \( \text{DH}^{\text{+/-}} \) line in **d**–**g** are two sets of three repeats each, with the latter done along with DLF16.1 mutants.
Extended Data Fig. 3 | See next page for caption.
Extended Data Fig. 3 | Generation and analyses of the D1–JH1+/− line and its mutant derivative lines. a, Coding and flanking D-RSS-UP and D-RSS-DN sequences and their RSS recombination information content (RIC) scores \(^{39,40}\) 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 DJ1 recombination on excision circle. Joining of J1 to D16 downstream of DFL16.1, which occurs on excision circles generated by primary joining between distal J1s (J12–J14) and distal D16s, is not subject to the same mechanistic constraints as chromosomal D-to-JH recombination\(^13\). To obviate such joins, we deleted J12–J14 in the D1–JH1+/− line to generate the D1–JH1+/− line. c, d, Southern blotting confirmation of D1–JH1+/− (done once after PCR confirmation; c) and intervening D1 inversion (done twice with similar results; d) lines. e, Utilization of D-RSS-UP and D-RSS-DN in the D1–JH1+/− line and its mutant derivatives with intervening D1 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 D1–JH1+/− cells with endogenous DFL16.1 deleted (n = 3 libraries for each genotype). e, f, Data represent mean ± s.d. from biologically independent samples, normalized to 70,000 total junctions for each library. g, Model of low level inversionsal RC-distal D joining involving loop-extrusion-based mechanism, which could bring distal upstream D-RSSs into diffusion radius of the RC. See Supplementary Discussion.
Extended Data Fig. 4 | See next page for caption.
Extended Data Fig. 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_{H}3-2 region (20%), with the latter probably resulting from secondary events on excision circles following primary J_{H}-to-distal D_{H} joins (left; also, see below). b, Southern blot confirmation of J_{H}\Delta lines (done once after PCR confirmation). c, Repeats of HTGTS V(D)J-seq experiments shown in Fig. 3a. 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_{H}3-2 in J_{H}\Delta lines, which is consistent with their excision circle origin. Also, unlike the D_{H}–J_{H}^{+/−} line with germline J_{H}s, robust RC downstream cryptic scanning activity is readily detected in the J_{H}\Delta lines. e, Repeats of GRO-seq experiments shown in Fig. 3d. Each library was normalized to a coverage of 10 million 100-nt reads for display. f–i, Model of cohesin loop-extrusion-mediated directional RAG scanning from RC DQ52-RSS-UP to upstream regions until reaching IGCR1 loop anchor. J–m, Model of extrusion-mediated directional RAG scanning from RC DQ52-RSS-DN to downstream regions until reaching the 3′ CBEs loop anchor. Transparent yellow rectangles in f, 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 Fig. 5 | See next page for caption.
Extended Data Fig. 5 | RAG cryptic targeting activity from DQ52-RSS-UP and DQ52-RSS-DN in JHΔ 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 D3-2 locales (grey transparent bars). Top, junctions plotted at 100-bp bin size. Bottom, examples of most robust peak near IGCR1 (I) and D3-2 (II) plotted at single-base-pair 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 eμ-Sμ, D3-2 and IGCR1 (grey transparent bars). Top, junctions plotted at 100-bp bin size. Bottom, examples of most robust Sγ2b (III) and 3′ CBEs (IV) locale peaks plotted at single-base-pair resolution. **c**, Low frequency DQ52-RSS-DN junctions upstream of RC detected by DQ52-RSS-DN bait. Top, expanded views of IGCR1 and D3-2 locales in b plotted at 20-bp bin size with representative junctions labelled (V–X). Bottom, single-base-pair resolution plot 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 CEs12, 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 or transcription impedes extrusion, allowing them to access RC-bound RAG by local diffusion15, analogous to diffusion-mediated DQ52-to-JH1 joining.
Extended Data Fig. 6 | See next page for caption.
Extended Data Fig. 6 | RAG targeting and transcriptional activity analysis in the DFL16.1J14-inv lines. a, Generation of the DFL16.1J14-inv line. Schematic shows two Igh alleles of DFL16.1J14 line and DFL16.1J14-inv line. In the DFL16.1J14 line, one Igh allele contains a nonproductive VDJ join involving VH1-2P and JH3, and the other allele harbours the DFL16.1J14 join. The DFL16.1J14-inv line was derived from the DFL16.1J14 line by inverting a 1-kb segment encompassing the DJH join using CRISPR–Cas9. b, Illustration of mechanism for RAG cryptic scanning activity from the RC DJH-RSS in DFL16.1J14 line (top), DFL16.1J14-inv line (middle) and DFL16.1J14-inv 3′ CBEs−/− line (bottom). c, Representative HTGTS V(D)J-seq profiles showing RAG cryptic scanning patterns of DFL16.1J14 line (top; n = 3 technical repeats), DFL16.1J14-inv line (middle; n = 3 biological replicates) and DFL16.1J14-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 three repeats of the DFL16.1J14-inv Rag2−/− line (n = 3 biological replicates).
Extended Data Fig. 7 | See next page for caption.
Extended Data Fig. 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 4-kb highly repetitive Sγ1 region on the C57BL/6 allele was introduced into the JHΔ-dCas9 line. b, Western blot confirmation of dCas9 expression in JHΔ-dCas9 lines but not the parental JHΔ line (done twice with similar results). c, RT–PCR confirmation of Sγ1-gRNA expression in the JHΔ-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 JHΔ-dCas9 lines and JHΔ-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, Expanded 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 JHΔ-dCas9 and JHΔ-dCas9-Sγ1-sgRNA lines. Each library was normalized to a coverage of 10 million 100-nt reads for display. Bar graph compares transcriptional activity of indicated regions (n = 3 libraries for each genotype). Data represent mean ± s.d. from biologically independent samples. P values were calculated via two-tailed paired t-test. NS, P ≥ 0.05. The modest decrease in Sγ2b transcription upon Sγ1–dCas9 binding is potentially due to impaired loop extrusion between Sγ2b and iEμ.
Extended Data Fig. 8 | See next page for caption.
Extended Data Fig. 8 | dCas9 binding impedes downstream loop formation in association with cohesin loading and accumulation at the impediment locale. a, Hi-C analysis of the 3′ Igh domain interaction of the JHΔ-dCas9 line compared with the JHΔ-dCas9-Sγ1-sgRNA line. We compared 1.3 billion contacts in the control line with 1.2 billion 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 the RC (B) and 3′ CBEs (F) locales, respectively, in the JHΔ-dCas9-Sγ1-sgRNA line. b, 3C-HTGTS repeats with iEμ bait (green stars) for JHΔ-dCas9 and JHΔ-dCas9-Sγ1-sgRNA lines shown in Fig. 3f. The iEμ bait primer strategy is shown above. Each library was normalized to 192,000 total junctions for analysis. While these lines retain downstream Cγ sequences on their 129/Sv allele (Extended Data Fig. 2b), the iEμ bait should have very low interactions in trans42. Blue and grey transparent boxes 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 activity43,44, was evident (see also Fig. 4d) and was accompanied by accumulation of RAD21 and NIPBL (see below) and a low level of RAG scanning activity (Extended Data Fig. 7d). c, RAD21 ChIP–seq profiles of JHΔ-dCas9 lines versus JHΔ-dCas9-Sγ1-sgRNA lines (n = 2 biological replicates). Each library was normalized to a coverage of 10 million 100-nt reads. d, NIPBL ChIP–seq profiles of JHΔ-dCas9 lines versus JHΔ-dCas9-Sγ1-sgRNA lines (n = 2 biological replicates). Each library was normalized to a coverage of 10 million 100-nt reads.
Extended Data Fig. 9 | See next page for caption.
Extended Data Fig. 9 | Working model for loop-extrusion-mediated RAG downstream scanning. 

**a–i**, Model for cohesin-mediated loop extrusion of chromatin past nascent *IgH* RC in JHΔ 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* subdomain. Illustrations show cohesin loading at the region downstream of RC. 

**b**, Cohesin-mediated extrusion promotes a linear interaction between the nascent RC and downstream regions. 

**c**, Robust transcription (green arrow) across the 1γ2b–Sγ2b impedes loop extrusion. 

**d**, In a subset of cells, loop extrusion proceeds past the 1γ2b–Sγ2b impediment to the 3′ CBEs loop anchor. 

**e–i**, Loop extrusion in JHΔ-dCas9-Sγ1-sgRNA lines is impeded, directly or indirectly, by dCas9-bound Sγ1. As dCas9 impediment is not a complete block, loop extrusion in a subset of cells proceeds downstream, allowing dynamic formation of sub-loops of RC with 1γ2b-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 1γ2b-promoter-driven transcription permits unimpeded RAG-bound RC extrusion to the 3′ CBEs anchor, increasing RAG scanning activity there. 

**q**, 3C-HTGTS analysis of RC interactions with DH and flanking regions in the JHΔ-dCas9 line and DH–JH+/- line. DpnII (n = 4, biological replicates) and NlaIII (n = 3, biological replicates) digestions are shown for the JHΔ-dCas9 line. The NlaIII digestion more clearly reveals an interaction peak near DH3-2 owing to a paucity of DpnII sites in that region. NlaIII digestion of the DH–JH+/- line shows a similar RC interaction pattern to that of the JHΔ-dCas9 line (r, n = 2 technical repeats). Bar graphs show relative RC interaction of the 25-kb intervening DH region (from DH2-3 to DH2-8) versus that of the same-sized neighbouring regions (n as indicated above). Data represent mean ± s.d. (q) or mean (r). P values calculated via two-tailed paired t-test.
Extended Data Fig. 10 | See next page for caption.
Extended Data Fig. 10 | Ectopic transcription of $\gamma_2b$-$S\gamma_2b$ region impedes downstream loop formation and RAG scanning. a, GRO-seq repeats for J$_{H\Delta}$-dCas9 lines ($\gamma_2b^{WT}$) and J$_{H\Delta}$-dCas9-$\gamma_2b$-del lines ($\gamma_2b^{\Delta/\Delta}$) shown in Fig. 4b. Each library was normalized to a coverage of 10 million 100-nt reads. b, HTGTS V(D)-seq repeats with DQ52-RSS-DN bait for $\gamma_2b^{WT}$ versus $\gamma_2b^{\Delta/\Delta}$ lines shown in Fig. 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 iE$\mu$ bait for $\gamma_2b^{WT}$ and $\gamma_2b^{\Delta/\Delta}$ lines for data shown in Fig. 4d. Each library was normalized to 150,000 total junctions for analysis. d, RAD21 ChIP–seq analysis for $\gamma_2b^{WT}$ and $\gamma_2b^{\Delta/\Delta}$ lines. Each library was normalized to a coverage of 10 million 100-nt reads for display. Bar graph shows comparison of RAD21 accumulation at the $S\gamma_2b$ region ($S\gamma_2a$ region as control) in $\gamma_2b^{WT}$ lines versus $\gamma_2b^{\Delta/\Delta}$ lines ($n = 3$ libraries for each genotype). Data represent mean ± s.d. from biologically independent samples. For bar graph presentation, the junction number recovered from the $S\gamma_2b$ region of $\gamma_2b^{WT}$ control samples was normalized to represent 100%, relative values of $S\gamma_2a$ region in the control and $S\gamma_2b$ and $S\gamma_2a$ regions in the $\gamma_2b^{\Delta/\Delta}$ samples are listed as a percentage of the control $S\gamma_2b$ values. P values were calculated by a two-tailed paired t-test. NS, $P \geq 0.05$. 
Corresponding author(s): Frederick W. Alt
Last updated by author(s): Aug 4, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
- Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

Next generation sequencing data were collected via Illumina sequencing platforms (MiSeq, NextSeq 550 or HiSeq 2500). Data generated from Hi-Seq 2500 were demultiplexed via bcl2fastq-1.8.3. Data generated from MiSeq and NextSeq 550 were demultiplexed via TranslocPreprocess.pl, a published pipeline available at (http://robinmeyers.github.io/transloc_pipeline/).

Data analysis

HTGTS V(D)J-seq and 3C-HTGTS data were processed through a 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, processed by samtools v1.8 and generated graph files via RSeQC tool v2.6. Statistical analysis in all assays were performed via GraphPad Prism6.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on 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; 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 Figs. 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...
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- [ ] Life sciences
- [ ] Behavioural & social sciences
- [ ] Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

**Sample size**
No statistical methods were used to predetermine sample size for all experiments. Sample sizes were chosen based on previous studies in this field that used similar sample sizes to generate reproducible results.

**Data exclusions**
No experiments were excluded from analysis.

**Replication**
Samples were analyzed with biological repeats. All attempts for replication were successful.

**Randomization**
Experiments were not randomized. Each experiment was performed with identified controls and mutant strains. Randomization was not relevant to the study as the study does not involve participant groups.

**Blinding**
Investigators were not blinded to allocation during experiments and outcome assessment. Blinding was not possible as investigators need to verify the control and matched mutant strains before each experiment. Also, based on previous studies in this field, these assays do not require blinding.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a | Involved in the study |
| --- | ---------------------- |
| [ ] | Antibodies |
| [ ] | Eukaryotic cell lines |
| [ ] | Palaeontology |
| [ ] | Animals and other organisms |
| [ ] | Human research participants |
| [ ] | Clinical data |

### Methods

| n/a | Involved in the study |
| --- | ---------------------- |
| [ ] | ChIP-seq |
| [ ] | Flow cytometry |
| [ ] | MRI-based neuroimaging |

### Antibodies

**Antibodies used**

- Rad21, Rabbit, Abcam, ab992, ChIP-Seq, 5 ug/ChIP
- Nipbl, Rabbit, Bethyl, A301-779A, ChIP-Seq, 5 ug/ChIP
- Cas9, Rabbit, Diagenode, C15310258-100, WB, 1:5,000
- β-Actin (8H10D10), Mouse, Cell Signaling Technology, 3700S, WB, 1:2,000

**Validation**

- Rad21 antibody (Abcam, ab992), manufacturer’s validation: [https://www.abcam.com/ rad21-antibody-chip-grade-ab992.html](https://www.abcam.com/rad21-antibody-chip-grade-ab992.html)
  This antibody was validated for ChIP-Seq by published papers including:
  Vian, L. et al. The Energetics and Physiological Impact of Cohesin Extrusion. Cell 173, 1165-1178 e1120, doi:10.1016/j.cell.2018.03.072 (2018).

- Nipbl antibody (Bethyl, A301-779A), manufacturer’s validation: [https://www.bethyl.com/ product/A301-779A/NIPBL+Antibody](https://www.bethyl.com/product/A301-779A/NIPBL+Antibody)
  This antibody was validated for ChIP-Seq by published papers including:
  Vian, L. et al. The Energetics and Physiological Impact of Cohesin Extrusion. Cell 173, 1165-1178 e1120, doi:10.1016/j.cell.2018.03.072 (2018).

- Cas9 antibody (Diagenode, C15310258), manufacturer’s validation: [https://www.diagenode.com/en/p/crispr-cas9-polyclonal-antibody](https://www.diagenode.com/en/p/crispr-cas9-polyclonal-antibody)
  This antibody was validated for ChIP in the following paper:
**Eukaryotic cell lines**

**Policy information about cell lines**

**Cell line source(s)**

The DH-JH+/− line was derived from a previously described RAG2+/− Eμ-bcl2+ v-Abl pro-B cell line from Dr. Barry P. Sleckman’s lab. Gapud, E.J., Lee, B.S., Mahowald, G.K., Bassing, C.H. & Sleckman, B.P. 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).

The DH-JH+/− line was used as the parental cell line to generate mutant derivatives including DFL16.1-RSS-DN-inv, DFL16.1-RSS-UP-inv, DFL16.1-inv, JHΔ and DH-JH1+/− lines. The DH-JH1+/− lines were used to generate mutant derivatives including DQ52-inv, DQ52Δ DFL16.1-to-DQ52, DFL16.1Δ DQ52-to-FL16.1, DFL16.1Δ DQ52-to-FL16.1-inv and intervening DHs inversion lines. The JHΔ lines were used to generate mutant derivatives including JHΔ-dCas9 and JHΔ-dCas9-sgRNA-S1 lines. The JHΔ-dCas9 lines were used to generate JHΔ-dCas9-γ1b lines. The DFL16.1-JH4 line was a published line from this lab. Jain, S., Ba, Z., Zhang, Y., Dai, H. & Alt, F.W. CTCF-Binding Elements Mediate Accessibility of RAG Substrates During Chromatin Scanning. Cell 174, 1-15 (2018).

The DH-JH+/−, JHΔ, DH-JH1+/−, intervening Ds inversion, DFL16.1-JH4-inv and DFL16.1-JH4-inv 3’CBE−/− and DFL16.1-JH4-inv RAG2−/− lines were verified via Southern blotting. The DFL16.1-JH4 line was used as the parental line to generate the DFL16.1-JH4-inv lines, which was used to further generate the DFL16.1-JH4-inv 3’CBE−/− and DFL16.1-JH4-inv RAG2−/− lines.

**Authentication**

The DH-JH+/−, JHΔ, DH-JH1+/−, intervening DHs inversion, DFL16.1-JH4-inv and DFL16.1-JH4-inv 3’CBE−/− lines were verified via Southern blotting. The DFL16.1-JH4 line was used as the parental line to generate the DFL16.1-JH4-inv lines, which was used to further generate the DFL16.1-JH4-inv 3’CBE−/− and DFL16.1-JH4-inv RAG2−/− lines.

**Mycoplasma contamination**

The cell lines were not tested for mycoplasma contamination.

**Commonly misidentified lines**

No commonly misidentified cell lines were used.

**ChIP-seq**

**Data deposition**

- Confirm that both raw and final processed data have been deposited in a public database such as GEO.

- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

**Data access links**

To review GEO accession GSE130224:

- Go to https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE130224
- Enter token erozikgorhctbil into the box

**Files in database submission**

- Rad21_dCas9Control_Rep1.bw
- Rad21_dCas9Control_Rep2.bw
- Rad21_dCas9Control_Rep3.bw
- Rad21_dCas9Control_Rep4.bw
- Nipbl_dCas9Control_Rep1.bw
- Nipbl_dCas9Control_Rep2.bw
- Nipbl_dCas9Control_Rep3.bw
- Nipbl_dCas9Control_Rep4.bw
- Rad21_dCas9Control_Rep1.bed
- Rad21_dCas9Control_Rep2.bed
- Rad21_dCas9Control_Rep3.bed
- Rad21_dCas9Control_Rep4.bed
- Rad21_dCas9Control_Rep1.broadpeak.bed
- Rad21_dCas9Control_Rep2.broadpeak.bed
- Rad21_dCas9Control_Rep3.broadpeak.bed
- Rad21_dCas9Control_Rep4.broadpeak.bed
Methodology

Replicates Two or three independent libraries were performed from at least two independent clones for each genotype analyzed.

Sequencing depth Read depth is > 50 million/sample

Antibodies Rad21 antibody (Abcam, ab992); Nipbl antibody (Bethyl, A301-779A)

Peak calling parameters Peak calling was performed using MACS2 with the following commands:

```
macs2 callpeak -t IP.bam -c Input.bam -n Result_directory --nomodel --extsize 51 --nolambda -B --SPMR -g mm --verbose 0 --broad
```

Data quality We relied on MACS2 algorithm to detect significant peaks, obtaining 15-30 thousand peaks at FRD5% and above 5 fold enrichment for Rad21 ChIP-Seq libraries and obtaining 4-12 thousand peaks at FRD5% and above 5 fold enrichment for Nipbl ChIP-Seq libraries.

Software ChIP-seq data was aligned to mm9 by bowtie 2.2.8., processed by samtools v1.8 and generated graph files via RSeQC tool v2.6. MACS 2.1.2 was used to call peaks.